Journal of Dynamics, Monitoring and Diagnostics, 2023, 2, 243-251
https://doi.org/10.37965/jdmd.2023.472

ISTP
RESEARCH ARTICLE

Tool Condition Monitoring Based on Nonlinear Output Frequency Response Functions and Multivariate Control Chart

Yufei Gui,[1] Ziqiang Lang,[1] Zepeng Liu,[1] and Hatim Laalej[2]
[1]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S10 2TN, UK
[2]Advanced Manufacturing Research Centre, University of Sheffield, Sheffield S10 2TN, UK

(Received 21 October 2023; Revised 16 November 2023; Accepted 29 November 2023; Published online 29 November 2023)

Abstract: Tool condition monitoring (TCM) is a key technology for intelligent manufacturing. The objective is to monitor the tool operation status and detect tool breakage so that the tool can be changed in time to avoid significant damage to workpieces and reduce manufacturing costs. Recently, an innovative TCM approach based on sensor data modelling and model frequency analysis has been proposed. Different from traditional signal feature-based monitoring, the data from sensors are utilized to build a dynamic process model. Then, the nonlinear output frequency response functions, a concept which extends the linear system frequency response function to the nonlinear case, over the frequency range of the tooth passing frequency of the machining process are extracted to reveal tool health conditions. In order to extend the novel sensor data modelling and model frequency analysis to unsupervised condition monitoring of cutting tools, in the present study, a multivariate control chart is proposed for TCM based on the frequency domain properties of machining processes derived from the innovative sensor data modelling and model frequency analysis. The feature dimension is reduced by principal component analysis first. Then the moving average strategy is exploited to generate monitoring variables and overcome the effects of noises. The milling experiments of titanium alloys are conducted to verify the effectiveness of the proposed approach in detecting excessive flank wear of solid carbide end mills. The results demonstrate the advantages of the new approach over conventional TCM techniques and its potential in industrial applications.

Keywords: intelligent manufacturing; multivariate control chart; Nonlinear Autoregressive with eXogenous Input modelling; Nonlinear Output Frequency Response Functions; tool condition monitoring

I. INTRODUCTION

Tool condition monitoring (TCM) is important in advanced manufacturing as cutting tool anomalies often compromise product quality and machining efficiency [1,2]. With the application of the Internet of Things techniques in industry, many researchers are dedicated to developing indirect TCM methods, which monitor tool conditions by analysing sensor signals such as vibration, force, acoustic, sound, etc [3–5]. A typical indirect TCM system consists of data collection, feature extraction, and tool condition identification [6]. The features that are associated with tool conditions can be obtained in the time domain, frequency domain, and time-frequency domain. However, a signal feature-based method has limited adaptability to variable machining environments. And it is even more complicated to determine what features should be used among numerous candidates.

Recently, Liu *et al.* [7] have proposed a novel TCM approach based on an innovative idea known as sensor data modelling and model frequency analysis. Instead of extracting features from the sensor signals directly, this approach builds nonlinear models that represent a dynamic relationship between two different vibration sensor signals. Then, the model's frequency domain properties are analysed and used as the features for TCM. Analysis shows that compared to conventional signal feature-based TCM, this approach can better adapt to external changes including tool replacement and process parameter variations and have better generalization capability. Condition monitoring based on sensor data modelling and model frequency analysis has been widely investigated in beam crack detection [8], structure health monitoring [9], and rotor system fault diagnosis [10]. The current study presents another practical application of this strategy and demonstrates, for the first time, how to apply this strategy in unsupervised TCM in advanced manufacturing.

In order to implement the proposed TCM system, multivariate statistical process control charts such as, Hotelling's T^2 chart [11] are applied. This method is easy to implement in a manufacturing setting and its performance in TCM based on signal features has been widely reported [5,12–14]. However, as signal features are oversensitive to external noises, it is often challenging to reliably identify anomalies induced by worn cutting tools. The proposed TCM system can resolve this problem as,

1. Instead of sensor data features, physically more meaningful model frequency features are used to build a TCM control chart that can, in principle, more accurately represent the tool wear state.
2. A moving window is introduced to build a subgroup observations-based control chart, which can further reduce the influence of external variations and reflect tool wear trend.

Corresponding author: Zepeng Liu (e-mail: z.lang@sheffield.ac.uk).

Overall, the proposed approach consists of sensor data modelling, model frequency analysis, and a control chart-based worn tool detection. A milling experiment is conducted to validate the performance of the proposed TCM system in detecting the working conditions of three cutters. The results demonstrate the effectiveness of the proposed approach and the advantage of the approach over traditional signal feature-based TCM techniques.

II. METHODOLOGY

The proposed methodology consists of sensor data modelling, model frequency analysis, and process monitoring. The sensor data modelling produces a NARX (Nonlinear AutoRegressive with eXogenous Input) model that represents the dynamic relationship between two different vibration sensor signals. Then, the frequency properties of the NARX model are extracted to represent physically meaningful features of the milling process. It is expected that cutting tool anomalies will lead to changes in process dynamics that can be represented by the NARX model's frequency domain features. Moreover, a multivariate control chart based on the NARX model's frequency features is constructed to monitor the cutting tool status.

A. SENSOR DATA MODELLING

Sensor data modelling here is also known as system identification, which is a data-driven modelling technique aiming to find a dynamic relationship between the input and output of a process or system. Let $u(t)$ and $y(t)$, $t = 1,2,\ldots,\Gamma$, be the vibration signals collected from a machining process by two accelerometers fitted on the spindle and supporting base of the workpiece, respectively. The number of samples is Γ and the sampling rate is f_s. Using the NARX model, the relationship between $u(t)$ and $y(t)$ can be expressed by

$$\begin{aligned} y(t) = F^{\ell}[&y(t-1), y(t-2), \ldots, y(t-\delta_y), \\ &u(t-1), u(t-2), \ldots, u(t-\delta_u)] + e(t) \end{aligned} \tag{1}$$

where $F^{\ell}[\cdot]$ is the polynomial function with the maximum polynomial degree $\ell \in \mathbb{Z}^+$, and $e(t)$ denotes the noise and unmodelled dynamics. Moreover, δ_u and δ_y are the maximum lags for the input $u(t)$ and output $y(t)$, respectively.

