Article title: Big Data Quantitative Risk Analysis Method for Machine Health Indicator Prediction
Authors: George Jordan[1], Allan Brimicombe [2], Yang Li[3]
Affiliations: DEKRA Organisational & Process Safety, PHI House, Southampton Science Park, Southampton SO16 7NS[1],
Centre for Geo-Information Studies School of Architecture, Computing and Engineering Room KD2.28, The Knowledge Dock,
Docklands campus University of East London, London, E16 2RD[2]
Orcid ids: 0000-0003-0471-3430[1]
Contact e-mail: george.jordan@gjanalytics.com
License information: This work has been published open access under Creative Commons Attribution License
http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, distribution, and reproduction in any
medium, provided the original work is properly cited. Conditions, terms of use and publishing policy can be found at
https://www.scienceopen.com/.
Preprint statement: This article is a preprint and has not been peer-reviewed, under consideration and submitted to
ScienceOpen Preprints for open peer review.
DOI: 10.14293/S2199-1006.1.SOR-.PP55DCW.v1
Preprint first posted online: 30 July 2021
Keywords: Data Mining, Change-point Detection, Data-driven Methods, Regression Modelling, Interaction Effect,
Process Fault Detection, Rolling Bearing, Bearing Fault
Big Data Quantitative Risk Analysis Method for Machine Health Indicator Prediction

George Jordan
Centre for Geo-Information Studies
School of Architecture, Computing and Engineering
Room KD2.28, The Knowledge Dock, Docklands campus
University of East London, London, E16 2RD

Professor Allan J. Brimicombe (Lead Supervisor)
Centre for Geo-Information Studies
School of Architecture, Computing and Engineering
Room KD2.28, The Knowledge Dock, Docklands campus
University of East London, London, E16 2RD

Dr Yang Li (Co)
Centre for Geo-Information Studies
School of Architecture, Computing and Engineering
Room KD2.28, The Knowledge Dock, Docklands campus
University of East London, London, E16 2RD

*George Jordan has since moved to DEKRA Organisational & Process Safety

Corresponding Author: George Jordan
DEKRA Organisational & Process Safety
PHI House Southampton Science Park
Southampton SO16 7NS

Email: george.jordan@dekra.org
HIGHLIGHTS

- Monitoring the health of mechanical equipment has always been a challenge for the highly hazardous process industries.
- Bearing is one of the basic components in process machinery so any risk associated with bearing deterioration may have a detrimental effect on the entire operation of the process system.
- Monitoring bearing health is applied for machine health indicator prediction used by the highly hazardous process industries.
- One such method is remaining useful life, which predicts the time after the onset of bearing deterioration.
- Big data quantitative risk analysis is an alternative approach which detects the onset of deterioration in a component bearing and can provide information about the contribution from the operation of other component bearings to the health issues of process machinery.

SUMMARY

Good health indicators of mechanical equipment have always been the desire of operation and safety managers because they help improve reliability, minimise operation cost and reduce downtime. One of the most basic components of a process machinery is the bearing, hence any risk associated with its deterioration may have a detrimental effect on the entire process operation. As a result, methods like risk analysis, preventative management, condition monitoring, prognostics and remaining useful life (RUL) are applied to monitor bearing health conditions to aid effective decision-making within process industries. We applied the Big Data Quantitative Risk Analysis (BDQRA) method to monitor bearing health conditions and verified the outcome with other bearing health indication method like the RUL which predicts the time to end of life of the bearing. We found that the BDQRA method provides good performance at detecting the onset of bearing deterioration. The BDQRA method also has a major advantage compared with other bearing health monitoring methods because it can provide information about the contribution to the deterioration from the operation of other components of a process machinery.

Abstract: Various data-driven methods have been applied to predict machine health indicators especially in the field of prognostics. Machine health indicators reveal the condition of equipment and/or its components including bearings by monitoring their operation data such as frequency vibration. To aid the prediction of the machine health indicators, this study applies the BDQRA method to monitor the health of bearings as a component of the machine. The BDQRA method involves applying data compression techniques like feature extraction to the bearing vibration data, to extract the most important features like time-domain, frequency domain, and time–frequency
domain features. Due to the complexity of the feature extraction process, this study proposes fast Fourier transformation for the data compression. This is followed by obtaining a time series profile of the bearing vibration data to analyse the health status of component bearing. It the uses change-point analysis to predict the period at which the bearing health deterioration is imminent. Since the bearing health deterioration could be due to the independent operation of a component bearing or through communication between the component bearing and other components (or bearings) within the process machinery, the method also applies the principle of interaction effect to investigate the contributions from the other components of the machinery to the health deterioration of the component bearing detected. The accuracy of the prediction of the point of imminent health deterioration of the component bearing is investigated by comparing the outcome of the BDQRA method with the outcome of other methods published in literature which have been applied to the dataset used in this study. The findings reveal the BDQRA method have comparative advantages to the methods used in the related studies.

**Keywords:** Data Mining, Change-point Detection, Data-driven Methods, Bearing Fault, Process Fault Detection, Regression Modelling, Rolling Bearing, Machine Breakdown, Interaction Effect.

1. **Introduction**

Health indicators in mechanical equipment have always been challenging to operation and safety managers over the years due to several reasons including improving reliability, minimising cost of operation, and reducing downtime (Heng et al., 2009; Lee et al., 2014). Owing to this, several methods has been applied to aid effective decision-making within industries. They include risk analysis, preventative management, condition monitoring, prognostics, and RUL (Qiu et al., 2003; Benkedjouh et al., 2013) so that appropriate action can be taken to prevent excessive downtime of the operation. All the aforementioned techniques have been applied for monitoring the health condition of mechanical components including bearings.

Bearing is one of the most basic components of most process machinery and plays a major role in issues affecting the overall operations of a process system. As a result, any risk associated with deteriorating bearing health condition may have a detrimental effect on the entire process operation of the facility. Many potential sources of bearing health issues including mechanical causes like overheating have been reported (CSB, 2009). As a result, numerous bearing health monitoring techniques including acoustic measurement, electrical effects monitoring, oil debris monitoring, power quality, temperature monitoring and vibration analysis have been investigated by researchers (de Azevedo et al, 2016).

