

# Tool wear monitoring and selection of optimum cutting conditions with progressive tool wear effect and input uncertainties

Sukhomay Pal · P. Stephan Heyns ·  
Burkhard H. Freyer · Nico J. Theron ·  
Surjya K. Pal

Received: 24 May 2009 / Accepted: 17 August 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** One of the big challenges in machining is replacing the cutting tool at the right time. Carrying on the process with a dull tool may degrade the product quality. However, it may be unnecessary to change the cutting tool if it is still capable of continuing the cutting operation. Both of these cases could increase the production cost. Therefore, an effective tool condition monitoring system may reduce production cost and increase productivity. This paper presents a neural network based sensor fusion model for a tool wear monitoring system in turning operations. A wavelet packet tree approach was used for the analysis of the acquired signals, namely cutting strains in tool holder and motor current, and the extraction of wear-sensitive features. Once a list of possible features had been extracted, the dimension of the input feature space was reduced using principal component analysis. Novel strategies, such as the robustness of the developed ANN models against uncertainty in the input data, and the integration of the monitoring information to an optimization system in order to utilize the progressive tool wear information for selecting the optimum cutting conditions, are proposed and validated in manual turning operations. The approach is simple and flexible enough for online implementation.

**Keywords** Tool wear · Monitoring · Neural network · Genetic algorithm · Wavelet packet analysis · Optimization · Turning operations

S. Pal · P. S. Heyns (✉) · B. H. Freyer · N. J. Theron  
Dynamic Systems Group (DSG), Department of Mechanical  
and Aeronautical Engineering, University of Pretoria,  
Pretoria 0002, South Africa  
e-mail: stephan.heyns@up.ac.za

S. K. Pal  
Department of Mechanical Engineering, Indian Institute of Technology  
Kharagpur, Kharagpur 721302, India

## Introduction

There has been significant interest in tool condition monitoring in the recent past (Byrne et al. 1995; Chao and Hwang 1997; Jemielniak et al. 1998; Sick 2002; Rehorn et al. 2005; Audy 2006; Wang et al. 2008; Purushothaman 2009). Various indirect methods for tool condition monitoring (TCM), which use a pattern in sensor data to detect a failure mode (Byrne et al. 1995; Scheffer and Heyns 2004; Heyns 2007), have been tried and tested by modeling the correlation between tool wear and sensory signals, namely the cutting force, torque, current, power, vibration, acoustic emission and airborne sound pressure acquired in machining processes. In fact, hundreds of research articles have been published on the subject of indirect TCM over the past decade, describing numerous methods of collecting process signals, the analysis and extraction of wear-sensitive features and modeling those features to correlate with tool wear. Reviews of current indirect TCM in turning can be found in Ref. (Byrne et al. 1995; Sick 2002; Rehorn et al. 2005; Heyns 2007). In general, cutting force is considered one of the most significant variables in the turning process (Sick 2002; Scheffer and Heyns 2004). Worn tools generally cause an increase in the static as well as dynamic components of the three orthogonal cutting force components, because of increased friction. Each of the three cutting force components, however, responds differently to varying machining parameters and different wear modes. Tool dynamometers are commonly used for recording these cutting forces (Rehorn et al. 2005). However, dynamometers are not suitable instruments for shop floor use due to their high cost, negative impact on machining system rigidity, the requirement for a wiring harness and the extra space for installation (Lee et al. 1995). Audy (2006) presents an overview of techniques and equipment used for measuring these cutting forces, using a strain gauge-based system

and Scheffer and Heyns (2004) reported the successful use of a simple sensor integrated tool holder using strain gauges. It was demonstrated that the system is robust, cost-effective and fit for use in an industrial TCM system.

The signal features extracted from only one sensor might not be sufficient to represent the complex relationship with tool wear (Dornfeld 1990). Therefore, it is preferable to employ multiple sensors instead of a single sensor to observe the same process. The most advantageous aspect of the multi-sensor system is perhaps the richness of information used in the decision making methodology. A current sensor is also considered one of the effective means of monitoring tool wear and adaptive control of machining processes (Mannan and Broms 1989), which does not hinder the machining process and is cost-effective (Byrne et al. 1995). This sensor signal individually or in combination with other sensors' signals can be used in a TCM system (Dornfeld 1990).

One difficulty with collecting wear-sensitive features from acquired signals is that these signals are typically affected by the process defects, process working conditions, process noise and sampling noise (Wu and Du 1996). This problem can be overcome by using advanced methods of signal processing. Recently, wavelet analysis has found important applications in signal analysis and processing in various science and engineering fields. It has also been successfully applied to TCM systems (Franco-Gasca et al. 2006). As an extension to the conventional wavelet transform, the wavelet packet transform decomposes both the approximate and detailed parts of a signal in order to extract more information from the signal. Wu and Du (1996) used the wavelet packet transform to extract the features for tool condition monitoring in turning and drilling processes. Berger et al. (1998), as well as Velayudham et al. (2005) used the wavelet transform as a tool to study the signal characteristics of cutting processes. However, these studies lack a detailed investigation of the selection of wavelet basis functions and decomposition levels. Furthermore, no information is given about the influence of the different wavelet basis functions on the signal features.

For proper classification, features that are sensitive to tool wear but insensitive to changing machining parameters are ideally required. In some cases, however, the machining parameters can be included in the wear model, then the sensitivity of the features to process parameters will become less important. Various techniques exist to select the most wear-sensitive features and to reduce the input feature matrix to a lower dimension. This is important, especially in multi-sensor based TCM systems, because of the computational expense of dealing with large input spaces and the fact that some of the data in the input space may be totally unrelated to the wear state. Including redundant data may reduce the accuracy of classification. Among the various techniques of feature selection and reduction, principal component

analysis (PCA) is considered one of the most popular (Jolliffe 1986).

Very often, artificial neural networks (ANNs) are used for classifying the tool condition, in which monitoring signal features are used to predict a tool condition. Various kinds of ANNs, such as Multilayer Perceptron (MLP), Recurrent Neural Network, Learning Vector Quantization, Time Delay Neural Network, Adaptive Resonance Theory, Self-Organizing Map, Radial Basis Function (RBF) network (Sick 2002) etc., have been tested in tool wear prediction problems. One of these, MLP, generally trained with a back-propagation error algorithm, is the most popular (Sick 2002). Dornfeld 1990, who pioneered the application of ANN to TCM, used a multilayer feed-forward neural network to integrate the features from multiple sensors (acoustic emission and force) in order to predict the tool wear in a turning operation. Scheffer and Heyns (2004) presented another approach to tool wear monitoring in turning operations, using a combination of static and dynamic ANNs. Much research has been done on TCM in turning operations using ANNs. The effectiveness and shortcomings of these studies have been reviewed by Sick (2002) and Rehorn et al. (2005).

The importance of TCM is implied by its potential economic advantages. It is possible to avoid the waste of product by exchanging worn tools in time, and tool costs can be reduced noticeably with precise exploitation of a tool's lifetime. Monitored information can be utilized for selecting the optimum cutting conditions, which further increases the economic advantages. However, most past research on optimizing the cutting conditions has been based on the critical assumption of machining with fresh cutting tools (Da et al. 1997). However, the machining performance may vary significantly with the progression of tool wear. Some works have been reported on including the effect of tool wear in the optimization process (Obikawa et al. 1996; Wang and Jawahir 2001). They used the Taylor tool life equation or an empirical relation to predict tool life, which is true only for small operating zone. Therefore, the use of this approach to the optimization problem may lead to suboptimal results.

