Bearings Ball Fault Detection Using Kullback Leibler Divergence in the EMD framework

Zahra Mezni, Claude Delpha, Demba Diallo, Ahmed Braham

Bearings Ball Fault Detection Using Kullback Leibler Divergence in the EMD framework

Zahra MEZNI*, Claude DELPHA†, Demba DIALLO‡,Ahmed BRAHAM*

*Laboratoire Matériaux, Mesures et Applications,INSAT,
Université de Carthage, Tunisia
†Laboratoire Des Signaux et Systèmes, UMR 8506
CNRS - CentraleSupélec - Univ. Paris-Sud, Gif sur Yvette, France
‡Group of Electrical Engineering of Paris, UMR 8507
CNRS -CentraleSupélec - Sorbonne Université - Univ. Paris-Sud, Gif sur Yvette, France

***Abstract*—In this work, we develop a fault detection methodology based on the use of selected narrow band time series signals. It is based on the Kullback Leibler Divergence (KLD) and selected components obtained from the Empirical Mode Decomposition (EMD) applied to non stationary time series vibration signals. The EMD decomposes the signal into narrow frequency bands components called Intrinsic Mode Functions (IMFs). A first selection of the most energised and consequently the most sensitive IMFs to fault occurrence is proposed. Thanks to the IMF to Noise Ratio computation, Thus, the retained components are timely analysed using the Kullback Kullback Leibler Divergence to proceed the fault detection. A quantitative sensitivity criteria is derived to evaluate the fault detection performances and confirmed by a probabilistic analysis. The proposed methodology is validated using an experimental dataset from the Case Western Reserve University with three different fault severities and operating load conditions. With this proposed methodology a 100% probability of detection is obtained with each of the first six selected IMFs, the best results being achieved with IMF_2.**

***Index Terms*—Fault Detection and Diagnosis; Empirical Mode Decomposition; Intrinsic Mode Function Selection; Feature selection; Statistical Analysis; Kullback Leibler Divergence; Bearing faults**

I. Introduction

Rolling bearings are key components of electrical machines that operate in various industrial applications. Unfortunately, these components are the major cause of the rotating machines failure with a varying rate ranging from 40% to 90% from small to large devices [1]. Therefore, a preventive or condition-based maintenance of the bearings is necessary to keep the process operating, increasing its availability and reliability. As a consequence this field of research is becoming intensively investigated with various approaches for fault detection and diagnosis. They can be broadly classified into physics-based, human experience-based and data driven-based methods [2], [3]. The latter is used in this work hereafter with the vibration signals collected from accelerometers as they contain huge amount of information on the system dynamics. Moreover it is undoubtedly one of the most used method for condition monitoring and diagnosing of rotatory machineries [4] compared to Motor Current Signature Analysis (MCSA) [5]. In this work, we propose to use the Empirical Mode Decomposition (EMD) in the preprocessing step as mechanical faults have significant signatures with multiple frequency components. Indeed, this technique which deals with the real physical behaviour of the system, consists of decomposing a non stationary signal into a finite number of modes called Intrinsic Mode Functions (IMFs) by means of numerical approximations [6], [7]. These IMFs will then be mono-component narrow frequency band signals containing a part of the original signal information. In the literature, this technique is widely used for mechanical type fault detection purpose (induction motors broken bars, bearings, gear box, ...) [8], [9]. It is a powerful and advantageous technique dealing with nonlinear and non-stationary signals for fault diagnosis. Most of the time, the extracted IMFs are studied in the frequency domain using a spectral analysis to characterise the faulty behavior. Nevertheless, all the extracted modes are not affected by the fault and without an a priori knowledge on the system characteristics (bearing dimensions,gear size, ...) it is difficult to estimate the frequency range and IMFs related to the presence of the fault. In this blind case, a selection technique for eliminating all the extracted modes not or less affected by the fault is necessary. The work proposed here concerns first the selection of the most relevant IMFs for the fault diagnosis. In second, we propose to analyse the IMFs in the time domain and then evaluate the fault detection in considering the selected features. A particular analysis of these features based on the Kullback Leibler Divergence (KLD) as a fault presence indicator because of its high sensitivity to incipient faults [10] is evaluated. For all this methodology, a validation is proposed using bearings data. For these kind of data, bearing ball faults are considered as the most difficult ones to be detected properly [11], [12] for bearings diseases. We propose then to focus on these bearing ball faults with small severity. The final detection performances of the methodology are evaluated through the Receiver Operating Curves (ROC) applied on the KLD iterations of the retained IMFs.

