

The Development of An Effective Time-Series Based Fault Identification Technique using Parametric-Distance Method

Pulung Nurprasetyo, Komang Bagiasna, Djoko Suharto, and Tegoeh Tjahjowidodo

Abstract— This paper presents the implementation of the combination of time-series modeling and nearest neighbor classification method in detecting common faults in rotating machineries. In this paper we propose the utilization of parametric distance as an instrument to diagnose faults. The parametric distance is defined as the Euclidean distance between the vector of parameters of an unknown fault and the vector of parameters of known faults obtained from the learning stage. Since the vectors are defined in a hyperspace spanned by the parameters of the identified time-series model, the parametric distance is definitely metric. The method has been successfully implemented in the laboratory using a simple vibration test rig.

Index Terms— Time series modeling, Euclidean distance, fault identification, rotating machinery

I. INTRODUCTION

In rotating machineries, it has been observed over the years that condition-based maintenance (CBM) plays a better role than the traditional run-to-breakdown and time-based preventive maintenance in terms of maximizing safety, security, and minimizing cost and downtime [1]. With the advent of new instruments and sophisticated electronic devices, the task of monitoring the condition of machineries becomes relatively trivial. However, without proper method in diagnostics which involves fault detection, isolation, and identification, finding the remedy for the unhealthy machine will become complicated and error-prone.

For rotating machineries, the most common CBM will involve vibration level measurement. The basic concept is that a healthy machine will produce minimum vibration. For some

simple equipment such as pumps and motors, there exist typical standards for the vibration levels of healthy and unhealthy machines [2, 3]. In addition, it is also known that any particular fault will produce a specific vibration waveform. The symptoms of common faults such as unbalance, misalignment, mechanical looseness, bearing failures, etc., have been well researched and documented [3, 4].

The most common method used in the industry is the frequency-domain analysis by means of fast Fourier transform (FFT). Despite its popularity, it is also known that the above method will perform poorly for low signal to noise ratio case. As an alternate and to complement the frequency domain analysis, Gersch et al. [5] proposed a method that combines time-series modeling (TSM) and pattern recognition. In their paper, Gersch et al. used Kullback distance which involves logarithmic and subtraction operation. As a result, the distance measure does not satisfy metric requirement since it might yield negative numbers. This fact has been verified in the preliminary stage of this research [6]. Other researchers, Hong et al. [7] used five different distance measures, i.e. the Euclidean, Residue, J-Divergence, Kullback-Leibler (KL), and Kurtosis ratio. The last two suffers the same problem as the Kullback distance, in the sense that they are not metric. Following Hong et al., Mechevske and Mathew [8] implemented Kullback distance in fault detection of roller bearings. Considering the formulation, the distance measure being used is prone to giving negative number and does not satisfy triangle inequality. A better approach was suggested by Staszewski et al. [9], in which Mahalanobis distance was employed. Since Mahalanobis distance is a quadratic form, the distance measure satisfies metric requirements. However, the feature vector was constructed from Wigner-Ville time-frequency distribution. Consequently, the size of the feature vector is normally large. In a more recent development, Pan et al. [10] proposed the use of Itakura distance. Because of the formulation, which includes logarithmic operation, the distance measure may not satisfy the metric requirements.

In this paper, a new and simple parametric distance is proposed. This method belongs to the family of nearest neighbor classification method (NNCM) of faults. Unlike the older Kullback distance which does not satisfy the metric distance requirements, i.e., that distance measure has to be positive-semi definite and satisfy the triangle inequality, the

Manuscript received September 10, 2011

Pulung Nurprasetyo is with the Faculty of Mechanical and Aerospace Engineering, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia (phone 62-22-2504243; fax: 62-22-2534099; e-mail: ipn@ftmd.itb.ac.id, pulungnur@gmail.com).

Komang Bagiasna is also with the Faculty of Mechanical and Aerospace Engineering, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia (e-mail: kb@dynamic.pauir.itb.ac.id).

Djoko Suharto is also with the Faculty of Mechanical and Aerospace Engineering, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia (e-mail: ds@labsurya.ms.itb.ac.id).

Tegoeh Tjahjowidodo is with the Division of Mechatronics and Design, School of Mechanical and Aerospace Engineering, College of Engineering, Nanyang Technological University, Singapore (e-mail: TTEGOEH@ntu.edu.sg).

newly proposed parametric distance is metric.

The rest of the paper is organized as follows: a comprehensive review of time-series modeling is presented in section 2, followed by an elaboration of the parametric distance method in section 3, description of the laboratory experiment and typical time and frequency domain characteristics of the data in section 4, construction of reference feature vector in the learning stage in section 5, implementation of the method in identifying new data set in section 6, and the conclusions are finally presented in section 7.

II. TIME-SERIES MODELING (TSM)

In TSM, any waveform signal will be regarded as the output of a linear system upon injecting Gaussian white noise into the system. Mathematically, the model may be expressed as the following stochastic difference equation [11 – 13]:

$$S_D : \sum_{i=0}^{n_a} a_i^o y[t - i] = \sum_{j=0}^{n_c} c_j^o w[t - j] \quad (1)$$

In the above equation, t stands for the discrete time index, the leading coefficients a_0^o and c_0^o are both equal to 1 (monic polynomial requirement), $w[t]$ is the independently and identically distributed (i.i.d.) Gaussian random sequence $N(0, \sigma_w^2)$, and the superscript o denotes the true parameters of the system.

Using the back shift operator q^{-1} , the model may be rewritten in the following AutoRegressive Moving Average, ARMA(n_a, n_c), form:

$$S_D: A(q^{-1}) y[t] = C(q^{-1}) w[t] \quad (2)$$

in which $A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a}$ is the AutoRegressive (AR) polynomial, $C(q^{-1}) = 1 + c_1 q^{-1} + \dots + c_{n_c} q^{-n_c}$ the Moving Average (MA) polynomial, $y[t]$ the output of the linear system, $w[t]$ Gaussian i.i.d. $N(0, \sigma_w^2)$ white noise input, and q^{-1} is the back shift operator, i.e., $q^{-1}y(t) \equiv y(t - 1)$.