TCM systems are not yet commonly used in industry, due at least partially to the lack of robustness of these systems regarding uncertainty in the input data. It has consequently been one of the goals of this research to test the robustness of the developed TCM systems against uncertainty in the input data, in this case exacerbated by the manual operation of the lathe. Monitoring information can be utilized for selecting the optimal cutting condition, which increases the economic advantages. Therefore, the integration of monitoring information into an optimization system, in order to utilize the progressive tool wear information for selecting the optimum cutting conditions, is also considered one of the goals. This proposed method is tested in manual turning operations.

## Experimental study

### Experimental set-up

A dedicated experimental study was conducted to provide the data required for this work. Turning operations on a manually operated lathe were considered and EN19 steel was chosen as the work piece material. A schematic diagram of the experimental set-up is shown in Fig. 1 with the equipment used for this work listed in Table 1. A tool holder instrumented with strain gauges was employed for the pick-up of cutting strain signals while pre-filtering for anti-aliasing and sampling was done by means of a data-acquisition system. The tool holder was instrumented so that it could sense bending caused by the elastic part of the force tangential to the work piece, bending caused by the elastic part of the force in the feed direction as well as longitudinal deformation caused by the elastic part of the thrust force in a radial direction to the work piece. The convention further is to use the terms tangential strain, feed strain and radial strain to describe the strain signals each of which is a linear combination of the force components acting at the tool tip. A 6667 Hz linear phase anti-aliasing filter was used as part of the data-acquisition system. Besides the cutting strain measurements, one phase of the lathe motor current was also acquired through a current transducer. All four signals were continuously acquired during turning operations at a sampling frequency of 20 kHz. Data storage and post-processing of the signals were then done on PCs.

### Experimental procedure

The experiment was designed so that measurements were taken and monitoring done for different machining conditions, consisting of all possible combinations of the cutting parameter settings considered, namely cutting speed ( $C_s$ ), depth of cut ( $a$ ), and feed rate ( $f$ ). A three-level, three-factor full factorial design was utilized. This design matrix required  $3^3$  i.e. 27 experimental runs. Cutting parameters with their notations, units and values at different levels are listed in Table 2. To increase the validation dataset, one more

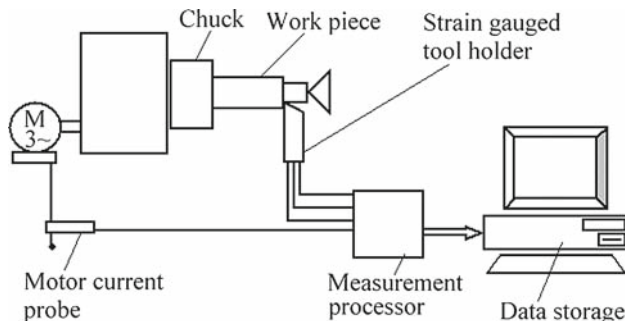


Fig. 1 Experimental set-up

Table 1 Experimental equipment

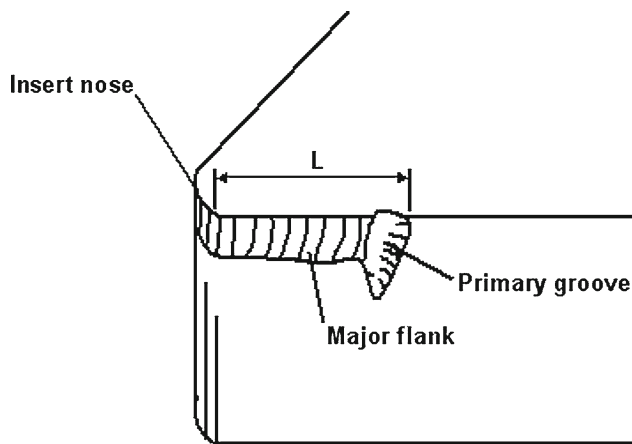
Machine, tools and work piece	
Lathe	Colchester student 1800 (manually operated)
Tool holder	SECO PTJNR-2020-16A
Insert	Mitsubishi TNGG 160408R
Work piece material	EN19 oil-quenched, tempered to T-condition, same batch
Work piece dimensions	Ø 100 to 89 mm × Length 500 mm
Instrumentation	
Strain sensing in feed direction	HBM strain gauges, 3/120ALY41, half-bridge
Strain sensing in tangential direction	HBM strain gauges, 3/120ALY41, half-bridge
Strain sensing in radial direction	HBM 90° rosette strain gauges, 1-XY31-6/120, full-bridge
Motor current sensing	Tektronix A622 AC/DC current probe with 100 mV/A setting
Flank wear measurement	Kyowa microscope, magnification: 40, 26 μm accuracy
Measurement processor	e-DAQ-lite by SoMat
Anti-aliasing	e-DAQ built-in 6667 Hz linear phase LP filter
Data storage	PC, Pentium 4, with SoMat test control environment (TCE) and InField software

Table 2 Cutting parameters and their different levels

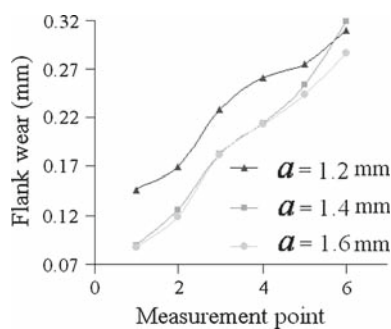
Parameter	Notation	Unit	Level 1	Level 2	Level 3
Cutting speed	$C_s$	m/mm	128.8	119.8	114.6
Depth of cut	$a$	mm	1.2	1.4	1.6
Feed rate	$f$	mm/rev	0.1	0.16	0.2

experiment was randomly performed. All other controllable parameters remained the same except for the abovementioned three cutting parameters.

The flank wear was measured from a photo as  $V_B = A/L$ , where  $V_B$  is the flank wear in mm,  $L$  is the length of the worn area along the upper edge on the side of the major flank and  $A$  is the area of the major flank inclusive of the area of the primary groove (see Fig. 2). At each experimental/cutting condition a fresh edge of an insert was used. Tool wear was measured after every two passes for Exp. No. 1–9, three passes for Exp. No. 10–27, and eight passes for Exp. No. 28 and 500 mm length of cut was used for each pass. Turning operations were halted when flank wear had reached close to the wear limit of 0.3 mm. Therefore, six measurements of



**Fig. 2** Flank wear measurement



**Fig. 3** Flank wear of tool insert with time ( $C_s$  is 128.8 m/min and  $f$  is 0.1 mm)

tool wear were taken for each cutting condition for the first 27 experiments, and five measurements of tool wear were taken for the last experiment. A comparison of flank wear with different machining parameter setting is shown in Fig. 3.

### Signal processing and feature selection

The signals measured during the monitoring of the tool wear process contain complex information and noise. To ensure the accuracy and reliability of the monitoring system, it is very important to extract characteristic features which capture this complexity sufficiently, without the noise. The force and current signals measured during a turning process are typically non-stationary (Sick 2002). Therefore, time-domain analysis of these signals may not be ideal for developing a TCM system for a turning process. More advanced signal processing must be used to remove unnecessary features from the signal and to extract the features that are more sensitive to tool wear. The wavelet packet transform is used here for this purpose.

### Wavelet packet analysis

Wavelet analysis has been demonstrated to be a flexible and powerful tool for the processing of non-stationary signals, because wavelet basis functions are usually irregular, asymmetric and of limited duration. The main feature that makes wavelets attractive is that the Wavelet Transform (WT) of a signal is a function of two parameters, namely time and scale, the latter being the key point of the WT. Unlike the windowed Fourier Transform (WFT), where the signal is filtered through a time window and then a Fourier transform is performed which works with a fixed window, the WT works with a scaled window, allowing the visibility of the entire frequency content.

