CHAPTER 6

6. Further investigations towards improvement of AI method

6.1 Introduction

In the previous chapters it was shown how a tool wear monitoring system can be designed using AI. The formulation of the AI method is such that it is not sensitive to the noisy shop floor conditions and other typical influences. The method was shown to be applicable to different turning operations and with different machining parameters. An important component of the research was that it was shown that a TCMS using AI could be implemented on the shop floor in a cost-effective way. The purpose of this chapter is to discuss some further results into the improvement of the TCMS. The investigations in this chapter are mainly concerned with improving the AI, signal processing and feature selection methods.

6.2 Signal processing

6.2.1 Coherence function

The coherence function is commonly used as a measure of the integrity of frequency response functions in vibration analysis. It can also be used to detect non-linearities in a system with an input-output relationship. Coherence is a function of frequency with values between 0 and 1 that indicate how well input corresponds to the output at each frequency. The coherence between an input X and an output Y is defined as:

$$\gamma^2 = \frac{\|P_{xy}\|^2}{P_{xx}P_{yy}} \qquad (6.1)$$

where:

P_{xx} = the power spectral density of x

P_{yy} = the power spectral density of y

P_{xy} = the cross spectral density of x and y

It follows that if x and y are completely correlated at a particular frequency the coherence will be 1 at that frequency. In practice, a value near one is reached for correlation and a value near zero is reached when there is no correlation. The coherence is typically calculated for a force input with a vibration output. However, the coherence can be calculated for any two sensors, and the coherence function can assist as a feature for condition monitoring. The coherence function between two acceleration signals was used by Li *et al.* [109] to detect tool wear and chatter during turning. They established that the value of coherence at certain frequencies could be used to trend tool wear or to detect the onset of chatter. Because the coherence is easy to calculate and interpret, it might be useful for on-line implementation, and it was decided to investigate the coherence function as a possible feature for TCM.

If force and vibration measurements are present, the coherence function is particularly useful to investigate the dynamics of the system. In Chapter 4, an accelerometer was attached to the tool holder dynamometer, as shown in Figure 6.1. The coherence function can now be used to determine the coherence in the response of the dynamometer over a range of frequencies. The coherence functions between force and acceleration in the three principal directions for various experiments with increasing tool wear was investigated over the range 8Hz – 2kHz.

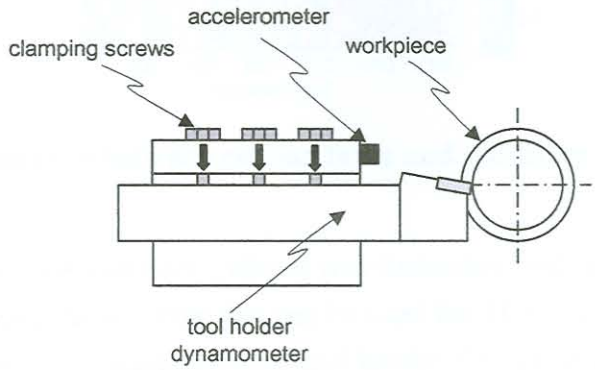

Figure 6.1: Sensor configuration (Chapter 4)

Examples of the coherence functions for experiment 2 are shown in Figure 6.2 and Figure 6.3. The vertical red lines on the figure indicate a high level of coherence. It is interesting to see that there are in fact quite a few areas in both figures where the coherence is very low, especially in the low frequency range. There are several possible reasons why coherence could be low:

- noise in the measurements
- resolution bias errors in the spectral estimates
- the output is due to other inputs besides the measured input
- non-linearity of the system

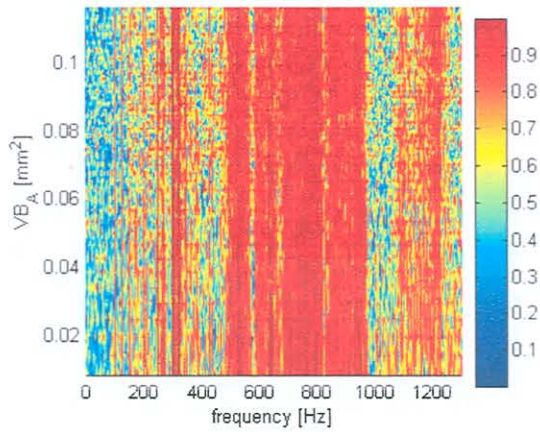

Figure 6.2: Coherence between thrust force and vibration (hard turning)