The remainder of this paper is organised as follows. Section II introduces the procedure and describes the preprocessing step and feature extraction used to sort the IMFs. In Section

III, bearing ball fault detection application is addressed through the KLD applied on the selected IMFs. The detection perrmances are then presented and discussed. Section IV draws the conclusion.

II. Preprocessing and feature extraction

The main steps of Fault Detection and Diagnosis (FDD) methodology are recalled in Figure 1. Because an accurate physics-based model for the faulty characteristics of the bearings is tedious to obtain, taking advantage of the available historical data, a data driven approach is adopted in the first step [13], [14].

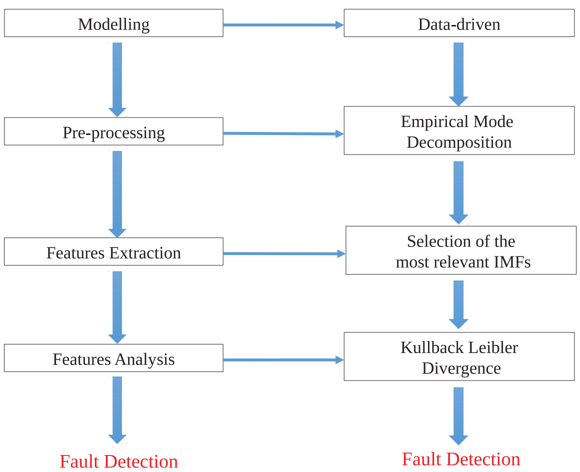


Fig. 1. Flowchart of the Fault Detection Methodology

~~Mechanical type~~ faults mostly modify the power spectrum of electrical and mechanical characteristics. Therefore the preprocessing (second step of the methodology) will consist of transforming the initial vibration signal from the time domain to the frequency domain. However because the operating conditions are variable, the signals are non stationary. In the literature, the Empirical Mode Decomposition (EMD) and its derivatives like Ensemble EMD, Complete Ensemble EMD are recognised as efficient methods for decomposing non stationary signals. Despite the limiting conditions (boundary effects and stopping criteria for sifting) we have adopted EMD to extract the fault features from the vibration signals in the frequency domain.

A. Experimental setup description

For this work, vibration signals which encompass all working conditions are obtained using the Case Western Reserve University (CWRU) Bearing Data Center [15]. According to the provider, the data were collected under the different following conditions:

- 2 hp Reliance Electric motor.
- 12 kHz for the sampling frequency.
- Acquisition using three accelerometers located respectively in the load zone, orthogonal to the load zone and in the bearing clearance zone.
- Three single point faults ranging from 0.007 to 0.021 inch at the balls.
- Experiments are repeated for the motor being unloaded (0%), 50% loaded, full loaded (100%) and over loaded (150%).

The principle of the test used for these data is given in Figure 2.

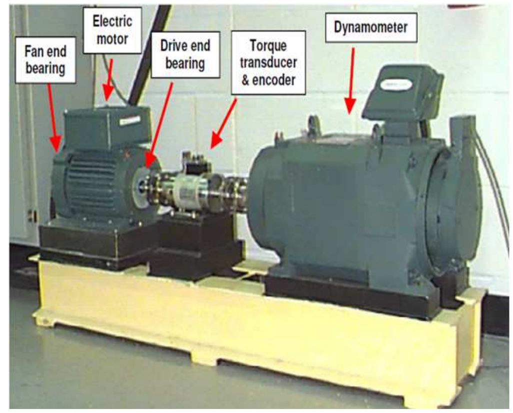


Fig. 2. CWRU Bearing test bed

B. Data preprocessing

As previously mentioned, the EMD is adopted for the preprocessing step.