Other big data techniques like alpha-stable probability distribution model (Li et al.,2016), artificial neural network (Ziani et al., 2012; Unala et al., 2014), discrete wavelet transformation (Kumar, 2013),
cluster analysis (Liu et al., 2014), wavelet decomposition analysis (Khanam et al., 2014; Li et al., 2013; Roulias et al., 2013; Sarvajith et al., 2013), data envelope analysis (Dalvand et al., 2014; Guo et al., 2014; Ming et al., 2015), local and RUL (Okoh et al., 2014; Ragab et al., 2014; Caesarendra & Tjahjowidodo, 2017) have been investigated. Time domain and frequency domain methods like root mean square (RMS), kurtosis and fast Fourier transformation (FFT) have also been investigated (Hsiung-Cheng et al., 2016; Cai et al., 2019; Dong et al., 2020).

This paper proposes a hybrid method for predicting the time at which imminent bearing health deterioration could occur. This hybrid method utilises a collection of methods including time-domain method like RMS, frequency domain method like FFT and big data methods like time series analysis, changepoint analysis, regression decision tree analysis, regression analysis, analysis of variance and interaction effects. Although each of the methods within the hybrid method can be applied for bearing health deterioration analysis on its own, we are of the view that when used separately, each method has its own strengths and weaknesses. Besides, some of the individual methods have already been investigated using the dataset applied in this research. In our view, the hybrid approach using a combination of all these methods will maximise the overall investigation than using the individual methods. The hybrid method was developed and tested using historical data but would be run in near real-time in actual situations. We called this method Big Data Quantitative Risk Analysis (BDQRA) because it was applied for quantitative risk analysis at the time of its development and testing (Jordan, 2019).

The proposed method differs from other methods which have been applied by previous researchers as it adopts the big data technique of change-point analysis (Basseville & Nikiforov, 1993; Killick et al., 2010) to capture the point at which the imminent bearing health deterioration could occur and applies the theory of interaction effect to determine the contribution of other mechanical components within the process machinery to the deterioration detected. Fig.1 is a schematic diagram of the method.
Fig. 1: Schematic diagram of the method

The method employs change-point analysis to detect the deterioration in condition of a bearing. It applies data compression technique (FFT) to the data after which time series profiles of the data is obtained, two change-point analysis techniques, regression decision tree analysis, regression analysis and interaction effect. The two changepoint techniques are (a) structural change, which determines changes in the time series and regression model and is therefore used to detect the onset of the deterioration condition and (b) change-point by changes in the variance which determines the point of the main deterioration condition. Any contribution from other components within the process system to the deterioration condition detected are then investigated.

The remainder of this paper is organized to include Section 2 which presents a description of the dataset used, including a description of data selection, storage, dominant features in bearing vibration data which depends on the health conditions of the bearing. Section 3 presents a description of change-point analysis and the reason for its selection as a deterioration condition detection method. Section 4 presents detail explanation on how the BDQRA method works. This is followed by a presentation of the various stages within the BDQRA method as applied in this...
2. **Data and Sources**

The data for the study was obtained from the NASA Prognosis data repository (https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/). A total of three bearing operation datasets labelled as Bearing Test 1, Bearing Test 2, Bearing Test 3, and Bearing Test 4 were obtained. The data labelled Bearing Test 2 used for this research. A publication by the donors of the data (Lee *et al.*, 2007) discloses that the process system has four bearings each of which was rotating at a constant speed of 2000 rpm by an AC motor coupled to the shaft via rub belts. Fig 2. Is an image of the process setup and schematic diagram of the arrangement of bearings within the process machinery.

![Process setup and schematic arrangements of bearings in the machine](image)

Fig. 2: Process setup and schematic arrangements of bearings in the machine (Source: Lee *et al.*, 2007)

Various researchers have suggested that the most dominant features in bearing vibration data is the level of the key vibration frequencies (Nistane & Harsha, 2016; Saruhan *et al.*, 2014). These features...
depend on the health conditions of the bearings. Table 1 is a summary of the bearing vibration features and how they are determined.

**Table 1: Summary of bearing vibration features.**

| Features                        | Mode of Determination                                                                                                                                 |
|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| Visually derived features       | A method is developed to capture patterns in the data.                                                                                                  |
| Statistical features            | Statistical methods are used to obtain statistical properties like the mean, median, standard deviation, root mean square, skewness, kurtosis, entropy, shape factor and crest factor. |
| Time-frequency representation   | One-dimensional time-domain signals are mapped to a two-dimensional function of time such as short-time Fourier transform, wavelet transform and Wigner-Ville distribution. |
| Complexity measurements        | Nonparametric tests (e.g. Kolmogorov-Smirnov and simple entropy tests) are used to measure the similarity of two cumulative distribution functions.             |
| Phase-Space Dissimilarity       | Dissimilarity measurements (e.g. Fractal Dimension and approximate entropy) are used to quantify signal complexity.                                      |
| Measurement                     |                                                                                                                                                      |
| Other features                  | Suitable methods (e.g. singular value decomposition, piecewise aggregate approximation and adaptive piecewise constant approximation) are used to quantify the periodicity of the time series data. |

The donors of the dataset provided the details of the four bearings (Qiu *et al.*, 2006) as:

- Ball diameter ($Bd$): 0.331 inch
- Number of rolling elements ($Nb$): 16
- Rotational speed ($Rs$): 2000 rpm = 33.3Hz
- Pitch diameter ($Pd$): 2.815 inch
- Contact angle ($\alpha$): 15.17° = 0.084227 radians

From these details, we calculated the rotational frequency ($RF$) from the rotational speed as 33.3 Hz (i.e., 2000/60) and the contact angle as 0.08427 rad (i.e. 15.17/180). We also calculated the characteristic bearing fault frequencies using descriptions proposed by other researchers (Granny & Starry, 2011; Kamaras & Dimitrakopoulos, 2016; Sohaib *et al.*, 2017). They include ball pass frequency for outer race (BPFO), ball pass frequency for inner race (BPFI), ball pass roller frequency (BSF), the fundamental train frequency (FTF) and statistical time-domain features such as shape factor (SF), crest factor (CF) and spectral kurtosis (SK). We adapted the description of the baring fault frequencies for the researchers into Table 2.
Table 2: Summary description of characteristic bearing fault frequencies.