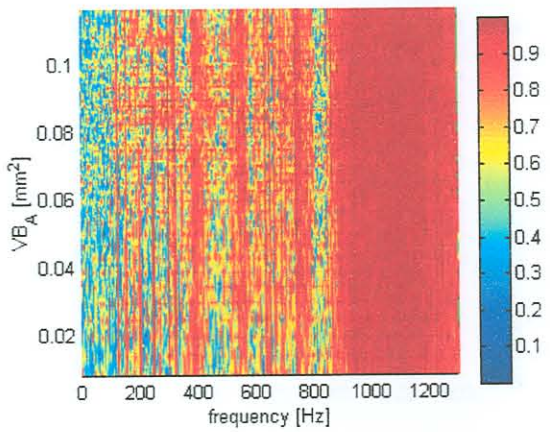

Figure 6.3: Coherence between cutting force and vibration (hard turning)

The areas where coherence is low probably indicate non-linearities and non-measured inputs. The areas with high coherence indicate frequencies that can be used for TCM because they are due to the actual dynamic force inputs and the dynamics of the tool holder. Closer investigations revealed that the **average** coherence function in the cutting force direction shows a slight increase due to tool wear, but would not suffice as a feature for TCM. There were also no particular frequencies that showed a significant correlation with tool wear.

The investigation was then moved to the coherence between two acceleration signals in the different directions, hence the method proposed by Li *et al.* [109]. Some of the results from experiment 2 are shown in Figure 6.4 and Figure 6.5. Again, there are certain regions where the coherence is very low, but in this case it is expected because different directions are used. However, in contrast with the findings in [109], there were no frequencies in the different coherence functions with a useful agreement with tool wear. From this it can be concluded that the usefulness of the coherence function is limited to the particular experimental setup, and is thus not a general solution for TCM.

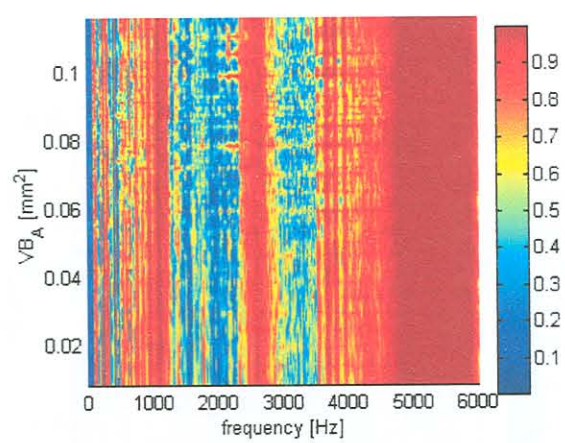

Figure 6.4: Coherence between feed and thrust vibration (hard turning)

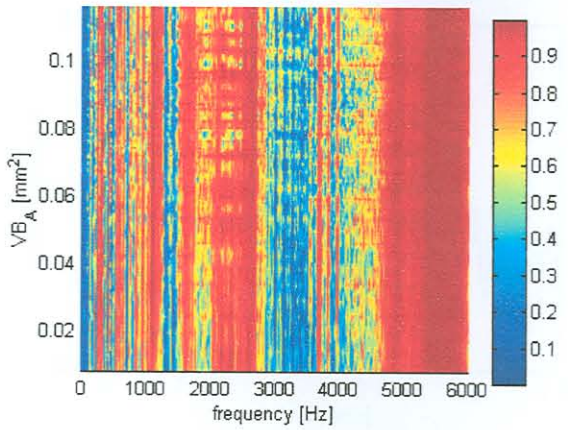

Figure 6.5: Coherence between cutting and thrust vibration (hard turning)

