A wavelet-based data-driven modelling for tool wear assessment of difficult to machine materials

Farbod Akhavan Niaki* and Lujia Feng

International Center for Automotive Research,
Clemson University,
4 Research Drive,
Greenville, SC, 29607, USA
Email: fakhava@clemson.edu
Email: lujiaf@clemson.edu
*Corresponding author

Durul Ulutan

Mechanical Engineering,
Bucknell University,
701 Moore Ave,
Lewisburg, PA 17837, USA
Email: du005@bucknell.edu

Laine Mears

International Center for Automotive Research,
Clemson University,
4 Research Drive,
Greenville, SC, 29607, USA
Email: mears@clemson.edu

Abstract: In this work, wavelet packet decomposition along with principle component analysis are utilised for feature extraction using two low cost sensing methods: vibration and power sensors, in end-milling of gamma prime-strengthened alloy. The high wear rate of this material induces a rapid transition from a sharp state to a dull state of the tool, and hence limits the number of available data for model establishment. To address this challenge, a neural network with Bayesian regularisation is designed and its performance is compared with regression and time-series models. A maximum of 4% estimation error for Bayesian regularisation neural network, compared to 33% and 17% estimation error of the latter models, shows the good potential of this technique when a limited dataset is available. In addition, the use of low cost measuring sensors in this paper aligned well with the industrial applications to detect and avoid unscheduled downtime in machining situations.

Keywords: condition monitoring; tool wear; wavelet packet decomposition; WPD; recurrent neural network; time series analysis.

Reference to this paper should be made as follows: Akhavan Niaki, F., Feng, L., Ulutan, D. and Mears, L. (2016) 'A wavelet-based data-driven modelling for tool wear assessment of difficult to machine materials', *Int. J. Mechatronics and Manufacturing Systems*, Vol. 9, No. 2, pp.97–121.

Biographical notes: Farbod Akhavan Niaki is a PhD candidate of Automotive Engineering at the Clemson University, International Center for Automotive Research (CU-ICAR). He received his Master degree from Sharif University of Technology specialising in modelling and simulation of dynamical systems. Since 2013, he joined Informed Control of Manufacturing lab at CU-ICAR working on tool condition monitoring and model-based control of manufacturing processes.

Lujia Feng is a PhD candidate at Clemson University, Automotive Engineering (CU-ICAR) Department. Her interest is in sustainable manufacturing systems, particularly on forecasting and optimisation of energy supply and consumption. She received her MSc and BSc in Environmental Engineering from Cornell University (2012) and Beijing Technology and Business University (2010) focusing on water conservation and renewable and low emission energy usage.

Durul Ulutan is an Assistant Professor at Bucknell University, Mechanical Engineering Department. He received his BS (2005) and MS (2007) from Koç University (Turkey), Mechanical Engineering Department, and PhD from the Department of Industrial and Systems Engineering at Rutgers University (2013). After his PhD studies, he worked at Clemson University, International Center for Automotive Research (CU-ICAR) as a post-doctoral researcher. He has over ten years of experience in analysing, modelling, and optimising manufacturing processes and systems.

Laine Mears is a Professor and founding faculty member in the Automotive Engineering Department at Clemson University, teaching and carrying out projects at the Clemson University International Center for Automotive Research. He teaches modelling and analysis of automotive manufacturing processes, quality systems and quality tools, and performs research in manufacturing quality systems, intelligent machining systems, manufacturing process design and control, and manufacturing equipment diagnostics. He has published over 100 peer-reviewed articles, and is the recipient of the NSF CAREER Award, SAE Ralph Teetor Educational Award, the South Carolina Governor's Young Researcher Award for Excellence in Scientific Research and the IMECHE George Stephenson Gold Medal.

This paper is a revised and expanded version of a paper entitled 'Wavelet based sensor fusion for tool condition monitoring of hard to machine materials' presented at the International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2015), San Diego, California, USA, 14–16 September 2015.

1 Introduction

In automated manufacturing systems, accurate estimation and monitoring of important process states is a critical factor to reduce downtime, avoid catastrophic tool failure, and preserve the quality of the final product. Tool wear in difficult-to-machine materials (those of: excessive strength compared to tooling material, poor thermal conductivity, or rapid work hardening) plays an important role in machining downtime and the quality of the resultant product. For these materials, the wear rate of inserts during machining is relatively high compared to conventional materials (Zhu et al., 2013). Such a high tool wear rate while machining makes establishing an accurate tool wear model a challenging task because only a limited number of experiments can be completed before tool failure; additionally, the aforementioned confounding effects also influence wear progression. Moreover, tool wear is a highly complex phenomenon that consists of different mechanisms such as abrasion, adhesion, chipping, and plastic deformation (Zhu et al., 2013; Akhtar et al., 2014). The model for each separate mechanism is either not developed yet or developed only under specific machining conditions for a particular type of material, which cannot be generalised for other materials or machining conditions (Ko and Koren, 1989).

Tool wear studies can be divided into three major categories: the first is the empirical or semi-empirical study based on the observed relationships of cutting conditions such as feed and cutting speed to tool wear. Taylor's tool life model is the most commonly used empirical model developed and still widely in use in the machining industry. While convenient to use, Taylor's tool life model is blind to the tool wear mechanisms and cannot be used for modern machining such as high speed machining (Yen et al., 2004). The second category is model-based studies that use pre-developed analytical models for certain mechanisms of tool wear. Such studies include Takeyama and Murata (1963) model of tool flank wear that considers mechanical abrasion and physicochemical type of wear, Usui et al. (1978) model of crater wear that uses a three-dimensional heat transfer model of the chip to derive the wear characteristic equation assuming only the diffusion mechanism exists, Koren (1978) model where only abrasion and diffusion mechanisms were considered for wear model development, and Rabinowicz et al. (1961) model where only the abrasive mechanism was investigated for modelling tool wear. Stochastic-based analysis of the mechanistic tool wear model is also considered by Akhavan Niaki et al. (2015a, 2015b). Numerical studies are another subcategory of model-based methods that are based on finite element analysis (FEA) of mechanical interactions between the tool and the workpiece material (Yen et al., 2004; Attanasio et al., 2008). While FEA is accurate in modelling tool wear, the major drawback is its time inefficiency which makes it impractical for industrial use. The third category of tool wear studies is data-driven modelling. As implied from its name, this method relies on input data and its correlation to tool wear. The advantages of data-driven techniques are mostly realised in situations where the process model is not available. This feature is particularly useful for studying tool wear in machining hard-to-machine materials due to the lack of proposed analytical models.

To use data-driven methods, three main decisions need to be made: The type of signal to use, the features of the signal to extract, and the method to select. Force data is the most widely used measurement signal in tool wear studies. Despite being precise, the measurement device (i.e., dynamometer) needs continuous calibration (Xiaoli, 1999), limits the workpiece size (Alaniz-Lumbreras et al., 2006), and is expensive, making it almost impossible to implement in industrial machine shops. Vibration and acoustic emission (AE) signals have shown good capability in studying tool wear; however, acquiring these signals requires powerful data acquisition systems with enough bandwidth and memory. In addition, the mounting location of the AE sensor or the accelerometer has significant influence on their performance (Xiaoli, 1999; Segreto et al., 2012). On the other hand, Hall effect sensors used for power sensing are relatively inexpensive, they can be easily mounted in the machine, and they do not limit the size of the workpiece, which make them suitable candidates for machining performance assessment in industrial applications. However, the power signal, which is representative of the resultant force, is not always as sensitive to the tool condition as cutting force (Teti et al., 2010). Due to the differences in the nature of sensors, each can capture different information from the system. Thus, by using sensor fusion methods, it is possible to extract more informative information from the signals. Wang et al. (2007) fused two sensors for flank wear estimation of ASSAB718 steel in dry milling, namely a vision system (direct measurement) and a force sensor (indirect measurement). Their results were promising for flank wear estimation. However, their method was only applicable for dry milling. The flow of coolant in machining titanium and nickel-based materials makes using the vision measurement systems almost impossible. Segreto et al. (2012, 2013) combined force, AE and vibration sensors for tool wear classification of Inconel 718 with a backpropagation neural network (NN) in wet machining. They could achieve a success rate of 98% for discrimination between a sharp and worn insert. They also reported small contribution of the vibration sensor due to the environmental noise in machining. While the use of classifiers, make combining the effect of multiple tool wear mechanisms possible, the output is less informative than estimators since no information will be provided on the evolution of particular tool wear mechanisms. Similar work on tool wear classification of AISI4340 steel was conducted by Cho et al. (2010). They compared the performance of different classifiers, multilayer perception NN, radial basis function NN, support vector machine, and machine learning ensemble, in fusing four different sensors (force, vibration, AE and power) with 11 different combinations. According to their results, machine learning ensemble had the highest accuracy with 97% success rate with the combination of vibration and force sensors. Vibration and power sensors, which are well suited for industrial applications, were fused in the work of Trejo-Hernandez et al. (2010). They created a field programmable gate array (FPGA) smart sensor for predicting tool wear area in turning of AISI 1045 and showed, by fusion of the aforementioned sensors, that estimation results improved three times as compared with a single sensor. In the recent work of Zhang et al. (2015), AE and cutting sound (microphone) sensors were utilised in turning super-alloy GH2135 for flank wear estimation with support vector regression for predicting tool wear and support vector machine for classifying the tool state. Based on their results the prediction accuracy of fused sensors is very close to the prediction result of a single AE sensor, which makes using sound sensors a redundant choice. Moreover, linear behaviour was observed in their result, which enables using simpler methods such as linear regression and time series modelling.