The NARX model is essentially a linear-in-the-parameter model that can be written as

$$y(t) = \sum_{m=1}^{M} \theta_m d_m(t) + e(t) \tag{2}$$

where θ_m is the parameter associated with the model term $d_m(t)$ and M indicates the total number of candidate terms. These terms are monomials composed of $y(t-1)$, $y(t-2),\ldots,y(t-\delta_y)$, $u(t-1), u(t-2),\ldots,u(t-\delta_u)$ such as $y(t-1)u(t-1)$ and $y^3(t-1)$.

Assume there are Γ samples in total, Eq. (2) can be further represented in the following matrix format:

$$\mathbf{y} = \mathbf{D}\Theta + \Xi \tag{3}$$

where $\mathbf{y} = [y(1),\ldots,y(\Gamma)]^{\mathrm{T}}$ is a model output vector, $\mathbf{\Theta} = [\theta_1,\ldots,\theta_M]^{\mathrm{T}}$ is a model parameter vector, and $\mathbf{\Xi} = [e(1),\ldots,e(\Gamma)]^{\mathrm{T}}$ is the error sequence. The model input matrix, $\mathbf{D} = [\mathbf{d}_1,\ldots,\mathbf{d}_M]$, also known as a dictionary matrix, has M column vectors with $\mathbf{d}_m = [d_m(1),\ldots, d_m(\Gamma)]^{\mathrm{T}}$.

As the matrix $\mathbf{D}$ usually contains redundant terms, the term selection is a key step in NARX model-based nonlinear system identification [15]. In this study, we use the Forward Regression with Orthogonal Least Squares (FROLS) algorithm to select the most important model terms from the dictionary matrix $\mathbf{D}$ [9]. At each step, the term with the strongest capability to represent the output $\mathbf{y}$ is selected. Let the selected terms consist of $\mathbf{W} = [\mathbf{d}_1,\ldots,\mathbf{d}_{M_0}]$ (generally $M_0 \ll M$), the final model becomes

$$\mathbf{y} = \mathbf{W}\alpha + \Xi \tag{4}$$

where $\boldsymbol{\alpha} = [\theta_1,\ldots,\theta_{M_0}]^{\mathrm{T}}$ is a FROLS parameter vector. In this study, $\boldsymbol{\alpha}$ is calculated by solving the following l_2-norm regularization problem.

$$\begin{aligned} \boldsymbol{\alpha} &= \arg\min_{\boldsymbol{\alpha}} \left\{ \|\mathbf{W}\alpha - \mathbf{y}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_2 \right\} \\ &= (\mathbf{W}^{\mathrm{T}}\mathbf{W} + \lambda \mathbf{I})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{y} \end{aligned} \tag{5}$$

where λ is the penalty parameter and $\mathbf{I}$ is an $M_0 \times M_0$ identity matrix. The penalty parameter controls the trade-off between bias and variance in the estimator. An evolutionary algorithm is exploited to tune λ such that the built model meets the stability, robustness, and accuracy requirements. Details about the parameter tuning procedure can be found in [7].

B. MODEL FREQUENCY ANALYSIS

After a NARX model has been identified, the NOFRFs (Nonlinear Output Frequency Response Functions) of the built NARX model can be extracted to investigate the frequency behaviours of the system [8]. Because the choice of λ guarantees the identified model is stable at zero equilibrium, the system can be described by the Volterra series in the discrete-time domain, see [9]. The corresponding frequency domain representation can be written as [16]

$$\mathcal{F}(y(t)) = Y(j\omega) \approx \sum_{n=1}^{N} Y_n(j\omega) = \sum_{n=1}^{N} G_n(j\omega) U_n(j\omega) \tag{6}$$

where $\mathcal{F}(\cdot)$ represents Fourier transform and ω is the frequency variable. $Y_n(j\omega) = \mathcal{F}(y_n(t))$ and $U_n(j\omega) = \mathcal{F}(u_n(t))$ are the nth order output and input frequency spectrum, respectively. $G_n(j\omega)$ is the nth order NOFRFs, which allows the system nth order output frequency response $Y_n(j\omega)$ to be described in a manner similar to the description for the output frequency response of linear systems.

In this study, a recently proposed Generalized Associated Linear Equations (GALEs) method is employed to calculate NOFRFs [17]. This method decomposes the NARX model into a series of linear difference equations such that the NOFRFs can be evaluated from the first order to an arbitrarily high order. Consider the general form of the polynomial NARX model [18]

$$\begin{aligned} y(t) = &\sum_{j=1}^{J} \sum_{p=0}^{j} \sum_{l_1,\ldots,l_{p+q}=1}^{L} c_{p,q}(l_1,\ldots,l_{p+q}) \\ &\times \prod_{i=1}^{p} y(t-l_i) \prod_{i=p+1}^{p+q} u(t-l_i) \end{aligned} \tag{7}$$

where $J, L \in \mathbb{Z}^+$, $p+q = j$, and $c_{p,q}(l_1,\ldots,l_{p+q})$ represents the model coefficient.

The GALEs of the NARX model are defined as

$$\begin{aligned} y_n(t) &= \sum_{l_1=1}^{L} c_{1,0}(l_1) y_n(t-l_1) \\ &+ \sum_{l_1,l_n=1}^{L} c_{0,n}(l_1,.,l_n) \times \prod_{i=1}^{n} u(t-l_i) \\ &+ \sum_{q=1}^{n-1} \sum_{p=1}^{n-q} \sum_{t_1,t_{p+q}=1}^{L} c_{p,q}(l_1,\ldots,l_{p+q}) y^{\mathbf{L}}_{n-q,p}(t) \\ &\times \prod_{i=p+1}^{p+q} u(t-l_i) + \sum_{p=2}^{n} \sum_{l_1,l_p=1}^{L} c_{p,0}(l_1,\ldots,l_p) y^{\mathbf{L}}_{n,p}(t) \end{aligned} \tag{8}$$

where $n = 1,\ldots,N$, $\mathbf{L} = (l_1,\ldots,l_n)$ and

$$\begin{cases} y^{\mathbf{L}}_{n,p}(t) = \sum_{i=1}^{n-(p-1)} y_i(t-l_p) y^{\mathbf{L}}_{n-i,p-1}(t) \\ y^{\mathbf{L}}_{n,1}(t) = y_n(t-l_1) \end{cases} \tag{9}$$

With the GALEs obtained, the NOFRFs can be calculated by evaluating the nth order system output response $y^*_n(t)$ to a specified input signal $u^*(t)$. Then, the nth order NOFRFs of the system, $G^*_n(j\bar{\omega})$, under the input excitation $u^*(t)$ can be calculated as

$$\begin{aligned} G^*_n(j\bar{\omega}) &= \frac{\mathscr{F}[y^*_n(t)]}{\mathscr{F}[u^*_n(t)]} = \frac{\mathscr{F}[y^*_n(t)]}{\mathscr{F}\{[u^*(t)]^n\}} \\ &= \frac{Y^*_n(j\bar{\omega})}{U^*_n(j\bar{\omega})}, \quad \begin{cases} n = 1,\ldots,N \\ \bar{\omega} \in \bar{\Omega}_n \end{cases} \end{aligned} \tag{10}$$

where $\bar{\Omega}_n$ indicates the frequency range of $U^*_n(j\bar{\omega})$.

