Data Depth-based Non-parametric Control Chart for Condition Monitoring of Rolling Element Bearings

I-Yen Wua, David Shan-Hill Wonga, Yu-Jeng Lina, Jia-Lin Kangb\*, Yuan Yaoa\*

a Department of Chemical Engineering, National Tsing Hua University, Hsinchu, 30013, Taiwan, ROC

b Department of Chemical and Materials Engineering, National Yunlin University of Science and Technology, Yunlin, 64002, Taiwan, ROC

\*Corresponding Authors: jlkang@yuntech.edu.tw (Kang); [yyao@mx.nthu.edu.tw](mailto:yyao@mx.nthu.edu.tw) (Yao)

Abstract

Rolling element bearings play a pivotal role in numerous industrial systems. Therefore, monitoring their condition is essential for ensuring safety and averting equipment failures. In this study, a non-parametric control chart rooted in data depth for the conditional monitoring of bearings was proposed, which transforms the multivariate bearing information into a univariate index by leveraging data sorting and the central tendency of the data cloud. We designate the well-behaved data as located at the centre of this data cloud. Subsequently, a statistic referred to as “rank” is computed for the operational data of the investigated bearing. Based on this rank calculation, we establish a health indicator for the rolling bearings. We have demonstrated the application and feasibility of this method through a case study, successfully using the indicators we created to assess the current condition of the bearings.

**Keywords**: data depth, condition monitoring, rolling element bearing, control chart

* 1. Introduction

Given their critical role in various industrial systems, it is essential to monitor the condition of rolling element bearings to ensure safety and prevent equipment breakdowns(Lei, Li et al. 2018). The Shock Pulse Method (SPM)(Zhang, Zhao et al. 2014) serves as a valuable diagnostic tool for rapidly assessing the operational state of these bearings. SPM extracts data from high-frequency vibrations to obtain two primary vibration features, namely LR (Low-Frequency Resonance) and HR (High-Frequency Resonance) which served as the basis for assessing equipment health. However, this health assessment lacks quantitative scaling and requires manual definition to confirm the current health status. In the analysis of bearings conditional monitoring (CM), various vibration characteristics are typically examined. However, it can be challenging to observe equipment health status with multiple variable indicators. Therefore, a reasonable and an intuitive metric should be provided as an indicator of equipment condition.

Principal Component Analysis (PCA) is a widely used tool for multivariate monitoring, including the condition monitoring of rolling element bearings (Ahsan, Mashuri et al. 2018). However, its reliance on the Gaussian distribution assumption often poses a significant challenge in practical scenarios, particularly when the data deviates from this assumed distribution, which is a common occurrence in the condition monitoring (CM) of rolling element bearings. This can lead to misleading results from control charts. To address this, we have developed a nonparametric control chart based on data depth, specifically Tukey depth, which does not require any assumptions about data distributions. This method allows for the creation of a health indicator that accurately quantifies the condition of rolling bearings over time. The effectiveness of this approach is demonstrated through a case study.

* 1. Methodology
     1. Data depth

The concept of data depth was first introduced by Tukey et al. (Tukey 1975), highlighting its significance in sorting and analyzing multivariate data. Data depth measures the centrality of multivariate data points in relation to the overall multivariate sample. It adheres to four key properties: 1. affine invariance, 2. maximum value at the center, 3. monotonicity around the deepest point, and 4. vanishing at infinity. Generally, data depth assumes that the data in statistical control follows a *p*-dimensional distribution function *G* (referred to as the reference distribution). When *G* is unknown, an empirical distribution is used, based on the reference sample set {,,…,} . The reference distribution and sample represent the process in control, i.e. under normal operating conditions.