1) Empirical Mode Decomposition brief description: Introduced by Huang et al [6], the EMD can decompose a non stationary signal $S(t)$ into a finite number N of narrow band oscillatory mono-components known as Intrisic Mode Functions (IMFs) and a non-zero mean residue $r(t)$ representing the central tendency of $S(t)$:

$$S(t) = \sum_{i=1}^{N} IMF_i(t) + r(t) \quad (1)$$

The main steps of the decomposition are summarized as below:

Step 1: Identifying all the local extrema of $S(t)$
Step 2: Estimating the Upper/Lower envelops through the cubic spline interpolation
Step 3: Computing the mean envelope by averaging the Upper/Lower envelops
Step 4: Generating the actual IMF, $h_k(t)$, by subtracting the mean envelop from $S(t)$
Step 5: Checking if $h_k(t)$ is an IMF
- If NO: Set $S(t) = h_k(t)$ and go to Step 1
- If YES:

Step 6: Set $h_k(t)$ to be the nth IMF and the residue function $r(t) = S(t) - \sum_{i=1}^{n} IMF_i(t)$
Step 7: Is $r(t)$ a monotonic trend?
- If NO: Let $S(t) = r(t)$ and go to Step 1
- If YES: The decomposition process is finished and $r(t)$ is the final residue

An IMF must satisfy the following two conditions:

1) In the whole data set, the number of extrema and the number of zero crossings should be equal or differ by one at most.
2) At any point, the local average of the upper and lower envelops is zero. This extraction is ensured by a sifting process.

C. Feature Extraction: Selection of the most relevant IMFs

The decomposition of the healthy signal using the EMD leads to a finite number of IMFs. In our application, the decomposition lead to 17 IMFs and a residue. Each IMF will contain a part of the energy in the original signal. The higher is this energy, the more sensitive will be the corresponding IMF. Therefore not all the extracted IMFs can be used to detect the fault occurrence. As a consequence, a selection of the most sensitive components is proposed with the method described in Figure 3.

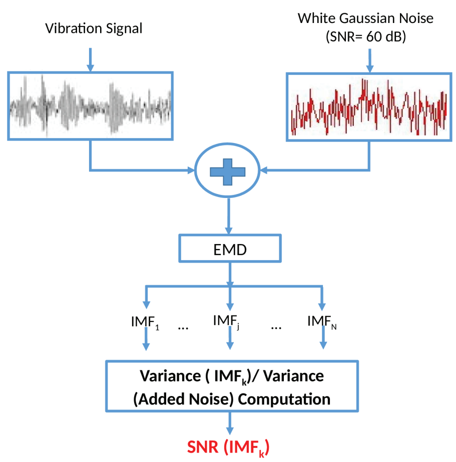


Fig. 3. IMF selection procedure

The methodology, consists of the following steps:

Ext 1: Addition of a weak white Gaussian Noise leading to a Signal to Noise Ratio SNR = 60dB in the original healthy signal. The EMD decomposition of the resulting noisy signal is very close to the original one because of the weak noise level.

Ext 2: Decomposition of the resulting signal to obtain the N IMFs.

Ext 3: The ratio of the variance of each IMF to the variance of the added noise (IMF to Noise Ratio denoted SNR_{IMF_i}) is calculated.

Figure 4 depicts the evolution of the SNR_{IMF_i} along with the IMF rank i under the different load conditions: as the variance of the noise is always constant ($SNR = 60dB$) the decrease of the IMF's Signal to Noise Ratio means that the IMF's variance or energy is degrading. In the following all the IMF with a SNR_{IMF_i} decreasing more than 20% from the maximum value will not be considered for fault detection. In our case it corresponds to the IMFs with a rank > 8 whatever the load conditions.