In wavelet packet analysis, a signal is split into a low-frequency component, known as approximate, and a high-frequency component, known as detail. The approximate and detail parts are then both split into a next level of approximation and detail, and this process is repeated. Wavelet packets are particular linear combinations of wavelets. They form the bases that retain the orthogonality, smoothness and locational properties of their parent wavelets (Keinert 2004).

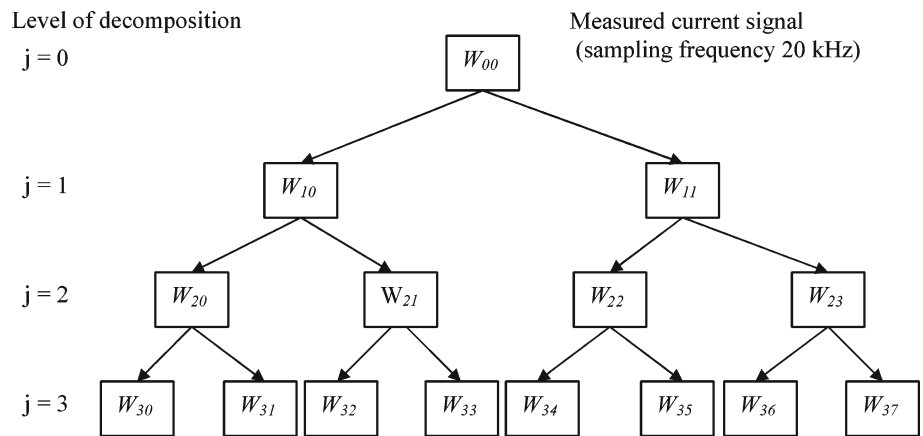
### Selection of optimal decomposition level

In this research, wavelet packet analysis was performed by using MATLAB 7.0 (Mathworks 2007) built-in functions for 1-D wavelet packet analysis. This is illustrated in Fig. 4 which shows the third level of the wavelet packet split tree of the current signal, acquired in this work at a 20 kHz sampling rate (i.e. 10 kHz useful bandwidth).  $W_{00}$  represents the measured current signal, and  $W_{30}$  through  $W_{37}$  are the eight wavelet packets (eight frequency bands) at the third level of decomposition. The decomposition may be continued down to the final level where there is only one element in each basis vector. In this case, the maximum resolution to which the binary wavelet transform can be performed is:

$$j = \log_2 N \quad (1)$$

i.e.  $j \geq 14$  where  $N$  is the sampling frequency and  $j$  is the decomposition level (Mathworks 2007). As this number is very large, the number of wavelet packets becomes very high and computational time will also be high, which is unsuitable for any monitoring system. There are some possible ways for determining the best wavelet packet decomposition levels, like, best on desired frequency band(s), viewing the decomposition signals at different levels, and minimum entropy energy criterion (Mathworks 2007). Among them the first two methods are not robust. Therefore, in this research the minimum entropy energy criterion was used to select the best wavelet packet decomposition level. According to the entropy criterion, a splitting is interesting if the entropy of a parent packet is more than that of the total entropy of the

**Fig. 4** Third level of wavelet packet decomposition tree of the measured current signal



child packets (Mathworks 2007). The Shannon entropy of a signal  $S$  can be computed by the following equation:

$$E(S) = - \sum_{k=1}^N S_k^2 \log(S_k^2) \tag{2}$$

where  $E(S)$  is the Shannon entropy of the signal  $S$  and  $S_k$  is the  $k$ th value of the signal. The entropy of all wavelet packets at a given level and the next level were computed using the abovementioned equation. Then the child entropies were compared with their corresponding parent entropy, using the following equation:

$$EI = E_j^i - E_{j+1}^{2i} - E_{j+1}^{2i+1} \tag{3}$$

where  $EI$  is the entropy increase and  $E_j^i$  is the entropy of the  $i$ th packet at the  $j$ th level. If more than 50% of packets at a level become unsuitable for next-level decomposition, i.e. if in 50% of cases  $EI$  became less than zero, the present level was considered as the optimal level. The optimal levels for strain in feed direction, strain in tangential direction, strain in radial direction and current signals were found at the 4th, 5th, 4th and 3rd levels, respectively.

The RMS values of the Wavelet Packet Coefficients (WPCs) in each frequency band were used as wear-sensitive features. The RMS value of WPCs in  $W_{3i}$  packet or frequency band of  $i \times 1250$  Hz to  $(i + 1) \times 1250$  Hz,  $i \in [0, 7]$ , is denoted as  $n_i$ .

### Selection of best mother wavelet function

In the common family of wavelet mother functions, there are Morlet, Haar, Shannon, Symmlets, Coiflets and Daubechies wavelets, etc. Among these, the most popular are the Daubechies, Symmlets and Coiflets wavelet. These mother wavelets give the best overall performance in respect of the mean squared error between reconstruction signal and original signal, and maximizing the signal-to-noise ratio improvement (Mittermayr et al. 1996). However, the performance of the

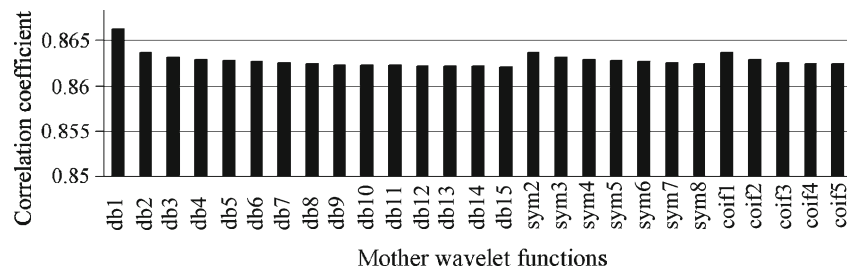
abovementioned wavelet mother functions in extracting the feature sensitive to wear quality may vary according to the mother wavelet properties (Mathworks 2007). Therefore, it is necessary to select the best mother wavelet function for processing the turning signals in the wavelet domain. The best mother wavelet function was selected, using the correlation coefficient between the average root mean square (RMS) values of WPCs and the tool wear. The correlation coefficient (Scheffer and Heyns 2004; Mathworks 2007) between the average RMS values of WPCs and tool wear can be represented as follows:

$$\rho = \left( \frac{\sum_i (V_{Bi} - \bar{V}_B) (F_i - \bar{F})}{\sqrt{\sum_i (V_{Bi} - \bar{V}_B)^2 \sum_i (F_i - \bar{F})^2}} \right) \tag{4}$$

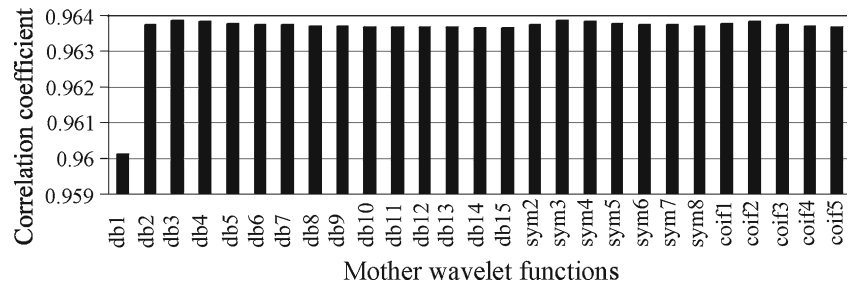
where  $V_B$  is the flank wear,  $F$  is the feature, and  $\bar{V}_B$  and  $\bar{F}$  are the average of  $V_B$  and  $F$ , respectively.