Various methods exist for calculating data depth, including Mahalanobis depth, Simplicial depth, and Tukey depth. Tukey depth, chosen for its higher breakdown point, measures data points in relation to the centre of the data distribution and considers the distribution across each dimension. The Tukey depth of a data point *x* under distribution *F*(·) is defined as follows (Yeh and Singh 1997, She, Tang et al. 2021):

|  |  |
| --- | --- |
| *TD*(*F*,*x*) = | (1) |

which represents the minimum probability mass carried by any closed halfspace containing the point. The sample version of *TD*(*F*,*x*) is defined by replacing F with , the empirical cumulative distribution function. In the univariate case, the formula simplifies to *TD*(*F*,*x*) = min{*F*(*x*), 1-*F*(*x*)}. In this study, the qcr package (Flores, Fernández-Casal et al. 2021) is used to calculate the multivariate Tukey depth. A high Tukey depth value implies that the data point is closer to the centre of the data cloud. Conversely, a lower value indicates that the point is more peripheral or potentially an outlier in the dataset.

* + 1. r-chart

A distribution-free control chart, as described by Bae et al. (2016), is used to monitor the Tukey depth of each data point. In this study, the r-chart, introduced by Liu in 1995 (Liu 1995), is adopted, which is defined mathematically as:

|  |  |
| --- | --- |
|  | (2) |

In this formula, *x* denotes operational data, *y* represents the reference or “golden” data, and *y*∼*G* indicates that *y* follows the distribution *G*. is the data depth, calculated based on (1). If *G* is unknown, the reference sample {,...,} is used instead. The monitoring statistic, or the *r* value, is then calculated as:

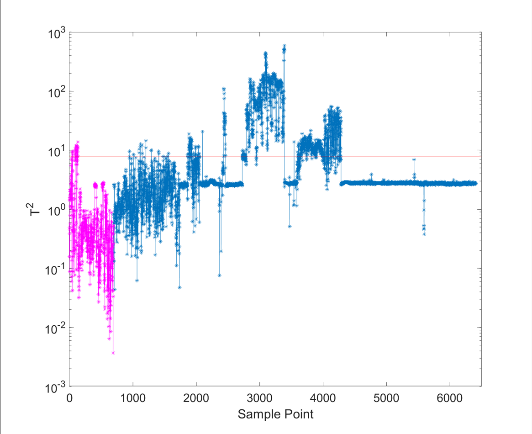
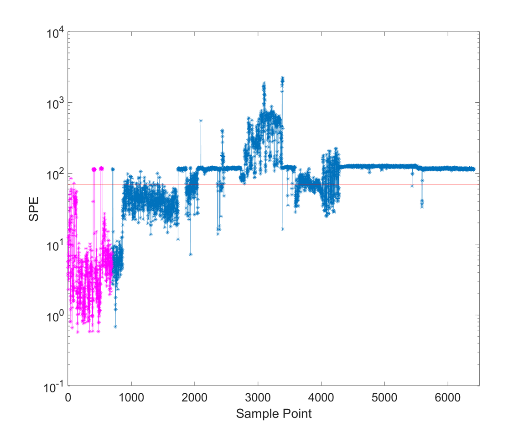
|  |  |
| --- | --- |
|  | (3) |

where represents the count function. The value assesses how closely vector *x* aligns with the centre of the data cloud formed by the reference sample, by comparing the depth values of *x* and each point in *y*. A high probability of *x* having a greater depth value than those in *y* suggests that *x* is likely near the data cloud’s centre, resulting in a higher *r* value and indicating a normal operation. Conversely, a lower probability implies *x* is likely on the data cloud’s edge, leading to a smaller *r* value and signifying out-of-control data. In the control chart, key thresholds include a centre line of 0.5 and a lower control limit set at *α*. A statistic below alpha signals an out-of-control process.

* 1. Case study

This study utilized real-world data from a rolling bearing mounted on an industrial machine. The data was gathered after applying the SPM technique, resulting in multiple variables relevant for the subsequent CM phase. In consultation with equipment experts, a period of “golden data”, representing normal operating conditions, was selected to construct control charts. Data collected after this period served as the test set.

Prior to implementing the data depth-based control chart, PCA was used for an initial comparison. The outcomes are illustrated in Fig. 1: Fig. 1(a) displays the Hotelling’s *T*2 control chart, and Fig. 1(b) shows the *SPE* chart. In these charts, the "golden data" is represented in pink, while other data points are in blue. Notably, a significant number of false alarms are observable during the “golden data” period, attributed to the non-Gaussian nature of the data distribution. Furthermore, a consistent pattern of alarms shortly after the golden data period is seen, contradicting the actual operating conditions of the rolling bearing. Neither the *T*2 chart nor the *SPE* chart effectively captures a progressive deterioration trend in the bearing’s data.

