

Impeller Fault Detection for a Centrifugal Pump Using Principal Component Analysis of Time Domain Vibration Features

Berli Kamiel^{1,2}, Gareth Forbes², Rodney Entwistle², Ilyas Mazhar² and Ian Howard²

¹ Department of Mechanical Engineering,
Universitas Muhammadiyah Yogyakarta, Indonesia

² Department of Mechanical Engineering,
Curtin University, Bentley, Western Australia
E-mail: b.kamiel@postgrad.curtin.edu.au

Abstract

In this paper, multi sensor data collection and Principal Component Analysis (PCA) are proposed to develop a framework for impeller fault detection. Vibration signatures of normal and faulty impellers were collected from a Spectra Quest Machinery Fault Simulator which was set up in the centrifugal pump mode. The impeller fault was introduced by cutting the blade at two locations for each blade, and four accelerometers were mounted on the pump on the volute case (channel 1), suction nozzle (channel 2), bearing (channel 3) and discharge nozzle (channel 4) respectively. A 1 HP electric motor was connected to the pump by a belt-pulley mechanism with 1:1 ratio and run at 1800 rpm. Vibration data was collected at sampling rate 8192 Hz with sampling duration of 10 seconds. Four statistical features (kurtosis, RMS, skewness, and variance) were extracted from time data which was previously separated into frequency bands or octaves. PCA fault monitoring method using Hotelling's T^2 statistic was developed to detect and diagnose the impeller fault. Thirty six features (9 octave bands x 4 statistical features) were used, and eighteen features were selected and retained in the model which was found to give 95% of the data variation, indicating that half of the total thirty six features were sufficient to detect the impeller fault. This study shows consistent results for four channels for detection of the faulty impeller which was indicated by T^2 values that exceed the control limit. It also showed that channel 1, with the accelerometer mounted on the volute casing, gave the most sensitive result for fault detection.

1 Introduction

A Centrifugal pump is one of the rotating machines that are widely used in various industries such as petrochemical, water treatment, power generation, agriculture, fertilizers, oil and gas, etc. During its operation, it can experience failures which can potentially cause disruption of production processes. Early detection of faults in centrifugal pumps can reduce energy consumption, service and maintenance cost, increase reliability, life-cycle and safety, therefore, significantly reduce through-life-time costs [1]. Various techniques for fault detection in a centrifugal pump have been developed by many researchers [2]. Application of statistical features combined with principal component analysis (PCA) is one of the techniques that can be used to monitor machine condition [3]. PCA is a widely used procedure which reduces the dimension of an original dataset into lower dimension by converting original possibly correlated variables into new sets of uncorrelated variables. Hotelling [4] was the first investigator who proposed using the PCA procedure to solve the problem of multivariate statistical data analysis.

PCA is recognised as a powerful tool of statistical process monitoring for fault detection and an extensive investigation based on PCA schemes have been made in various fields [5]. Sun et al [6] investigated improved PCA for boiler water and steam leak fault detection. Standard PCA based on the new test statistics has been proposed by Ding et al [7] and Ahmed et al [4] who applied the PCA technique to vibration data from reciprocating compressor for fault detection. Some researchers used nonlinear PCA instead of standard PCA, for instance, Nguyen and Golival [8] used kernel PCA to address the nonlinear

problem of fault detection in mechanical systems and Cui et. al [9] proposed an improved kernel PCA for fault detection.

In this study, statistical features were not extracted directly from the raw time data; instead they were extracted from time data which was previously separated into a set of frequency ranges or octave bands.

The use of octave bands to separate the vibration data into different frequency components for fault analysis is still rarely investigated when using PCA. In the octave band approach, the whole frequency range is divided into a set of frequency bands. The octave band is originally a technique used in sound and noise analysis, and only few applications in the vibration area have been developed. Kang and Kaizu [10] investigated the working efficiency and comfort of operation of tractors by measuring vibration acceleration. They used 1/3 octave band analysis to calculate the power spectrum and root mean square of vibration acceleration in each frequency band. Nishiyama et al [11] used four equally effective 1/3 octave bands in defining multiple effect of hand-arm vibration on the temporary threshold shift in vibratory sensation.

In the proposed method, raw vibration time data was separated into octave bands and then statistical features were extracted from each band. PCA was then used to reduce the dataset dimension. This paper is organized as follows. Section 2 discusses an overview of PCA for fault detection using Hotelling's \mathbf{T}^2 method. Section 3 describes the octave bands. Statistical time domain vibration features are discussed in section 4. Vibration data acquisition and detection result and discussion are presented in section 5 and section 6 respectively.

2 Principal Component Analysis

Fault detection based on vibration often deals with a large number of variables which creates difficulties in analysis. PCA provides a method for reducing dimension or data compression to facilitate efficient data analysis. The goal of PCA is to find which variables in the system are more important, which are redundant, and which are noises. This goal is basically achieved by projecting original correlated variables into a new set of uncorrelated variables, which are called, Principal Components (PCs).