Furthermore, the ARMA(n_a, n_c) system is assumed to satisfy the following standard requirements:

1. The roots of the $A(q^{-1})$ and $C(q^{-1})$ polynomials lie strictly inside the unit disk in the complex plane (stationarity and invertibility assumptions, respectively).
2. The signal $y[t]$ is wide sense stationary with auto Power Spectral Density (auto PSD) that is a rational function of frequency.

Based on the above, the measured signal $y[t]$ may be modeled as the following ARMA(n_a, n_c):

$$\begin{aligned} M_D: \sum_{i=0}^{n_a} a_i y[t - i] &= \sum_{j=0}^{n_c} c_j w[t - j] \\ E[w(s)w(t)] &= \sigma_w^2 \delta(s - t) \end{aligned} \quad (3)$$

A. Pre-modeling stage

In order to obtain an appropriate ARMA model, i.e., a model that is linear, stable, and invertible, and having Gaussian white noise $N(0, \sigma_w^2)$ as the input, we need to perform stationary test. The test consists of qualitative, as well as quantitative test. Qualitative test is performed by direct observation of time-frequency distribution of the auto power spectra. In this case, the following periodogram is utilized as an estimate of the power spectral density (PSD) [13]:

$$\hat{P}_{PER}(f) = \frac{1}{N} \left| \sum_{t=0}^{N-1} x[t] e^{-j2\pi f t} \right|^2 \quad (4)$$

In equation (3), $\hat{P}_{PER}(f)$ denotes the periodogram of the signal $x[t]$, N number of data, f signal frequency [Hz], t discrete time index, and j the complex operator. After initial mean removal, the data is divided into several independent segments. Periodogram is calculated for all segments of the data and plotted in three-dimensional space with time, frequency, and periodogram as the x , y , and z axis, respectively. For stationary signal, the periodogram will appear to be roughly similar for the entire data set.

As for the quantitative test, we split the data into two independent segments and perform the following chi-square (χ^2) test [14]:

$$\chi^2 = \left[\frac{1}{n_{d1}} + \frac{1}{n_{d2}} \right]^{-1} \sum_{i=1}^{n/2} \left[\log \frac{\hat{P}_1(f_i)}{\hat{P}_2(f_i)} \right]^2 \leq \chi_{\frac{n}{2}, \alpha}^2 \quad (5)$$

In this case, $\hat{P}_1(f_i)$ stands for the average auto-PSD of the k -th segment, ($k = 1, 2$), n_{d1} and n_{d2} are the numbers of the independent sub-segments used in calculating the average auto-PSD, n number of data in each sub-segment, and $\chi_{\frac{n}{2}, \alpha}^2$ is the $(1-\alpha) \times 100\%$ quantile of the chi-square distribution for $n/2$ degrees of freedom (dof). Based on the above test, the data will be considered stationary if the auto-PSDs of the two segments are statistically equivalent.

B. Modeling Stage

Next comes the modeling. In this stage, we determine the model order and estimate the polynomial coefficients of the ARMA model. First, we need to come up with a guess of the order of the AR polynomial. For this purpose, the following averaged periodogram is calculated:

$$\hat{P}_{avper}(f) = \frac{1}{K} \sum_{m=0}^{K-1} \hat{P}_{per}^m(f) \quad (6)$$

in which $\hat{P}_{avper}(f)$ is the averaged periodogram (time segment average), $\hat{P}_{per}^m(f)$ the periodogram of the m -th segment, and K is the number of independent segments (no overlap) used in the calculation.

The averaged periodogram will be the estimator of the following theoretical auto-PSD [13]:

$$P_{ARMA}(f) = \left| \frac{C(e^{j2\pi f})}{A(e^{j2\pi f})} \right|^2 \sigma_w^2 \quad (7)$$

where $P_{ARMA}(f)$ denotes the auto-PSD of the output signal of an ARMA system, $C(e^{j2\pi f})$ and $A(e^{j2\pi f})$ are the MA and AR polynomials, respectively, evaluated at $e^{j2\pi f}$, f stands for frequency [Hz], and j the complex operator.

From equation (7), it may be observed that a pair of complex conjugate poles of the transfer function will produce a peak in the auto-PSD. Since the poles are also the roots of the AR-polynomial, the easiest way to estimate the order of the AR-polynomial is by simply counting the number of peaks in the auto-PSD curve or its estimator. Therefore, upon calculating the averaged periodogram according to (6) and plot it against frequency, we may count the number of peaks and this number will be the order of AR-polynomial, n_a .

Once the AR-order is determined, the next step is the AR parameter estimation. The coefficients of the polynomial and the variance of the residual are calculated using standard code for parameter estimation [15]. The most suitable ARMA(n_a, n_c) may be obtained through step-by-step increment of the MA-order. As an indicator, the following Bayesian Information Criterion (BIC) is used [16]:

$$BIC(k) = \ln \hat{\sigma}_e^2 + \frac{k \ln N}{2N} \quad (7)$$

In the above expression, N is the number of data, k total number of parameters = ($n_a + n_c$), and $\hat{\sigma}_e^2$ is the variance of the residual. The best ARMA(n_a, n_c) should produce minimum BIC value. In the case that the BIC value saturates, the parameter estimation iteration should be terminated and we pick the smallest ARMA(n_a, n_c) model, in accordance to the parsimony principle [13, 20]. Once the model is obtained, we need to check the poles and zeros of the transfer function to ensure that they lie strictly within the unit circle in the complex plane.

To slightly elaborate the parameter estimation, the standard code implements the one-step-ahead prediction error method [12]. Mathematically, the one-step-ahead prediction may be expressed as:

$$\hat{y}[t|t-1] = y[t] - \hat{e}[t|t-1] \quad (8)$$

with $\hat{y}[t|t-1]$ being the one-step-ahead prediction of $y[t]$. The one-step-ahead prediction error or the residual $\hat{e}[t|t-1]$ may be obtained from the following Infinite Impulse Response (IIR) filter relation:

$$\hat{e}[t|t-1] = \frac{A(q^{-1})}{C(q^{-1})} y[t] \quad (9)$$

The estimated parameters consisting of the AR and MA-polynomials and the variance of the residual are obtained as a result of minimizing a second order quadratic function of the prediction error. The above prediction error method will be equivalent with the maximum likelihood estimation method if

and only if the residual sequence is normally distributed. This condition, known as statistical efficiency requirement, will guarantee that the estimator possess minimum covariance matrix (Cramer Rao bound [12]).