| FAULT FREQUENCY | DESCRIPTION |
|-----------------|-------------|
| BPFO            | The rate at which a ball or roller passes a point on the outer race \[0.5Rf(1 - \text{Ratio})\] = 14.69 |
| BPFI            | The rate at which a ball or roller passes a point on the inner race \[0.5Nb \times a(1 - \text{Ratio})\] = 0.595 |
| BSF             | The rate at which a point on a ball or roller will contact the inner OR outer race \[\frac{p_d}{b_d} \times 0.5a(1 - \text{Ratio}^2)\] = 0.353 |
| FTF             | Rate at which bearing cage travels around the bearing \[0.5a(1 - \text{Ratio})\] = 0.0149 |
| VHF             | Very high frequency (> 6 kHz) |
| HF              | High frequency (2.6 – 6 kHz) |
| MF              | Medium frequency (1.5 – 2.6 kHz) |
| LF              | Low frequency (0- 1.5 kHz) |
| CF              | Calculate the magnitude of impact due to rolling and raceway contact, appropriate for spiky signals. It is the standard deviation divided by the RMS \(\frac{\sigma}{\text{RMS}}\). |
| Entropy         | A measure of the degree of randomness in the data. |
| Kurtosis        | Quantifies the peak value of the PDF. The value is approximately 3 for a healthy bearing (Eftekharnejab et al., 2011). |
| RMS             | Values help identify differences between vibration signals. The same applies to the mean and SD. |
| SF              | The RMS to mean ratio. The value depends on an object’s shape but independent of dimensions. \(\frac{\text{RMS}}{\mu}\) |
| Sk              | Quantifies symmetry of data, value approximates to 0 for healthy bearing, shifts to positive or negative when a fault develops. |
| Var             | Measures the dispersion of the data around the mean. |

Where,

\[
\text{Ratio} = \frac{p_d}{b_d} \cos(a)
\]

We applied frequency domain analysis using fast Fourier transform (FFT) algorithm to decompose the signals into their Nyquist frequency (Nf). The Nf is half the sampling rate of the of the signal (Seeber & Ulrici, 2016). Thus, the FFT help remove noise from the observations without losing key features. As a result, the data is condensed to a size that could be handled by system memory for
easy analysis using R programming language. There is data for the four bearings which became defective during the operation. An image of the defects is provided (Cerrada et al, 2018) as Fig.3.

According to the metadata, attributes of temperature, vibration, and flow of lubricant were monitored as part of the process (Qui et al., 2011). The usability of the data was investigated by a review of peer-reviewed research publications which involves the use of the datasets (Jordan, 2019).

3. Change-point Analysis for Risk of Bearing Health Deterioration Detection

We selected change-point analysis for risk of bearing health deterioration detection after a review of literature on the use of big data for bearing health analysis (Jordan, 2019), together with considerations like the size of the data and the data being time stamped. Because the data was recorded over given time periods, we hope to extract important information such as descriptive and explanatory variables which can be processed and used with data mining models including change-point analysis. We also considered the application of Weibull distribution or time series analysis as part of the big data technique to incorporate into the change-point analysis for risk of bearing health deterioration detection.
We found that both Weibull distribution and time series analysis has been extensively applied in engineering research like; fatigue and endurance life in engineering devices and materials (Herbert et al., 2010), failure analysis (Rajkumar et al., 2011; Zhai et al., 2013), traffic modelling (Ageyev & Qasim, 2015) and survival analysis, ball bearing, communication systems engineering, energy from wind turbines and wind speeds (Salh, 2014). However, it has been reported that most Weibull models only use time series data which are normally distributed (Kadhem et al., 2017). Owing to these, we decided to combine time series analysis and change-point analysis for the risk of bearing health deterioration detection.

Change-point analysis focuses on sequential detection of change in trend in a time series data and has been applied in many areas of research including oceanography (Killick et al., 2010), biological systems such as DNA copy (Killick & Eckley, 2014), process behaviour in engineering (Sharma et al., 2016) and meteorology (Arif et al., 2017). The method can be (a) sequential (online), where the analysis may be performed to identify abrupt changes as soon as the real-time data becomes available, or (b) retrospective (offline), where one off analysis may be performed on historical time series datasets (Jordan, 2019). It can also handle data of different dimensions as well as detecting single and multiple events (Aminikhanghahi & Cook 2017; Fisher & Jensen, 2018).

The change-point detection techniques include change-point analysis, quality control, anomaly detection, breakout detection, segmentation, structural change, and event detection. Although these techniques appear similar, they are different. For instance, anomaly detection aims at detecting outliers, quality control focuses on the stability of mean and standard deviation, while change-point detection only interprets abrupt changes in the data. Some researchers refer to change-point techniques as change-point mining because the process involves using data mining techniques to acquire knowledge (Boettcher, 2011). The various change-point techniques have been investigated, categorized, and compared (Zarenistanak et al., 2014). Reviews of change-point methods have also been extensively covered (Reeves & Chen, 2007; Aminikhanghahi & Cook, 2017, Truong et al., 2018).

4. **How the BDQRA Method Works**

Our method combines the two change point techniques (change-point and structural change) to detect the onset of deterioration of bearing health to help avoid catastrophic events in process machinery. We the apply regression decision trees to determine predictors and moderators for the regression models for the interactions up to the point at which the risk of deterioration of the bearing health is imminent. We are of the view that regression decision tree in the BDQRA method is applicable because the study involves continuous variables (De Cock et al., 2017, Ahmad et al., 2018, Pekel, 2020). We apply linear regression models to determine the interactions up to the point of the eminent risk of deterioration. To enable a side-by-side comparison of the linear regression models,
we use the ready-made regression tables from the Stargazer package on the R-language platform to provide summaries of the linear regression analysis results (Hlavac, 2014). Any statistically significant interactions detected in the linear regression analysis will be further investigated using analysis of variance (ANOVA) type II. If the significant interactions are confirmed, effect plots would be produced to visually express the relationships and any uncertainties within the model measurements.

Interaction effect applies to a situation where the operation of one of the components within the process machinery affects another. We are of the view that significant effect of any interactions up to the point of imminent risk of deterioration of the bearing health detected by the change-point technique would help provide insights of the type of systems being exhibited within the process machinery. This may help us to provide some theory arguments and explanations to the statistically significant effect of the interactions between the bearing and other components of the process machinery (Andersson et al., 2014). This is because the components of a process machinery can exhibit organised simplicity or organise complexity (Goerlandt & Reniers, 2018) with the bearing. In an organised simplicity, each of the component of the process machinery including the bearing may be operating independently so any risk of deterioration of the bearing’s health detected in the bearing operation data may not be due to contributions from the other components of the process machinery (i.e. no interaction effect). In organized complexity, the operations of the components of the process machinery including the bearing may have an effect on one another through non-linear interactions and feedback loops. As a result, the risk of deterioration of the bearing health detected may be due to contributions from the operation of one or more of the other components (or combination of their contribution) through interaction effect.