Consider a $m \times n$ data matrix, \mathbf{X} , which contains, vibration data from n measured variables and m experimental trials as illustrated in equation (1).

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mj} & \dots & x_{mn} \end{pmatrix} \quad (1)$$

Each row vector (\mathbf{x}_i) represents measurements from all variables at the specific time instants; meanwhile each column vector (\mathbf{w}_j) represents the measurement of one variable over the whole experimental time.

Since the variables may have different units, then the original data has to be scaled before further analysis. One of the scaling techniques is auto scaling which re-scales the original data to have zero mean and unity variance. This is achieved by transforming column vector \mathbf{w}_j as follows,

$$\mu_{w_j} = \frac{1}{n} \sum_{i=1}^m x_{ij}, \quad (2)$$

$$\sigma_{w_j} = \sqrt{\frac{1}{n-1} \sum_{i=1}^m (x_{ij} - \mu_{w_j})^2}, \quad (3)$$

$$\bar{x}_{ij} = \frac{x_{ij} - \mu_{w_j}}{\sigma_{w_j}}, \quad (4)$$

where μ_{w_j} and σ_{w_j} are the mean and standard deviation of variable w_j respectively, whereas \bar{x}_{ij} is the re-scaled data. For convenience, in the rest of the paper, the re-scaled data is considered without bar notation. PCA decomposes the $m \times n$ matrix \mathbf{X} into a set of new directions \mathbf{p}_i as [5],

$$\begin{aligned}\mathbf{X} &= \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_n\mathbf{p}_n^T, \\ &= \mathbf{TP}^T,\end{aligned}\quad (5)$$

where \mathbf{p}_i is an $n \times 1$ column vector of eigenvectors of the covariance matrix \mathbf{X} which is defined as the principal component loading matrix. \mathbf{t}_i is a $m \times 1$ column vector, which is a projection of original data onto the new basis vector \mathbf{p}_i which is defined as the score matrix of principal components. For each eigenvector \mathbf{p}_i , there is a value of eigenvalue λ_i which is an indicator of the covariance of the matrix \mathbf{X} in the direction \mathbf{p}_i .

After applying PCA to the data set, one often finds that only the first few eigenvectors contribute to the systematic variation in the data (important information in the data) while the rest of the eigenvectors relates to the noises in the data. The PCA model is developed by retaining only the eigenvectors which have the most contribution to the variation in the data. For a reduced model of PCA where only the first k PC's are retained in the model, then the equation becomes,

$$\begin{aligned}\mathbf{X} &= \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_k\mathbf{p}_k^T + \mathbf{E}, \\ &= \mathbf{TP}^T + \mathbf{E},\end{aligned}\quad (6)$$

where \mathbf{E} is the residual error matrix.

Hotelling's \mathbf{T}^2 statistic (\mathbf{T}^2) can be used for fault detection modelling as it measures unusual variability within the normal subspace. For a new measurement of vector \mathbf{x}_i , \mathbf{T}^2 based on the first k PC's can be calculated from [5],

$$\mathbf{T}_i^2 = \mathbf{x}_i\mathbf{P}\boldsymbol{\lambda}^{-1}\mathbf{P}^T\mathbf{x}_i^T, \quad (7)$$

where $\boldsymbol{\lambda}$ is a $k \times k$ diagonal matrix composed of eigenvalues λ_i ($i=1,2,\dots,k$). The control limit for \mathbf{T}_α^2 at a confidence level $(1-\alpha)$ corresponds to the F-distribution as follows [4],

$$T_\alpha^2 = \frac{k(m-1)}{m-k} F(k, m-1; \alpha), \quad (8)$$

where m is the number of samples used to develop the PCA model, k is the number of PC's retained in the model, and $F(k, m-1; \alpha)$ is an F-distribution at the level of significance α with k and $(m-1)$ degrees of freedom. The calculation of \mathbf{T}^2 given by Equation (7) for a new data set is used for fault identification in the system. If the value of T_α^2 is within the threshold given by Equation (8), it indicates that statistical distribution in the new data is the same as the one in the data matrix \mathbf{X} . On the contrary, if the new data set has different statistical distribution with the one in the data matrix \mathbf{X} , it indicates a fault at that time.

The algorithm for fault detection based on PCA model and \mathbf{T}^2 is represented as follows [12],

- Scaling the data set baseline to have zero mean and unity variance,
- Calculating the covariance matrix,
- Calculating the eigenvalues and eigenvector,
- Determining the number of principal components to be retained in the model,
- Calculating the \mathbf{T}^2 control limit,
- Scaling the new data set,

- Projecting the new data set onto the retained principal components,
- Calculating the T^2 statistic for each measurement or sample set,
- Detecting the fault if the T^2 measure exceeds the control limit.