In addition to available data driven tools, several works in the literature studied feature selection techniques. The features of the acquired signal related to tool wear are categorised as time domain, frequency domain, and time-frequency domain features. Scheffer et al. (2005) used correlation and the statistical overlap factor of statistical features of force signals in the time domain, and showed that the standard deviation of the force signal is the most sensitive feature in turning to tool wear. In another effort, Segreto et al. (2012, 2013) used time domain features of AE, vibration, and force signals in addition to principal component analysis (PCA) and linear predictive analysis for tool wear classification of Inconel 718 in turning. Fang et al. (2012) studied the frequency domain features of the fast Fourier transform of the vibration signal for tool edge wear assessment of Inconel 718. Time-frequency decomposition of the signal using wavelet transforms has been considered by different researchers. As mentioned by Xiaoli (1999), a successful feature selection method should be as sensitive as possible to tool wear and insensitive to changes in cutting conditions. Xiaoli used continuous wavelet transform (CWT) for decomposing AC motor current and discrete wavelet transform (DWT) for decomposing AC servomotor current for small drill bit breakage detection. Choi et al. (2004) studied the relationship between root mean square (RMS) of DWT of a force signal in ramp cuts of AISI 1018, and showed that the RMS value of the wavelet coefficient exhibits a linear relationship to tool wear in different cutting conditions. However, they proposed a separate model for each cutting condition and could not develop a generalised tool wear model. Chuangwen and Hualing (2007) used the wavelet packet decomposition (WPD) of the vibration signal and introduced a new feature selection procedure for eliminating the effect of cutting conditions. In another effort, Jemielniak et al. (2011) introduced two new correlation coefficients for selecting the statistical features of wavelet coefficients of WPD for rough turning of Inconel 625. The variability of different feature selection methods is shown in Table 1.

The focus of the majority of literature in the field of tool condition monitoring (TCM) is on conventionally available material such as steel. Recently the attention is shifted toward nickel-based alloys, but mainly on classification purposes, which does not provide information on tool wear evolution. Nickel-based alloys exhibit a unique combination of high strength and high corrosion resistance at elevated temperatures, which lead to high tool wear rate, and consequently a limited amount of available data for training wear recognition methods. Therefore, in this work a Bayesian regularisation NN is chosen as the tool wear predictor, which previously was shown to perform well where the size of the training set is limited (Burden and Winkler, 2009). Moreover, as confirmed by Zhang et al. (2015), difficult to machine materials exhibit a linear evolution of tool wear, so it is feasible to evaluate the performance of the machine learning method to linear regression and times series models. To demonstrate the industrial applicability, two available and low-cost sensing technologies, spindle power and vibration, were used to predict wear in an end-milling machining operation. The remaining organisation of this work is as follows: the theoretical background of WPD, principle component analysis, recurrent NN, and time series models are briefly introduced in Section 2. The experimental setup and selected cutting conditions are explained in Section 3. In Section 4, the processing of the data and feature reduction is explained. Results and conclusions are provided in Sections 5 and 6.

Table 1 Methods for feature selection

Authors	*Signal*	*Method*	*Feature*	*Material*
Li and Patri (1999)	AE	WPD	Root mean square	40Cr Steel
Kamarthi et al. (2000)		DWT	Energy of wavelet coefficient	AISI 6150
Wu and Du (1996)	Vibration	WPD	Peak to valley/crest factor/mean/variance/ kurtosis	AISI 45
Chuangwen and Hualing (2007)		WPD	Normalised energy	-
Antić et al. (2013)		Time domain	Mean/variance/ skewness/kurtosis	-
He et al. (2013)		Time domain/WPD	Root mean square	AISI 45
Hong et al. (1996)	Force	DWT	Mean/variance of local maxima	ASSAB 760
Choi et al. (2004)		DWT	RMS	AISI 1018
Scheffer and Heyns (2004)		Time domain	Variance/energy around certain frequency	Aluminum alloy
Scheffer et al. (2005)		Time domain – frequency domain	Variance/energy around certain frequency	Aluminum alloy
Wang et al. (2007)		SOM	Average/variance over one revolution	ASSAB 718HH
Wang and Wang (2012)		WPD	Energy of signal	ASSAB 718HH
Alaniz-Lumbreras et al. (2006)	Current	DWT	Low frequency in wavelet coefficients	AISI 1018
Patra et al. (2007)		WPD	Mean	Mild steel
Li et al. (2007)	Image (vision)	WPD	Energy	25Cr3Mo3NiNb
Fang et al. (2012)	Force/ vibration	WPD	Root mean square	Inconel 718
Jemielniak et al. (2011)	Force/ AE	WPD	Root mean square/ log-energy/skewness/ kurtosis/ring down counts/pulse width	Inconel 625
Segreto et al. (2012)	Force/ AE/ vibration	Time domain	Coefficients of principal component analysis	Inconel 718
Segreto et al. (2013)		Time domain	Linear predictive analysis	Inconel 718
Xiaoli (1999)	Motor/ axis current	CWT/DWT	-	AISI 45 quench steel

2 Theoretical background

2.1 Wavelet analysis

Depending on the application, spectral information of a measured signal may be useful. Time series analysis of a given signal only provides information in specific time horizons while spectral content is lost. The Fourier transform is suggested for converting the signal from the time domain into frequency domain. It is powerful in studying stationary signals, but does not provide information about the exact *time* of a change. This is particularly important when studying non-stationary signals where sudden peaks or anomalies may occur. To overcome the shortcomings of Fourier transform, the wavelet transform was proposed as an alternative (Morlet et al., 1982). In this transformation, a wavelet (small wave) function, which should be oscillatory with finite integral, is chosen as a windowing function, and is applied to the signal. wavelet functions can be stretched to different frequencies and also shifted in time to synchronise with a particular event. Assuming the existence of a continuous signal, the wavelet transformation is defined as equation (1) (Fugal, 2009). Where ψ is the wavelet function called the *mother wavelet*, the superscript '*' denotes the complex conjugate of the mother wavelet, τ is the shift in time, s is the scale factor and X_{WV} is the magnitude of transformed signal in time-frequency domain. The term $1/\sqrt{s}$ is a normalising constant for different scales.

$$X_{WV}(\tau, s) = \frac{1}{\sqrt{s}} \int x(t) \psi^* \left(\frac{t-\tau}{s} \right) dt \tag{1}$$

The mother wavelet can be chosen as any integrable, zero mean and zero norm function that satisfies the admissibility criteria [equation (2)] (Fugal, 2009).

$$\int_0^\infty \frac{|\psi(t)|^2}{t} dt < \infty \tag{2}$$

There are some pre-defined mother wavelets proposed such as Daubechies, Coiflet, Haar, Morlet, symlet and many others which satisfy the above criteria and have been shown to work well in different problems. The choice of the mother wavelet is dependent on the type of problem, but usually a trial and error method to find the best mother wavelet is useful (Fugal, 2009). In CWT, the parameters s and τ change continuously. In practical applications, sweeping these two parameters in the domain of real numbers is time consuming, and hence not practical. Therefore, positive integer values are selected instead to reduce the computational burden.