For clarity, we adopt the terminology proposed by Grace-Martin to provide clarity of the use of ‘moderators’ and ‘predictors’ in the principle of interaction effect within the context of this study (Grace-Martin 2018). By ‘predictors’, we refer to any component whose operations may have a potential to produce an effect on the risk of deterioration of the health of the bearing which develop health issues, without any real distinction between their roles. We also refer to the components whose operations can make some contributions to the effect of the operation of the ‘predictor component’ on the the risk of deterioration of health of the bearing detect as ‘moderators.’

5. **The Stages within the Proposed Approach**

As mentioned in Section 1, the BDQRA method is a hybrid method that combines a collection of big data techniques. The BDQRA method was developed and tested as a quantitative risk analysis technique to determine deterioration in the health conditions of a process machinery using data from
the operations of the machinery. Generally, the deterioration of the health of a component of the operation machinery is deemed to have occurred when the operation of the machine fails or some issues of abnormality within parts of the operation of the machine occurs. The deterioration could be caused independently or caused by a remote event including interaction of the various components in the form of a loop. We apply the BDQRA method to bearing operation dataset because the bearings forms part of the operation of many machines used for industrial application. Sections 4.1 to 4.5 detail the stages within the BDQRA method.

5.1. **Data Selection and Data Quality Investigation Stage**

The initial step of the method involves obtaining information about the process machinery for which the method would be applied. Data from the operation of the process machinery are then obtained for analysis. This is followed by assessing quality and attributes of the data using available information on the data detailed in the meta data. Other attributes of the data such as sampling rate or sampling times are also investigated using at least 10% of the data selected at random.

5.2. **Data Mining and Exploration Stage**

The second stage involves application of descriptive statistics to investigate the completeness of data in the data files by looking for missing observations such as nulls and N/A's. Where missing observations are observed, we apply appropriate techniques such as (a) filling the missing observation with the last observation, (b) applying linear interpolation to fill in missing observation, (c) applying polynomial interpolation to fill the missing observation. These are followed by visual investigation using a plot. The plots are then compared. The best fit plot is deemed as the appropriate technique to fill the missing observations. Descriptive statistics is used to investigate other parameters including the magnitude of the mean, maximum, minimum values, and standard deviation of the observations.

Visual plots such as histogram, box plots, and quantile plots are used to help establish the underlying distribution of the data. Sequence and lag plots are used to investigate randomness and time effect of the observations, and any potential bias in the operation data. The normality of the distribution of the data are investigated using Anderson-Darling multivariate normality test. Pearson’s correlation analysis was used to investigate any potential relationships which may exist between the operation of the components.

5.3. **Data Condensing Stage**

At this stage, the data obtained from the operations of the process machinery are condensed into data sizes which can be handled within the memory of a PC. Since data obtained from the
operation of some of the components of process machinery e.g. bearings are generally suppressed by noise due to the small magnitudes of bearing vibration frequencies (Shirong et al., 2014), fast Fourier transform (FFT) is used to obtain key statistical and time-domain features. Other techniques like calculation methods are also used to obtain feature whose information may be missing from the data. For instance, the speed and contact angles of the component bearing were described in ‘rpm’ and ‘degrees’, so appropriate equation was applied to convert the speed and contact angle into units of Hz and rad. The data is also formatted according to the time sequence from the source files, to create bearing-specific data files which were timestamped according to the time on each of the test file in the dataset. The bearing-specific data files are then combined into a new data frame and stored in a storage media.

5.4. Investigation of Bearing Health Condition Stage

We explore the bearing-specific datasets using the approach discussed in Section 4.2 - data mining and exploration stage. Time series profile plot and plot of the RMS are applied to the bearing-specific datasets to investigate the health condition of the component bearing. Statistical RMS has been successfully applied in a previous research to investigate the health condition of bearings (Wu et al., 2017). We use the two change-point analysis techniques - change-point by structural change (strucchange) and change-point by changes in the variance (changepoint), to determine the point of imminent bearing health deterioration. This approach is based on the assumption that if a bearing develops or have any health-related issues like a deterioration, there would be an abrupt change or a disturbance in the time series profile of the vibration data. The onset and offset of this deterioration can be detected by strucchange and changepoint respectively. However, if the bearing has already suffered a deterioration of health at the onset of the operation or is healthy but not properly secured within the operation machinery before the operation of the process machinery was started, the time series profile of the vibration data will show a disturbance from the onset of the bearing operation. Additionally, the time index of the onset and offset of the bearing health deterioration determined by strucchange and changepoint would be relatively similar.

We then use the plot the RMS of the data which has been successfully applied by Wu et al., 2017 to confirm the outcome of our investigation by comparing the time indices detected as the onset and offset of the bearing health deterioration detected by strucchange and changepoint. Wu et al. reveal that bearing health-related issues appears as changes in the trends of the RMS profile plot. As a result, we expect the onset and offset of the bearing health deterioration detected by our method to approximate the time indices of any changes in the trend in the RMS profile plot of the data.
5.5. **Investigation of Contributions from Other Components Stage**

We investigate contribution from other components of the process machinery to the deterioration of health suffered by the component bearing. This contribution may be due to relationships between the operation bearing which suffered the health deterioration and that of the other components of the process machinery through interactions up to the point of the imminent deterioration. We investigate this relationship with Pearson’s correlation plots and significance tests.

A statistically insignificant correlation at $p > 0.05$ means there is no contribution from the operations of the other components of the process machinery to the deterioration suffered by the component bearing. A significant correlation at $p < 0.05$ means there is some contribution from the operations of the other components process machinery to the deterioration suffered by the component bearing through interactions up to the point of imminent deterioration hence further investigation is required. As a result, the correlation plots are reviewed to ascertain whether the significance is due to a relationship between the operation bearing which suffered the deterioration and that of the other components of the process machinery.

The significant correlations are further investigated with regression decision tree models to establish (a) the extent of contribution from the operation of the other components of the process machinery to the deterioration of the bearing, (b) the component whose operation is the makes the most contribution to the deterioration of the bearing (i.e. the predictor), (c) the components whose operations somehow influences the contributions from the predictor to the deterioration of the bearing. This is followed by regression modelling to investigate the contributions being made through interaction of the operation of the components up to the point of imminent deterioration of the bearing. Here, the component whose operation was found have made more contribution to the deterioration of the bearing us used as the predictor, and the components making less contribution as moderators. Significant interaction effects determined from the regression modelling are further investigated using Type II ANOVA test.