In this study, the data collected from different tool conditions are used to build corresponding NARX models. The same input excitation signal is used to evaluate the NOFRFs of these NARX models. It is expected that the changes in tool conditions can be reflected by the evaluated NOFRFs. As a result, the NOFRFs G^*_n, $n=1,2,\ldots$ can be used as representative features for TCM.

C. NOFRFs FEATURE DIMENSION REDUCTION

In the resulting NOFRFs $G^*_n(j\bar{\omega})$, $\bar{\omega}$ indicates the frequency variable belonging to the frequency range of $\bar{\Omega}_n$, which can be calculated based on $\bar{\Omega}_1$ as shown in [19]. $G^*_n(j\bar{\omega})$ may have multiple frequency ranges. The present study selects the magnitude of NOFRFs over characteristic frequency range as representative features, i.e., $\mathbf{f}_n = [|G^*_n(j\bar{\omega}^*_1)|,\ldots, |G^*_n(j\bar{\omega}^*_{\bar{K}_n})|]$ where $\bar{\omega}^*_k \in \bar{\Omega}^*_n$ indicates the kth frequency point in the nth order NOFRFs and $k = 1,\ldots,\bar{K}_n$, $\bar{\Omega}^*_n \subseteq \bar{\Omega}_n$. The determination of $\bar{\Omega}^*_n$ usually depends on the physical process of interest. Concatenating the feature vectors extracted from up to Nth order NOFRFs produces a final feature vector $\mathbf{x} = [\mathbf{f}_1,\ldots,\mathbf{f}_N]$, which has a dimension of $1 \times R$, and $R = \bar{K}_1 + \ldots + \bar{K}_N$.

Assuming there are P available data segments in total, a feature vector $\mathbf{x}_p = [\mathbf{f}_{1,p},\ldots,\mathbf{f}_{N,p}]$ can be extracted from the pth data segment. Let the features of the first P_1 segments consist of the training data set. The feature matrix can be written as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{P_1} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_{1,1},\ldots,\mathbf{f}_{N,1} \\ \vdots \\ \mathbf{f}_{1,P_1},\ldots,\mathbf{f}_{N,P_1} \end{bmatrix} \tag{11}$$

where the dimension of $\mathbf{X}$ is $P_1 \times R$.

The multivariate process monitoring scheme assumes that the samples from in-control processes follow a multivariate normal distribution. The process behaviour is reflected by the shift in the mean of these variables. Thus, the control chats' capability of timely detecting mean shifts will decrease if the number of variables is very large [11]. Besides, the existence of multi-collinearity sometimes can lead to numerical instability. It is, therefore, necessary to reduce the dimension of variables and eliminate multi-collinearity before implementing the control chart.

To achieve this, in the present study, the principal component analysis (PCA) is applied to $\mathbf{X}$[20]. Firstly, $\mathbf{X}$ is standardized to obtain $\mathbf{X}'$ with zero-mean and one-standard deviation variables. Then the covariance matrix is computed as $\mathbf{C} = 1/(P_1 - 1)\mathbf{X}'^T\mathbf{X}'$. The eigendecomposition of $\mathbf{C}$ is $\mathbf{C} = \mathbf{P}\Lambda\mathbf{P}^T$, where $\mathbf{P} = [\mathbf{p}_1,\ldots,\mathbf{p}_R]$ consists of all eigenvectors and $\Lambda = diag(\lambda_1,\ldots,\lambda_R)$ is a diagonal matrix with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_R$. This indicates the first component $\mathbf{p}_1$ explains the largest amount of variance in $\mathbf{X}$. The second component $\mathbf{p}_2$ is orthogonal to $\mathbf{p}_1$ and explains the largest amount of the remaining variance, and so on. The first R' (generally $R' \ll R$) components are selected as principal ones when $(\sum_{i=1}^{R'} \lambda_i / \sum_{i=1}^{R} \lambda_i) \times 100\%$ exceeds a pre-set threshold, e.g., 99%. Finally, the feature vector is transformed into a new subspace spanned by the first R' eigenvectors,

$$\mathbf{T} = \mathbf{X}\mathbf{P}' = \mathbf{X}[\mathbf{p}_1,\ldots,\mathbf{p}_{R'}] \in \mathbb{R}^{P_1 \times R'} \tag{12}$$

For any subsequent feature vector $\mathbf{x}_p$, the transformed vector is calculated by $\mathbf{t}_p = \mathbf{x}_p\mathbf{P}'$, with $p = P_1 + 1$, $P_1 + 2$, $\ldots$. These transformed feature vectors are then used to build multivariate control charts.

D. TCM

The multivariate control chart-based condition monitoring involves both training and monitoring stages. The training stage aims to construct a baseline description of in-control processes. The mean vector and covariance matrix are estimated based on "normal" samples. The control limits are determined as well. Then, the subsequent process is monitored. The control limit determined in the training stage is used to judge if any operation condition falls outside these limits, which could indicate an out-of-control process.

Machining is a very complicated process involving variations induced by inconsistent material properties, changing process dynamics, increasing tool wear, and so on. All these factors result in uncertainty in the collected signals, reflected by the extracted features. Hence, a monitoring solution purely relying on these features will inevitably suffer from either false alarms or missing faults. However, the long-term shift of feature vectors is still dominated by the increase of tool wear. A natural way to grasp the trend of a series of data and overcome local variations is to use the moving window strategy.