3 Octave Band

In the octave band approach, the whole frequency range is divided into a set of narrow frequency bands. A frequency band is said to be an octave width if the upper band frequency is twice the value of the lower band frequency. The central frequency of an octave band is defined as $\sqrt{2}$ times the lower cut-off frequency. Therefore,

$$f_l = \frac{f_c}{\sqrt{2}} \quad (9)$$

$$f_u = \sqrt{2}f_c \quad (10)$$

$$f_u = 2f_l \quad (11)$$

where f_l, f_c , and, f_u are the lower band, central and upper frequency bands respectively. For instance, if the central frequency of an octave band is 30 Hz, then it has a lower frequency of 21.2 Hz and an upper frequency of 42.2 Hz. In this study, the central frequency for the second octave band corresponds to the shaft frequency of the impeller whereas the central frequencies of the other octave bands correspond to its harmonics. This separation technique will give more detailed information across the frequency spectrum.

4 Statistical Time Domain Vibration Features

There are many features that can be extracted from vibration signals for fault detection. They have been investigated in previous studies by many researchers [13]. In this study, the features extracted from the octave band time domain vibration signals include, kurtosis, RMS, skewness, and variance. Those features can be represented mathematically as follows,

$$Kurtosis = \frac{\frac{1}{N} \sum_{i=1}^N y_i^4}{\sigma^4}, \quad (12)$$

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}, \quad (13)$$

$$Skewness = \frac{\frac{1}{N} \sum_{i=1}^N y_i^3}{\sigma^3}, \quad (14)$$

$$Variance = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N-1}, \quad (15)$$

where $y = y_1, y_2, \dots, y_N$ represents the octave band time domain signals and σ is the standard deviation.

The procedure for impeller fault detection as shown in Figure (1) commences by converting the raw vibration waveform to the frequency domain using the Fast Fourier Transform (FFT). The frequency domain is then divided into octave bands such that the central frequency of each octave band corresponds to the shaft frequency and harmonics. This step will produce n octave bands from the whole frequency range. The



Figure 1: Fault Detection Procedure

inverse FFT is used to convert the frequency domain data (in n octave bands) to the time domain. Furthermore, four statistical features are extracted from each of the n octave bands which yield $n \times 4$ set of features. This set of features is then used to develop the PCA model.

5 Vibration Data Acquisition

Vibration signatures of normal and faulty impellers were collected from a Spectra Quest Machinery Fault Simulator which was set up with a centrifugal pump. An impeller fault was introduced by cutting the impeller blade at two locations for each blade as shown in Figure 2. Four accelerometers were mounted on the pump respectively on the volute case (channel 1), suction nozzle (channel 2), bearing (channel 3) and, discharge nozzle (channel 4). A 1 HP electric motor was connected to the pump by a belt-pulley mechanism with 1:1 ratio and run at 1800 rpm. The centrifugal pump test rig set up can be seen in Figure 3. Acceleration vibration data was collected at a sampling rate of 8192 Hz with sampling duration of 10 seconds.



Figure 2: Faulty impeller

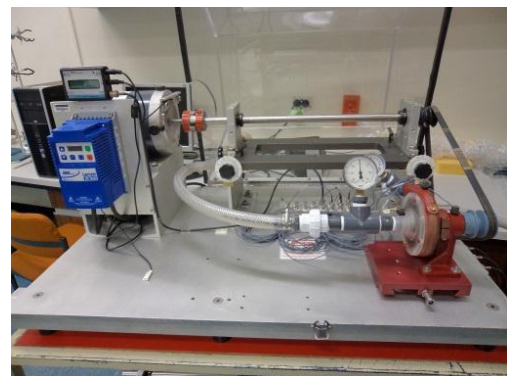


Figure 3: Centrifugal pump test rig

6 Detection Result and Discussion

Before being processed, the 10 second data segment was further divided into 1 second segments so that there were 10 intervals of data for each channel. That division was performed in order to accomplish more detailed observations on the vibration signal. Nine octave bands as shown in Table 1 were used to extract statistical features for normal and faulty impeller.

Lower cut-off Frequency (Hz)	Centre Frequency (Hz)	Upper cut-off Frequency (Hz)
10.6	15	21.2
21.2	30	42.4
42.4	60	84.8
84.8	120	169.7
169.7	240	339.4
339.4	480	678.8
678.8	960	1357.6
1357.6	1920	2715.3
2715.3	3840	5430.6

Table 1: Octave bands used to extract statistical features

For each channel, there were 36 features (9 octave bands \times 4 statistical features) used to develop the PCA model. Each of the features has 8192 samples taken from each one second of data. For stable results, the statistical features were extracted for every two rotations of the impeller with an overlap of 3/4 of a shaft rotation to the following segment. Eighteen features were selected, and retained in the model which gave 95% of the variance, as depicted in the Table 2. This means that the PCA model, which consists of only half of the total features extracted, is sufficient to detect the impeller fault of a centrifugal pump.