Type II ANOVA test was preferred to Type I and Type III for this study because we observed that in Type I ANOVA, the order of the variables generally matter. As a result, the position of the components being used as the predictor and moderators in the model (i.e. first or second) makes a difference in that the first variable in the regression model is compared to a model with just the intercept, and the second variable in the regression model is compared to a model with the first variable and the intercept. As a result, a change in the order of the variable within the regression model causes variable outcomes due to correlation of the predictors. We also found that Type III ANOVA gets around the issue observed with Type I ANOVA by assessing each predictor in the
regression model including the interactions, against a regression model which include every variable but the predictor. However, dealing with interactions without one of the main effects could lead to some extent a meaningless outcome because of its sensitivity to the main effects and any missing observations. The study therefore applied Type II ANOVA because it does not suffer from the issues observed with Type I and Type III ANOVA since the main effects are tested with other main effects in the regression model but not the interactions. Thus, each effect is easily interpreted as the unique contribution to the prediction of the model.

We then use effect plots to produce a visual representation of the predicted values of the outcome for given values for the component used as the predictor in the regression models. This helps to us to provide a visual presentation to help explain how we select the appropriate model fit of the analysis. We also explore the nature of the significant interaction to understand the form taken by the interaction using Johnson Neyman plots and simple slope analysis. This helps to describe the relationship between the operation of the component bearing which suffered the health deterioration and low, medium, and high operation values of the component of the machinery whose operations made more contribution to the health deterioration of the bearing.

6. Research Findings and Discussion

Fig. 4 is the plot of the spectrum obtained for all four bearing-specific data of the four component bearings in the process machinery after the data compression with FFT. The first spectrum on the frequency axis (around 10) in the plot shows the fundamental train frequency (FTF). This is followed by the ball pass frequency for inner race (BPFI) around 140, the ball pass frequency for outer race (BPFO) around 240, then the ball pass frequency for the roller (BSF) around 300. It was observed from the plot that none of the four characteristic deterioration frequencies have been affected.
Fig. 4: FFT profile plot of Bearings
Fig. 5 are the plots obtained when we applied *changepoint* to the data after the bearing-specific data. The plots represent the time series profile of the vibrations of each of the component bearing, with the time index of the bearing health deterioration detected (indicated by a red line). The plots reveal that a bearing health deterioration was detected at time index of 968 in the data of Bearing 1. However, no bearing health deterioration was detected in the data of Bearing 2, Bearing 3, or Bearing 4. Since each time index represent a test file, the corresponding test file for time index 968 was found to be the file with the time stamp “2004.02.19.03.42.39”.

Fig. 6 are the plots obtained when *strucchange* was applied to the bearing-specific data for each bearing. This plot also reveals that a bearing health deterioration was detected in the data of Bearing 1 at time index 837. Again, no bearing health deterioration was detected in the data of Bearing 2, Bearing 3, or Bearing 4.
Fig. 6: A plot of deterioration detected with change-point by structural change. Note: the y-axis for each plot is at a different scale to accommodate data range.

We compared the time indices of the bearing health deterioration detected by the two change-point techniques in our method with a plot of the profile of RMS of the data of Bearing 1. We create a plot of the RMS profile of only Bearing 1 because both change-point techniques in our method detected bearing health deterioration its operations data. The plot (Fig. 7) shows a slight increase in trend from the beginning of the operation of the Bearing 1. There is a break in the slope of the trend at time index of about 700 where a sharp increase occurred. This is followed by a fall back to about time index 800. After this point, a slight increase occurred in the trend up to around the time index 850, followed by large fluctuation. This period is around the time index of the lower threshold of the bearing health deterioration detected by the change-point technique \textit{strucchange}. An exponential rise occurred from about time index 900 which continue to the end of the life of the bearing. This period is around the time index of the upper threshold of the bearing health deterioration detected by change-point technique \textit{changepoint}. The RMS plot thus revealed the two change-point detection techniques within our hybrid method provides a good detection of the deterioration of the bearing health.
We investigate the contribution from the operation of the other bearings in the process machinery to the health deterioration of Bearing 1 by investigation of the interactions of their operation in the process machinery. We focus on interaction of the operation of the components up to the time indices 837 and 968 detected by the method. We begin the investigation using Pearson’s correlation analysis. Fig. 8 is the plots of the histogram of the and correlation of the distributions of the bearing-specific data, the correlation coefficients, and their corresponding statistical significance of Pearson’s correlation (indicated by stars). The plots reveal that the distribution of the operation data of all four bearings are positive skewed. However, the data of the data from the operations of Bearing 2 and Bearing 3 have a significant positive Pearson correlation with that of Bearing 1. This may be due to their contribution of the operations of Bearing 2 and Bearing 3 to the health deterioration of Bearing 1. We also noted that the data from the operations of Bearing 4 has no significant correlation with that of Bearing 1. Although there appear to be a significant correlation between Bearing 2 and Bearing 4, a careful observation of the correlation plot led to the conclusion that the significant correlation may be attributed to the large size $n$ and not because the two distributions are significantly correlation.
Big Data Quantitative Risk Analysis Method for Machine Health Indicator Prediction

We then proceed with the analysis using regression decision trees to investigate the significant Pearson's correlation of the interactions up to time indices 837 and 968. The plot of the regression decision tree Fig.9, reveals that the operations of Bearing 3 is the main component in the process machinery whose operations made a dominant contribution to the operations of Bearing 1. This is followed by the operations of Bearing 2 whose operations appears to influence the contributions made by the operations of Bearing 3 to that of Bearing 1.

Fig.8: Distribution, and correlation of data of all four component bearings in operation and their interactions up to the time indices 837 and 968 of the health deterioration detected in data of Bearing 1.

Fig.9: Decision tree for interactions up to time indices 837 and 968 detected in Bearing 1.
As a result, we proceed with linear regression modelling to help predict the effect of interaction of the operation of the component bearings using Bearing 1 as the dependant component, Bearing 3 as the predictor, and Bearing 2 as the moderator. The summary output of the linear regression modelling is presented in Table 3.