In this study, we adopt a subgroup observations-based Hotelling's T^2 control chart [11]. The subgroup contains the samples within a moving window. The mean of the samples in each subgroup is used to calculate the statistics of the

control chart such that the influence of external noises is eliminated. To be specific, the mean and covariance matrix of the kth subgroup samples are calculated by

$$\begin{cases} \mathbf{z}_k = \frac{1}{k}\sum_{p=1}^{k} \mathbf{t}_p \\ \mathbf{C}_k = \frac{1}{k}\sum_{p=1}^{k} (\mathbf{t}_p - \mathbf{z}_k)^T(\mathbf{t}_p - \mathbf{z}_k) \end{cases}, \quad 1 \le k \le w$$
$$\begin{cases} \mathbf{z}_k = \frac{1}{w}\sum_{p=k-w+1}^{k} \mathbf{t}_p \\ \mathbf{C}_k = \frac{1}{w}\sum_{p=k-w+1}^{k} (\mathbf{t}_p - \mathbf{z}_k)^T(\mathbf{t}_p - \mathbf{z}_k) \end{cases}, \quad w < k \tag{13}$$

where w denotes the window length. As the training data set contains P_1 samples, the mean and covariance matrix representing in-control processes are given as

$$\bar{\mathbf{z}} = \frac{1}{P_1}\sum_{k=1}^{P_1} \mathbf{z}_k, \quad \bar{\mathbf{C}} = \frac{1}{P_1}\sum_{k=1}^{P_1} \mathbf{C}_k \tag{14}$$

Therefore, the Hotelling's T^2 statistic of $\mathbf{z}_k$ is calculated by

$$T_k^2 = R'(\mathbf{z}_k - \bar{\mathbf{z}})\bar{\mathbf{C}}^{-1}(\mathbf{z}_k - \bar{\mathbf{z}})^T \tag{15}$$

Regarding the determination of control limits, if the assumption of multivariate normality of the measurements is true, the control limits can be determined theoretically as the statistics should follow a standard distribution. However, in practice, it is more reliable to determine the control limit according to the distribution of available statistic values. This study uses kernel density estimation (KDE) to estimate the control limit [5]. This approach treats the T^2 statistic as a random variable and estimates the probability density function $pdf(T^2)$ of T^2 statistic based on kernel smoothing and available samples $\{{T_1}^2, \ldots, {T_{P_1}}^2\}$. Then, the upper control limit (T_{UCL}^2) is determined via the following equation

$$P(T^2 < T_{UCL}^2) = \int_{-\infty}^{T_{UCL}^2} pdf(T^2)\mathrm{d}T^2 = \delta \tag{16}$$

where $\delta \in [0,1]$ and $1-\delta$ is the significance level, indicating the probability of the T^2 statistic falling beyond the upper control limit when the process is "in control."

The control limit is obtained in the training stage. In the monitoring stage, the condition is monitored by comparing T^2 statistic with the limit T_{UCL}^2. Even though the moving window method has been used to cope with local variations, this system cannot fully avoid false alarms. To further improve the reliability of the detection result, in the decision-making step, we set the alarm-triggering condition as the occurrence of several successive samples exceeding the control limit.

Figure 1 shows the flowchart of the proposed TCM system. The whole procedure contains two stages. The feature extraction processes are the same for the training and monitoring stages, including NARX model identification, NOFRFs extraction, and PCA dimension reduction. Then, a moving window is used to create subgroup samples, whose mean vector is used to determine the monitoring T^2statistic. In offline training, the mean, covariance matrix, and control limit are obtained to represent the normal process conditions. In the monitoring stage, the T^2 statistics of new samples are calculated and compared with the

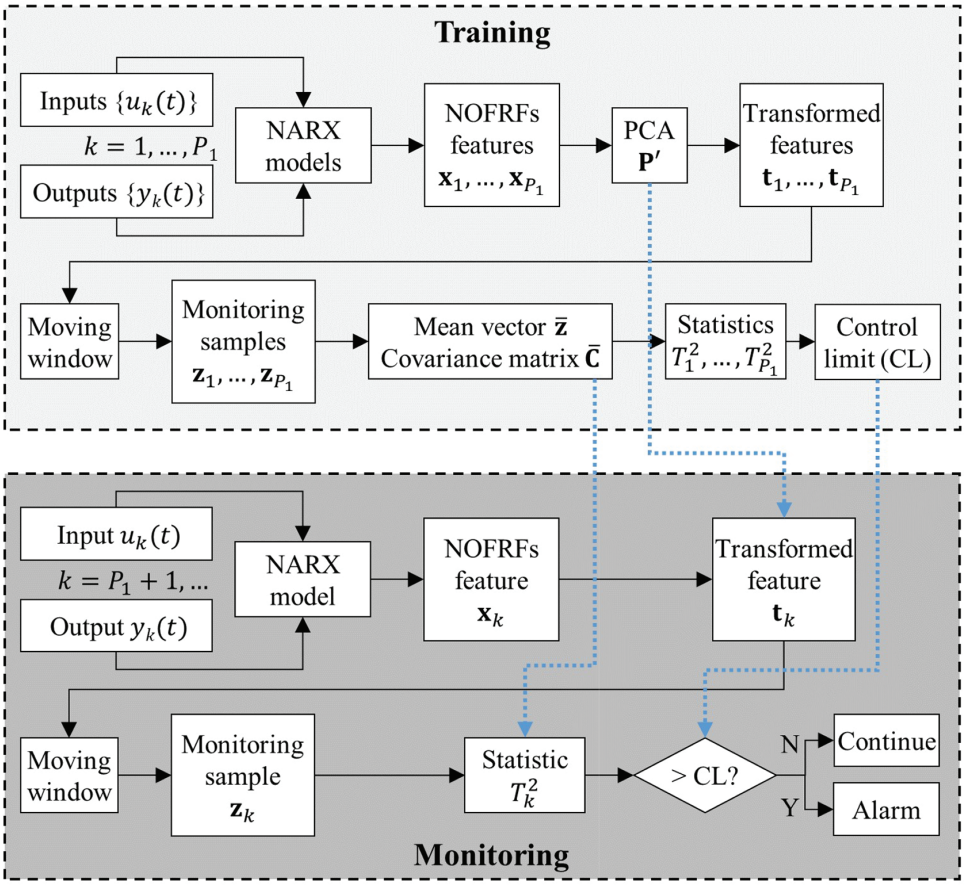

Fig. 1. The flowchart of the proposed TCM system based on NOFRFs and multivariate control chart.

control limit. The system will trigger alarms when several successive samples are identified as out-of-control processes. Otherwise, the machining process continues.