Eigenvector	Eigenvalue	% Variance	% Total Variance
1	5.2	14.4	14.4
2	4.7	13.0	27.3
3	4.4	12.2	39.5
4	3.6	9.9	49.4
5	2.4	6.6	56.1
6	2.2	6.1	62.1
7	1.8	5.1	67.2
8	1.6	4.3	71.6
9	1.3	3.7	75.3
10	1.1	3.2	78.5
11	1.1	2.9	81.4
12	1.0	2.7	84.1
13	0.9	2.5	86.5
14	0.8	2.3	88.9
15	0.8	2.2	91.0
16	0.6	1.7	92.7
17	0.5	1.3	94.0
18	0.4	1.2	95.1

Table 2: Variance of PCA model

In order to detect the impeller fault, the control limit for the T^2 measure must first be determined using Equation (8). The PCA model can then be used to test each new set of vibration signals from a centrifugal pump and investigate the presence of the impeller fault. For model evaluation, the T^2 measure for the good impeller (non- faulty) was calculated, and the obtained values can be shown in Figures (4) to Figure (7) for channel 1 to 4 respectively. The control limits are marked with the dashed red line while the T^2 value obtained from the consecutive shaft rotations is marked with the blue line.

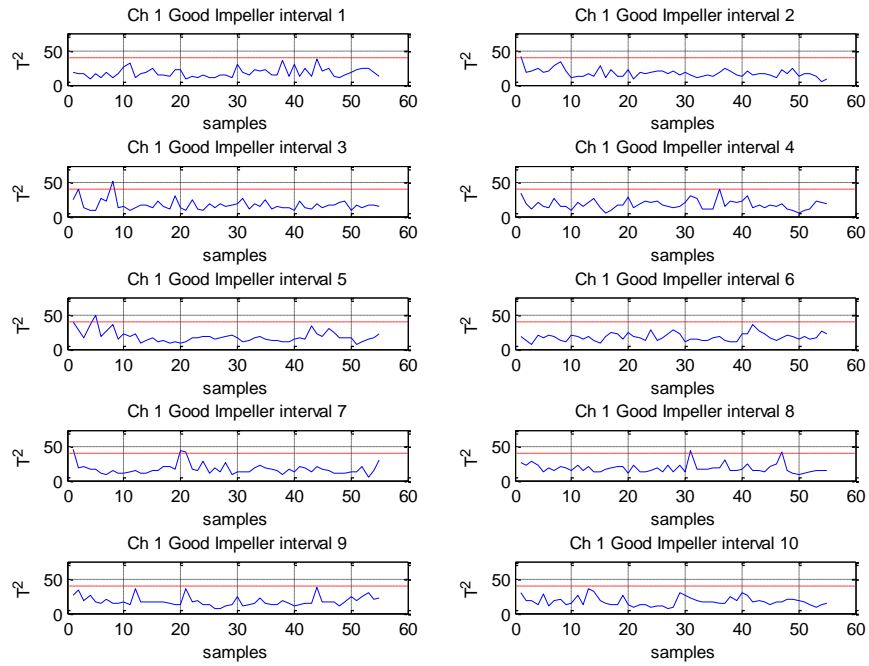


Figure 4: T^2 value of good impeller (channel 1)

From figure (4), interval 1 to 10, it can be seen that most of the values of T^2 for channel 1 are below the threshold except in the interval 3, 5 and, 7 where there is one sample for each of them that slightly exceeded the threshold. This is due to the non-stationary behaviour of the water in the volute casing of the impeller.

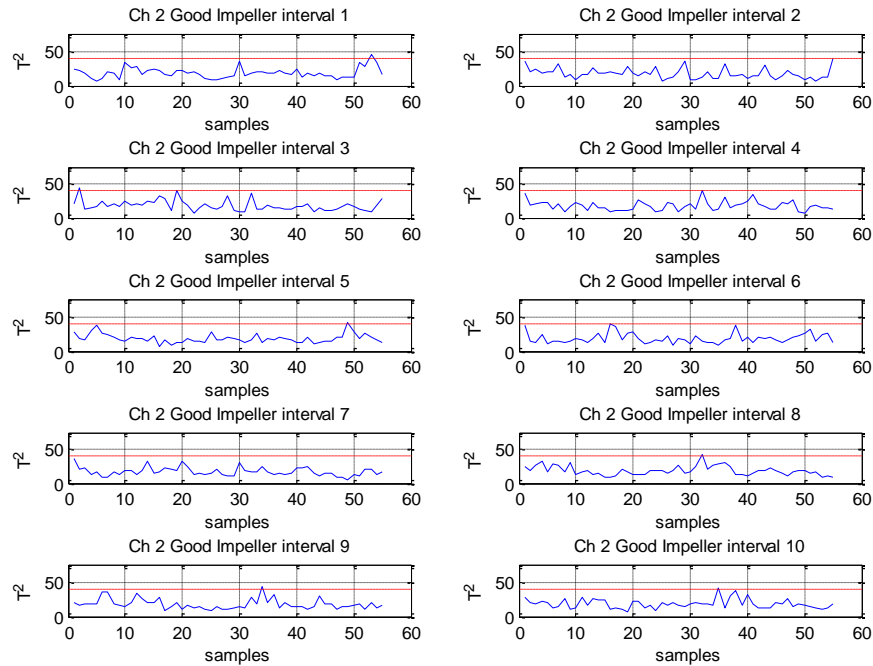


Figure 5: T^2 value of good impeller (channel 2)