In the MATLAB library (Mathworks 2007), a maximum of 15 mother wavelets or wavelet bases is available for the Daubechies family (dbN, where N is the order), 7 for the Symmlets family (symN), and 5 for the Coiflets family (coifN) wavelets. The correlation coefficient between the average RMS values of WPCs for different signals with various mother wavelets (available in MATLAB library) and tool wear are shown in Figs. 5, 6, 7, 8. These figures show that the degree of correlation of db1, sym3, db1 and sym2 mother wavelet functions for radial strain, tangential strain, feed strain and current signal processing, respectively, with tool wear are higher than the other mother wavelet functions. Therefore, to extract features from the radial strain, tangential strain, feed strain, and current signals db1, sym3, db1 and sym2 mother wavelet functions, respectively, were considered the best mother wavelet functions. Since the variation in the correlation coefficients (Figs. 5, 6, 7, 8) among the different mother wavelet functions is insignificant. Therefore, any of these mother wavelet functions may be used for feature extraction.

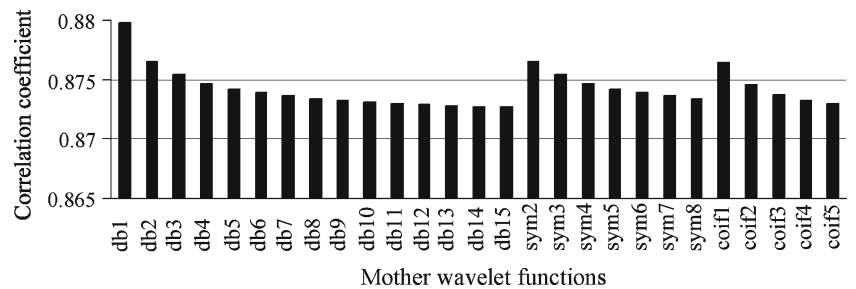
**Fig. 5** Comparison of correlation coefficients of different wavelet mother functions for strain in radial direction signal



**Fig. 6** Comparison of correlation coefficients of different wavelet mother functions for tangential strain signal



**Fig. 7** Comparison of correlation coefficients of different wavelet mother functions for feed strain signal



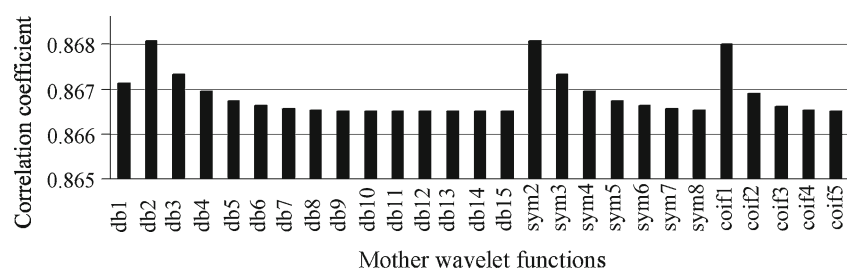
Feature selection

Once a list of possible features i.e. the WPCs of different signals have been generated, the next step is to select the most reliable features for the TCM system. The number of features selected for radial, tangential and feed direction strain, and current signal are 32, 64, 32 and 8, respectively. As it is inadvisable to handle such a large number of features in a monitoring system, principal component analysis (PCA) was used to reduce the dimension of the signal features and to select the most wear-sensitive uncorrelated features.

Principal component analysis

PCA is one of the most popular multivariate statistical methods. It is a simple, non-parametric method of extracting relevant information from confusing datasets without much loss of information. PCA is mathematically defined (JolliHe 1986) as an orthogonal linear transformation that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for the greatest possible statistical variability (or entropy) in the data and the

**Fig. 8** Comparison of correlation coefficients of different wavelet mother functions for current signal



second-greatest variance on the second principal component, and so on. PCA is theoretically the optimum transform for the given data in least-square terms.

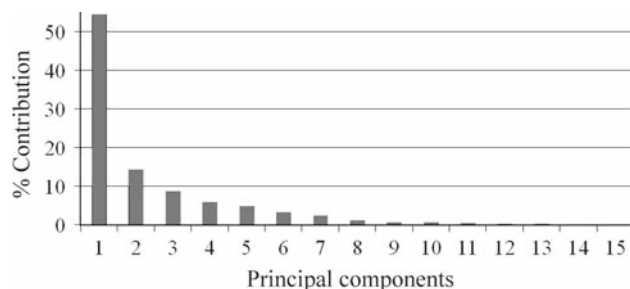
#### Calculation of principal components

Principal components are found by extracting the eigenvectors and eigenvalues of the covariance matrix of the data, and are calculated efficiently via singular value decomposition. These eigenvectors describe an orthonormal basis that is effectively a rotation of the original cartesian basis. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data (Jolliffe 1986).

The principal components are created in order of decreasing variance, and each eigenvalue,  $\lambda_k$  ( $k = 1, 2, \dots, N_f$ ), represents the relative variance contribution of each principal component. The percentage of the total variability explained by each principal component is calculated by using following equation:

$$P_k^\lambda = \frac{\lambda_k}{\sum_{j=1}^{N_f} \lambda_j} 100 \quad (5)$$

where  $P_k^\lambda$  is the percentage of total variability explained by the  $k$ th principal component, and  $N_f$  is the number of features. The percentage variability explained by the first 15 principal components is shown in Fig. 9. The first, and largest, eigenvalue represents 54.53% of the overall variance, whereas the first eight components represent some 95.58% of the variance. The individual and total contributions of all other components are less than 1 and 4.42%, respectively. The largest differences occur between the first few principal components (up to eight), after which differences become quite small, indicating that the first few components hold the largest fraction of the variance. Principal components ranked higher than eight are small and probably represent the noise of the data. Therefore, the first eight principal components were selected to represent the data dimension.



**Fig. 9** Contribution of different principal components to the overall variance

#### Modeling of tool wear using an artificial neural network

A standard back-propagation neural network (BPNN) and an RBF network (RBFN) were used for modeling the turning tool wear. Computer programs for these ANN models were developed using the C programming language. In general there is no exact way of determining the size of a training set. However, too few training samples will result in the neural network not being able to learn the input–output relationships, and too many training samples will result in a neural network that overemphasizes the training, jeopardizing the generalization capabilities. Generally, for small data sets some 15–20% data are used for testing (Wang et al. 2008; Jemielniak et al. 1998). Because of the costs of the experiments 15% data was used as testing data in this case. To ensure the generality of the developed models, three (training, testing, and validation) datasets were generated randomly. Among the full factorial design dataset, 142 and 15 patterns were randomly selected for training and testing the models, respectively. The remaining ten patterns (five patterns from full factorial design data set and five patterns from Exp. No. 28) were used for the validation of the developed models. Over-fitting of the network was avoided by periodically checking the testing dataset in between the training set-based learning iterations. Only if both the training error and the test error continued to drop, was the learning process continued. The first eight principal components and cutting conditions such as cutting speed, depth of cut, and feed rate were considered as inputs to the ANN model.