(a) (b)

Figure 1. PCA monitoring results

Subsequently, the proposed method was implemented in the following manner.

1. Reference Sample Selection: The data collected when the bearings are functioning normally and have not experienced failure is identified as “golden data”, serving as the reference sample.

2. Tukey Depth Calculation: The Tukey depth for each data point in the dataset was then calculated.

3. r-chart Construction for Condition Monitoring: An r-chart was subsequently constructed for CM. A process is deemed out of control if the *r* value falls below the threshold of *α* = 0.05.

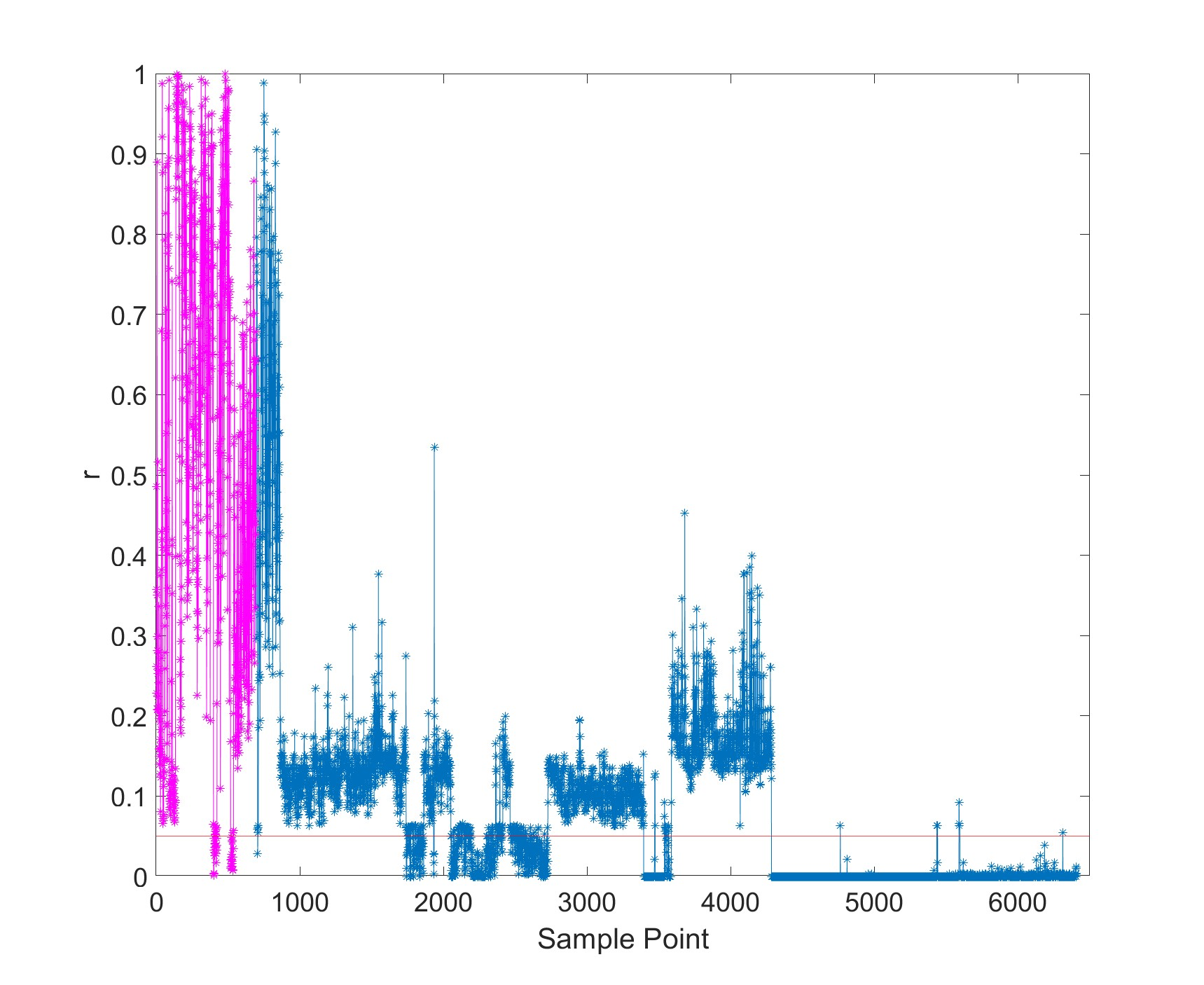
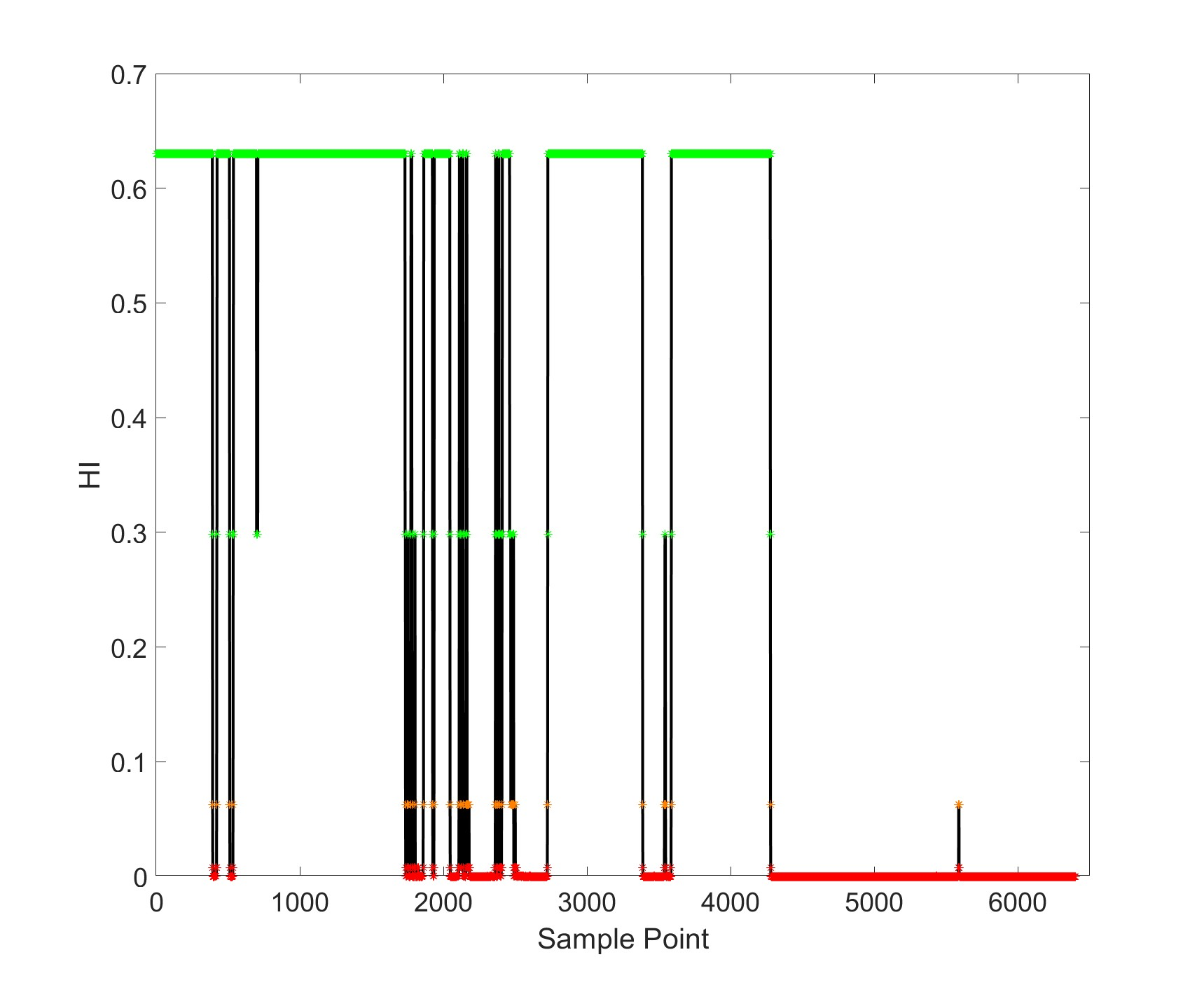
4. Minimizing False Alarms with a Moving Window: To further reduce the incidence of false alarms, a moving window approach was adopted. This window consists of *n* consecutive data points. Based on this window, a health indicator (HI) was formulated using a binomial distribution as shown in equation (4):