Since any processing algorithm and any acquired signals are in digitised format, discretisation should be implemented for CWT. The discretisation (often called dyadic sampling) is based on using the powers of 2 for shifting (s) and scaling parameters (τ). In that case, equation (1) can be written as equation (3), where s_0, $\tau_0 > 1$, j is a positive integer number, and k is constant. Discretising CWT makes the digital calculation possible, although it cannot still eliminate the redundancy and the high computational cost.

$$X_{WV}(j,k) = s_0^{\frac{-j}{2}} \int x(t)\psi^* \left(s_0^{\frac{-j}{2}} t - k\tau_0 \right) dt \tag{3}$$

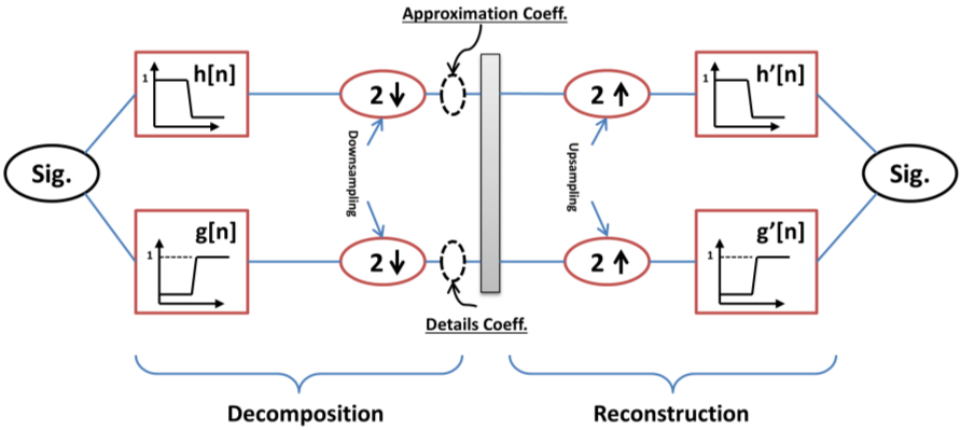
To overcome this limitation, the DWT is proposed. In DWT, special digital filters (high and low pass filters) combined with downsampling and upsampling of the filtered signal are used for decomposition or reconstruction. To decompose the signal, it is first passed through the low/high pass filters, and then every other sample will be dropped to avoid aliasing (Fugal, 2009). Output of the downsampled signal from the low pass filter is a denoised version of the actual signal called the *approximation* and the output of the downsampled signal from the high pass filter containing high frequency information of the actual signal is called the *detail*. To reconstruct the signal, it is first upsampled and then passed through another set of low/high pass filters (Figure 1). The combination of all four filters used for decomposition and reconstruction is called the *filter bank* or quadrature mirror filters (QMF). The main feature of QMF is that the high pass filter is dependent on the low pass filter as in equations (4) to (6), where L is the length of the filter and g and h are high pass and low pass digital filter coefficients respectively (Fugal, 2009).

$$g[L-1-n] = (-1)^n h[n] \tag{4}$$

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k] \tag{5}$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \tag{6}$$

Figure 1 1-level DWT decomposition and reconstruction with up/down sampling of QMF (see online version for colours)



In DWT, it is assumed that information in the detail coefficients is not as important as the information in the approximation coefficients. Therefore, in each level, decomposition is applied to the approximation coefficients and is continued for multiple levels of decomposition as in Figure 2. While DWT does not emphasise the significance of the

detail coefficients, wavelet packet decomposing (WPD) considers the detail coefficients to be as important as the approximation coefficients, and decomposition is applied on the detail coefficients as well. This decomposition tree is shown in Figure 3. WPD is particularly useful when an anomaly (e.g., tool breakage or chipping) occurs at high frequency. In that case, high frequency coefficients are as important as low frequency coefficients. Since chipping is one of the dominant tool wear mechanisms in difficult-to-machine materials (Akhtar et al., 2014; Thakur and Gangopadhyay, 2016), and it occurs in short instances during machining, it is worthwhile to consider WPD for analysing and extracting related features of the measured signal.

Figure 2 3-level DWT cascade diagram (see online version for colours)

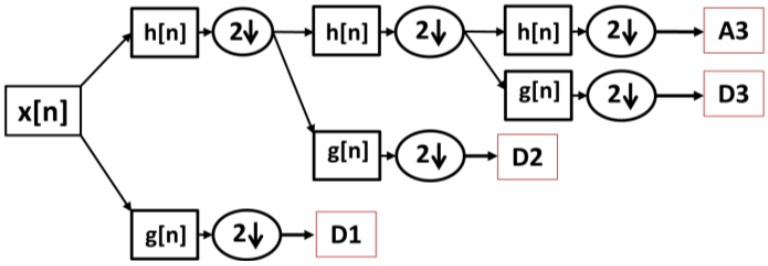
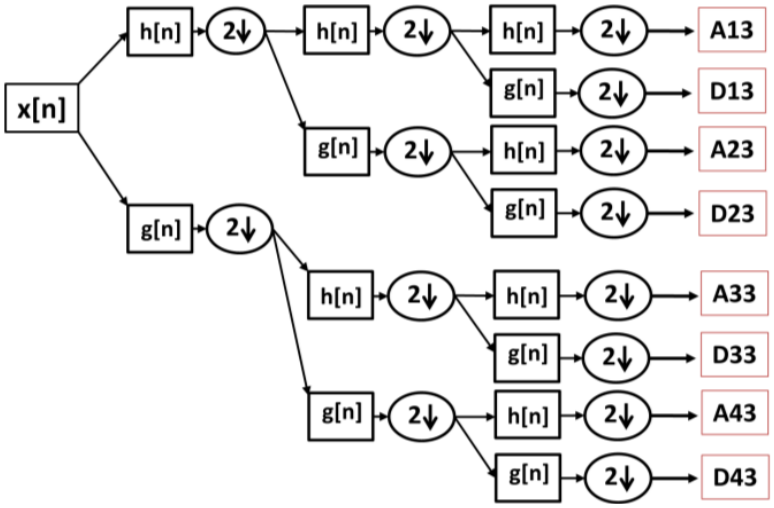


Figure 3 3-level WPD cascade diagram (see online version for colours)



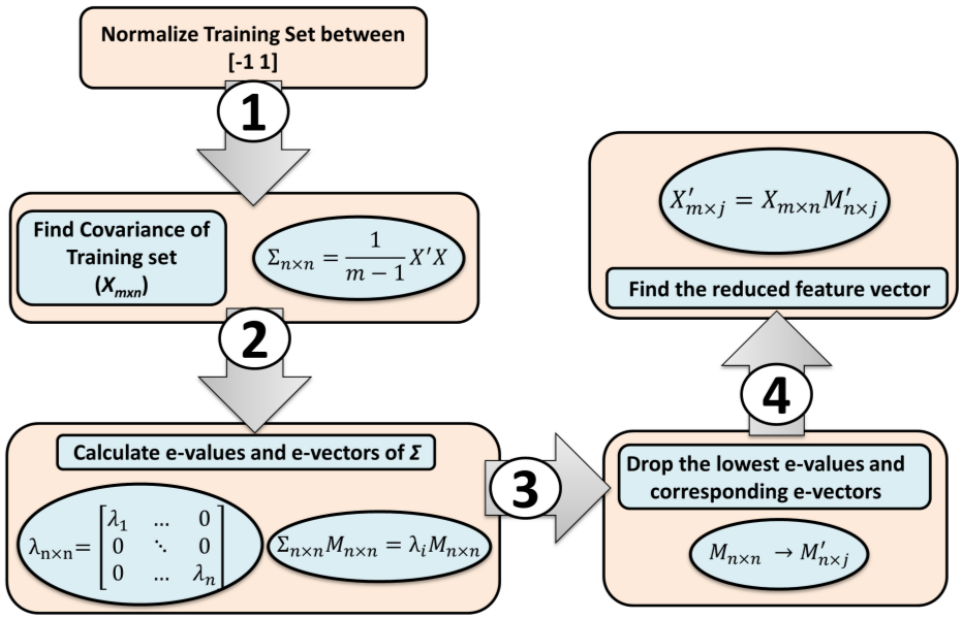
2.2 Principal component analysis

Considering the fact that a large number of samples are measured during experiment, it is not possible to use all the samples as the training set. Therefore, new parameters are required to shrink a large sample set, and at the same time capture the essential characteristics of the system. A review on the variability of selected features is shown in Table 1. In this work, the statistical features of the WPD coefficients were considered as the features of interest. Selected features do not necessarily bring additional information,

and may be highly correlated, and therefore redundant. The undesired effect of the large feature set, which is known as the curse of dimensionality (Bishop, 2006), can be overcome by implementing a feature reduction method called PCA that can identify the most related features with the highest degree of variation.

The idea behind PCA is first transforming the training set from the current coordinate system into a new orthogonal coordinate system (i.e., principal coordinates), a procedure that decouples the features, and then eliminating the coordinates that have the lowest contribution to the variation of sample set. The procedure for running PCA is shown in Figure 4.