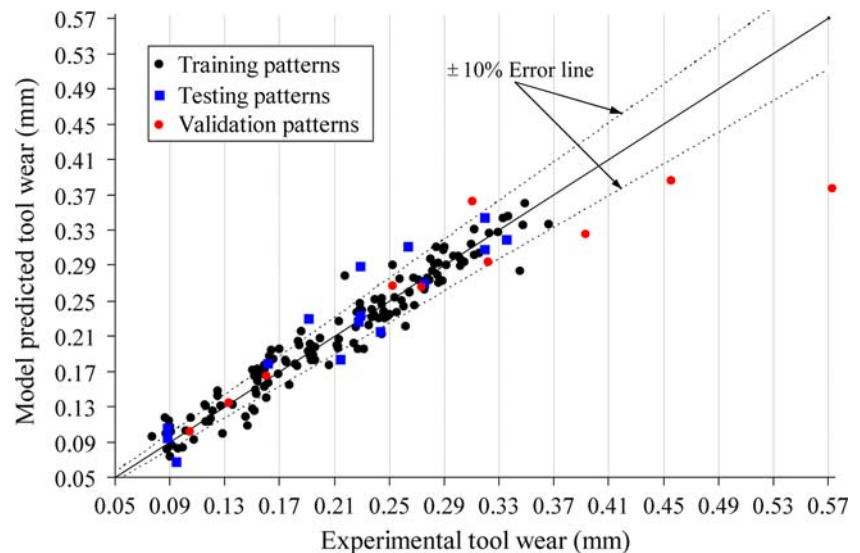
#### Prediction of tool condition using BPNN model

The performance of a BPNN model depends on the number of hidden layers, the number of neurons in the respective hidden layers, the learning rate, and momentum coefficient. Therefore, several combinations should be tried out to choose an optimal combination. Here a single hidden layer was assumed. The number of neurons in the hidden layer was varied from 2 to 30 in steps of 1. The learning rate and momentum coefficient were varied between 0.05 and 1 in steps of 0.04. Initial weight values were chosen randomly between  $\pm 0.9$ , and the bias value at the input layer was taken as 0 and those of the hidden and output layers as 1.0. All the input and output variables were normalized between 0.1 and 0.9. The training objective was mean square error (MSE) minimization by updating the weights through the gradient descent method (Haykin 2003).

$$\text{MSE} = \frac{1}{2N} \sum_i^N (T_i - O_i^m)^2, \quad (6)$$

**Table 3** Prediction error statistics of the ANN model

Pattern	BPNN model				RBFN model			
	Minimum	Maximum	Average	Standard deviation	Minimum	Maximum	Average	Standard deviation
% Error on training patterns	0.197	-36.87	7.53	6.67	0.001	-37.09	5.03	6.71
% Error on testing patterns	-0.088	28.20	10.01	9.18	-0.32	36.48	7.63	9.57
% Error for validation patterns	-1.79	34.18	10.76	10.29	0.004	60.74	22.63	23.35

**Fig. 10** Scatter diagram of BPNN model

where  $N$  is the total number of training datasets,  $T_i$  is the target output of the  $i$ th dataset i.e. experimental output of the  $i$ th dataset, and  $O_i^m$  is the output from the ANN model on the  $m$ th iteration when the  $i$ th dataset is considered as the network input. The best architecture was found to be 11-15-1, with learning rate and momentum coefficient values of 0.95 and 0.6, respectively. The prediction error statistics of the training, testing, and validation patterns of the BPNN model are shown in Table 3. The scatter diagram of training, testing, and validation patterns is shown in Fig. 10. This scatter diagram (Fig. 10) shows that the prediction accuracy at higher tool wear is higher than that at lower tool wear.

#### Prediction of tool condition using RBFN model

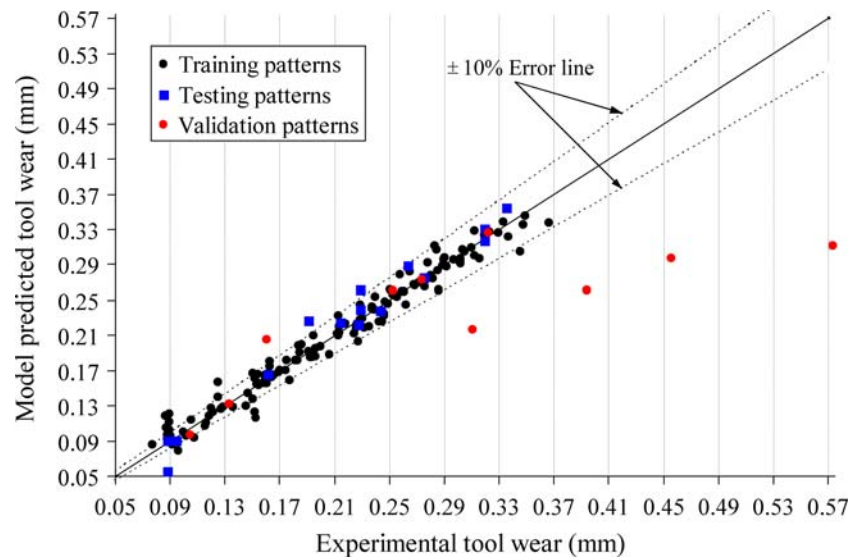
A special class of multi-layer feed-forward network is the radial basis function neural network (Haykin 2003). This network provides an alternative to BPNN as a tool for classification and function approximation. This technique often proves more accurate than other network models (Pal et al. 2007). The RBFN is trained by adaptively updating the free parameters, i.e. center and width of the basis function, and the weight between the hidden and output neurons of the

network. Two types of training algorithms (Pal et al. 2007) can be used, namely half training and full training. In the case of half training, the hidden layer parameters, i.e. center vector and width of basis function, were updated in an unsupervised manner whereas the connection weights between the hidden and the output layer were adjusted in a supervised manner. By contrast, in full training algorithms, all free parameters (center vector, width of basis function, and weights between the hidden and the output layer) were updated in the supervised manner. A detailed description of this network is available in (Pal et al. 2007). Among the various combinations of activation function at the output layer (linear, pseudo linear, and sigmoid) and training algorithms (half and full training), the sigmoid activation function for the output layer with full training algorithm, is the best (Pal et al. 2007). Therefore, in this research the sigmoid activation function at the output layer with full training algorithm was used for developing the RBFN model.

To select an optimal RBFN model, the number of neurons in the hidden layer was varied from 2 to 30, and the learning rate was varied between 0.05 and 0.5. The initial basis function centers were chosen randomly from the input space, and the initial weight values were chosen randomly



**Fig. 11** Scatter diagram of RBFN model



between  $\pm 0.9$ . The same normalized datasets as those used in the BPNN model, were used for the training, testing, and validation of the RBFN model. The best network was found to be one having 26 basis functions with a learning rate of 0.9, and 0.05 for center and weight, respectively. The prediction error statistics of the training, testing, and validation patterns of the RBFN model are shown in Table 3. The scatter diagram of the training, testing, and validation patterns is shown in Fig. 11. The prediction errors of the validation patterns are larger because these patterns are outside the training space.

### Evaluation of neural network robustness against uncertainty

Neural network predictions exhibit uncertainty due to inaccuracies in the training data and to the limitations of the model. The training set contains intrinsically noisy and incomplete (not all possible input-output examples are available) patterns. In reality, data noise may arise from inaccurate measurements or human error. For the manually operated lathe considered here, slight inconsistencies in the process can be expected to be important. Data noise is not the only source of uncertainty. The neural network introduces uncertainty due to model misspecification and the inefficiencies of the training method (Papadopoulos and Edwards 2001). A network trained on a given dataset forms a better representation of the data in regions of high input data density (Papadopoulos and Edwards 2001). Moreover, because of the nature of the training algorithm, there is no guarantee that the weight values correspond to the global minimum of the error function. Even if the global minimum is found, the solution will not necessarily be optimal because the finite training set does not fully describe the true data-generating mechanism (Papadopoulos and Edwards 2001).

Having investigated the network prediction performance to crisp (i.e. single-valued) values, the next step was to quantify the robustness of the models to variations in the input data. Several approaches have been developed for output confidence interval predications (Papadopoulos and Edwards 2001); they all adopt a probabilistic standpoint and therefore suffer from the common drawback that, since the probability distributions are usually estimated from the low-order moments of the data (typically mean and standard deviation), there is often no validation of the extremes of the distributions. An alternative approach to this issue is the application of an interval-based non-probabilistic (Papadopoulos and Edwards 2001) technique or the Monte-Carlo technique to evaluate the robustness of the neural network to uncertainty in input data. The interval-based technique is posited on the theory of information-gap uncertainty and its advantage lies in presenting both crisp data, and interval (Ben-Haim 2001) data to a number of neural networks under evaluation. We do not consider the computationally expensive Monte-Carlo technique, which involves randomizing the input data (within defined bounds) a large number of times, and evaluating the change in output error for each discrete set of inputs. This technique has a significant drawback, especially when applied to MLP networks, namely that it is impossible to be sure of mapping all possible combinations of variation in input space to output space, unless an unfeasibly large number of sample points are used.