The result for channel 2 depicted in Figure (5) shows that all of the samples in all of the intervals have T^2 values below the threshold but there is only one sample in interval 1 that exceeds the limit. This indicates that the mounting location of channel 2 has a better stationary behaviour of the vibration signal compared to channel 1.

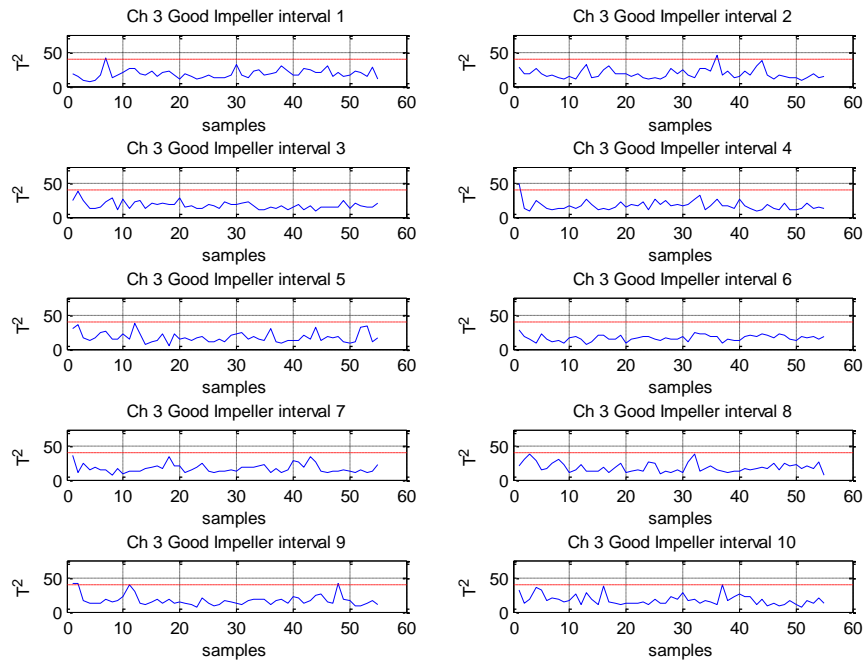


Figure 6: T^2 value of good impeller (channel 3)

Figure (6) shows that for channel 3, there are two samples in the interval 2 and 4 which have T^2 values that exceed the control limit. This channel also has a relatively stationary vibration signal since it is mounted on the bearing casing and, therefore, relatively far from the rotating water in the volute.

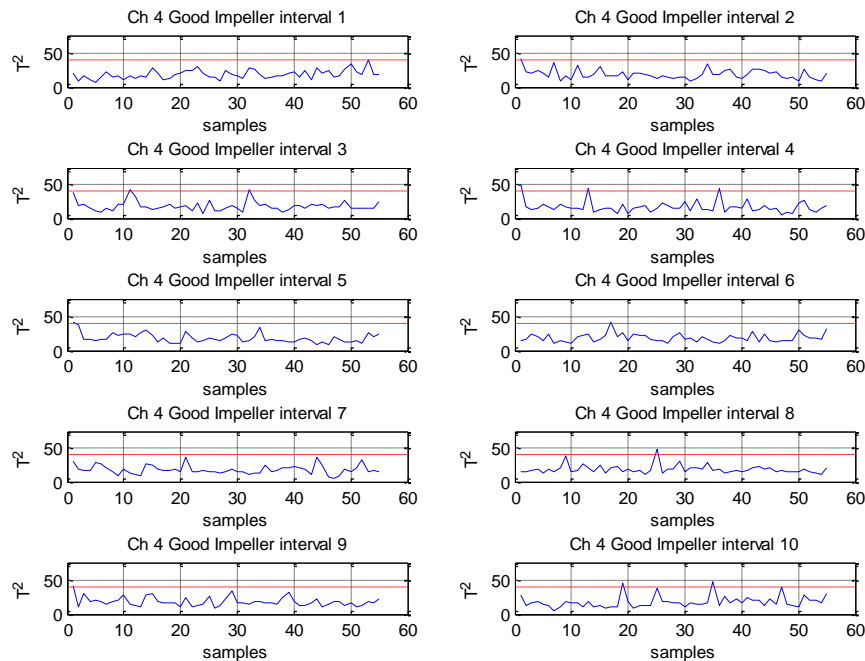


Figure 7: T^2 value of good impeller (channel 4)

The T^2 value for channel 4 shown in Figure (7) crosses the control limit several times in intervals 4, 8 and, 10. This is caused by the turbulent water in the discharge nozzle resulting in the non-stationary vibration signal.