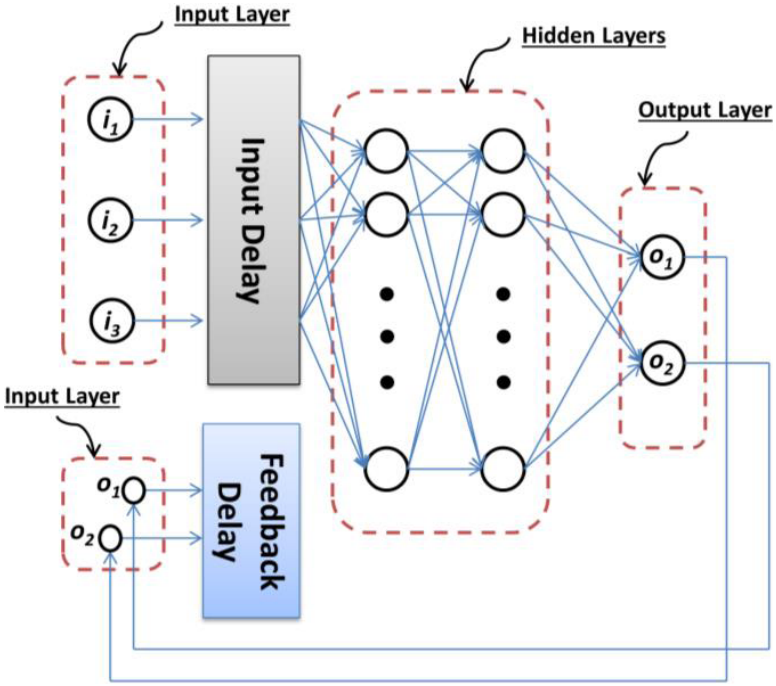
Figure 4 Principle component analysis flowchart (see online version for colours)



2.3 Dynamic NN architecture

NN algorithms are designed based on the learning behaviour of neurons in human brains (Bishop, 2006). Recognition with NNs has the advantage that it eliminates the need for a model. In this paper, the nonlinear auto regressive with exogenous (NARX) input architecture is selected as a NN topology. In the NARX model, there is a feedback with time delay from the outputs to the inputs of the NN, so that outputs are added as new features to the inputs. As shown in Figure 5, the NARX model consists of three layers: The first layer is called the input layer, the middle layer is called the hidden layer, which may consist of more than one layer with different number of neurons, and the last layer is the output layer. The number of hidden layers and the number of hidden neurons are design parameters, and they should be determined by trial and error to reach the minimum error without overfitting the training set.

Figure 5 NARX NN schematic (see online version for colours)



The backpropagation method is the most widely used training technique for adjusting the weights of a NN. However, the likelihood of overfitting the data using backpropagation is high. The Bayesian regularisation training method was proposed to avoid overfitting (Baba and Kozaki, 1992). This method of training is more robust than backpropagation and can eliminate or reduce the need for a lengthy cross-validation set (Burden and Winkler, 2009). This is specifically beneficial when a large dataset is not available for training the NN. As mentioned in Section 1, due to high wear rate and short life span of the tool in machining gamma-prime strengthened alloys, a limited number of experiments can be performed before reaching failure. Therefore, Bayesian regularisation is selected for training the NN. In this method, an additional term related to the sum of squares of NN weights is added to the sum-of-squared error of outputs and targets as shown in equation (7), where E_T is the total sum of errors weighted by parameters α and β. E_1 is the sum-of-squared errors of targets and outputs, and E_2 is the sum-of-squared of NN weights. The objective is to minimise the new error term E_T. In the Bayesian regularisation algorithm, it is assumed that NN weights are random variables with a known prior, which is updated at each iteration. For a detailed discussion of Bayesian NNs, the authors encourage readers to review the studies from Baba and Kozaki (1992) and Burden and Winkler (2009).

$$E_T = \alpha E_1 + \beta E_2 \tag{7}$$

2.4 Time series analysis

Time series analysis is a powerful tool in modelling of historical data. The objective of this analysis is to extract the underlying hidden structure in the data by finding the trend, seasonality or the autocorrelation. Auto-regressive moving-average (ARMA) modelling is a class of time series analysis used where weakly stationary behaviour is observed in the dataset (Brockwell and Davis, 2006). ARMA contains two linear parts; the auto-regression (AR) and the moving average (MA). The AR model estimates the output X at time t, based on p previous outputs and current measurement noise denoted as Z_t, as shown in equation (8). On the other hand, the MA model estimates the output X_t, based on q previously available measurement noise points [equation (9)]. By combining these two parts the ARMA(p, q) models can be written as equation (10). The constants θ and Φ can be estimated through many different algorithms such as least square estimation, Yule-Walker, or maximum likelihood (the latter is used in this work) (Brockwell and Davis, 2006).

$$X_t = Z_t + \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} \tag{8}$$

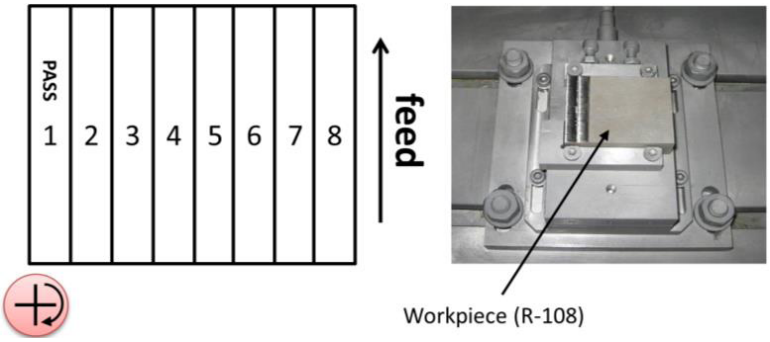
$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_p Z_{t-q} \tag{9}$$

$$Z_t + \Phi_1 X_{t-1} - \ldots - \Phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_p Z_{t-q} \tag{10}$$

3 Experimental setup

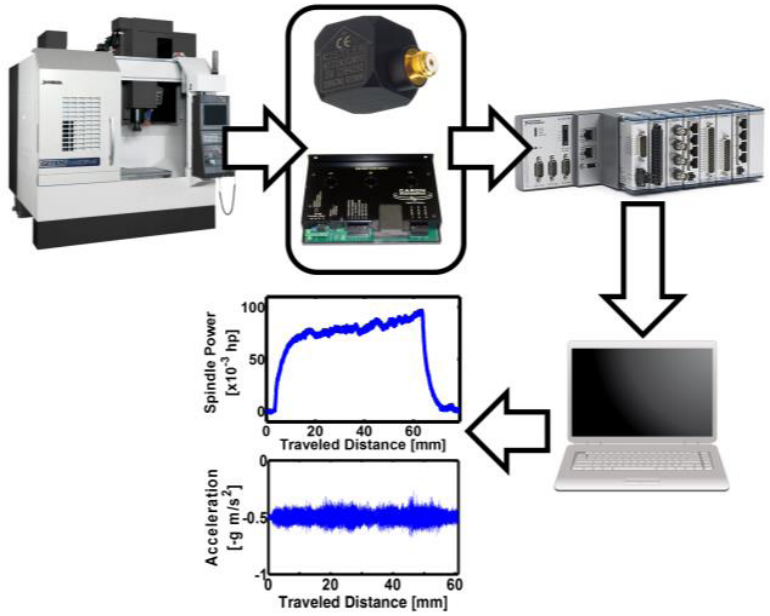
As shown in Figure 6, a total of eight experiments with three replications were conducted to predict the tool wear when machining the gamma-prime strengthened alloy Rene-108 (R-108). Among these three replications, the first two were chosen as training sets, and the last replication was chosen for validating the results (i.e., the testing set). An OKUMA GENOS M460-VE 3-axis CNC machine was used to end mill (in the down-milling direction) rectangular blocks with dimensions of 60 mm × 80 mm × 25 mm, using an 8% concentrated water-soluble coolant. A 2-flute indexable tool holder with a diameter of 15.875 mm was used, and the width of cut was chosen to be 9.5 mm, which corresponds to 60% tool engagement, as this was the maximum manufacturer recommendation for the particular tool. The full length of the blocks (60 mm) was utilised for machining, which was considered as a ‘pass’. Depth of cut, cutting speed, and feed are kept constant for all passes at 0.5 mm, 25 m/min, and 0.1 mm/rev, respectively. The cutting conditions are determined based on the industrial applications targeted by this study, and the decision to keep the cutting conditions constant is due to the fact that a change in the cutting conditions can induce abrupt changes (appears in terms of excessive chipping and early failure) in the behaviour of hard-to-machine alloys and subsequent wear progression. Two different sensors were utilised in this work. A Hall effect sensors was mounted on the electric panel of the CNC machine to capture the change in current consumed by the spindle, and an accelerometer is mounted on the spindle case to capture the induced vibration between tool and workpiece during the milling operation.

Figure 6 Schematic of milling experiments (see online version for colours)



A data acquisition device (DAQ) was programmed to capture the spindle power consumption and vibration at a sampling rate of 10.24 kHz during cutting. To measure spindle power and vibration, the outputs of the Hall effect sensor and the Kistler 8772A accelerometer were fed into NI9215 and NI9234 analogue input modules, respectively. Those were mounted on an NI-cRIO9103 chassis programmed with LabVIEW (Figure 7).

Figure 7 Data acquisition with NI-cRIO9103 (see online version for colours)



Inserts used in this work were Sandvik Coromill R390-11 T3 08M-PM 1030, hereafter referred to as '1030'. The 1030 grade (TiAlN PVD coated) is recommended by Sandvik for milling R-108 and similar difficult-to-machine materials due to its resistance to material build-up on the cutting edge and plastic deformation (Sandvik, 2006). Fresh unworn inserts were used at the beginning of each set of tests (eight passes), with Olympus optical microscope used to measure flank wear on the bottom edge of insert.