Table 3: Output of regression for the of deterioration through interaction effect

| Dependent variable | model 1 (1) | model 2 (2) | model 3 (3) | model 4 (4) |
|--------------------|-------------|-------------|-------------|-------------|
| Constant           | 3.475***    | 5.611***    | 4.004***    | 6.118***    |
| (0.314)            | (0.569)     | (0.309)     | (0.559)     |
| BPF0.B3            | 0.112***    | 0.021       | 0.099***    | 0.008       |
| (0.013)            | (0.024)     | (0.013)     | (0.024)     |
| BPF0.B2            | 0.174***    | 0.002       | 0.165***    | -0.065      |
| (0.023)            | (0.045)     | (0.023)     | (0.045)     |
| BPF0.B3:BPF0.B2    | 0.006***    | 0.005***    |
| (0.002)            | (0.001)     |

For the interactions up to the time index 968, the general model (model 3) reveals that, when Bearing 2 and Bearing 3 are not in operation up to time index of 968, the average vibration of Bearing 1 is 4.00 Hz. Statistically, the vibration of Bearing 1 increases significantly by 0.165 Hz or 0.099 Hz per unit increase in the vibrations of Bearing 2 or Bearing 3 respectively. The interaction model (model 2) also reveals that, the vibration of Bearing 1 is 5.61 Hz when the Bearing 2 and Bearing 3 are not in operation. Statistically, the vibration of Bearing 1 increases by 0.002 Hz or 0.021 Hz per unit increase in the vibrations of Bearing 2 or Bearing 3. These values are small that they can be assumed statistically to be 0. Although the interaction effect is statistically significant, it is too small and can be ignored to all intents and purposes. These small values could also be due to the small magnitude of the value of the vibrations from the onset of the operation of the bearings in the machinery up to the onset of the bearing health deterioration detected. The output also reveals a statistically significant Bearing 2- Bearing 3 interaction effect which causes the vibration of Bearing 1 to increase by 0.006 Hz per unit increase in Bearing 2- Bearing 3 interaction. Thus, there appear to be evidence of moderation by the operations of Bearing 2 to the effect of the operations of Bearing 3 on that of Bearing 3.
unit increase in the vibrations of Bearing 2 or Bearing 3. The interaction model (model 4) reveals that when Bearing 2 and Bearing 3 are not in operation, the vibration of Bearing 1 is 6.12 Hz. This decreases by 0.008 Hz or increases by a similar magnitude per unit increase in the vibrations of Bearing 2 or Bearing 3 respectively. However, these values are not statistically significant. These values are too small and can be ignored to all intents and purposes. The magnitude of the values could also be due to the small magnitude of the vibrations of the bearings. However, there is a significant Bearing 2 - Bearing 3 interaction effect which causes the vibration of Bearing 1 to increase by 0.006 Hz per unit increase in the Bearing 2 – Bearing 3 interaction vibration. This led to the suspicion that there is some moderation by the operations of Bearing 2 to the effect of the operations of Bearing 3 on that of Bearing 1.

The summary output in Table 3 also shows low values of $R^2$ and adjusted $R^2$ for the interaction models (model 2 and model 4) of the vibration from the interaction of the operation of the bearings up to the time indices 837 and 968. The is suspected to have resulted from suppression of the bearing vibration frequencies by nonlinear characteristic features due to the small magnitude of the bearing vibration frequencies extracted during the data condensation stage (Shirong et al., 2014). However, the statistically significant $p$-values means that the vibrations from the operations of Bearing 2 and Bearing 3 made some contributions to the vibration observe in Bearing 1. The $R^2$ and adjusted $R^2$ of the interaction models are also comparatively higher than that of the general models (model 1 and model 2). This also suggests that more of the variables explains the outcome of the interaction models than that of the general models. Additionally, the interaction models have lower residual standard errors than the general models.

The model matrices (Table 4) reveals that the AIC and the BIC of the interaction models have comparatively lower significant $p$-values than those of the general models.

### Table 4: Model Matrices

| Model | adj.r.squared | sigma | AIC    | RMSE  | MAE   | BIC    | p.value |
|-------|---------------|-------|--------|-------|-------|--------|---------|
| 1     | 0.243         | 3.81  | 4618.  | 3.800066 | 2.960519 | 4637.  | 1.43e-51 |
| 2     | 0.260         | 3.76  | 4600.  | 3.755169 | 2.918988 | 4624.  | 9.76e-55 |
| 3     | 0.191         | 4.06  | 5464.  | 3.807111 | 2.97546  | 5484.  | 1.70e-45 |
| 4     | 0.207         | 4.02  | 5446.  | 3.761114 | 2.935438 | 5470.  | 8.95e-49 |
Investigation of the significant interactions with ANOVA type II, confirms that statistically significant interactions observed in the linear regression models (Table 5). Thus, there appears to be some contribution from the operations of Bearing 3 to the health deterioration of Bearing 1. This contribution is influenced by moderations from the operations of Bearing 2.

Table 5: Output of ANOVA type II test table for the interaction models

| Analysis of Variance Table |
|----------------------------|
| Model 2: BPFO.B1 ~ BPFO.B3 * BPFO.B2 |
| Model 4: BPFO.B1 ~ BPFO.B3 * BPFO.B2 |
| Res.Df | RSS  | Df Sum of Sq | F     | Pr(> F) |
| 2   | 833 11803 | 1 | 283.91 | 20.03 8.65e-06 *** |
| 4   | 964 15567 | 1 | 329.25 | 20.39 7.09e-06 *** |

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

We therefore infer:

a) the health deterioration detected in the data of Bearing 1 occurred through its independent operation and that of the operations of Bearing 2 and Bearing 3,

b) the operation of Bearing 3 made the largest contribution to the health deterioration detected,

c) the contribution of the operations of Bearing 3 to the health deterioration of Bearing 1 is moderated by some contributions from Bearing 2,

d) the contribution of the operations of Bearing 3 to the health deterioration of Bearing 1 is also influenced by a Bearing 2 – Bearing 3 interaction effect,

e) there is no contribution from the operations of Bearing 4 to the health deterioration of Bearing 1 within operation of the process machinery.

To help select appropriate model fit of the regression analysis, we produce a visual representation of predicted values of the outcome for given values of data from the operation of Bearing 2 and Bearing 3 in the regression models for the interactions up to the time indices 837 and 968 of the deteriorations detected in the data for Bearing 1 as effect plots (Fig.10). The plots show that at 95% confidence interval, the model fits better when the moderation from the operation of Bearing 2 approaches the one standard deviation above the mean, than the mean and one standard deviation below the mean.
Fig. 10: Effect plots for interactions up to time indices (837 and 968) of the health deterioration detected in the data of Bearing 1.