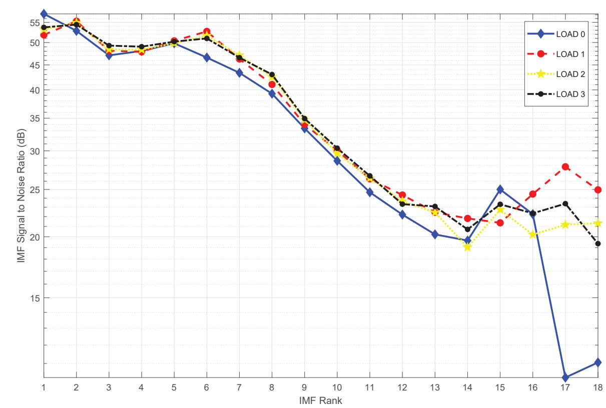


Fig. 4. IMF selection procedure

This methodology has been successfully applied by Z. Mezni et al [16] when using the first two statistical moments Mean and Variance as detection criteria for bearings ball fault detection problem. For this work, we intend to improve the detection in the case of incipient faults and retrieve more IMFs able to lead to an accurate fault detection. In this work, the proposed detection criteria is the Kullback-Leibler Divergence (KLD).

III. Ball Fault Detection using Kullback Leibler Divergence (KLD)

A. KLD basics

The KLD is an informational measure and an instance of f-divergence family. It has been used in many signal processing applications such as fault diagnosis, pattern recognition, anomaly detection [17], [18]. It computes the dissimilarity between two Probability Density Functions (PDF) $m(z)$ and $n(z)$ of a continuous random variable z through the Kullback Leibler Information (KLI) from m to n as [19]:

$$KLI(m||n) = \int_{-\infty}^{+\infty} m(z) \log \frac{m(z)}{n(z)} \mathrm{d}z \qquad (2)$$

The KL Divergence is then defined as the symmetric version of the KL Information [20] denoted by:

$$KLD = KLI(m||n) + KLI(n||m) \qquad (3)$$

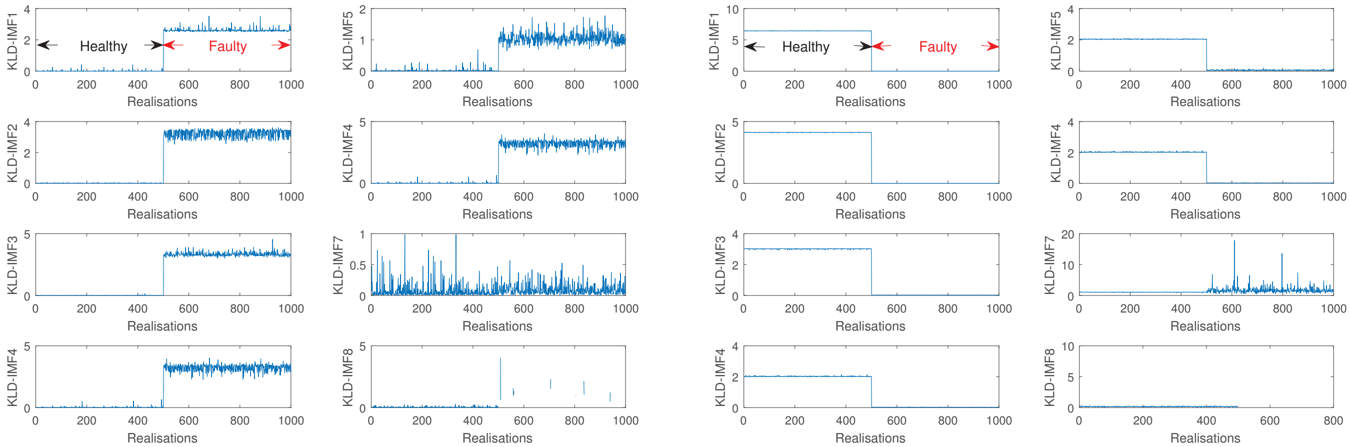
Previous works have compared the Hotelling T^2, the Squared Prediction Error SPE and Q traditional statics [10], [21] to the KLD. The results have shown that the KLD is most sensitive to incipient fault. It is therefore adopted to analyse the fault features extracted from the EMD.