C. Post-modeling Stage

In the last stage, we need to verify the ARMA model. The validation test consists of whiteness test of the residual and one-step-ahead prediction test.

In the first validation test, in order to see the whiteness of the residual, one needs to calculate and plot the following normalized autocorrelation sequence [11, 17]:

$$r_{ee}[k] = \frac{1}{N-k+1} \sum_{t=0}^{N-k} e[t]e[t+k] \quad (10)$$

$$r_{ee}^N[k] = \frac{r_{ee}[k]}{r_{ee}[0]} \quad (11)$$

In equations (10) and (11), $r_{ee}[k]$ is the autocorrelation sequence, $r_{ee}^N[k]$ the normalized autocorrelation sequence, k time lag, $e[t]$ measured signal, and N the number of data. Any signal $e[t]$ may be considered white noise if the normalized autocorrelation sequence for lag $k \geq 1$ lies within the 95% confidence interval which is equal to $\frac{1.96}{\sqrt{N}}$ [17]. Consequently, after obtaining the model and calculating the one-step-ahead prediction error sequence, $\hat{e}[t|t-1]$, according to (9), we need to compute the normalized autocorrelation sequence, $r_{ee}^N[k]$, plot it against the lag k , and confirm that the values are all within the 95% confidence interval for lag $k \geq 1$.

The second validity check is through the evaluation of the one-step-ahead prediction. In this case, one need to use data set that has not been used in the parameter estimation process. The one-step-ahead prediction of the signal, $\hat{y}[t|t-1]$, may be calculated as [10]:

$$\hat{y}[t|t-1] = \left[1 - \frac{A(q^{-1})}{C(q^{-1})} \right] y[t] \quad (14)$$

where $A(q^{-1})$ is the AR-polynomial, $C(q^{-1})$ the MA-polynomial, and $y[t]$ measurement data. If by visual inspection it is observed that $\hat{y}[t|t-1]$ is 'close' to $y[t]$ or it is within the 95% confidence interval from $y[t]$ then we may say that the estimated ARMA model is satisfactory. The 95% confidence interval for $y[t]$ are defined by $\pm 1.96 \hat{\sigma}_e$, in which $\hat{\sigma}_e$ is the standard deviation of the residual [17].

III. PARAMETRIC DISTANCE METHOD

From practical point of view, fault diagnosis involves pattern recognition since we must detect and identify the common pattern of the measured signal, and finally recognize the pattern in order to classify the fault [18]. A viable technique is by measuring the 'distance' between an unidentified time-series and some previously known classified time-series. In this paper, we propose a simple parametric distance defined as follows.

Suppose that the result of the ARMA modeling of the known fault is given by $\theta(m)$ and the parameters of the

unidentified time series is denoted by $\theta(o)$. The $N\theta \times 1$ column vectors of the estimated parameters of ARMA(n_a, n_c), which are called feature vectors, may be expressed as:

$$\theta^{(m)} = \begin{bmatrix} a_1^{(m)} \\ a_2^{(m)} \\ \vdots \\ a_{n_a}^{(m)} \\ c_1^{(m)} \\ c_2^{(m)} \\ \vdots \\ c_{n_c}^{(m)} \\ \hat{\sigma}_W^2(m) \end{bmatrix} \text{ and } \theta^{(o)} = \begin{bmatrix} a_1^{(o)} \\ a_2^{(o)} \\ \vdots \\ a_{n_a}^{(o)} \\ c_1^{(o)} \\ c_2^{(o)} \\ \vdots \\ c_{n_c}^{(o)} \\ \hat{\sigma}_W^2(o) \end{bmatrix} \quad (15)$$

The above vectors which form the features of the time signal are basically concatenation of the ARMA parameters.

The proposed parametric distance is defined as:

$$d(\theta^{(o)}, \theta^{(m)}) = \|\theta^{(m)} - \theta^{(o)}\|_2 \quad (16)$$

In the above definition, $\theta^{(o)}$ stands for the feature vector of the new or unidentified time-series, $\theta^{(m)}$ the reference feature vector of previously known fault, and $\|\cdot\|_2$ is the ℓ_2 matrix norm or Euclidean distance. Since the parameter vectors are defined in a hyperspace of dimension N_θ , which is the number of the estimated parameters of the ARMA model, by virtue of the matrix norm properties [19], the above parametric distance clearly satisfies the metric distance requirements stated as [18]:

$$\left. \begin{aligned} d(a, b) &= d(b, a) \\ d(a, b) &\leq d(a, c) + d(b, c) \\ d(a, b) &\geq 0 \\ d(a, b) &= 0 \text{ if } a = b \end{aligned} \right\} \quad (17)$$

Another issue that needs to be addressed is regarding the reliability of the method. From asymptotic properties of the estimated parameters, it is known that θ will converge to the true value θ^* with probability 1 as the number of data, N , goes to infinity [12, 20]. Based on this fact, it is theoretically guaranteed that $d(\theta^{(o)}, \theta^{(m)})$ will approach 0 as the unidentified time-series comes from the same faulty condition as the previously known fault.

IV. LABORATORY EXPERIMENT

In order to show the effectiveness of the method, a laboratory experiment was performed using the available vibration test rig. The rig consisted of a continuous beam with pin-pin supports and an unbalanced-mass exciter as depicted in Fig. 1. The rotating disk is made of aluminum alloy, the diameter is 165 mm, thickness 5 mm, and it weighs 288.5 g. To create unbalance condition, additional mass of 10.44 g is added at a radius equals to 65 mm, giving a combination of

mass times eccentricity, me , equals to 67.9 g-cm.