We further investigate the significant interactions with Johnson Neyman plots (Fig. 11) to help describe the relationship between the operation of Bearing 1, and low, medium, and high operation values of Bearing 2 and Bearing 3. The plots reveal that the slope of Bearing 3 is statistically significant when the range values of the vibration from the operations of Bearing 2 is greater than 3.09 Hz for the interaction up to time index 837, or 4.50 Hz for the interactions up to time index 968. However, the vibration from the operations of Bearing 2 have no effect on the contributions from the vibration from the operations of Bearing 3 to the health deterioration of Bearing 1 detected, except when the vibration from the operations of Bearing 2 are higher than 3.09 Hz up to the time index of 837, or 4.50 Hz up to the time index of 968.
Data from the simple slope analysis (Fig. 12) reveals that the changes in the conditional slope of the data from the vibration from the operations of Bearing 3 increases by 0.14 Hz per unit increase in the vibration from the operations of Bearing 2. The conditional intercept also reveals that while the slope associated with the data from the vibrations of Bearing 3 increase when the data from the vibration from the operations of Bearing 2 increases, the conditional intercept also increases when the vibration from the operations of Bearing 2 increases. This suggests that any increase in the vibration from the operations of Bearing 3 for high vibration from the operations of Bearing 2 will turn towards being equal to the vibration from the operations of Bearing 1. Thus, there is evidence of contribution from the operations of Bearing 2 and Bearing 3 to the health deterioration of Bearing 1 through interaction of their operations up to the time indices 837 and 968. The health deterioration of Bearing 1 is therefore not just due to its independent operation but also the contribution from the operations of Bearing 2 and Bearing 3, and a combined effect of Bearing 2-Bearing 3 interaction.
In the absence of other methods for detecting the time for which the bearing health deterioration is eminent, we compared the performance of our hybrid BDQRA method with RUL predictions in previous research (Widodo & Yang, 2011; Wu, Li & Qiu, 2017). We perform this comparison because while the BDQRA method determines the point at which deterioration of the bearing health is imminent, RUL determines the duration from the point of the deterioration to the end of the bearing’s life. Thus, by subtracting the RUL from the data, we should obtain the time of imminent health deterioration of the bearing. To achieve this, we first calculate the proportional difference in time from the number of observations in each test file and the sampling time. From the sampling rate of 20 kHz/sec, as described in the meta data accompanying the dataset, each data file must contain 20,000 observations. However, 20480 observations were found in each data which means the sampling time is 1.024 secs. We calculate the proportional difference in time using the formula:
Proportional difference \((t_{sec})\) = \(\frac{\text{Observations}_{\text{found}} - \text{Observations}_{\text{expected}}}{\text{Observations}_{\text{expected}}}\) \hspace{1cm} (1)

This gave a difference of 23.616 seconds for the entire duration of the bearing’s lifespan. The actual RUL was provided as 55.3 sec by Widodo & Yang, 2011 and Wu, Li & Qiu, 2017. From this, we calculate the accuracy of our prediction using the formula

\[
\text{Expected duration} = \text{Files}_{\text{total}} + \text{Time}_{\text{extra}} - \text{RUL}_{\text{actual}}
\] \hspace{1cm} (2)

This gave expected time for the risk event as 952.316 seconds. We then calculated the accuracy of our prediction using the formula

\[
\text{Prediction} \%(\%) = 100 - \left( \frac{\text{Detected} - \text{Actual}}{\text{Actual}} \right)
\] \hspace{1cm} (3)

Where \text{Detected} is the time of the bearing health deterioration detected by the BDQRA method and \text{Actual} is the expected time of the bearing health deterioration. This gave a prediction of 99.98\% which indicates a good performance of the BDQRA method in predicting the time of the health deterioration of the bearing.

7. Conclusion

Accurate predictions of the point of imminent health deterioration of a component of a bearing in a process machinery is crucial for health monitoring of the machinery and helps prevent conditions which could lead to fatal incidents. Most literature concentrating on health monitoring employs RUL prediction approach, that is to predict the remaining useful life of the bearing. This paper applied the BDQRA method which was developed and tested with historical data for quantitative risk analysis but can be used to predict health issues of bearing components of a machinery. The method can also be used to analyse real-time data.

The advantage of this method is that it is a hybrid method which uses two change-point analysis techniques to detect the time at which imminent bearing health deterioration could occur. As a hybrid method, it incorporates a collection of big data method techniques including time series analysis, change-point analysis, RMS, FFT, regression decision tree analysis, regression analysis ANOVA and interaction effect. Each of these methods could be applied on its own to determine bearing health issues from bearing vibration data including bearing health deterioration.
We were inspired by a different approach for predicting bearing health deterioration and using our BDQRA method would help in future research. Owing to this, we verify our findings by comparing them with findings of other bearing health indication methods which have used the data from this research and published in peer reviewed journals. The results reveal that our hybrid BDQRA method is very effective and has some advantage over some of the published methods and can be applied as effective tools for practical industrial application.

References
Ahmad, M.W., Reynolds, J. and Rezgui, Y. (2018) Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. Journal of cleaner production, 203: 810-821.

Basseville, M. and Nikiforov, I.V., (1993) Detection of abrupt changes: theory and application (Vol. 104). Englewood Cliffs: Prentice hall.

Benkedjouh, T., Medjaher, K., Zerhouni, N. and Rechak, S. (2013) Remaining Useful Life Estimation Based on Nonlinear Feature Reduction and Support Vector Regression, Engineering Applications of Artificial Intelligence, 26(7): 1751–1760.

Cai, G., Wang, L., Yang, D., Sun, Z. and Wang, B., (2019) Harmonic detection for power grids using adaptive variational mode decomposition. Energies, 12(2): 232.

Caesarendra, W. and Tjahjowidodo, T. (2017) A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing. *Machines*:2-28.

Cerrada, M, Sanchez, R-V, Li, C, Pacheco, F., Cabrera, D., de Oliveira, J.V. and Vásquez, R.E (2018) A review on data-driven fault severity assessment in rolling bearings, Mechanical Systems and Signal Processing 99: 169-196.

CSB (2009) Investigation Report: Sugar dust Explosion and Fire, Imperial Sugar Company, Port Wentworth, Georgia, CSB Investigation Report No. 2008-051-GA, http://www.csb.gov/assets/1/19/Imperial_Sugar_Report_Final_updated.pdf, Last Access Date: 8th September 2017.

Dalvand, F., Shokohi, S., Kalantar, A. and Bevrani, H. (2014) Time-Domain Bearing Condition Monitoring in Induction Motors Using Instantaneous Frequency of Motor Voltage Conference: Smart Grid Conference, AtIran,Tehran.
de Azevedo, H. D. M., Araújo, A. M. and Bouchonneau, N. (2016) A Review of Wind Turbine Bearing Condition Monitoring: State of the Art and Challenges, Renewable and Sustainable Energy Reviews, 56(C): 368 - 379.