From all of the results, it can be seen clearly that most of the samples have T^2 values below the control limit. Some of the samples slightly exceed the control limit, but that is due to a non-stationary behaviour of

water in the volute. It is shown that the PCA model successfully recognises the samples as from the vibration signal of a good impeller. The new data set from the faulty impeller was then examined by the PCA model and the results are provided in Figure (8) to Figure (11).

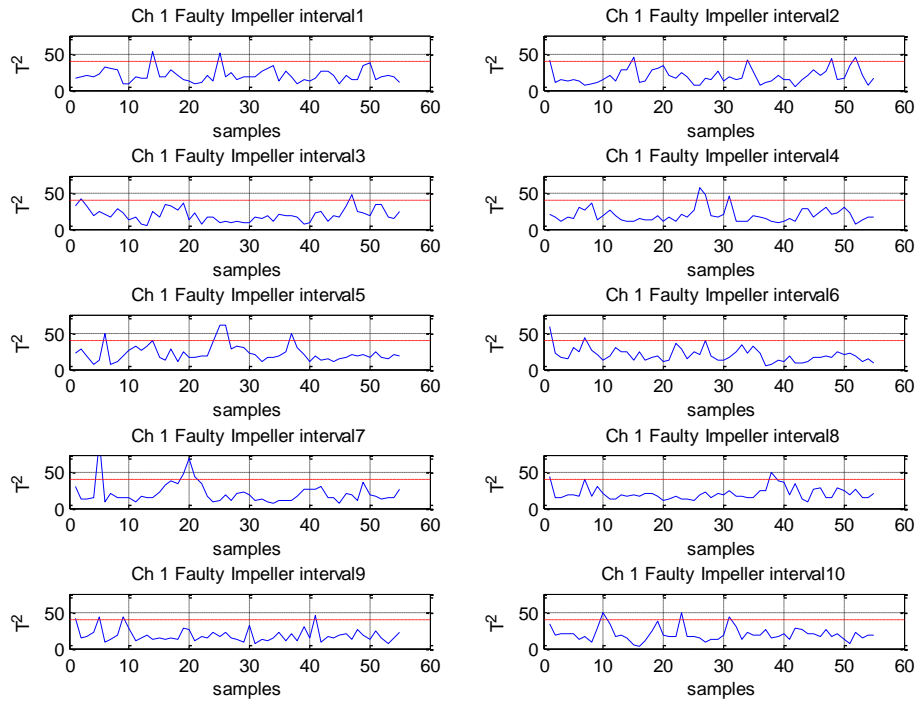


Figure 8: T^2 value of faulty impeller (channel 1)

Figure (8) shows that T^2 values of samples from channel 1 exceed the control limit many times. There are around 18 samples from interval 1 to 10 that cross the threshold, and several of them have large amplitudes, which indicate that the fault condition has occurred.

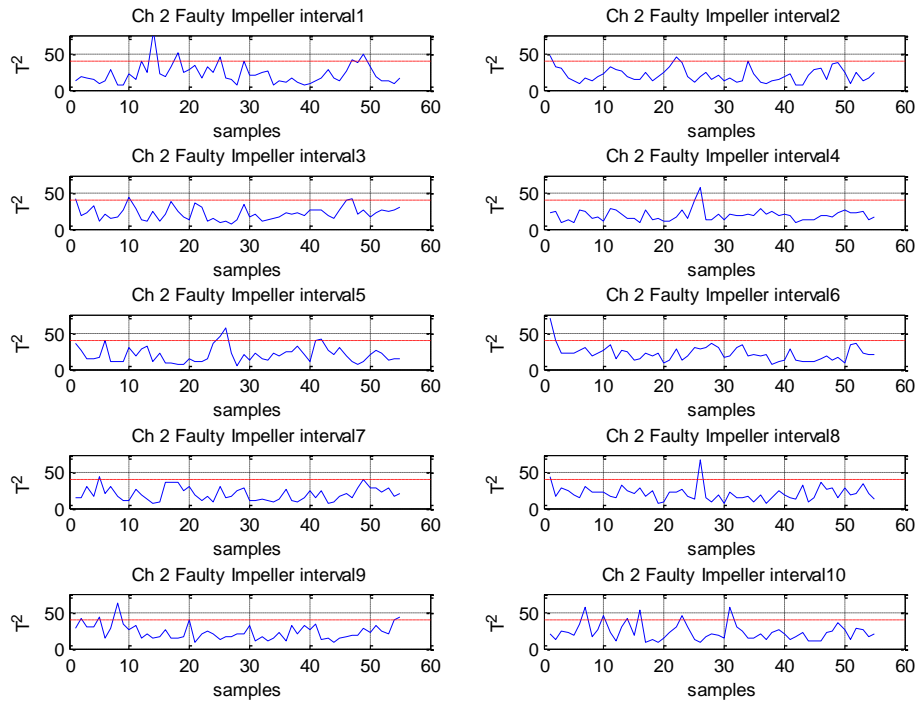


Figure 9: T^2 value of faulty impeller (channel 2)