### Formulation of interval-based confidence estimation technique

In this research, an interval-based technique was used for the evaluation of neural network robustness to uncertainty in the input data. The main advantages of the interval approach are

that it is possible to establish worst-case error predictions for a specified degree of input uncertainty to the network and also to remove the dependence on probabilistic-based estimates of confidence bounds. This approach can be used as an alternative to selecting a particular ANN model which is intrinsically more robust to uncertainty on the input data than network solutions obtained by conventional training paradigms.

The dataset was made uncertain by applying an interval expansion of size  $\alpha$  in all dimensions of the datasets. Interval numbers (Ben-Haim 2001; Papadopoulos and Edwards 2001) occupy a bounded range of the number line, and can be defined as an ordered pair of real numbers  $[a,b]$  with  $a < b$  such that

$$[a, b] = \{x \mid a \leq x \leq b\} \tag{7}$$

Therefore, each input dimension of the dataset was internalized by a parameter  $\alpha$  using the following equation:

$$[x_{ij}^+, x_{ij}^-] = [(x_{ij} + \alpha), (x_{ij} - \alpha)] \tag{8}$$

where  $x_{ij}$  is the  $i$ th dimension of the  $j$ th pattern, and  $x_{ij}^+$  and  $x_{ij}^-$  are the upper and lower interval bounds of the  $x_{ij}$ , respectively. This interval input dataset was then forward-propagated through the developed networks, and the interval output values  $[y^+, y^-]$  were compared with the target values to calculate worst case (WC) and best case (BC) (Ben-Haim 2001; Papadopoulos and Edwards 2001) average percentage errors. The WC and BC error were calculated using the following relations:

$$WC = \frac{100}{N} \sum_{i=1}^N \left\{ \begin{array}{l} \left| \frac{(y_i^+ - t_i)}{t_i} \right| \text{ if } |y_i^+ - t_i| > |y_i^- - t_i| \\ \left| \frac{(y_i^- - t_i)}{t_i} \right| \text{ if } |y_i^+ - t_i| \leq |y_i^- - t_i| \end{array} \right\} \tag{9}$$

$$BC = \frac{100}{N} \sum_{i=1}^N \left\{ \begin{array}{l} 0 \text{ if } y_i^+ > t_i > y_i^- \\ \left| \frac{(y_i^+ - t_i)}{t_i} \right| \text{ if } |y_i^+ - t_i| < |y_i^- - t_i| \\ \left| \frac{(y_i^- - t_i)}{t_i} \right| \text{ if } |y_i^+ - t_i| \geq |y_i^- - t_i| \end{array} \right\} \tag{10}$$

where  $t_i$  is the target values of the  $i$ th pattern, and  $y_i^+, y_i^-$  are the upper and lower interval bound of the  $i$ th output pattern, respectively. The worst-case errors corresponded to the furthest output bound from the target, and the best-case errors to the closest, with the proviso that the best case errors would be zero if the targets were contained wholly within the output interval.

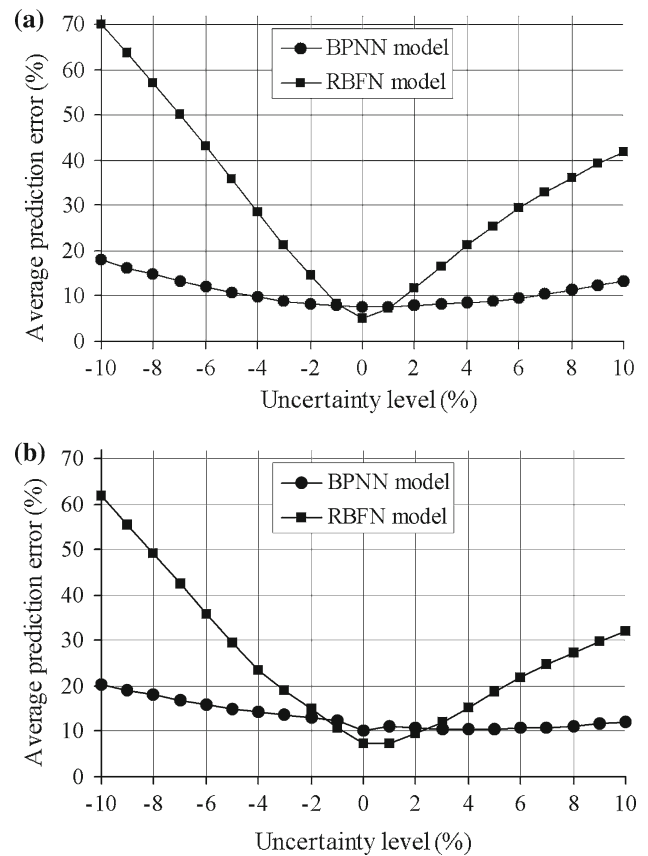


Fig. 12 The variation in the average prediction error under different uncertainty levels

### Forward propagation of intervals through networks

Using conventional network training and evaluation, the network architectures of 11-15-1 and 11-26-1 were found to be the best architectures for the BPNN and RBFN models, respectively. The responses of these two networks to uncertainty in the input data were investigated. Discrete values of uncertainty 1–10% in steps of 1% were applied to the training and testing patterns, and the resulting intervalized sets propagated through the abovementioned two networks. The variation in the average prediction error of training and testing patterns under different uncertainty levels for both the models is shown in Fig. 12a, b. WC and BC errors were calculated using Eqs. (9) and (10), and these errors were plotted as functions of interval size, as shown in Fig. 13. Such a plot is useful, as it allows a designation of a permitted level of input data fluctuation to guarantee a particular absolute worst-case average error on the output. For example a designated allowed average worst-case error of 21% corresponds to an uncertainty in input parameters of 7% and 3% for the BPNN and RBFN models (shown in Figs. 12 and 13), respectively.

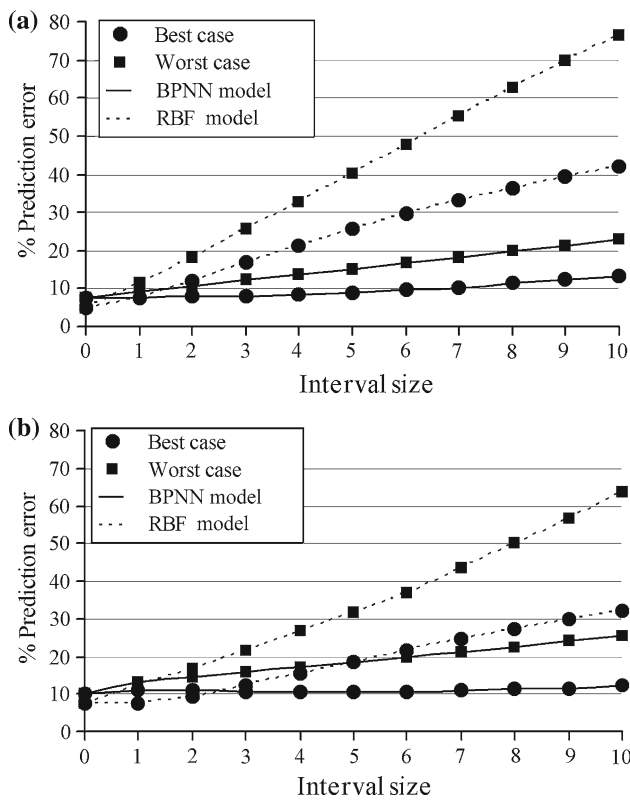


Fig. 13 Worst-case and best-case errors at different interval size a Training, b Testing

Selection of robust model using interval-based techniques

From the tabulated data (see Table 3) and figures (see Figs. 10, 11) it may be argued that the prediction performance of RBFN model is better than the BPNN model under normal conditions. But the robustness of the BPNN model to uncertainty in the input data is better than that of the RBFN model. Figure 13 shows the worst-case and best-case errors at different interval sizes of both the models. The worst-case error of a RBFN model with 1% or more uncertainty in the input data is more than that for the corresponding BPNN model’s worst-case error (see Fig. 13). It can also be seen in Fig. 13 that the average prediction performance of the RBFN model is worse than the BPNN model when the uncertainty level is close to more than 1%.