The progress of tool flank wear is shown in Figure 8 for tests 1.2, 1.4, 1.6 and 1.8 (i.e., the first replication). Measured power and tool flank of all three replication sets are shown in Table 2.

Figure 8 Optical measurement of the flank wear for (a) test 1.2, (b) test 1.4, (c) test 1.6, and (d) test 1.8 (see online version for colours)

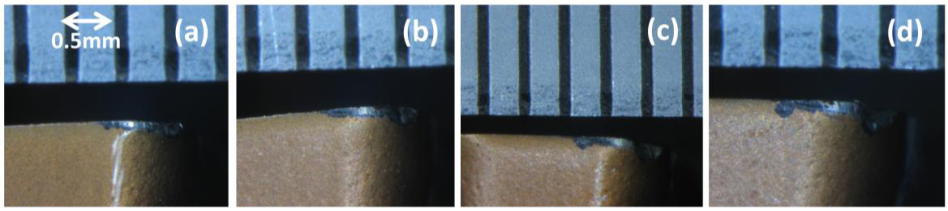


Table 2 Flank wear measurement for (a) training sets, and (b) testing set

(a) Training sets				*(b) Testing set*	
Replication 1		*Replication 2*		*Replication 3*	
Test #	*VB (μm)*	*Test #*	*VB (μm)*	*Test #*	*VB (μm)*
1.1	84	2.1	83	3.1	81
1.2	89	2.2	87	3.2	87
1.3	100	2.3	103	3.3	99
1.4	108	2.4	107	3.4	103
1.5	111	2.5	109	3.5	109
1.6	116	2.6	116	3.6	115
1.7	119	2.7	125	3.7	116
1.8	125	2.8	127	3.8	120

4 Data processing with WPD

The spindle power and vibration signals for each pass are divided into five regions, each near 10, 20, 30, 40, and 50 mm cutting distance within a pass, containing 2^{11} sample points (i.e., 11 levels of WPD is feasible), as shown in Figures 9 and 10. Note that tool wear measurements were only available after each pass. Therefore, to compensate the inconsistency between the number of data points, extra data points were interpolated between each tool wear measurement, and Gaussian noise with zero mean and variance equal to 1.36×10^2 μm^2 were added to the interpolation to include measurement uncertainties. Variances of tool wear were selected based on maximum variance between each pass in the 1st and 2nd sets of experiments [equation (11)].

$$\sigma^2_{\text{tool wear}} = \max\left\{\text{var}\begin{pmatrix}1.1\\2.1\end{pmatrix}, \text{var}\begin{pmatrix}1.2\\2.2\end{pmatrix}, \ldots, \text{var}\begin{pmatrix}1.8\\2.8\end{pmatrix}\right\}_{\text{tool wear measurement}} \tag{11}$$

Figure 9 Tool holder vibration (see online version for colours)

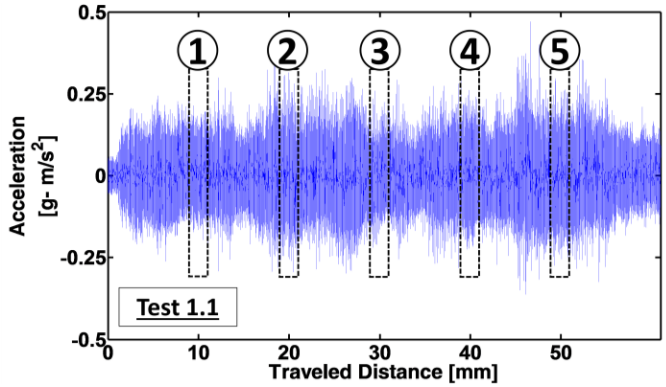
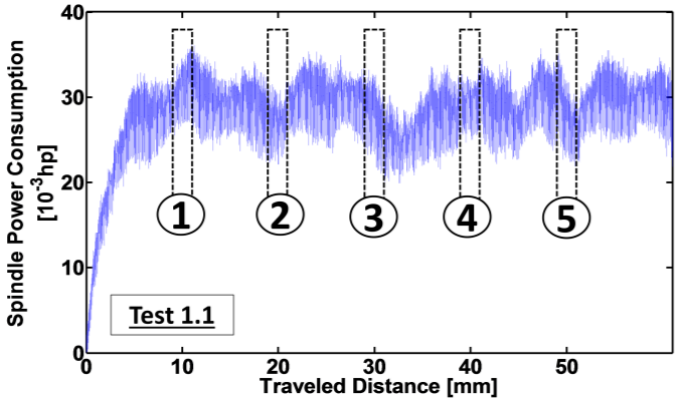
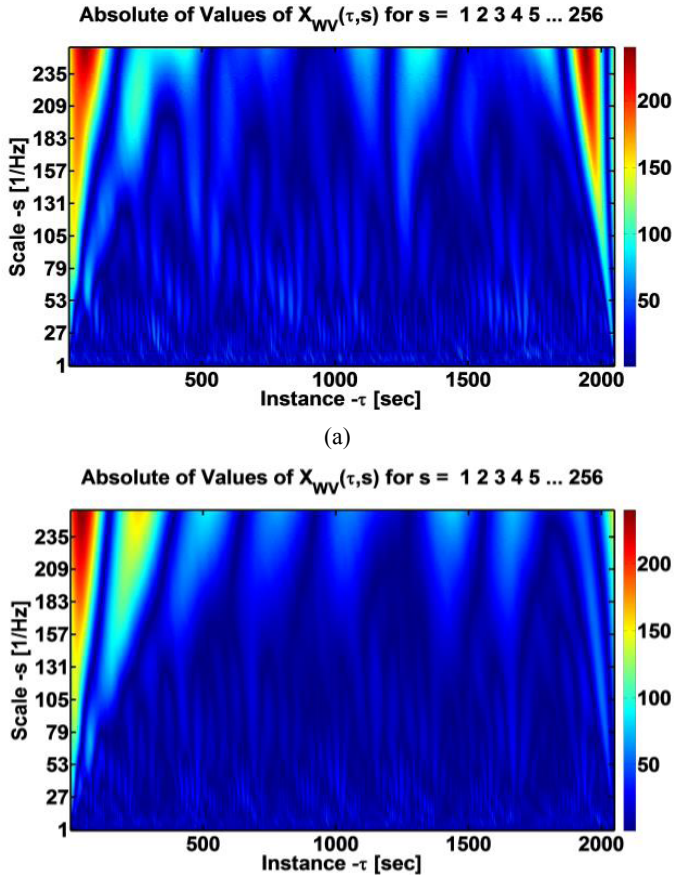


Figure 10 Spindle power consumption (see online version for colours)



With the progress of tool wear and subsequent change in the condition of insert from sharp to dull, a change in frequency amplitude with respect to time was observed. This change is shown in Figures 11 and 12, as the first region of test 1.1 (sharp insert) and its corresponding scalogram (top figure) is compared with the last region of test 1.8 (dull insert) and its corresponding scalogram (bottom figure). According to these figures, the amplitude of the lower frequencies (equivalent to higher scales in *y*-axis of scalogram) decreased when the insert became duller. To easily compare the change in frequency of the spindle power and the tool holder vibration over time, CWT was originally applied to the signals; however, due to the computational cost and redundancy of CWT, a non-redundant method (i.e., WPD) has been used in this work. Thus, the wavelet scalograms in Figures 11 and 12 are presented solely for clarifying the concept.