|  |  |
| --- | --- |
|  | (4) |

Here, *k* is the count of data points within the window indicating normal operation, and *p* (set at 1 - *α* = 0.95) is the probability of successful operation. *P*(*k*) is utilized as the HI, which is interpreted as follows.

* Green Light (Normal Operation): A *P*(*k*) value greater than 0.1 suggests normal operating conditions.
* Yellow Light (Caution): A *P*(*k*) value between 0.05 and 0.1 acts as a cautionary signal.
* Red Light (Alarm): A *P*(*k*) value below 0.05 indicates an alarm, signaling potential issues in the operation.

The r-chart and HI are displayed in Fig. 2(a) and Fig. 2(b), respectively. It is noticeable that the *r* values of the golden data predominantly converge around an average of 0.5. As time advances and the bearings experience gradual wear, these *r* values correspondingly decline. This contrasts with the PCA results, where data anomalies appear immediately after the golden data phase. The non-parametric approach, grounded in data depth, facilitates earlier detection of bearing health degradation while effectively minimizing false alarms. Employing this method to develop a health indicator empowers proactive measures against imminent bearing faults.

(a) (b)

Figure 2. Data depth-based CM charts

* 1. Conclusions

In this study, we presented a methodology based on data depth analysis to develop a health indicator for rolling bearings. This method proves particularly effective when dealing with data that does not follow a normal distribution, where traditional PCA for CM may result in inaccuracies. By employing data depth calculation in control charts through a non-parametric approach, this method is not limited by the data’s distribution, effectively avoiding the inaccuracies inherent in PCA. The non-parametrically derived health indicator enables earlier detection of bearing health deterioration and significantly reduces false alarms. This innovative approach offers manufacturing facilities a proactive means to foresee and address potential equipment malfunctions.

Acknowledgment

This work was supported in part by the National Science and Technology Council under project number NSTC 112-2221-E-007-105.

References

Ahsan, M., M. Mashuri, H. Kuswanto and D. D. Prastyo (2018). "Intrusion detection system using multivariate control chart Hotelling’s T2 based on PCA." Int. J. Adv. Sci. Eng. Inf. Technol **8**(5): 1905-1911.

Flores, M., R. Fernández-Casal, S. Naya and J. Tarrío-Saavedra (2021). "Statistical quality control with the qcr package." R Journal **13**(1): 194-217.

Lei, Y., N. Li, L. Guo, N. Li, T. Yan and J. Lin (2018). "Machinery health prognostics: A systematic review from data acquisition to RUL prediction." Mechanical systems and signal processing **104**: 799-834.

Liu, R. Y. (1995). "Control charts for multivariate processes." Journal of the American Statistical Association **90**(432): 1380-1387.

She, Y., S. Tang and J. Liu (2021). "On Generalization and Computation of Tukey's Depth: Part I." arXiv preprint arXiv:2112.08475.

Tukey, J. W. (1975). Mathematics and the picturing of data. Proceedings of the International Congress of Mathematicians, Vancouver, 1975.

Yeh, A. B. and K. Singh (1997). "Balanced confidence regions based on Tukey’s depth and the bootstrap." Journal of the Royal Statistical Society Series B: Statistical Methodology **59**(3): 639-652.

Zhang, X., J. Zhao, J. Kang, H. Li and H. Teng (2014). "Bearing prognostics with non-trendable behavior based on shock pulse method and frequency analysis." Journal of Vibroengineering **16**(8): 3963-3976.