III. EXPERIMENTAL STUDY

To validate the effectiveness of the proposed TCM method in TCM, a run-to-failure experiment has been carried out at the Advanced Manufacturing Research Centre, University of Sheffield. The proposed method is applied to the collected vibration signals. As a comparison, traditional signal features are extracted and used for TCM. The experimental details and results are presented in this section.

A. EXPERIMENTAL SETUP

The dynamic milling strategy was adopted in the experiment. As shown in Fig. 2(a), the machining was run on a 5-axis milling machine (DMG MORI's DMU 40evo). The material of the workpiece is TC4 titanium alloy. The type of cutting tool is Sandvik CoroMill Plura 1630 solid carbide square shoulder end mill with a diameter of 16 mm. This cutting tool has 4 flutes. In the experiment, we used three cutters in total (referred to as T1, T2, and T3). Fig. 2(b) shows the machining process. The milling of one cylinder workpiece was conducted layer by layer. Each layer consists of 6 round cuts (referred to as R1∼R6) from outer to inner. To coincide with the machining condition in a real manufacturing process, the rotational speed was set as 2586.3 rpm, the feed rate was 1055.2 mm/min, and the radial and axial cutting depth were 1.6 mm and 20 mm, respectively.

Two accelerometers were mounted on the spindle and workpiece supporting base to collect vibration signals,

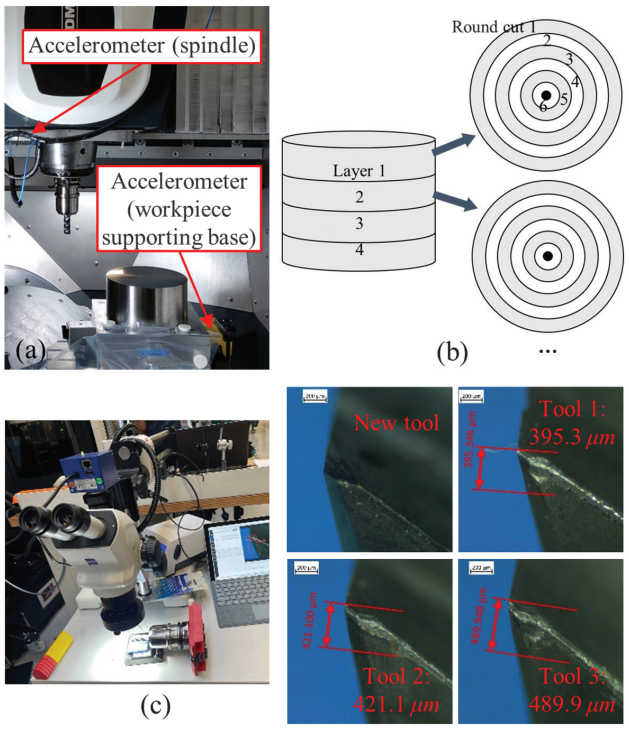

Fig. 2. (a) Experimental setup, (b) illustration of the machining process, (c) measurement of tool wear (right: the images of flutes before and after machining).

Table I. List of sensors and specifications

Sensor	Type	Sensitivity	Frequency range
Accelerometer (spindle)	PCB, 356A02	10 mV/g	1–5000 Hz
Accelerometer (base)	PCB, 356A17	500 mV/g	0.5–3000 Hz

which were used as input and output for sensor data modelling. This setting coincides with the transmission of power from the spindle to the workpiece. Thus, the built models are able to represent the dynamics of milling processes. Table I lists the sensor types and specifications. A relatively high-sensitivity sensor was mounted on the workpiece supporting base so as to record the base vibration better. The signals were acquired by NI CompactDAQ systems with a sampling rate of 51200 Hz.

After each round cut, the tool wear was measured by a microscope as shown on the left of Fig. 2(c). As the experiment was a run-to-failure one, each tool was used to complete one and a half workpieces, i.e., six layers (referred to as L1∼L6). As shown in Fig. 2(c), the maximum flank wear of all tools reached 395.3 μm, 421.1 μm and 489.9 μm finally. In practice, 300 μm is typically used as the threshold for excessive tool wear. During the experiment, the flank wear of all three tools exceeded 300 μm at Layer 5.

B. MONITORING RESULTS

From the signals collected for each round cut, a 1-second snapshot is taken every 5 seconds. After pre-processing, the signal segments are used for sensor data modelling. Considering the tooth passing frequency is 172.41 Hz (= 2586.27/60×4), to investigate the nonlinear characteristics at this frequency, an input excitation with the frequency range from 167 Hz to 177 Hz is designed to evaluate the corresponding NOFRFs of each identified NARX model. The designed input excitation is

$$u^*(t) = \frac{3}{2\pi}\frac{\sin(2\times 177\pi t) - \sin(2\times 167\pi t)}{t} \tag{17}$$

where $-1 \le t \le 1$. Therefore, the features of the NOFRFs under input (15) are determined to evaluate the status of cutting tools. The magnitudes of the first three order NOFRFs $|G_1^*(j\bar{\omega})|$,$|G_2^*(j\bar{\omega})|$,$|G_3^*(j\bar{\omega})|$, covering 41, 81, and 121 frequency points, respectively, are used as the NOFRFs features. The frequency range is determined according to the new frequency generation phenomenon occurring with a nonlinear system [16]. Hence, the dimension of one NOFRFs feature vector is 1×243. Since each tool has completed 36 round cuts and 30 NARX models are identified from each round, the total number of feature vectors over the tool's lifecycle is 1080.