Optimization of cutting condition with consideration of the progressive effect of tool wear

The optimization of cutting conditions in turning operations plays an important role in process planning. It is generally done by assuming that the cutting tools are fresh, and that all passes, or all passes except the finishing pass, have equal cutting conditions (Da et al. 1997; Wang and Jawahir

2001). This usually gives suboptimal results in real machining operations. In an actual machining process, the machining performance may vary significantly as overall tool wear progresses (Wang and Jawahir 2001). In this paper, a genetic algorithm (GA) based optimization method is developed for selecting the optimum cutting conditions for each pass of a turning operation, with consideration of the effect of overall progressive tool wear on the machining performance.

Objective function and constraints on optimization

In the constructed optimization problem, three cutting conditions, namely cutting speed, feed, and depth of cut are considered the decision variables. Two different mutually conflicting objectives are optimized in this research. The first objective is the minimization of wear in the tool used, considered as the part of the whole tool wear, described as follows:

$$\xi = \frac{V_B^i}{V_B^{\max} - \text{initial tool wear}} \tag{11}$$

$$V_B^i = f(\text{initial tool wear}, C_s, f, a) \tag{12}$$

where  $V_B^i$  is the amount of tool wear used at the  $i$ th pass, and  $V_B^{\max}$  is the maximum acceptable tool wear. The second objective is the machining time ( $\tau$ ), measured as the time required to carry out the process:

$$\tau = \frac{V}{\text{MRR}} \tag{13}$$

$$\text{MRR} = f(C_s, f, a) \tag{14}$$

where  $V$  is the amount of material to be removed, and MRR is the material removal rate. Hence, the optimization problem becomes the minimization of

$$f(C_s, f, a, \text{initial tool wear}) = \xi + \tau \tag{15}$$

subject to

$$R_a < R_a^{\max} \tag{16}$$

$$C_s^{\min} \leq C_s \leq C_s^{\max} \tag{17}$$

$$f^{\min} \leq f \leq f^{\max} \tag{18}$$

$$a^{\min} \leq a \leq a^{\max} \tag{19}$$

where  $R_a^{\max}$  is the maximum allowable surface roughness, and  $C_s^{\min}$  and  $C_s^{\max}$  are the minimum and maximum cutting speeds, respectively,  $f^{\min}$  and  $f^{\max}$  are the minimum and maximum feed rates, and  $a^{\min}$  and  $a^{\max}$  are the minimum and maximum depths of cut.

Prediction of machining performance

The prediction of machining performance is prerequisite for the optimization process. Most of the earlier works used the extended Taylor tool life equation to predict tool life (Da et al. 1997; Wang and Jawahir 2001). But this equation is true only

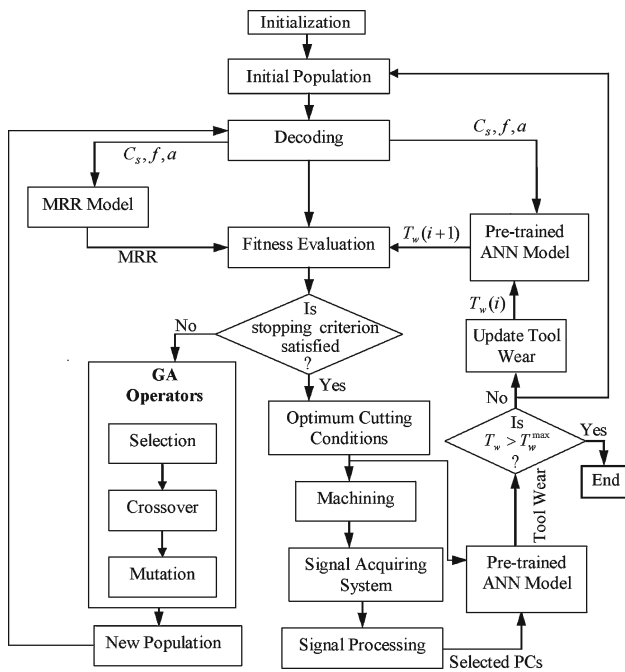


Fig. 14 Flow chart of genetic algorithm-based optimization method

for a small operating zone, and an accurate estimation of the empirical constants of this equation is very difficult. Another important point is that tool wear is a dynamic phenomenon. Use of this equation to solve the optimization problem gives a suboptimal result. Therefore, an ANN-based monitoring system was developed, as described above, for the prediction of tool wear. Surface roughness was calculated using the following empirical relationship (Wang and Jawahir 2001)

$$R_a = \left( \frac{0.125 f^2}{r_E} \right) \times (1 + 1.6103 T_w)^{0.7315} \quad (20)$$

where  $r_E$  is the tool nose radius. And MRR was calculated using the following equation:

$$\text{MRR} = 1000 C_s f a \quad (21)$$

#### Development of optimization technique using genetic algorithm

The schematic diagram of the developed GA-based optimization method is shown in Fig. 14, and the working principle is elaborated below. In order to use GAs to find the optimal decision variables, one first needs to represent them in binary strings. In this research, 7, 4 and 5 bits were used to represent cutting speed, feed, and depth of cut, respectively. Depending on the population size, a few hundred such strings were generated for the initial population, i.e. the initial possible solution. However, these binary strings had to be converted into real numbers, i.e. decoding in order to compute the fitness function. The binary strings were decoded using the following equation:

Table 4 Range of different parameters with optimum cutting condition

Parameter	Minimum	Maximum	Optimal cutting condition	
			Fresh tool	0.22 mm initial tool wear
Cutting speed (m/min)	117.41	137.38	117.41	117.41
Feed (mm/rev)	0.105	0.212	0.1692	0.1549
Depth of cut (mm)	1.14	1.65	1.65	1.65

$$x_i = x_i^{\min} + \frac{b_i}{(2^{n_{bi}} - 1)} (x_i^{\max} - x_i^{\min}), \quad (22)$$

where  $x_i$  is the  $i$ th input parameter (variable),  $x_i^{\min}$  and  $x_i^{\max}$  are the lower and upper bounds of the  $i$ th input parameter, respectively,  $n_{bi}$  is the string length used to code the  $x_i$  parameter, and  $b_i$  is the decoded value of the string  $s_i$  (where the complete string is the  $s = \cup_{i=1}^n s_i$ ,  $n$  is total number of variables). The ranges of the cutting conditions are shown in Table 4.

Two ANN models were used in this optimization process. The first one was used for predicting the amount of current tool wear, which takes cutting condition and different signals features as inputs. By contrast, the second one was used for predicting the expected tool wear for the next pass, which takes the current tool wear and cutting condition as inputs. The tool wear of a fresh tool bit was considered zero. Constraints were handled by penalty function approach, which converts a constrained problem to an unconstrained form by modifying the search space.