Fig. 1. Experimental set-up, which shows the simple vibration test rig, the Kyowa accelerometer mounted on the continuous beam, and other related instrumentations.

The pin support at the left end is fixed, while the one at the right end may be alternately replaced using a tight roller support or a slack support to simulate mechanical-looseness. The measurement was conducted using Kyowa AS-2C CD-4659 piezoresistive accelerometer. The accelerometer was cross-calibrated using Lutron VB-8202 Vibration Meter. Data acquisition was carried out by NI USB-6009 external card, and the data acquisition was governed by LabVIEW SignalExpress which was set at 1 kHz sampling frequency.

As for the experiment, two cases were simulated, namely the pure unbalance case, and the combination of unbalance and mechanical looseness. For the first case, the beam was supported tightly using pin-type supports at both ends, while for the second case, the housing of the right roller was replaced, resulting in a loose support. In all cases, the rotational velocity of the disk is fixed at 1200 rpm.

Typical measurement signals in the time domain are depicted in Fig. 2 and 3, for the unbalance and the combination of unbalance and mechanical looseness, respectively

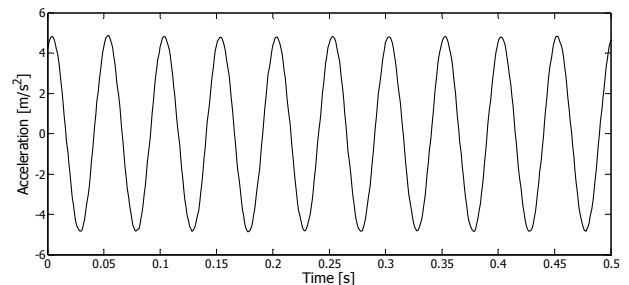


Fig. 2. Acceleration signal in the time domain for the pure unbalance case. Data is captured for 5 s at 1 kHz, but only fraction of the data is shown in the graph.

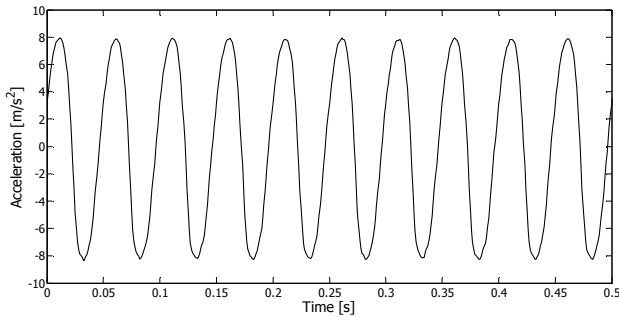


Fig. 3. Acceleration signal in the time domain for the combined unbalance and mechanical looseness case. Data is captured for 5 s at 1 kHz, but only fraction of the data is shown in the graph.

Typical frequency profiles of the simulated cases are shown in Fig. 4 and 5, expressed in terms of the averaged periodogram of the measured signals which was calculated using (6). Fig. 4 is for the unbalanced case, and Fig. 5 is for the combination of unbalanced and mechanical looseness case. In the calculation of the averaged periodogram, 30 independent data sets for both the unbalanced and combined unbalanced and mechanical looseness are used. The length of each independent data set is equal to 5000.

For brevity, in the rest of the paper, the unbalance case will be denoted as UB, and the unbalance and looseness case will be denoted as UBL.

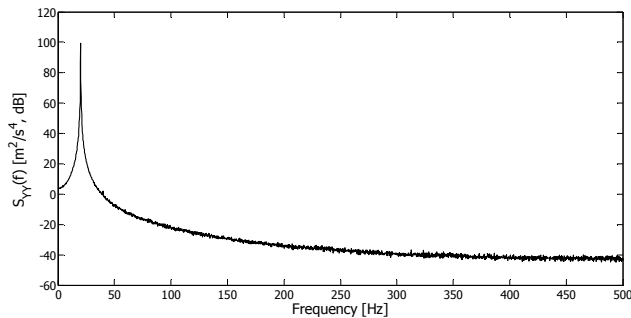


Fig. 4. Averaged periodogram of the acceleration signal for the pure unbalance case. Thirty independent data sets are used, and the calculation follows (6).

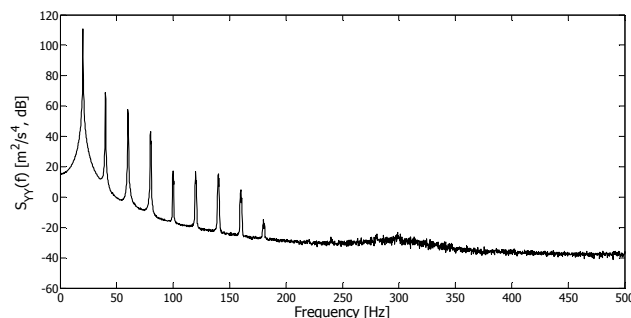


Fig. 5. Averaged periodogram of the acceleration signal for the combined unbalance and mechanical looseness case. Thirty independent data sets are used, and the calculation follows (6).

V. LEARNING STAGE VIA TIME-SERIES MODELING

As has been outlined in Section 3, to implement parametric distance, reference feature vectors are needed. To obtain the above reference, in the learning stage, the experiment was repeated 30 (thirty) times to collect independent data sets for both the unbalance and combined unbalanced and looseness cases. Reference feature vectors will be constructed from the average time-series coefficients of the 30 data sets.

Following the procedure outlined in Section 2.1., the data need to be checked for stationarity. This part of the work, which is considered elementary, has been performed beforehand to avoid loss of focus. Shortly, it may be mentioned that all data sets satisfies both the qualitative as well as quantitative stationary tests [21]. Owing to this result, we may proceed with the time-series modeling of the data. Each independent set consists of 5000 data as the result of sampling at 1 kHz for 5 s. We split the 5000 data into two segments. The first 4000 data is used in the modeling stage, while the remaining 1000 data is used for prediction purposes. The calculation of the ARMA coefficients are performed with MATLAB® [15].