B. Bearing ball fault detection

The KLD will be evaluated for each of the retained IMF according to the following method:

A1 : Compute the EMD on the original healthy signal.

A2 : Compute the probability density functions (PDF) for each selected IMF.



(a) Load 0%

(b) Load 100%

Fig. 5. KLD evolution for each retained IMF

A3 : Repeat Step A1 and Step A2 500 times by generating each time a weak White Gaussian Noise with $SNR = 60dB$ and add it to the original signal.

A4 : Compute the EMD on the resulting noisy signal signal.

A5 : Compute the PDF for each retained IMF and then:

1) Compute the corresponding PDF
2) Evaluate the KLD for the 500 realisations

A6 : Repeat Step A1 to Step A5 for the faulty signal.

The results of this evaluation are presented in Figure 5 for the different ~~loaded~~ operating conditions and for the smallest ball fault severity 0.007 inch:

- The realisations are made in such a way that $R = [1 : 500]$ are dedicated to the healthy state and $R = [501 : 1000]$ to the faulty one.
- Loads conditions : 0%, 50%, 100% and 150%

Figure 5 displays only the unloaded case (a) and the full one (b). From the depicted results, we can draw the conclusion that only the first six IMFs exhibit a significant sensitivity to the fault occurrence. Therefore IMFs 7 and 8 will not be such efficient for fault detection. In this case, these IMFs can be removed and ~~not being~~ considered for the ~~rest of study for the~~ fault detection ~~process~~.

C. KLD fault detection efficiency

In this section, we evaluate the sensitivity of the KLD to the fault severities and load variations. The sensitivity ~~criteria~~ is defined as in [13]:

$$S(KLD) = \frac{\overline{KLD_f} - \overline{KLD_h}}{\max(KLD_h - \overline{KLD_h})} \tag{4}$$

Where:
$\overline{KLD_h}$: Mean of the KLD iterations of the healthy signal.
$\overline{KLD_f}$: Mean of the KLD iterations of the faulty signal.
The value of the sensitivity is related to the fault detection performances through the Probability of Detection (PD) as follows:

- If $\mid (S(KLD) \mid > 2$: $PD = 1$; the fault is detected in 100% of the cases.
- If $\mid (S(KLD) \mid < 2$: $PD < 0.5$; the fault is detected in less than 50% of the cases.

In this work, the sensitivity is evaluated according to the flowchart in Figure 6.

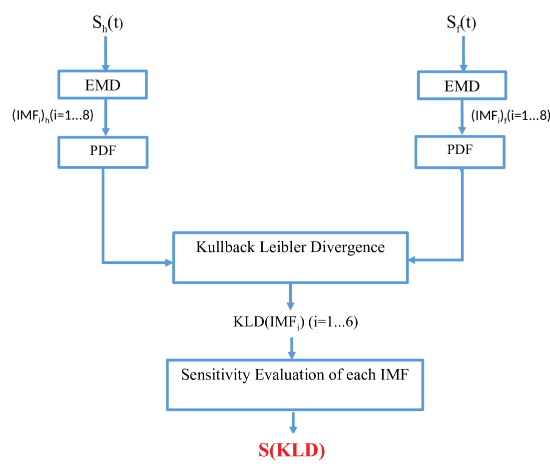


Fig. 6. Flowchart of the KLD sensitivity evolution

In fact, the methodology is carried out through the following steps :

B1: EMD decomposition of the healthy signal $S_h(t)$;
B2: PDF computation for each retained IMFs;
B3: Step B1 and Step B2 are ~~conducted for~~ 500 times;
B4: Repeat Step B1 to Step B3 for the faulty signal $S_f(t)$;
B5: KLD computation;
B6: Sensitivity evaluation for the first six IMFs;