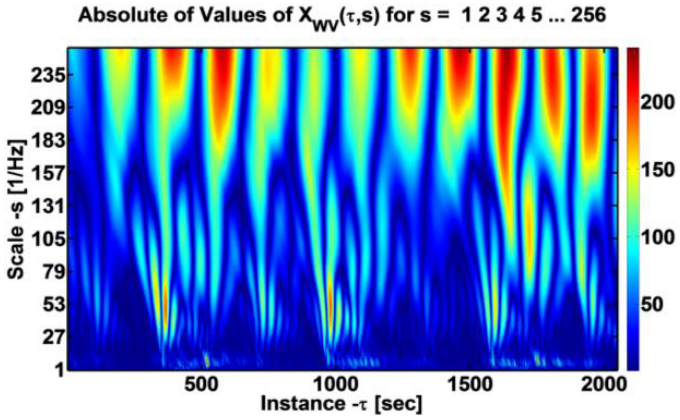
Figure 11 Scalogram of (a) sharp and (b) dull insert using Db4 wavelet of spindle power (see online version for colours)



(a)

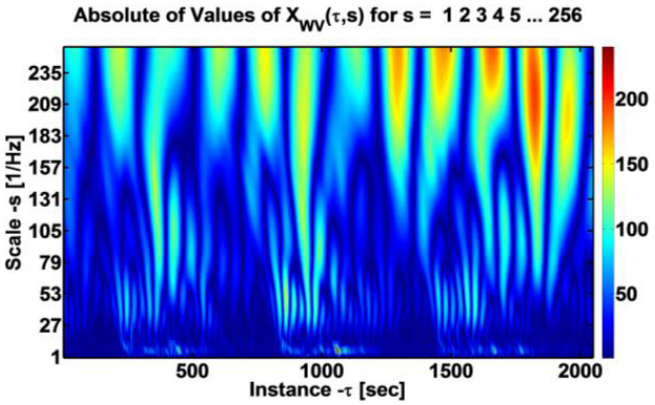
(b)

Figure 12 Scalogram of (a) sharp and (b) dull insert using Db4 wavelet of accelerometer (see online version for colours)



(a)

Figure 12 Scalogram of (a) sharp and (b) dull insert using Db4 wavelet of accelerometer (continued) (see online version for colours)



(b)

4.1 Feature selection

Statistical features of the wavelet packet coefficients: mean, standard deviation, RMS, skewness, kurtosis factor, crest factor, peak to valley, and energy of the coefficients were selected as candidate features. Depending on the level of decomposition, packets consisting of approximation and detail coefficients were generated. For example, wavelet decomposition of level 4 produces $2^4 = 16$ packets including eight approximation and 8 detail coefficients, from which each of the eight features can be extracted. Therefore, the challenges are determining the type of wavelet to select, the number of decomposition levels to use, and the choice of packet. As noted by Teti et al. (2010), different researchers in the past have chosen different wavelets and different level of decomposition; however, there is no clear explanation why the particular wavelet was selected. Therefore, in this work, four different wavelets, namely Daubechies 4 (Db4), Daubechies 6 (Db6), Coiflet2 (Coif2), and Coiflet3 (Coif3) are selected, and corresponding features of three- and four-level decomposition were calculated. In Scheffer and Heyns (2004), two factors were introduced to find the most correlated features to the output, correlation factor, and statistical overlap factor, which determines the amount of difference between sharp and dull inserts. By computing the two above factors for features of each packet, the most correlated features to tool wear, corresponding packets, decomposition level, and mother wavelet were determined. For spindle power signal, features derived from the approximation coefficients (A31 and A33 i.e., packets 1 and 3 of 3rd level decomposition) of 3-level WPD with Db3 as the mother wavelet showed the highest correlation to the tool wear. The same process was repeated for the vibration signal, and features derived from detail coefficients (D46 and D410, i.e., packets 4 and 10 of 4th level decomposition) of a 4-level WPD with Db4 as the mother wavelet were selected. From these findings, it can be concluded that the features in lower frequency ranges of spindle power signal and features in higher frequency ranges of vibration signal are most highly correlated to tool wear.

4.2 Feature reduction with PCA

Following the above procedure, a total of 32 different features were available. As stated in Section 2.2, redundancy exists between some of the features; on the other hand, some other features are not as correlated to tool wear as others. Therefore, to avoid redundancy and decrease the run-time of the NN, the use of feature reduction is inevitable. After applying PCA, the number of features decreased from 32 to 6. The cumulative sum of the first six highest eigenvalues in equation (12) shows that 85% of feature variations in the training set are captured by six of the transformed features. In Figure 13, these six features for the first replication are plotted. The p-value from analysis of variance (ANOVA) in Table 3 shows the significance of each coefficient to the tool wear. According to the results shown in Table 3, none of the six features significantly contributed to the outputs and the R^2 of linear regression based on these features are only 19%. This suggests that input delays and time series analysis are important.

$$\begin{aligned}&\text{Cumulative sum of e-values}\\&= [0.29 \quad 0.49 \quad 0.63 \quad 0.72 \quad 0.80 \quad 0.85 \quad 0.89 \quad 0.931 \quad \ldots \quad 1]^{T}_{32\times 1}\end{aligned} \tag{12}$$

Figure 13 Selected features of 1st replication for training (see online version for colours)

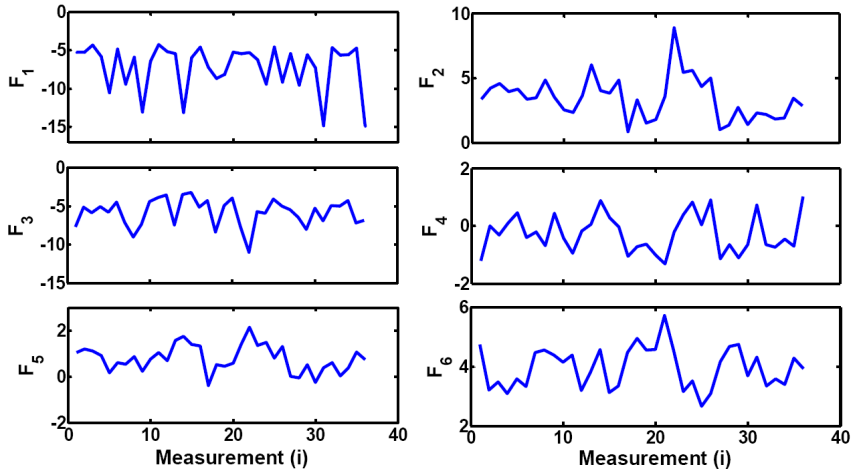


Table 3 ANOVA table and p-value significance test of features

Features	*Coefficients (× 10^{-2})*	*Standard error (× 10^{-3})*	*t-stat*	*p-value*
F1	0.1	7.7	0.15	0.88
F2	–1.2	8.4	–1.44	0.16
F3	–1	4.9	–2.01	0.05
F4	1.7	41.3	0.40	0.68
F5	0.6	15.9	0.41	0.68
F6	–0.9	22.7	–0.41	0.68

4.3 Training the ARMA and NARX model

Before training, all six of the features and target values were normalised to avoid inconsistency between feature dimensions. The mean squared error (MSE) between outputs of the NN and targets are selected as the performance metric in training. Moreover, to avoid overfitting, cross-validation in the training set was conducted with 30% of the training inputs. Training iterations continue until the mean squared error of outputs in the validation set start to increase in six consecutive iterations (a sign for overfitting the training set) or iterations exceed the predetermined threshold (maximum of 1,000 iterations). There are four parameters for the NARX model that need to be tuned in order to minimise the NN error: input delay (τ_i), feedback delay (τ_f), number of hidden layers (H), and number of hidden neurons (h) in each hidden layer. In this work, only one hidden layer (H = 1) was assumed for training the network, and the other three parameters were tuned by running the NARX several times to reach the lowest MSE. The same trial and error method was repeated for the ARMA model with maximum likelihood estimation for calculating the coefficients in addition to Akaike information corrected criterion (AICC). By increasing the model order p and q, the residual variance of data decreases, so AICC penalises the model when a new parameter is added to it. This way the best model order p and q will be selected without overfitting the data.

5 Results and discussion

After training the NARX, input delay τ_i = 6, feedback delay τ_f = 1, and hidden neurons h = 1 were selected as the best NN parameters to minimise the testing set error. The NN weights and biases are updated in a direction to minimise the gradient error until a convergence is reached. The descent-of-error gradient and mean square error while training are shown in Figures 14 and 15. As shown in these figures, after about 20 iterations, the error gradient reached a convergence point. The performance of ARMA modelling is also shown in Table 4, where the best ARMA model with the lowest MSE was selected by choosing p, q = 2.

Figure 14 Gradient descent while training NARX (see online version for colours)

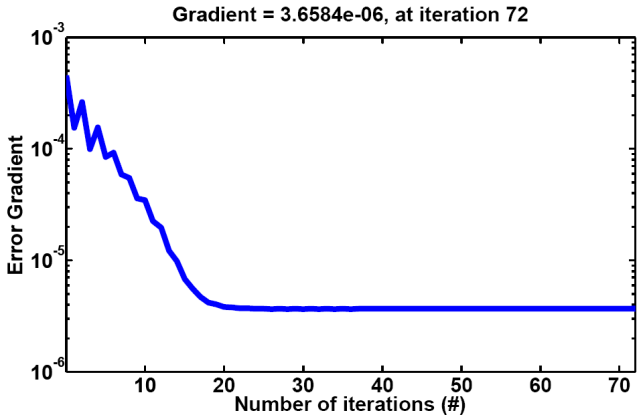


Figure 15 Mean squared error of targets and NARX output while training (see online version for colours)