Following the procedure outlined in Section 2.1., the data need to be checked for stationarity. This part of the work, which is considered elementary, has been performed beforehand to avoid loss of focus. Shortly, it may be mentioned that all data sets satisfies both the qualitative as well as quantitative stationary tests [21]. Owing to this result, we may proceed with the time-series modeling of the data. Each independent set consists of 5000 data as the result of sampling at 1 kHz for 5 s. We split the 5000 data into two segments. The first 4000 data is used in the modeling stage, while the remaining 1000 data is used for prediction purposes. The calculation of the ARMA coefficients are performed with MATLAB® [15].

As the next step, in line with Section II.B, order determination using BIC is performed. From Fig. 2 – 5, and by the nature of the excitation, we know from literature [11, 13] that for harmonic signals, the structure of the ARMA model should be of the form of ARMA(2m,2n), since every frequency or every peak in the periodogram requires a pair of complex conjugate poles. Therefore, we should attempt to find the best or optimum AR order by fitting AR(2m) starting at $m = 1$, calculate the BIC's, and plot them. Once we found the optimum AR-order m_o , then we may embark in the MA direction by fitting ARMA(2m_o,2n) starting at $n = 0$. The BIC plots for UB along the AR and MA directions are shown in Fig. 6 and 7, respectively. From Fig. 6, the BIC values saturate starting at AR(4). Therefore, the optimum AR order is equal to four. Next, from Fig. 7, after fitting various ARMA(4,2n) models to UB data, the lowest BIC value is obtained for $n = 0$. Based on the BIC, we may conclude that the best time-series model for UB data is ARMA(4,0) or AR(4).

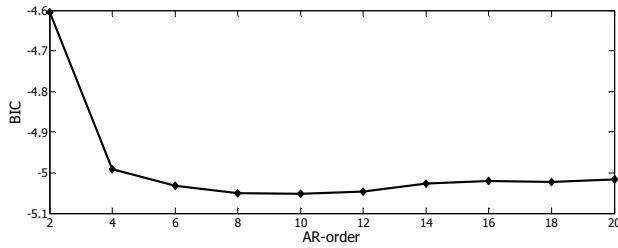


Fig. 6. Typical BIC-plot for UB along the AR direction. The AR order is varied starting from 2 up to 20. In this case, only even AR order is chosen to accommodate the harmonic feature of the periodogram [11, 13]. From the figure, it may be observed that the best AR order is equal to four.

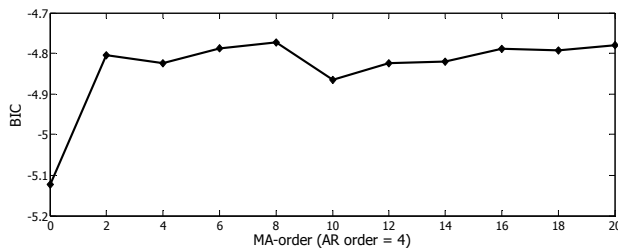


Fig. 7. Typical BIC-plot for UB along the MA direction. The ARMA(4,n) order is varied starting from ARMA(4,0) up to ARMA(4,20). Based on the BIC-plot, it may be observed that the best model for UB is ARMA(4,0) or AR(4).

Likewise, we perform the same procedure to UBL data sets. The BIC-plots are shown in Fig. 8 and 9 for order determination in the AR and MA directions, respectively. From Fig. 8, the BIC values saturate starting at AR(12). Therefore, the optimum AR order should be equal to 12. In the next step, following the previous process, after fitting various ARMA(12,2n) models to UBL data, since the minimum BIC value in Fig. 9 settles starting ARMA(12,0), it is clear that the best time-series model for UBL data is ARMA(12,0) or AR(12).

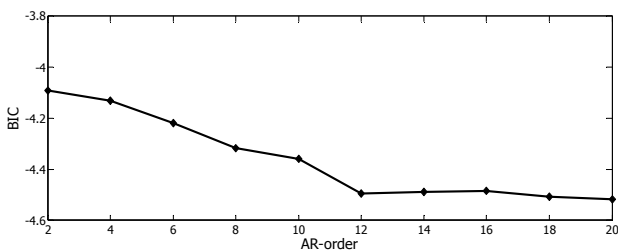


Fig. 8. Typical BIC-plot for UBL along the AR direction. The AR order is varied starting from 2 up to 20. In this case, only even AR order is chosen to accommodate the harmonic feature of the periodogram [11, 13]. From the figure, it may be observed that the best AR order is equal to 12.

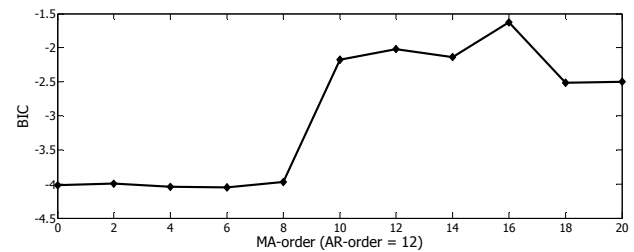


Fig. 9. Typical BIC-plot for UB along the MA direction. The ARMA(4,n) order is varied starting from ARMA(4,0) up to ARMA(4,20). The lowest BIC values settles starting at ARMA(12,0). Therefore, the best model for UBL is AR(12).

To validate the time-series model and ensure that it fits the data, we should proceed with the calculation of the one-step-ahead prediction error sequence according to (9) and plot the normalized autocorrelation sequence. Typical result for UB, given that we use the AR(4) model obtained from the modeling, is depicted in Fig. 10. Since all values of the normalized autocorrelation $r_{ee}^N[k]$ for lag $k \geq 1$ fall within the 95% confidence interval, the model passed the whiteness test, which is the first validation test.

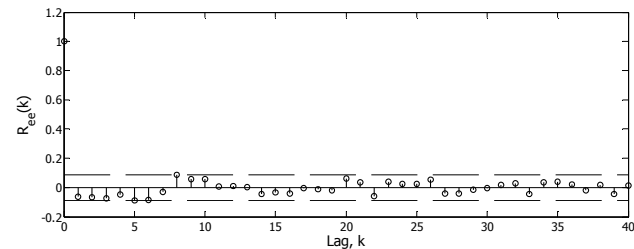


Fig. 10. Typical normalized autocorrelation sequence of the residual as the result of fitting AR(4) model to the UB data sets. The 95% confidence interval, which are $\pm 1.96/\sqrt{N}$, N : length of the data used in calculating the autocorrelation sequence, are drawn in dashed lines.