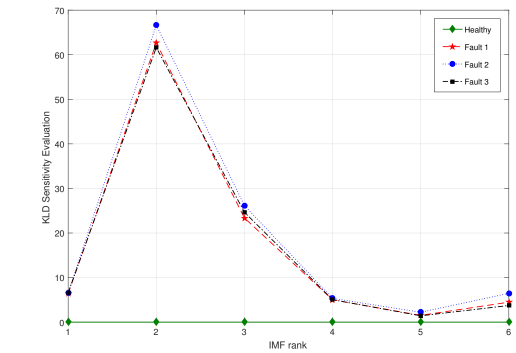


Fig. 7. KLD sensitivity evaluation for no load case

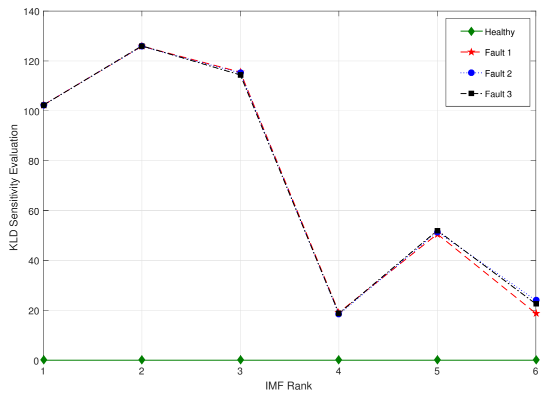


Fig. 8. KLD sensitivity evaluation for full load case

For the no load case, the results depicted displayed in Figure 7 are consistent with the data in Table I. For all the remaining IMFs the sensitivity is higher than 2 with the best results obtained for IMF_2. In the full load case, the results are even better as it can be deduced from both graphically in the Figure 8 and Table II where the poorest value of sensitivity is equal to 18.382. For all the operating conditions and the three fault severities, the fault is detected with no error.

TABLE I
KLD SENSITIVITY EVALUATION FOR EACH IMF IN UNLOADED CASE AND WITH THREE FAULT SEVERITIES

	Healthy	Fault 1	Fault 2	Fault 3
IMF_1	0	6.375	6.546	6.602
IMF_2	0	62.644	66.654	61.61
IMF_3	0	23.381	26.162	24.663
IMF_4	0	5.014	5.338	5.039
IMF_5	0	1.526	2.223	3.74
IMF_6	0	4.483	6.465	3.74

D. Fault detection and diagnosis performances

The present part of work aimed to investigate the performances of the fault detection based on the KLD, using the Receiver Operating Characteristic (ROC) curves analysis.

TABLE II
KLD SENSITIVITY EVALUATION FOR EACH IMF IN FULL LOAD CASE AND WITH THREE FAULT SEVERITIES

	Healthy	Fault 1	Fault 2	Fault 3
IMF_1	0	102.341	102.319	102.301
IMF_2	0	125.803	125.803	126.062
IMF_3	0	115.512	115.209	1114.379
IMF_4	0	19.119	18.382	18.737
IMF_5	0	50.578	51.473	52.051
IMF_6	0	18.795	24.008	22.578