The results from the coherence function investigations with the data from Chapter 4 also raised an important question – does the dynamometer measure the actual dynamic force inputs or is it measuring its response to these inputs, which may contain non-linearities within the measuring range? This can only be answered if proper dynamic tests are conducted. However, the use of resistance strain gauges (Chapter 5) offered a cost-effective alternative that had no effect on the dynamic characteristics of the system. The dynamic properties of the sensor-integrated tool holder were characterised in Section 5.2.3. Consequently, coherence functions were also calculated on the force data from the strain gauges for the Aluminium turning experiments. A result for the coherence between forces *Fx* and *Fy* (refer to Chapter 5) is shown in Figure 6.6. It can be seen that in this case, the coherence is high for most of the frequencies. However, the coherence is influenced by the tool wear in the lower frequency range up to about 800Hz, and also at the tool holder natural frequency near 3kHz. In this case, increasing tool wear generally cause lower coherence values, except for the very low frequencies near 0 Hz where an increase is present. The mean value of the coherence over the whole frequency with respect to tool wear is shown in Figure 6.7. In this case the coherence could be used as a feature for TCM. Further investigations revealed that the features chosen for wear monitoring in Chapter 5 are more reliable than the coherence function.

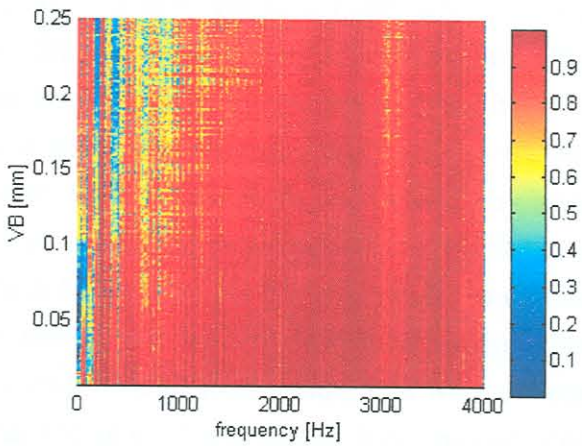

Figure 6.6: Coherence between Fx and Fy (Aluminium turning)

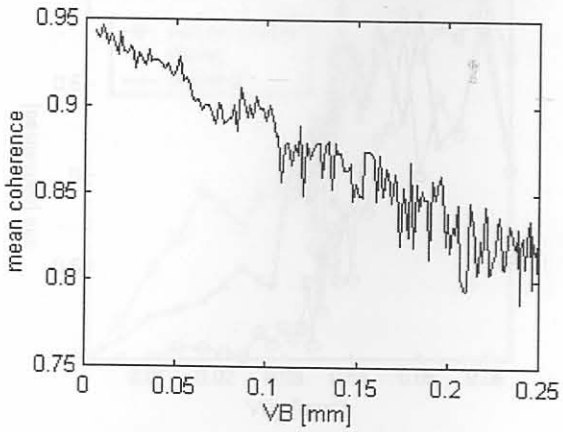

Figure 6.7: Mean of the coherence function between *Fx* and *Fy* (Aluminium turning)

6.2.2 Wavelet analysis

As a further step towards generating more reliable features for wear monitoring, wavelet analyses were conducted on the data sets. Some authors state that wavelet analysis is the key to successful TCM (refer to Section 3.4.2). However, the usefulness of wavelet analysis in TCM applications is debatable. Wavelet analysis cannot be used effectively to detect temporal information in data because they are not time invariant. In the case of TCM, this is not always important, depending on the aim of monitoring system. They only way that wavelet analysis can really assist in TCM applications is to act as a filter to enhance the signal-to-noise problem common to tool wear data.

The reader is referred to Appendix F for a more detailed discussion on wavelet analysis. During this research, wavelet packet analyses were performed with various types of wavelets to act as digital filters. With the wavelet packets calculated, certain packets may contain useful information on tool wear with little noise present. A simple method is to select the packets containing the most energy by means of an energy function (*e.g.* rms or entropy) for a second phase of processing. The second phase is generally a repeat of the initial signal processing in the time and frequency domains, such as rms, variance and crest factor values.