As a second validation test, we compute the one-step-ahead prediction using (8) and compare it with the segment of the data that is not used in the modeling stage. The typical comparison of the prediction and the original data for UB is depicted in Fig. 11. In this regard, the original series is presented in solid line, the one-step-ahead prediction sequence in '+', and the 95% confidence interval band in dashed lines. The 95% confidence interval is calculated as $\pm 1.96 \times \sigma_{\hat{e}}$ deviations from the predicted value, in which $\sigma_{\hat{e}}$ stands for the standard deviation of the one-step-ahead prediction error sequence $\hat{e}[t|t-1]$. Based on Fig. 11, we may observe that the prediction tightly tracks the original series with very small 95% confidence interval. Based on the two validation tests, we conclude that AR(4) is indeed the best model for UB data sets.

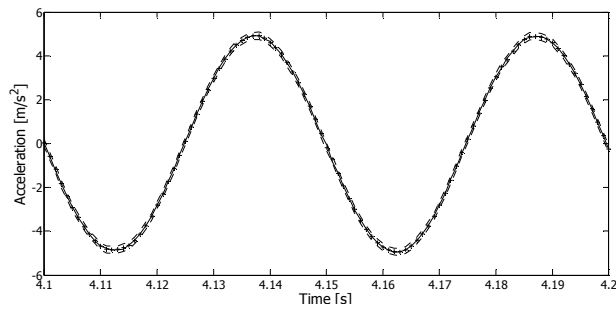


Fig. 11. One-step-ahead prediction of the UB data set using AR(4) model obtained previously. The original series is shown in solid line, '+' denotes prediction, and the 95% confidence interval drawn in dashed lines.

Similarly, the same validation plots are computed for UBL data sets, and typical results of fitting an AR(12) are depicted in Fig. 12 and 13. Based on the above plots, the same conclusion may be drawn, i.e., AR(12) is the best model for UBL data sets.

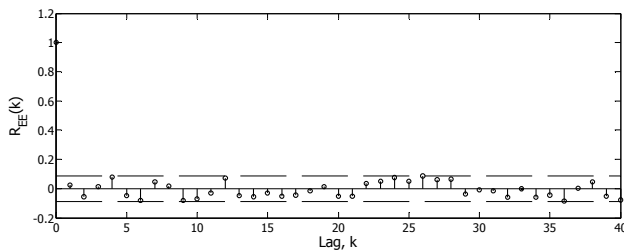


Fig. 12. Typical normalized autocorrelation sequence of the residual as the result of fitting AR(12) model to the UBL data sets. The 95% confidence interval, which are $\pm 1.96/\sqrt{N}$, N : length of the data used in calculating the autocorrelation sequence, are drawn in dashed lines.

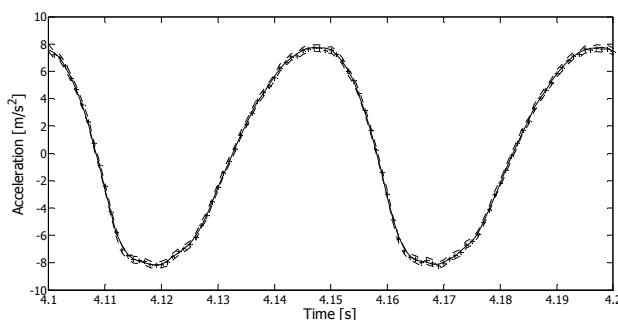


Fig. 13. One-step-ahead prediction of the UBL data set using AR(12) model obtained previously. The original series is shown in solid line, '+' denotes prediction, and the 95% confidence interval drawn in dashed lines.

Since the best models for UB and UBL have been obtained, i.e., AR(4) and AR(12), respectively, we perform modeling to the entire 30 data sets available in the learning stage to form the reference feature vectors. These vectors are basically the average value of the parameter vectors of the time-series model. The reference feature vector for UB and UBL cases are shown in Table I. and Table II., respectively.

TABLE I.
ENTRIES OF THE REFERENCE FEATURE VECTOR FOR THE UB CASE

Parameter	Value
$a_1^{(m)}$	-1.113
$a_2^{(m)}$	-0.311
$a_3^{(m)}$	0.044
$a_4^{(m)}$	0.416
$\hat{\sigma}_w^2(m)$	0.005

TABLE II.
ENTRIES OF THE REFERENCE FEATURE VECTOR FOR THE UBL CASE

Parameter	Value
$a_1^{(m)}$	-2.133
$a_2^{(m)}$	1.253
$a_3^{(m)}$	-0.231
$a_4^{(m)}$	0.568
$a_5^{(m)}$	-0.330
$a_6^{(m)}$	-0.527
$a_7^{(m)}$	0.308
$a_8^{(m)}$	0.380
$a_9^{(m)}$	-0.109
$a_{10}^{(m)}$	-0.251
$a_{11}^{(m)}$	-0.087
$a_{12}^{(m)}$	0.186
$\hat{\sigma}_w^2(m)$	0.033

Regarding the time-series modeling results, an important remark needs to be addressed for the UBL case. Strictly speaking, the loose support makes the system non-linear due to the nature of the stiffness variation. However, the ARMA model that is employed in the modeling will only accommodate linear, stable, and invertible systems. As a result, the ARMA model for UBL will 'linearize' the system, and thus will be suboptimum. For the purpose of fault identification, this suboptimum model will not pose any problem, since the reference feature vector should bear the specific characteristics of the UBL case.