De Cock, M., Dowsley, R., Horst, C., Katti, R., Nascimento, A.C., Poon, W.S. and Truex, S. (2017) Efficient and private scoring of decision trees, support vector machines and logistic regression models based on pre-computation. IEEE Transactions on Dependable and Secure Computing, 16(2): 217-230.

Dong H., Hongrui C., Shibin, W., and Xuefeng C. (2020) Time-reassigned synchrosqueezing transform: The algorithm and its applications in mechanical signal processing, Mechanical Systems and Signal Processing 117: 255-279.

Guo, Y., Na, J., Li, B. and Fung R. F. (2014) Envelope Extraction Based Dimension Reduction for Independent Component Analysis in Fault Diagnosis of Rolling Element Bearing. J Sound Vib, 333(13): 2983–94.

Heng, A., Zhang, S., Tan, A. C. C. and Mathew, J. (2009) Rotating Machinery Prognostics: state of the art, challenges and opportunities, Mechanical Systems and Signal Processing 23(3): 724 – 739.

Hsiung-Cheng, L., Yu-Chen, Y., Bo-Jyun, H. and Jia-Lun, S. (2016) Bearing vibration detection and analysis using enhanced fast Fourier transform algorithm, Advances in Mechanical Engineering, 8(10): 1–14.

Jordan, G. (2019) Quantitative Risk Analysis using Real-time Data and Change-point Analysis for Data-informed Risk Prediction. Prof Doc Thesis University of East London School of Architecture, Computing and Engineering https://doi.org/10.15123/uel.87789

Kumar, H. S. (2013) ANN Based Evaluation of Performance of Wavelet Transform for Condition Monitoring of Rolling Element Bearing Conference: ICoNDM-2013, Chennai. Volume:64: 805–14.

Khanam, S., Tandon, N. and Dutt, J. K. (2014) Fault Size Estimation in the Outer Race of Ball Bearing Using Discrete Wavelet Transform of the Vibration Signal. Procedia Technol 2014; 14:12–9. http://dx.doi.org/10.1016/j.protcy.2014.08.003 2nd International Conference on Innovations in Automation and Mechatronics Engineering, ICIAME2014.
Killick, R., Fearnhead, P. and Eckley, I.A., (2012) Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association, 107(500): 1590-1598.

Lee, J., Qiu, H., Yu, G. and Lin, J. (2007). IMS, University of Cincinnati. "Bearing Data Set", NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L. and Siegel, D. (2014) Prognostics and Health Management Design for Rotary Machinery Systems - reviews, methodology and applications, Mechanical Systems and Signal Processing, 42(1-2): 314 – 334.

Li, C., Zhu, X., Li, X., Zhu, Q. and Yu, G. (2016) The Bearing Health Condition Assessment Method Based on Alpha-stable Probability Distribution Model, 2016 Prognostics and System Health Management Conference (PHM-Chengdu), Chengdu, 2016: 1-5.

Liu, T., Lee, T., Singh, P., and Liu, G. (2013) Using Acceleration Measurements and Neuro-fuzzy Systems for Monitoring and Diagnosis of Bearings, Proc SPIE8916,2013, http://dx.doi.org/10.1117/12.2035880.10.

Li, W., Huang, Z., Lin, H, and Ding, K. (2013) Envelope Analysis by Wavelet-filter Based Spectral Kurtosis for Bearing Health Monitoring. Conference: Instrumentation and Measurement Technology Conference (I2MTC), IEEEnInternational.2013.

Ming, A. B., Zhangb, W., Qina, Z. Y. and Chua F. L. (2015) Envelope Calculation of the Multicomponent Signal and its Application to the Deterministic Component Cancellation in Bearing Fault Diagnosis. Mech Syst Signal Process 50–51: 70 – 100.

Okoh, C., Roy, R., Mehnen, J. and Redding, L. (2014) Overview of Remaining Useful Life Prediction Techniques in Through-Life Engineering Services. Procedia CIRP 16: 158 – 163.

Palaniswami, S. and Muthiah, K. (2018) Change Point Detection and Trend Analysis of Rainfall and Temperature Series over the Vellar River Basin. Pol Journal of Environmental Studies, 27(4: 1673-1681.

Pekel, E. (2020) Estimation of soil moisture using decision tree regression. Theoretical and Applied Climatology, 139(3): 1111-1119.

Qiu, H., Lee, J., Lin, J. and Yu, G. (2003) Robust Performance Degradation Assessment Methods for Enhanced Rolling Element Bearing Prognostics, Advanced Engineering Informatics, 17(3-4): 127–140.
Ragab, A., Ouali, M-S., Yacout, S. and Osman, H. (2014) Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan–Meier estimation. Journal of Intelligent Manufacturing: 1- 16.

Roulias, D., Loutas, T. and Kostopoulos, V. (2013) A Statistical Feature utilising Wavelet Denoising and Neighblock Method for Improved Condition Monitoring of Rolling Bearings. Italian Assoc Chem Eng 2013. http://dx.doi.org/10.3303/CET1333175.

Sarvajith, M., Shah, B., Kulkarni, S. and Jana, S. (2013) Condition Monitoring of Rolling Element Bearing Using Wavelet Transform and Support Vector Machine Conference: NCCM2013.

Shaojiang, Dong, Shirong Y., Baoping T., Lili C., and Tianhong Luo (2014) Bearing Degradation Process Prediction Based on the Support Vector Machine and Markov Model, Shock and Vibration (2014): 1 – 15.

Witten, I. H., Frank, E., and Hall, M. A. (2011) Data mining: Practical machine learning tools and techniques. Second Edition, Los Altos, CA: Morgan Kaufmann; ISBN: 0-12-088407-0:1-524.

Wu, B., Li, W. and Qiu, M (2017) Remaining Useful Life Prediction of Bearing with Vibration Signals Based on a Novel Indicator. Shock and Vibration 2017:1-10.

Unala, M., Onatb, M., Demetgulc, M. and Kucukc, H. (2014) Fault Diagnosis of Rolling Bearings using a Genetic Algorithm Optimized Neural Network. Measurement, 58:187–96.

Ziani, R., Zegadi, R., Felkaoui, A. and Djouada, M. (2012) Bearing Fault Diagnosis using Neural Network and Genetic Algorithms with the Trace Criterion, In Condition Monitoring of Machinery in Non-Stationary Operations: 89 - 96).