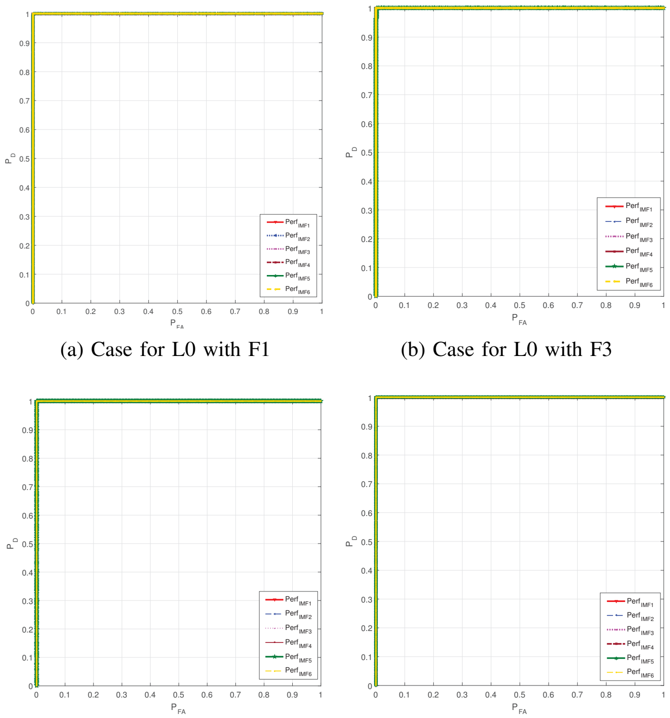
In fact, ROC methodology is pertaining to signal diagnosis field [22] where it was used to ascertain if an electronic receiver is capable to sufficiently differentiate between signal and noise. It has been applied in several areas such as medical imaging and radiology, non-destructive testing [23].
For this performance evaluation technique, the most commonly used global index of diagnostic accuracy is the Area Under the ROC curve (AUC) [24] which is widely recognized as the measure of a diagnostic test's discriminatory power. Considering the ROC curves, the Probability of Detection (P_D) along with the Probability of False Alarm (P_{FA}) will be displayed according to the AUC variation between non-informative $(AUC = 0.5)$, less accurate $(0.5 < AUC \leq 0.7)$, moderately accurate $(0.7 < AUC \leq 0.9)$, highly accurate $(0.9 < AUC < 1)$ and perfect performances $(AUC = 1)$ [25].

For our case, in order to evaluate the KLD ball fault detection performances, the ROC curves will be plotted for the IMF_1 to IMF_6 as shown to be the most relevant ones. Figure 9 displays the results for the two operating points (no load L0 and full load L2) and the two fault levels (0.007 inch denoted F1 and 0.021 inch denoted F3).

From the above results displayed on Figure 9 we notice that all the AUC values for all the 6 selected IMFs and under the different work conditions have reached the maximum value 1. This leads to point out the high performance and proves the sensitivity of the proposed method towards fault detection even if with small severity at an early stage of development.

IV. CONCLUSION

In this paper, we have evaluated the performances of the Kullback Leibler Divergence (KLD) for bearing ball fault detection in specifically selected narrow frequency band signals. The KLD was applied on the fault features extracted with the Empirical Mode Decomposition (EMD) used to process the non stationary time series vibration signals components. First a selection of the most energized components denoted as Intrinsic Mode Functions (IMFs) is performed. Applied to Bearing ball fault detection, 6 IMFs over 17 are selected. Then in a second time, the fault detection capability is performed with the application of the KLD on these time series selected



(a) Case for L0 with F1 (b) Case for L0 with F3

(c) Case for L2 with F1 (d) Case for L2 with F3

Fig. 9. KLD ROC curves

components. Results are computed for several conditions of load and faults severities ~~on the ball bearing~~. The sensitivity of the method has been quantitatively evaluated and correlated to the Probability of Detection. Finally the fault detection performances are examined using the Receiver Operating Curves (ROC) of the KLD iterations for the six first IMFs and the results show that the ball fault is perfectly detected ($AUC = 1$) with zero false alarms. For this work, the whole methodology has been applied to experimental raw data obtained from the Case Western Reserve University database. The results presented ~~highlights~~ that from the first six selected IMFs, the second one denoted IMF_2 is the most sensitive one for all the operating points and for all the fault severities.

REFERENCES

[1] M. Riera-Guasp, J. A. Antonino-Daviu, and G. A. Capolino, "Advances in electrical machine, power electronic, and drive condition monitoring and fault detection: State of the art," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1746–1759, March 2015.

[2] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[3] R. Kothamasu, S. H. Huang, and W. H. VerDuin, "System health monitoring and prognosticsa review of current paradigms and practices," *The International Journal of Advanced Manufacturing Technology*, vol. 28, no. 9-10, pp. 1012–1024, 2006.