VI. THE IMPLEMENTATION OF THE PARAMETRIC DISTANCE IN FAULT IDENTIFICATION

Having the reference feature vectors for the UB and UBL case, we are ready to test the effectiveness and reliability of the method. As a test case, we generate a separate data consisting of 30 independent data sets for UB and also UBL. More over, we assign an operator to perform the modeling and obtain the feature vectors of the test data. During modeling, the identity of the model is not revealed to the operator [21]. Since the identity is not revealed, for discussion purposes, we label the unidentified data sets as set A and B. Behind the scene, we hold the key that set A is actually UB and set B is in fact UBL. To keep the unbiasedness of the operator, we did

not allow any pre-modeling such as plotting of the time and frequency domain features of the data. The complete procedure for the operator is outlined as follows:

1. Get individual data from set A and B
2. Model each data using AR(4), which is the best model for UB, and AR(12), which is the valid model for UBL
3. Concatenate the AR parameters and the variance of the residual sequence to form feature vectors $\theta^{(o)}$ as in (15)
4. Calculate the parametric, i.e., Euclidean distances, according to (16) for each feature vectors $\theta^{(o)}$ relative to UB and UBL reference feature vectors $\theta^{(m)}$
5. Draw box plots [22] of the parametric distances of set A and B relative to the reference feature vectors of UB and UBL. Thus, there will be two sets of box plots, one set consisting of two box plots depicts the distribution of the parametric distances of set A and B relative to UB reference feature vector, and another set consisting of also two box plots that shows the distribution of the parametric distances of set A and B relative to UBL reference feature vector.
6. Calculate the average feature vector of set A and B, according to the postulated model, i.e., AR(4) for UB and AR(12) for UBL
7. Perform Hotelling's T^2 hypothesis testing [24] to check whether the average feature vector of set A and B are the same or different than the reference feature vectors for UB and UBL. This final step will substantiate the qualitative test of eyeballing the box plots.

The box plots, showing the distribution of parametric distances of set A and B relative to UB reference feature vector, are depicted in Fig. 14. Consequently, we force AR(4) models to fit all data. From the figure, it is clear that set A is essentially taken from UB set up since the values of the parametric distances relative to UB reference feature vector are smaller than set B.

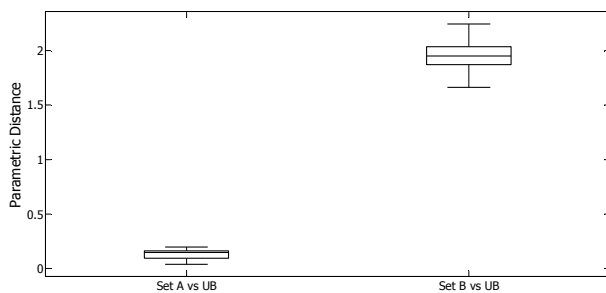


Fig. 14. Boxplots showing the distribution of parametric distances of feature vectors of set A and B upon fitting AR(4) models, relative to UB reference feature vector. Box plot on the left is for set A, and box plot on the right is for set B.

For the hypothesis testing, we formulate the following hypotheses:

- H_0 : The mean of the feature vectors = UB reference feature vector (null hypothesis)
- H_1 : The mean of the feature vectors \neq UB reference feature vector (alternate hypothesis)

To compare two sample mean vectors, we implement Hotelling's T^2 test, with the following test statistic [23]:

$$F = \frac{(N_1 + N_2 - N_\theta - 1)T^2}{\{(N_1 + N_2 - 2)N_\theta\}} \sim F(N_\theta, N_1 + N_2 - N_\theta - 1) \quad (18)$$

in which,

$$T^2 = N_1 N_2 (\theta^{(m)} - \bar{\theta}^{(o)})^T C^{-1} (\theta^{(m)} - \bar{\theta}^{(o)}) / (N_1 + N_2) \quad (19)$$

and,

$$C = \{(N_1 - 1)C_1 + (N_2 - 1)C_2\} / (N_1 + N_2 - 2) \quad (20)$$

In (18) – (20), F is the $(1 - \alpha)\%$ quantile of the F distribution with N_θ and $(N_1 + N_2 - N_\theta - 1)$ degree of freedoms, $N_1 = 30$ stands for the number of data sets used in forming the reference feature vector, $N_2 = 30$ is the number of data in set A or B, N_θ the length of the reference feature vectors (which is five for UB and 13 for UBL), $\theta^{(m)}$ the reference feature vector for UB, $\bar{\theta}^{(o)}$ the average feature vector of set A or B, C_1 the covariance matrix of the feature vector set in the UB case, and C_2 the covariance matrix of the feature vectors in set A or B, depending which data set is being tested.

Upon performing the calculation, it is found that the value of the test statistic for set A equals to $F_A^{UB} = 2.301$, whereas the test statistic for set B equals to $F_B^{UB} = 45.546$, as compared to the threshold $F_{0.05(5,54)} = 2.386$. Therefore, based on the hypothesis testing, we accept set A as data from UB and reject set B as data from UB at level $\alpha = 5\%$. Viewing at the conclusion of the box plot analysis and hypothesis testing, we conclude that the method correctly identifies set A as being UB and set B as being something other than UB.

After completing the analysis with UB as the reference feature vector, we continue with the box plot analysis and hypothesis testing of set A and B relative to UBL reference feature vector. The box plots, showing the distribution of parametric distances of set A and B relative to UBL reference feature vector, are depicted in Fig. 15. Accordingly, we fit AR(12) models to all data. From the figure, it is clear that set B basically comes from UBL case since the values of the parametric distances relative to UBL reference feature vector are smaller than set A.

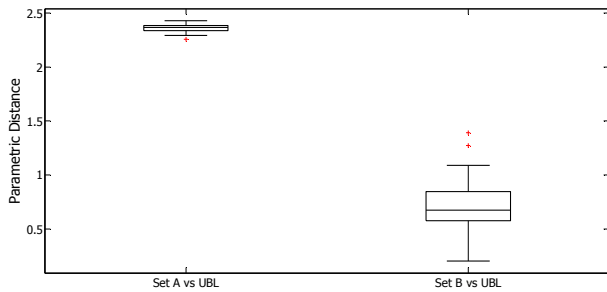


Fig. 15. Boxplots showing the distribution of parametric distances of feature vectors of set A and B upon fitting AR(12) models, relative to UBL reference feature vector. Box plot on the left is for set A, and box plot on the right is for set B.