Figure 3 shows all the extracted NOFRFs-based feature vectors of T1, T2, and T3. One line corresponds to one feature vector consisting of $|G_1^*(j\bar{\omega})|$, $|G_2^*(j\bar{\omega})|$, and $|G_3^*(j\bar{\omega})|$. As can be seen, the NOFRFs magnitudes become larger as the order of nonlinearity increases. To compare the difference of NOFRFs features under different tool wear conditions, the feature vectors from L1-2 and L3-6 are drawn separately in Fig. 3. In all three figures, the green lines on the left-hand show the distribution of NOFRFs

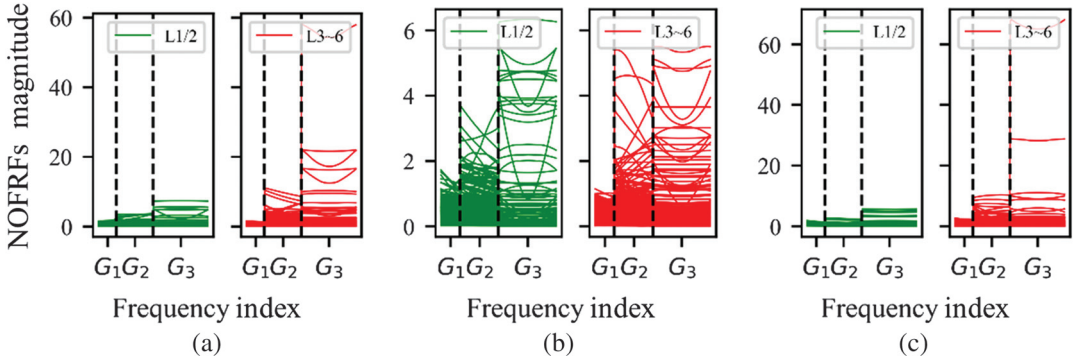

Fig. 3. The magnitudes of the first 3 order NOFRFs versus frequency index, (a) tool 1, (b) tool 2, and (c) tool 3; in each figure, left: feature vectors from Layers 1–2, right: feature vectors from Layers 3–6.

features from L1 and L2 representing slightly wear condition, and the red lines on the right are from L3 to L6 corresponding to severe tool wear. Intuitively, one can find that the NOFRFs magnitudes are larger when the tool gets worn, indicating that the NOFRFs can reflect the change in tool conditions.

TCM was carried out using both NOFRFs-based features and signal features for comparison. In the proposed TCM framework, the data collected from the initial stage when the tool is not severely worn was used for training. As the number of finished layers is 6 during the experiment, it is reasonable to select the data from Layers 1 and 2 for training and the remaining for monitoring. It should be noted that the selection of training samples is more difficult when the lifecycle of one tool is unknown. In this case, a priori knowledge or expert experience is necessary, which is out of the scope of this paper.

For the proposed NOFRFs-based method, in the dimension reduction step, the components whose cumulative variance percentage exceeds 99% are determined as principal ones. To create subgroup samples, the length of the moving window is set to 20. The significance level in KDE-based control limit determination is 0.1%, which means there is a very tiny possibility for the T^2 statistic of an in-control state exceeding the control limit. If 7 successive samples exceed the control limit, the system will trigger alarms.

For the signal features-based method, 8 statistics are extracted from one vibration signal, including average, variance, skewness, kurtosis, entropy, median, range, and crest factor. The dimension reduction is not performed in this case as the dimension of signal feature vectors is relatively small. Except that, the subsequent procedures such as control chart construction and parameter setting are the same as those applied for the NOFRFs feature-based condition monitoring.

Figures 4–6 present the monitoring results for the three cutters. As can be seen, the statistic of NOFRFs is overall stationary in the initial monitoring stage and tends to increase when the tool wear reaches a value of around 300 μm. In other words, the proposed NOFRFs features do not change a lot when the tool is slightly worn, compared with the distribution of features in the training stage. The alarms are triggered at the last several round cuts of Layer 4 when tool wear is between 268 μm and 285 μm. However, when using the signal feature-based control chart, the monitoring statistic is always increasing, making it difficult to identify whether tool wear exceeds the critical value. Hence, the alarms are triggered in a very early stage during monitoring. Especially for T2 and T3, after the first round cut of Layer 3, the tool condition is identified as worn. The result shows that the signal features are too sensitive to the process change, tending to cause more false alarms, and cannot make reliable decisions in practice.

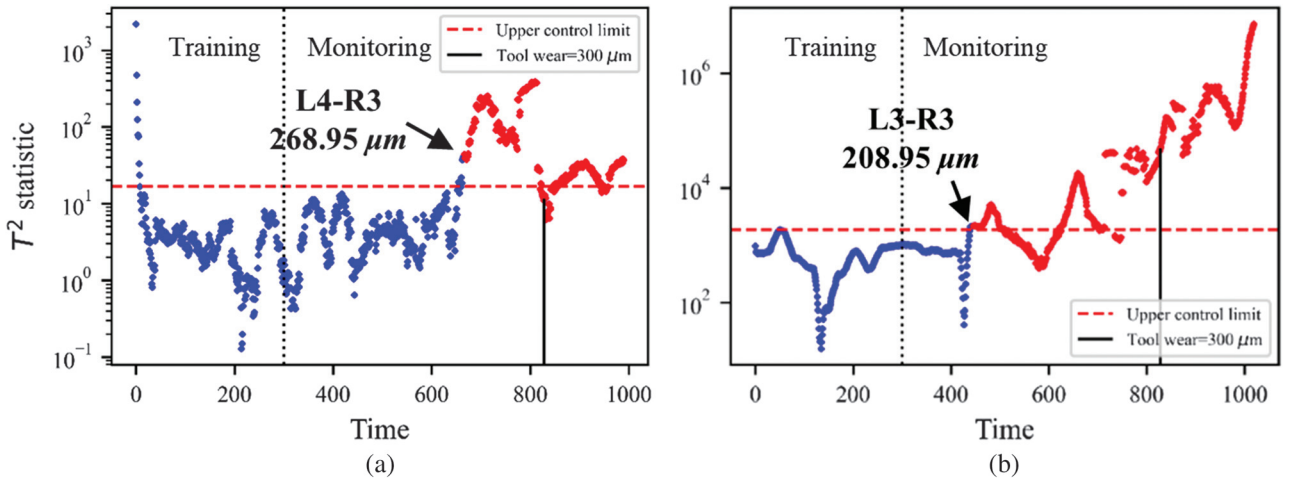

Fig. 4. TCM results for T1 based on (a) NOFRFs-based features and (b) signal features.