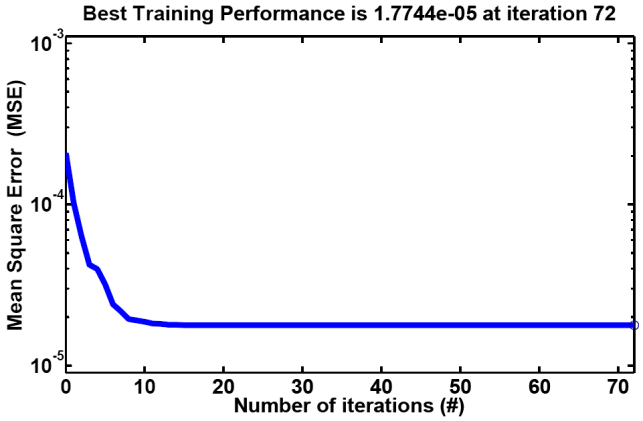


Table 4 Mean squared error of different ARMA output and measured tool wear

Coefficients	*ARMA(2,0)*	*ARMA(0,2)*	*ARMA(1,1)*	*ARMA(2,2)*	*ARMA(3,3)*
Φ_1	0.44	0	0.26	0.25	0.27
Φ_2	0.14	0	0	–0.78	0.80
Φ_3	0	0	0	0	0.34
Θ_1	0	0.53	0.16	0.22	0.24
Θ_2	0	0.31	0	1.00	1.00
Θ_3	0	0	0	0	0.04
MSE	3.06×10^{-5}	2.99×10^{-5}	3.10×10^{-5}	2.18×10^{-5}	2.19×10^{-5}

The estimation results using linear regression, ARMA, and NARX models for both the testing set and training set are shown in Figures 16 and 17. In Table 5, the measured and estimated tool wear and corresponding errors of testing set are compared. The maximum error of 33% for linear regression and 17% for time series model compare with 4% error of NN model, which demonstrates the usefulness of NNs for modelling nonlinear/non-stationary tool wear process. Moreover, one of the most important factors in model evaluation is the extent of the similarity between a model's output and the realistic/physical process. Tool wear is a monotonically increasing phenomenon, so any model that estimates tool wear should be able to generate a monotonically increasing output. Considering Figures 16 and 17, the NARX model is capable of estimating a monotonically increasing tool wear trend while the other models fail to do so. Note that due to measurement error (human error), even measured tool wear is not a monotonically increasing function. Finally, the RMS and average error were selected as two measures of performance and are compared in Figure 18 for linear regression, ARMA, and NN modelling methods.

Figure 16 NN and ARMA results for training set 1st and 2nd (see online version for colours)

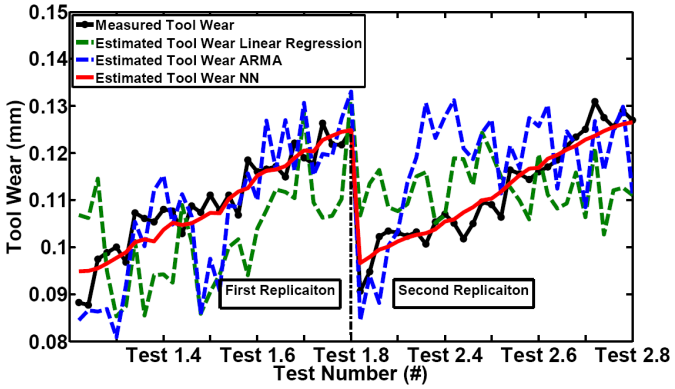


Figure 17 NN and ARMA results for testing set (see online version for colours)

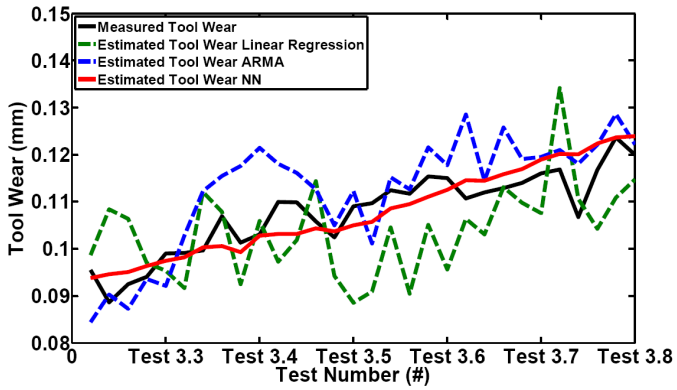


Figure 18 Error comparison of different modelling methods (see online version for colours)

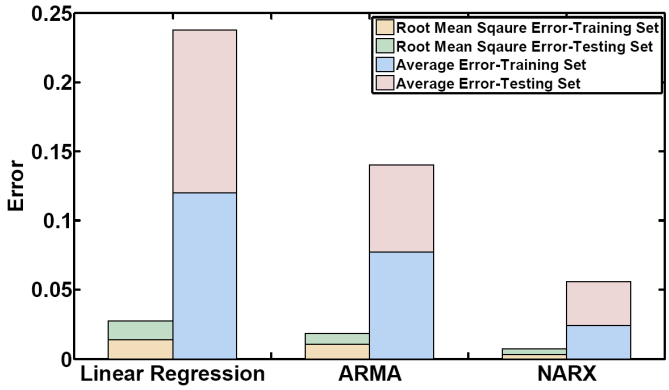


Table 5 Comparison of the tool wear measurement and the tool wear estimation in different method

Test	$VB_{measured}$	$V_{Bestimated}$ *Error*		$V_{Bestimated}$ *Error*		$V_{Bestimated}$ *Error*	
		Linear regression		*ARMA*		*NN*	
#	*(μm)*	*(μm)*	*%*	*(μm)*	*%*	*(μm)*	*%*
3.1	81	108	33	90	11	N/A	-
3.2	87	103	18	87	0	N/A	-
3.3	99	95	–4	92	–7	97	–2
3.4	103	106	3	121	17	103	0
3.5	109	88	19	112	3	105	–4
3.6	115	95	17	117	2	112	–3
3.7	116	107	7	119	3	119	3
3.8	120	114	–5	122	1	124	3

6 Conclusions

In this study, an auto-regressive moving average model and a recurrent NN with statistical features of wavelet packets were used for tool wear estimation during machining gamma-prime strengthened difficult-to-machine alloys. Tracking tool wear in modern manufacturing processes is critical due to the fact that it can simultaneously reduce machine downtime and increase productivity. Spindle power and acceleration were used as measured signals and decomposed with the WPD method to extract proper signal features. Since the Hall effect sensor and accelerometer that were used in this work are relatively inexpensive, this combined power and vibration method can be considered a cost effective method of measurement for monitoring manufacturing processes. The following major conclusions are drawn from this study:

- WPD was used to analyse measured signals in time-frequency domain. Eight different statistical features of the wavelet coefficients were extracted, and it was shown that decomposition of level 3 with Daubechies3 (Db3) wavelet and level 4 with Daubechies4 (Db4) wavelet produce the most significantly correlated features to tool wear.
- PCA was utilised to reduce the dimensionality of extracted feature vectors from 32 features to six new features that are the most significantly correlated ones to tool wear.
- ARMA time series model was trained with a maximum likelihood estimation method, and the best model with input/output delay of two was selected.
- The recurrent NN with NARX architecture was selected and trained with a Bayesian regularisation backpropagation method to avoid overfitting of noise in training set. It was shown that NN architecture with one input delay, six feedback delays, and one hidden neuron is able to estimate tool wear within a maximum of 72 iterations.

- To validate the performance of NN and time series model with unseen data, new set of data was fed after training ARMA and NARX models, and it was shown that the NN approach is capable of estimating with 43 μm RMS error and 4% maximum error. In addition, the realistic estimation of tool wear with different methods was discussed and it was shown that the NARX model is the best method for estimating the monotonically increasing tool wear function.

Data-driven modelling is appealing where models of the process are not available. Taking into account the important fact that the NN operates as a black box that only finds the mathematical relation between inputs and outputs, the physical realisation of the process is ignored when using this technique. Moreover, although tool wear estimation techniques are beneficial for modern manufacturing plants, they work as passive methods, which only warn before catastrophic failure and cannot be used to prevent such failure modes. Active methods, which can be used to identify the state of the tool, are at the same time capable of preventing the tool reaching into catastrophic failure regions, and/or slowing down the tool wear rate progress through adaptive control approaches. Such methods should be identified and investigated for active TCM, which is the subject for future studies.

References

Akhavan Niaki, F.A., Ulutan, D. and Mears, L. (2015a) 'Stochastic tool wear assessment in milling difficult to machine alloys', *International Journal of Mechatronics and Manufacturing Systems,* Vol. 8, Nos. 3–4, pp.134–159.

Akhavan Niaki, F.A., Ulutan, D. and Mears, L. (2015b) 'Parameter estimation using Markov chain Monte Carlo method in mechanistic modeling of tool wear during milling',in *ASME International Manufacturing Science and Engineering Conference,* June 8–12, pp.1–8.

Akhtar, W., Sun, J., Sun, P., Chen, W. and Saleem, Z. (2014) 'Tool wear mechanisms in the machining of Nickel based super-alloys: a review', *Frontiers of Mechanical Engineering,* Vol. 9, No. 2, pp.106–119.