From Figure (9), T^2 values from channel 2 exceeds the threshold many times (around 13 points) in several intervals. The largest value can be seen in interval 1 and interval 8. This indicates that faulty impeller condition has occurred.

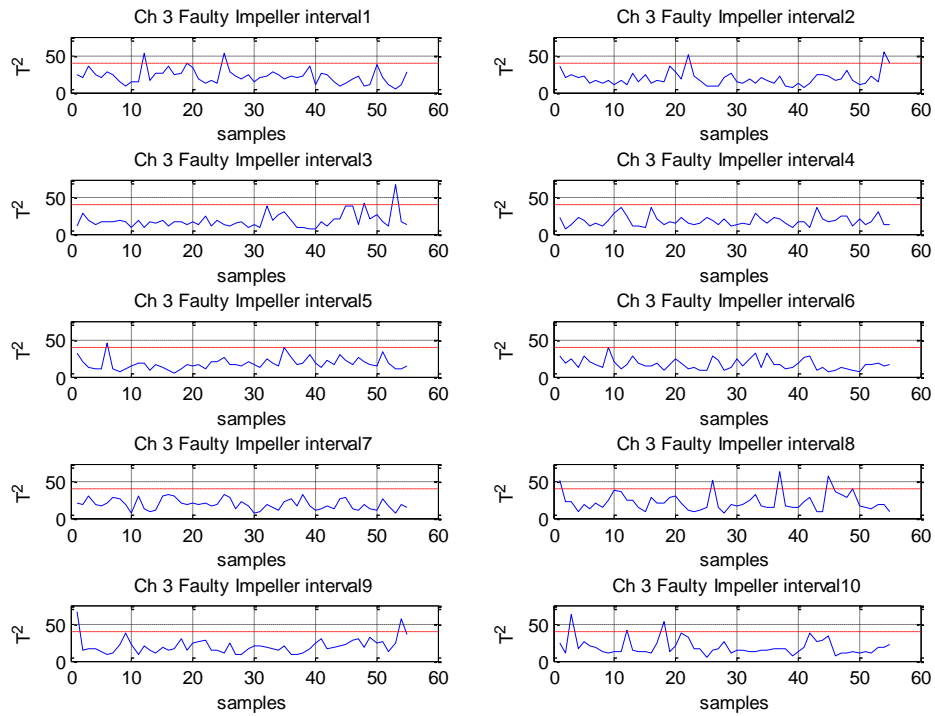


Figure 10: T^2 value of faulty impeller (channel 3)

Twelve samples exceed the control limit for channel 3 as depicted in Figure (10). From the obtained result, it can be seen that the large amplitude was observed in intervals 3, 8 and, 10. These results also confirm the occurrence of the faulty impeller condition.

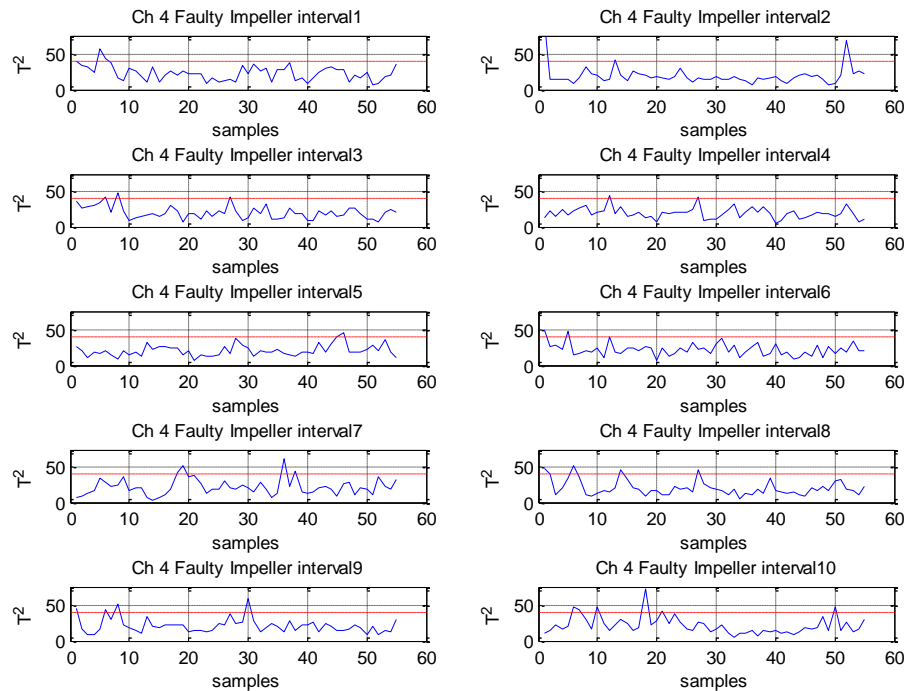


Figure 11: T^2 value of faulty impeller (channel 4)