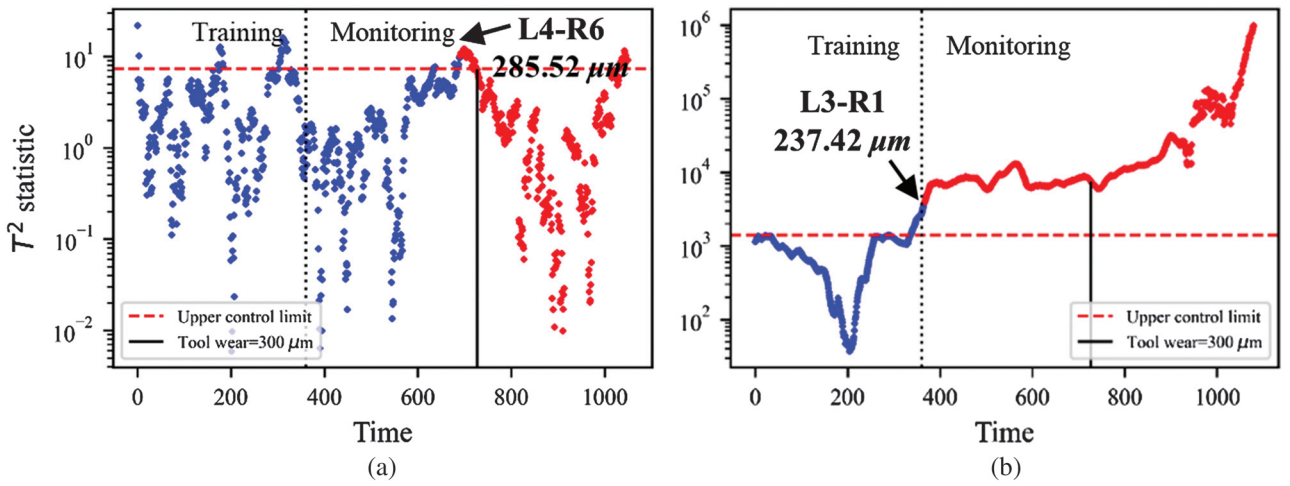

Fig. 5. TCM results for T2 based on (a) NOFRFs-based features and (b) signal features.

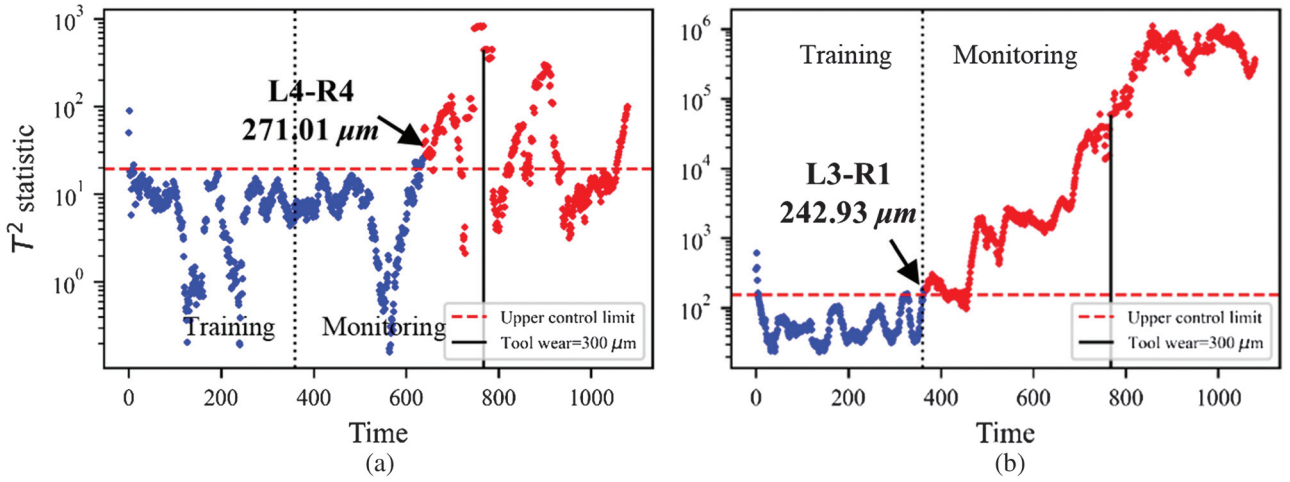

Fig. 6. TCM results for T3 based on (a) NOFRFs-based features and (b) signal features.

To further illustrate the effect of moving window, Fig. 7 shows the control charts based on NOFRFs features with individual samples, which means the transformed feature vectors $\mathbf{t}_p$ are directly used to calculate monitoring statistics. Most of the statistics are randomly scattered within a range. Although there is an increasing trend as tool wear reaches 300 μm, the system does not trigger alarms as there are no consecutive samples exceeding the control limit. The use of a moving window highlights the trend of these statistics, leading to successful detection results. One issue with this method is how to select a reasonable window length. The consideration involves

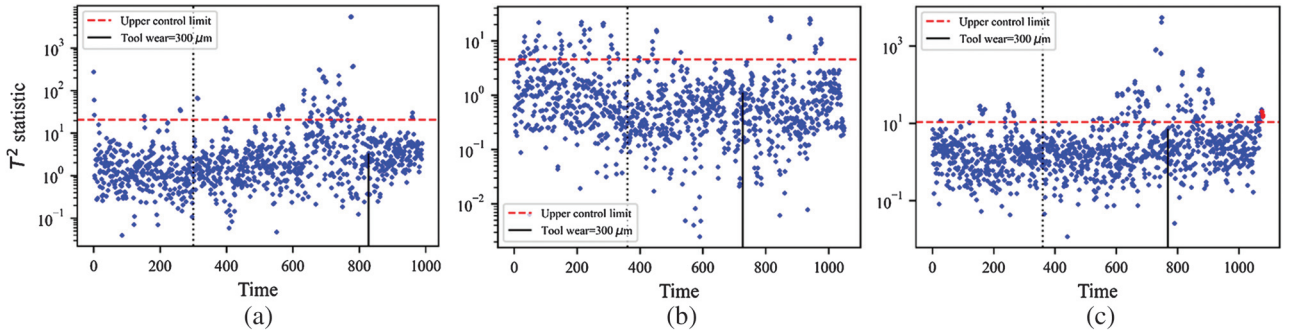

Fig. 7. TCM results using NOFRFs-based features without moving average applied for (a) T1, (b) T2, and (c) T3.

Table II. Tool condition monitoring results using NOFRFs and signal features, respectively

Tool	Tool wear $\geq 300\ \mu m$	NOFRFs-based features	Signal features
T1	L5-R4	L4-R3: 268.95 μm	L3-R3: 208.95 μm
T2	L5-R1	L4-R6: 285.52 μm	L3-R1: 237.42 μm
T3	L5-R2	L4-R4: 271.01 μm	L3-R1: 242.93 μm

the data sampling rate, the pattern of the machining process, and the wearing rate of tools. A further discussion will be given in future studies.