Alaniz-Lumbreras, P.D., Gómez-Loenzo, R.A., Romero-Troncoso, R.d.J., Peniche-Vera, R.d.R., Jáuregui-Correa, J.C. and Herrera-Ruiz, G. (2006) 'Sensorless detection of tool breakage in milling', *Machining Science and Technology,* Vol. 10, No. 2, pp.263–274.

Antić, A., Šimunović, G., Šarić, T., Milošević, M. and Ficko, M. (2013) 'Model sustava za klasifikaciju trošenja alata pri obradi tokarenjem', *Tehnički vjesnik,* Vol. 20, No. 2, pp.247–254.

Attanasio, A., Ceretti, E., Rizzuti, S., Umbrello, D. and Micari, F. (2008) '3D finite element analysis of tool wear in machining', *CIRP Annals-Manufacturing Technology,* Vol. 57, No. 1, pp.61–64.

Baba, N. and Kozaki, M. (1992) 'An intelligent forecasting system of stock price using neural networks' in *IEEE International Joint Conference on Neural Networks,* p.371.

Bishop, C.M. (2006) *Pattern Recognition and Machine Learning,* Springer, USA.

Brockwell, P.J. and Davis, R.A. (2006) *Introduction to Time Series and Forecasting,* Springer Science & Business Media, USA.

Burden, F. and Winkler, D. (2009) 'Bayesian regularization of neural networks', in *Artificial Neural Networks,* pp.23–42, Springer, USA.

Cho, S., Binsaeid, S. and Asfour, S. (2010) 'Design of multisensor fusion-based tool condition monitoring system in end milling', *The International Journal of Advanced Manufacturing Technology,* Vol. 46, Nos. 5–8, pp.681–694.

Choi, Y., Narayanaswami, R. and Chandra, A. (2004) 'Tool wear monitoring in ramp cuts in end milling using the wavelet transform', *The International Journal of Advanced Manufacturing Technology,* Vol. 23, Nos 5–6, pp.419–428.

Chuangwen, X. and Hualing, C. (2007) 'A research of tool wear recognizing based on wavelet packet pretreated and neural network', *Journal of System Design and Dynamics,* Vol. 1, No. 4, pp.760–770.

Fang, N., Pai, P.S. and Edwards, N. (2012) 'Tool-edge wear and wavelet packet transform analysis in high-speed machining of Inconel 718', *Strojniški vestnik-Journal of Mechanical Engineering,* Vol. 58, No. 3, pp.191–202.

Fugal, D.L. (2009) *Conceptual Wavelets in Digital Signal Processing: An In-depth, Practical Approach for the Non-mathematician,* Space & Signals Technical Pub., USA.

He, K., Jia, M.P. and Zhao, Z.Z. (2013) 'Quality monitoring of surface roughness and roundness using hidden Markov model' in *Transaction of Technical Publication in Advanced Materials Research,* p.308.

Hong, G., Rahman, M. and Zhou, Q. (1996) 'Using neural network for tool condition monitoring based on wavelet decomposition', *International Journal of Machine Tools and Manufacture,* Vol. 36, No. 5, pp.551–566.

Jemielniak, K., Kossakowska, J. and Urbański, T. (2011) 'Application of wavelet transform of acoustic emission and cutting force signals for tool condition monitoring in rough turning of Inconel 625', *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* Vol. 225, No. 1, pp.123–129.

Kamarthi, S., Kumara, S. and Cohen, P. (2000) 'Flank wear estimation in turning through wavelet representation of acoustic emission signals', *Journal of Manufacturing Science and Engineering,* Vol. 122, No. 1, pp.12–19.

Ko, T. and Koren, Y. (1989) 'Cutting force model for tool wear estimation', *Ann Arbor,* Vol. 1001 p.48109.

Koren, Y. (1978) 'Flank wear model of cutting tools using control theory', *Journal of Manufacturing Science and Engineering,* Vol. 100, No. 1, pp.103–109.

Li, P., Hao, C. and Zhu, S. (2007) 'Machining tools wear condition detection based on wavelet packet' in *IEEE International Conference on Machine Learning and Cybernetics,* p.1559.

Li, X. and Patri, K.V. (1999) 'Wavelet packet transforms of acoustic emission signals for tool wear monitoring', *Journal for Manufacturing Science & Technology,* Vol. 1, No. 2, pp.89–93.

Morlet, J., Arens, G., Fourgeau, E. and Glard, D. (1982) 'Wave propagation and sampling theory-Part I: complex signal and scattering in multilayered media', *Geophysics,* Vol. 47, No. 2, pp.203–221.

Patra, K., Pal, S.K. and Bhattacharyya, K. (2007) 'Application of wavelet packet analysis in drill wear monitoring', *Machining Science and Technology,* Vol. 11, No. 3, pp.413–432.

Rabinowicz, E., Dunn, L. and Russell, P. (1961) 'A study of abrasive wear under three-body conditions', *Wear,* Vol. 4, No. 5, pp.345–355.

Sandvik, C. (2006) *Steel milling stars, GC4240 and GC1030, C-1140, A Pair of Grades that Won't Crack under Pressure*, Sandvik Coromant, Sweden.

Scheffer, C. and Heyns, P. (2004) 'An industrial tool wear monitoring system for interrupted turning', *Mechanical Systems and Signal Processing,* Vol. 18, No. 5, pp.1219–1242.

Scheffer, C., Engelbrecht, H. and Heyns, P. (2005) 'A comparative evaluation of neural networks and hidden Markov models for monitoring turning tool wear', *Neural Computing & Applications,* Vol. 14, No. 4, pp.325–336.

Segreto, T., Simeone, A. and Teti, R. (2012) 'Sensor fusion for tool state classification in nickel superalloy high performance cutting', *Procedia CIRP,* Vol. 1 pp.593–598.

Segreto, T., Simeone, A. and Teti, R. (2013) 'Multiple sensor monitoring in nickel alloy turning for tool wear assessment via sensor fusion', *Procedia CIRP,* Vol. 12 pp.85–90.

Takeyama, H. and Murata, R. (1963) 'Basic investigation of tool wear', *Journal of Manufacturing Science and Engineering,* Vol. 85, No. 1, pp.33–37.

Teti, R., Jemielniak, K., O'Donnell, G. and Dornfeld, D. (2010) 'Advanced monitoring of machining operations', *CIRP Annals-Manufacturing Technology,* Vol. 59, No. 2, pp.717–739.

Thakur, A. and Gangopadhyay, S. (2016) 'State-of-the-art in surface integrity in machining of nickel-based super alloys', *International Journal of Machine Tools and Manufacture,* Vol. 100 pp.25–54.

Trejo-Hernandez, M., Osornio-Rios, R.A., Romero-Troncoso, R.d.J., Rodriguez-Donate, C., Dominguez-Gonzalez, A. and Herrera-Ruiz, G. (2010) 'FPGA-based fused smart-sensor for tool-wear area quantitative estimation in CNC machine inserts', *Sensors,* Vol. 10, No. 4, pp.3373–3388.

Usui, E., Shirakashi, T. and Kitagawa, T. (1978) 'Analytical prediction of three dimensional cutting process – Part 3: cutting temperature and crater wear of carbide tool', *Journal of Manufacturing Science and Engineering,* Vol. 100, No. 2, pp.236–243.

Wang, M. and Wang, J. (2012) 'CHMM for tool condition monitoring and remaining useful life prediction', *The International Journal of Advanced Manufacturing Technology,* Vol. 59, Nos. 5–8, pp.463–471.

Wang, W., Hong, G., Wong, Y. and Zhu, K. (2007) 'Sensor fusion for online tool condition monitoring in milling', *International Journal of Production Research,* Vol. 45, No. 21, pp.5095–5116.

Wu, Y. and Du, R. (1996) 'Feature extraction and assessment using wavelet packets for monitoring of machining processes', *Mechanical systems and signal processing,* Vol. 10, No. 1, pp.29–53.

Xiaoli, L. (1999) 'On-line detection of the breakage of small diameter drills using current signature wavelet transform', *International Journal of Machine Tools and Manufacture,* Vol. 39, No. 1, pp.157–164.

Yen, Y., Söhner, J., Lilly, B. and Altan, T. (2004) 'Estimation of tool wear in orthogonal cutting using the finite element analysis', *Journal of Materials Processing Technology,* Vol. 146, No. 1, pp.82–91.

Zhang, K., Yuan, H. and Nie, P. (2015) 'A method for tool condition monitoring based on sensor fusion', *Journal of Intelligent Manufacturing,* Vol. 26, No. 5, pp.1011–1026.

Zhu, D., Zhang, X. and Ding, H. (2013) 'Tool wear characteristics in machining of nickel-based superalloys', *International Journal of Machine Tools and Manufacture,* Vol. 64 pp.60–77.