The results in Figure (11) show T^2 values for channel 4 which crosses the control limit around 14 times with the largest amplitude occurring in intervals 2 and 10. These results confirm the three previous results which indicate the occurrence of the faulty impeller condition.

From all of the four channels of acceleration data, the PCA model shows consistent results in detection of the faulty impeller condition which is indicated by T^2 values that exceed the control limit. However, channel 1 gives the best results having the greatest number of samples that exceed the limit with large amplitude observed in intervals 5 and interval 7. The mounting location of this channel is on the volute casing of the impeller and hence can more sensitively record the change of fluid dynamics due to the impeller fault than the other locations. The faulty impeller makes the fluid in the volute change to have a more turbulent state. Since the impeller rotates the water in the radial direction, then its effect is best observed in the radial location of the volute as shown by the results using sensor on channel 1.

7 Conclusion

Octave band analysis was used to separate the measured pump accelerometer time domain data into nine octaves, where the octave bands were designed to coincide with the pump shaft speed and harmonics. Four statistical features along with nine octaves gave thirty six potential features for each channel. A PCA model has been developed, which retained eighteen out of the thirty six PC's. The model had 95% of data variation, which was sufficient to detect the impeller fault. For a good impeller, most of the samples had T^2 values below the threshold whereas only few had T^2 values that slightly exceeded the threshold; this was due to the non-stationary behaviour of water in the volute. Many T^2 values of samples obtained from the impeller fault's time waveform exceeded the confidence limits, which indicated that the model successfully detected the failure of the impeller. Channel 1 with the sensor location on the volute casing showed a very clear result for fault detection compared to other sensor locations.

References

- [1] Hellman, D.-H., *Early fault detection - an overview*, in *World Pumps* 2002, Elsevier Ltd.
- [2] Sakthivel, N.R., V. Sugumaran, and S. Babudevasenapati, *Vibration based fault diagnosis of monoblock centrifugal pump using decision tree*. *Expert Systems with Applications*, 2010. **37**(6): p. 4040-4049.
- [3] Mujica, L., et al., *Q-statistic and T2-statistic PCA-based measures for damage assessment in structures*. *Structural Health Monitoring*, 2011. **10**(5): p. 539-553.
- [4] Ahmed, M., et al. *Fault detection and diagnosis using Principal Component Analysis of vibration data from a reciprocating compressor*. in *Control (CONTROL), 2012 UKACC International Conference on*. 2012.
- [5] Liying, J., et al. *Improved confidence limits of T^2 statistic for monitoring batch processes*. in *Control and Decision Conference (CCDC), 2012 24th Chinese*. 2012.
- [6] Sun, X., et al., *An improved PCA method with application to boiler leak detection*. *ISA Transactions*, 2005. **44**(3): p. 379-397.
- [7] Ding, S., et al., *On the Application of PCA Technique to Fault Diagnosis*. *Tsinghua Science & Technology*, 2010. **15**(2): p. 138-144.
- [8] Nguyen, V.H. and J.-C. Golinval, *Fault detection based on Kernel Principal Component Analysis*. *Engineering Structures*, 2010. **32**(11): p. 3683-3691.
- [9] Cui, P., J. Li, and G. Wang, *Improved kernel principal component analysis for fault detection*. *Expert Systems with Applications*, 2008. **34**(2): p. 1210-1219.
- [10] Kang, T.-H. and Y. Kaizu, *Vibration analysis during grass harvesting according to ISO vibration standards*. *Computers and Electronics in Agriculture*, 2011. **79**(2): p. 226-235.
- [11] Nishiyama, K., et al., *The temporary threshold shift of vibratory sensation induced by hand-arm vibration composed of four one-third octave band vibrations*. *Journal of Sound and Vibration*, 1997. **200**(5): p. 631-642.
- [12] Roskovic, A., R. Grbic, and D. Sliskovic. *Fault tolerant system in a process measurement system based on the PCA method*. in *MIPRO, 2011 Proceedings of the 34th International Convention*. 2011.

- [13] Ahmed, M., F. Gu, and A. Ball, *Feature Selection and Fault Classification of Reciprocating Compressors using a Genetic Algorithm and a Probabilistic Neural Network*. Journal of Physics: Conference Series, 2011. **305**(1): p. 012112.