#### Determination of optimum cutting condition

The initial set of parameter settings or the initial population was chosen randomly inside the selected experimental parameter space. After making the selection, the response characteristics of individuals in the population were computed using a previously trained ANN model and Eq. (13) for tool wear and machining time, respectively, and fed into the GAs. Then the GAs performed different genetic operations to generate a new population. The response characteristics of this population were again computed and fed into the GAs. The process continued until the optimum was found, i.e. the population is assumed to have converged when 70% of the population has the same objective function value.

In GA, the population size, crossover rate (probability), and mutation rate are important factors that influence the performance of the algorithm (Deb 1996). Therefore, a large number of combinations were tried to select a suitable value of these parameters. The parameters of the GA computations used in this work are shown in Table 5. After having tried out the abovementioned parameters, a population size of 100,

**Table 5** Parameters of GA computations

Parameters/operators	
Population size	50–500 in steps of 50
String length	16 bits
Selection process	Tournament selection
Crossover	One-point crossover with probability of 0.8, 0.85, 0.9, 0.95
Mutation	Probability of mutation was 0.001, 0.005, 0.01, and 0.05

a crossover probability of 0.9, and a mutation rate of 0.001 were chosen as the best parameters for this GA analysis. Using these selected parameters, the GA model was run to determine the optimal cutting conditions for each pass. The optimum cutting condition is shown in Table 4.

## Conclusion

This paper describes a neural network-based sensor fusion model for a tool wear monitoring system in a manual turning operation. In addition, the monitoring information was integrated to an optimization system in order to utilize the progressive tool wear information for selecting the optimum cutting conditions. Wavelet packet analysis has shown that the optimal decomposition levels are not the same for all the signals. However, as mother wavelet function has little influence on the selection of a wear-sensitive feature, any of the considered mother wavelet functions may be used for signal analysis. PCA has been proved very effective for selecting a small number of uncorrelated wear-sensitive features. Taking the conventional training and evaluation approach, it was found that the prediction performance of the RBFN model is better than that of the BPNN model. However, an interval-based uncertainty technique shows that the BPNN model is more robust to uncertainty in the input data than the RBFN model. This approach can be used as an alternative for selecting a particular ANN model, which is intrinsically more robust to uncertainty in the input data than the network solutions obtained by conventional training paradigms. It was found that the proposed monitoring technique is highly effective for the prediction of tool wear even with the consideration of uncertainty in the input data; and the optimization process showed that the cutting parameter settings were not the same for all passes when tool wear was considered in the process.

## References

- Audy, J. (2006). An appraisal of techniques and equipment for cutting force measurement. *Journal of Zhejiang University Science A*, 7, 1781–1789.
- Ben-Haim, Y. (2001). *Information-gap decision theory: Decisions under severe uncertainty*. Academic Press.
- Berger, B. S., Minis, I., Harley, J., Rokin, M., & Papadopoulos, M. (1998). Wavelet based cutting state identification. *Journal of Sound and Vibration*, 213(5), 813–827.
- Byrne, G., Dornfeld, D., Inasaki, I., Ketteler, G., König, W., & Teti, R. (1995). Tool condition monitoring (TCM)—The status of research and industrial application. *Annals of the CIRP*, 44(2), 541–567.
- Chao, P., & Hwang, Y. D. (1997). An improved neural network model for the prediction of cutting tool life. *Journal of Intelligent Manufacturing*, 8, 107–115.
- Da, Z. J., Sadler, J. P., & Jawahir, I. S. (1997). Prediction of optimum cutting conditions for turning operations at varying tool-wear states. *Transaction of NAMRI/SME*, XXV, 75–80.
- Deb, K. (1996). *Optimization for engineering design*. New Delhi: Prentice Hall of India (p) Ltd.
- Dornfeld, D. A. (1990). Neural network sensor fusion for tool condition monitoring. *Annals of the CIRP*, 39(1), 101–105.
- Franco-Gasca, L. A., Herrera-Ruiz, G., Peniche-Vera, R., Romero-Troncoso, R. D. J., & Leal-Tafolla, W. (2006). Sensorless tool failure monitoring system for drilling machines. *International Journal of Machine Tools and Manufacture*, 46, 381–386.
- Haykin, S. (2003). *Neural Networks: A comprehensive foundation*. Delhi: Pearson Education.
- Heyns, P. S. (2007). Tool condition monitoring using vibration measurements—a review. *Insight*, 49(8), 447–450.
- Jemielniak, K., Kwiatkowski, L., & Wrzosek, P. (1998). Diagnosis of tool wear based on cutting forces and acoustic emission measures as inputs to a neural network. *Journal of Intelligent Manufacturing*, 9, 447–455.
- Jolliffe, I. J. (1986). *Principal Component Analysis*. New York: Springer.
- Keinert, F. (2004). *Wavelets and multiwavelets*. USA: Chapman and Hall/CRC.
- Lee, J. M., Choi, D. K., Kim, J., & Chu, C. N. (1995). Real-time tool breakage monitoring for NC milling process. *Annals of the CIRP*, 44(1), 59–62.
- Mannan, M. A., & Broms, S. (1989). Monitoring and adaptive control of cutting process by means of motor power and current measurements. *Annals of the CIRP*, 38(1), 347–350.
- Mathworks Inc. (2007). User manual of wavelet toolbox. MATLAB 7.0.
- Mittermayr, C. R., Nikolov, S. G., Hutter, H., & Grasserbauer, M. (1996). Wavelet denoising of Gaussian peaks: a comparative study. *Chemometrics and Intelligent Laboratory Systems*, 34, 187–202.
- Obikawa, T., Kaseda, C., Matsumura, T., Gong, W. G., & Shirakashi, T. (1996). Tool wear monitoring for optimization cutting conditions. *Journal of Material Processing Technology*, 62, 374–379.
- Pal, S., Pal, S. K., & Samantaray, A. K. (2007). Radial basis function neural network model based prediction of weld-plate distortion due to pulsed metal inert gas welding. *Science and Technology of Welding and Joining*, 12(8), 725–731.
- Papadopoulos, G., & Edwards, P. J. (2001). Confidence estimation methods for neural networks: a practical comparison. *IEEE Transactions on Neural Networks*, 12(6), 1278–1287.
- Purushothaman, S. (2009). Tool wear monitoring using artificial neural network based on extended Kalman filter weight updation with transformed input patterns. *Journal of Intelligent Manufacturing*. doi:10.1007/s10845-009-0249-y.
- Rehorn, A. G., Jiang, J., & Orban, P. E. (2005). State-of-the-art methods and results in tool condition monitoring: a review. *International Journal of Advanced Manufacturing Technology*, 26, 693–710.
- Scheffer, C., & Heyns, P. S. (2004). An industrial tool wear monitoring system for interrupted turning. *Mechanical Systems and Signal Processing*, 18, 1219–1242.

- Sick, B. (2002). On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. *Mechanical Systems and Signal Processing*, *16*, 487–546.
- Velayudham, A., Krishnamurthy, R., & Soundarapandian, T. (2005). Acoustic emission based drill condition monitoring during drilling of glass/phenolic polymeric composite using wavelet packet transform. *Material Science and Engineering A*, *412*, 141–145.
- Wang, X., & Jawahir, I. S. (2001). Optimization of multi-pass turning operations using genetic algorithms for the selection of cutting conditions and cutting tools with tool-wear effect. In *IFSA world congress and 20th NAFIPS international conference (IEEE)* (vol. 5, pp. 3093–3100). Canada: Vancouver, BC.
- Wang, X., Wang, W., Huang, Y., Nguyen, N., & Krishnakumar, K. (2008). Design of neural network-based estimator for tool wear modelling in hard turning. *Journal of Intelligent Manufacturing*, *19*, 383–396.
- Wu, Y., & Du, R. (1996). Feature extraction and assessment using wavelet packets for monitoring of machining processes. *Mechanical System and Signal Processing*, *10*, 29–53.