[4] M. Cerrada, R.-V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. V. de Oliveira, and R. E. Vásquez, "A review on data-driven fault severity assessment in rolling bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, 2018.

[5] F. B. Abid, S. Zgarni, and A. Braham, "Bearing fault detection of induction motor using swpt and dag support vector machines," in *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, Oct 2016, pp. 1476–1481.

[6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[7] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 01, no. 01, pp. 1–41, 2009.

[8] H. Akcay and E. Germen, "Identification of acoustic spectra for fault detection in induction motors," in *AFRICON, 2013*. IEEE, 2013, pp. 1–5.

[9] C. Zhang, J.-J. Chen, and Y.-L. Xu, "A bearing fault diagnosis method based on emd and difference spectrum theory of singular value," *Journal of vibration engineering*, vol. 24, no. 5, pp. 539–545, 2011.

[10] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on kullback–leibler divergence using principal component analysis: Part I," *Signal Processing*, vol. 94, pp. 278–287, 2014.

[11] M. Cerrada, R. Sànchez, C. Li, F. Pacheco, D. Cabrera, J. de Oliveira, and R. Vasquez, "A review on data-driven fault severity assessment in rolling bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, 01 2018.

[12] J. Harmouche, C. Delpha, and D. Diallo, "Improved fault diagnosis of ball bearings based on the global spectrum of vibration signals," *IEEE Transactions on Energy Conversion*, vol. 30, no. 1, pp. 376–383, 2015.

[13] J. Harmouche, C. Delpha, D. Diallo, and Y. Le Bihan, "Statistical approach for nondestructive incipient crack detection and characterization using kullback-leibler divergence," *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1360–1368, 2016.

[14] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.

[15] U. Case Western Reserve University, "Bearing data centre," http://csegroups.case.edu/bearingdatacenter/pages/download-data-file/, 2018, [Online; accessed March-2018].

[16] Z. Mezni, C. Delpha, D. Diallo, and A. Braham, "Bearing fault detection using intrinsic mode functions statistical information," in *IEEE International Multi-Conference on Systems, Signals and Devices (SSD'18)*, 2018.

[17] M. Xie, J. Hu, S. Guo, and A. Y. Zomaya, "Distributed segment-based anomaly detection with kullback–leibler divergence in wireless sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 101–110, 2017.

[18] N. Saleem and G. Ijaz, "Low rank sparse decomposition model based speech enhancement using gammatone filterbank and kullback–leibler divergence," *International Journal of Speech Technology*, pp. 1–15, 2018.

[19] C. Delpha, D. Diallo, and A. Youssef, "Kullback-leibler divergence for fault estimation and isolation: Application to gamma distributed data," *Mechanical Systems and Signal Processing*, vol. 93, pp. 118–135, 2017.

[20] M. Raitoharju, Á. F. García-Fernández, and R. Piché, "Kullback–leibler divergence approach to partitioned update kalman filter," *Signal Processing*, vol. 130, pp. 289–298, 2017.

[21] A. Anderson and H. Haas, "Kullback-leibler divergence (kld) based anomaly detection and monotonic sequence analysis," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*. IEEE, 2011, pp. 1–5.

[22] J. A. Nevin, "Signal detection theory and operant behavior: A review of david m. green and john a. swets' signal detection theory and psychophysics." *Journal of the Experimental Analysis of Behavior*, vol. 12, no. 3, pp. 475–480, 1969.

[23] C. Nockemann, H. Heidt, and N. Thomsen, "Reliability in ndt: Roc study of radiographic weld inspections," *NDT & E International*, vol. 24, no. 5, pp. 235–245, 1991.

[24] D. Faraggi and B. Reiser, "Estimation of the area under the roc curve," *Statistics in medicine*, vol. 21, no. 20, pp. 3093–3106, 2002.

[25] M. Greiner, D. Pfeiffer, and R. Smith, "Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests," *Preventive veterinary medicine*, vol. 45, no. 1-2, pp. 23–41, 2000.