As for the hypothesis testing, after the calculation, it is found that the value of the test statistic for set A equals to $F_A^{UBL} = 534.02$, whereas the test statistic for set B equals to $F_B^{UBL} = 1.78$, as compared to the threshold $F_{0.05(13,46)} = 1.94$. Therefore, based on the hypothesis testing, we reject set A as data from UBL and accept set B as data from UBL at level $\alpha = 5\%$.

Looking at the conclusion of the box plot analysis and hypothesis testing, we conclude that again the method correctly identifies set A as being non-UBL and set B as data coming from the UBL case.

VII. CONCLUSION

In this paper, we have presented a comprehensive technique in fault diagnosis using parametric distance. Unlike the previously known Kullback distance, the proposed parametric distance satisfies the metric conditions. Laboratory experiment using a simple vibration test rig has been performed to evaluate the effectiveness of the method. Experimental results confirms that the method works extremely well in the sense that it correctly identifies the simulated faults, namely the pure unbalance and the combined unbalance and mechanical looseness. Therefore, the method is deemed satisfactory to be used as a complement to the traditional frequency domain method.

Lastly, it is also necessary to mention that due to the parsimony principle, although not shown in this paper, the predicted autopower spectrum of the signal does not exactly match the periodogram, especially in the UBL case. The limited order of the ARMA(n_a, n_c) will not allow us to match the numerous peaks in the periodogram, except for the most significant frequency components. This minor flaw gives way to the improvement of the method, which will be taken care of in our future work.

ACKNOWLEDGMENT

The authors gratefully acknowledge partial support from Institut Teknologi Bandung through the Institutional-Fortification Research Grant. In addition, the authors would like to express their gratitude to Mrs. Dianvivianthi who

participated in the early stage of this research work, and to Mr. Freddy Wijaya and Mr. Gabriel Pramudita Saputra for their assist in setting up the test-rig, conducting the experiment, and perform the initial data processing.

REFERENCES

- [1] A.K.S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics condition-based maintenance", *Mechanical Systems and Signal Processing* 20 (2006) 1483-1510.
- [2] J.T. Broch, *Mechanical Vibration and Shock Measurements*, Bruel and Kjaer, Naerum, Denmark, 1984.
- [3] C. Scheffer, and P. Girdhar, *Practical Machinery Vibration Analysis and Predictive Maintenance*, Newnes, Oxford, 2004
- [4] S. Braun, *Mechanical Signature Analysis – Theory and Application*, Academic Press, London, 1986.
- [5] W. Gersch, T. Brotherton, and S. Braun, "Nearest Neighbor – Time Series Analysis Classification of Faults in Rotating Machinery", *Trans. ASME, J. of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 105, April 1983, pp. 178 – 184.
- [6] Dianvivianthi and I.P. Nurprasetio, "The Implementation of Time-Series Analysis in Fault Detection of Rotating Machineries", *Proceedings, Experimental and Theoretical Mechanics*, Bandung, June 20-21, 2000, pp. IV-1.1 – IV-1.6 (in Indonesian)
- [7] D. Hong, G. Xiuwen, and Y. Shuzi, "An Approach to State Recognition and Knowledge-Based Diagnosis for Engines", *Mechanical Systems and Signal Processing*, Vol. 5, No. 4, 1991, pp. 257 – 266
- [8] C.K. Mechefske and J.Mathew, "Fault Detection and Diagnosis in Low Speed Rolling Element Bearings – Part II: The Use of Nearest Neighbor Classification", *Mechanical Systems and Signal Processing*, Vol. 6, No. 4, 1992, pp. 309-316
- [9] W.J. Staszewski, K. Worden, and G.R. Tomlinson, "Time-Frequency Analysis in Gearbox Fault Detection using the Wigner-Ville Distribution and Pattern Recognition", *Mechanical Systems and Signal Processing*, Vol. 11, No. 5, pp. 673 – 692, 1997
- [10] M-Ch Pan, P. Sas, and H. Van Brussel, "Machine Condition Monitoring using Signal Classification Techniques", *Journal of Vibration and Control*, Vol. 9, 2003, pp. 1103 – 1120
- [11] S.M. Pandit and S.M. Wu, *Time Series and System Analysis with Applications*, Wiley, New York, 1983.
- [12] L. Ljung, *System Identification – Theory for The User*, Prentice Hall, Englewood Cliffs, N.J., 1987.
- [13] S.M. Kay, *Modern Spectral Estimation – Theory and Application*, Prentice Hall, Englewood Cliffs, N.J., 1988.
- [14] J.S. Bendat and A.G. Piersol, *Random Data – Analysis and Measurement Procedure*, 2nd ed., Wiley, New York, 1986.
- [15] L. Ljung, *System Identification Toolbox™ User's Guide – R2011b*, MathWorks, 2011
- [16] H. Akaike, "On Entropy Maximization Principle", in *Application of Statistics* (P.R. Krishnaiah, ed.), North Holland, Amsterdam, 1977
- [17] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Method*, Springer-Verlag, N.Y., 1987.
- [18] G.S. Sebestyen, *Decision-Making Processes in Pattern Recognition*, The Maxmillan Company, New York, 1962.
- [19] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 2nd ed., The Johns Hopkins University Press, 1989
- [20] T. Soderstrom and P. Stoica, *System Identification*, Prentice Hall, 1989
- [21] G.P. Saputra, *Robustness Evaluation of Combined Time-Series Modeling, Parametric Minimum-Distance Classification, and Hypothesis Testing in Fault Identification of Rotating Machineries*, Undergraduate Thesis, Mechanical Engineering Department, Faculty of Mechanical and Aerospace Engineering, Institut Teknologi Bandung, 2011 (in Indonesian).
- [22] E. Kreyszig, *Advanced Engineering Mathematics*, 9th ed., Wiley, 2006, Ch. 24, Sect. 24.1
- [23] B.F.J. Manly, *Multivariate Statistical Methods – A Primer*, Chapman and Hall, 1986