Table II summarizes the TCM results in the experimental study. For all three tools, the detected tool wear is closer to 300 μm when the NOFRFs-based features are applied. Besides, the consistent monitoring results demonstrate that the proposed NOFRFs-based method is able to overcome the impact of varying working environments, such as workpiece and tool material properties and coolant concentration variations. These demonstrate the advantage of the proposed NOFRFs-based TCM over traditional signal feature-based approaches.

IV. CONCLUSION

In this study, a novel TCM method based on NOFRFs and the multivariate control chart is proposed. The vibration signals measured from the spindle and workpiece supporting base are collected and used for identifying a NARX model that reveals the dynamic relationship between the vibration signals. From the identified NARX model, the NOFRFs are calculated to provide physically meaningful features of the milling process. Then the PCA technique is applied to reduce the dimension of the NOFRFs features, and the multivariate control chart with a moving window of observations is applied to monitor tool wear conditions. An industrial-scale milling experiment is carried out to validate the performance of the proposed method. Traditional signal features are also extracted and used for TCM for comparison. The results show that the NOFRFs features are more stationary when the tool is not severely worn and can trigger alarms just in time when the tool wear reaches a critical threshold. This makes the NOFRFs a better choice for designing a TCM system in practice. Besides, in theory, the proposed framework can be extended to condition monitoring of a wide range of industrial processes due to its advantages of low costs and high reliability. The potential applications deserve more research.

However, the NOFRFs feature-based method is more time-consuming and requires a significant amount of computation to guarantee the accuracy of the identified model. In our experiment, 6 to 10 seconds are needed to yield a detection result from raw data, which is acceptable for tool wear monitoring, but may not be acceptable for scenarios requiring emergency response [21]. As the most time-consuming step, the parameter estimation for building the NARX model can be implemented in parallel, in the future, we will try to improve the computational efficiency using parallel computing techniques.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

[1] W. Xiao, J. Huang, B. Wang, and H. Ji, "A systematic review of artificial intelligence in the detection of cutting tool breakage in machining operations," *Measurement*, vol. 190, p. 110748, 2022.

[2] G. Serin, B. Sener, A. M. Ozbayoglu, and H. O. Unver, "Review of tool condition monitoring in machining and opportunities for deep learning," *Int. J. Adv. Manuf. Technol.*, vol. 109, pp. 3–4, 2020.

[3] X. Qin, W. Huang, X. Wang, Z. Tang, Z. Liu, "Real-time remaining useful life prediction of cutting tools using sparse augmented lagrangian analysis and gaussian process regression," *Sensors*, vol. 23, p. 1, 2022.

[4] W. Dai, K. Liang, T. Huang, and Z. Lu, "Tool condition monitoring in the milling process based on multisource pattern recognition model," *Int. J. Adv. Manuf. Technol.* vol. 119, pp. 3–4, 2022.

[5] W. J. Lee, "Monitoring of a machining process using kernel principal component analysis and kernel density estimation," *J. Intell. Manuf.*, vol. 31, pp. 1175–1189, 2020.

[6] Y. Zhou and W. Xue, "Review of tool condition monitoring methods in milling processes," *Int. J. Adv. Manuf. Technol.*, vol. 96, pp. 5–8, 2018.

[7] Z. Liu, Z.-Q. Lang, Y.-P. Zhu, Y. Gui, H. Laalej, and J. Stammers, "Sensor data modeling and model frequency analysis for detecting cutting tool anomalies in machining," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 53, p. 5, 2022.

[8] Z. K. Peng, Z. Q. Lang, and S. A. Billings, "Crack detection using nonlinear output frequency response functions," *J. Sound Vib.*, vol. 301, pp. 3–5, 2007.

[9] S. A. Billings, Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains, Chichester, West Sussex, UK: John Wiley & Sons, Inc., 2013.

[10] L. Chen, Z. Zhang, J. Cao, and X. Wang, "A novel method of combining nonlinear frequency spectrum and deep learning for complex system fault diagnosis," *Measurement*, vol. 151, p. 107190, 2020.

[11] D. C. Montgomery, *Introduction to Statistical Quality Control*, 8th ed. Hoboken, NJ: John Wiley & Sons, Inc., 2020.

[12] L. Bernini, P. Albertelli, and M. Monno, "Mill condition monitoring based on instantaneous identification of specific force coefficients under variable cutting conditions," *Mech. Syst. Signal Process.*, vol. 185, p. 109820, 2023.

[13] G. Wang, Y. Zhang, C. Liu, Q. Xie, and Y. Xu, "A new tool wear monitoring method based on multi-scale PCA," *J. Intell. Manuf.* vol. 30, p. 1, 2019.

[14] P. Orth, S. Yacout, and L. Adjengue, "Accuracy and robustness of decision making techniques in condition based maintenance," *J. Intell. Manuf.*, vol. 23, p. 2, 2012.

[15] L. Zhang and K. Li, "Forward and backward least angle regression for nonlinear system identification," *Automatica*, vol. 53, pp. 94–102, 2015.

[16] Z. Q. Lang and S. A. Billings, "Energy transfer properties of non-linear systems in the frequency domain," *Int. J. Control* vol. 78, p. 5, 2005.

[17] Y.-P. Zhu, Z. Q. Lang, H.-L. Mao, and H. Laalej, "Nonlinear output frequency response functions: a new evaluation approach and applications to railway and manufacturing systems' condition monitoring," *Mech. Syst. Sig. Process.*, vol. 163, p. 108179, 2022.

[18] S. A. Billings and J. C. Peyton Jones, "Mapping non-linear integro-differential equations into the frequency domain," *Int. J. Control*, vol. 52, p. 4, 1990.

[19] Z.-Q. Lang and S. A. Billings, "Output frequency characteristics of nonlinear systems," *Int. J. Control*, vol. 64, p. 6, 1996.

[20] B. De Ketelaere, M. Hubert, and E. Schmitt, "Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data," *J. Qual. Technol.*, vol. 47, p. 4, 2015.

[21] Y. Gui, Z.-Q. Lang, Z. Liu, Y. Zhu, H. Laalej, and D. Curtis, "Unsupervised detection of tool breakage: a novel approach based on time and sensor domain data analysis," *IEEE Trans. Instrum. Meas.*, vol. 72, p. 1–13, 2023.