The purpose of the wavelet analysis is to act as a filter, which automatically identifies certain frequency ranges and bandpass filter the signal through these ranges. When the signal is reconstructed with the wavelet coefficients from the selected wavelet packets, it can be compared with the original signal. During this research, it was often observed that the reconstructed signal is just a filtered representation of the original. Hence, the same result can be obtained by using an appropriate digital filter. In fact, the wavelet packet type of filtering is a "black box" type filter because there is no indication beforehand or afterward which frequencies are attenuated. In the case of digital filtering, this can be controlled and explained. A result from the wavelet packet analysis from experiment 4 is plotted in Figure 6.8. In this figure, the rms trend with increasing tool wear are compared for a selected wavelet packet, a digital filtered signal and an unfiltered signal. It can be seen from this figure that although the wavelet packet rms trend is more consistent than the unfiltered signal, the digital filter yielded the best result. A similar result can be observed in Figure 6.9, where the trends of the variance with increasing tool wear are compared.

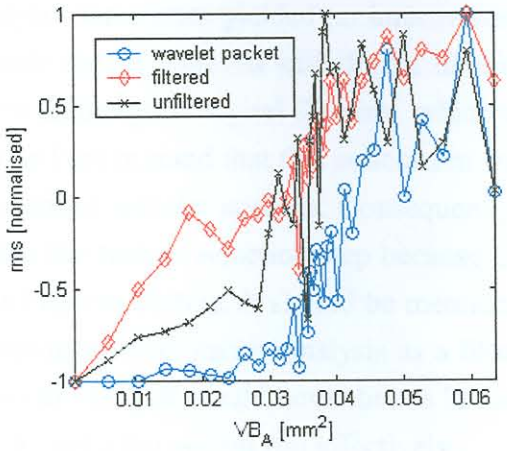

Figure 6.8: Comparison of rms trend with tool wear (hard turning)

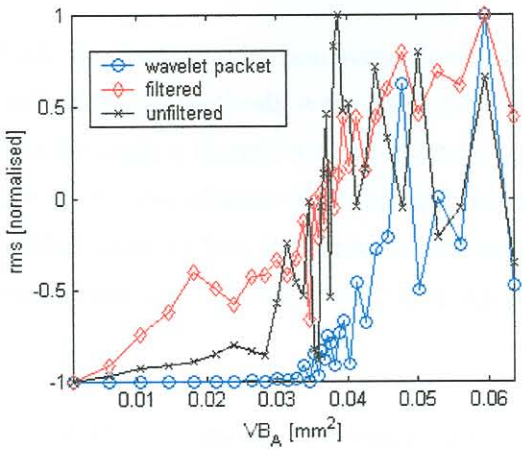

Figure 6.9: Comparison of variance trend with tool wear (hard turning)

The wavelet analysis was also conducted on the TCM data for Aluminium turning. Exhaustive investigations were conducted with different types of wavelets and threshold values. One of the results is shown in Figure 6.10, where the rms values of the best wavelet packets are compared to the rms of the digital highpass filtered signal and unfiltered signals. In this case, it is again clear that the digital filter yields the best results.

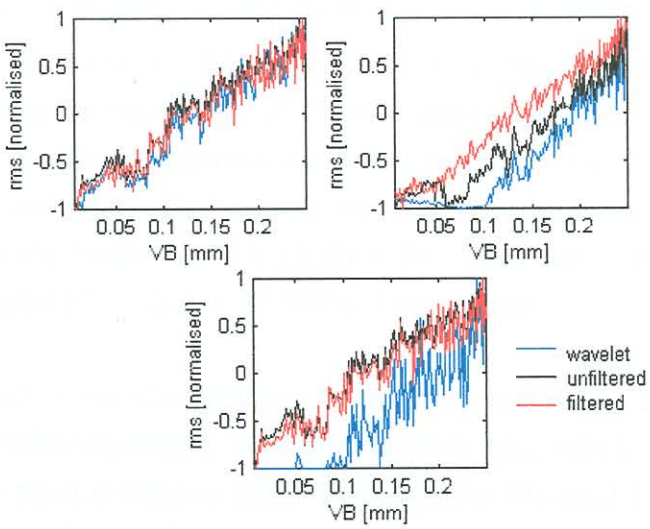

Figure 6.10: Wavelet packet rms comparison (Aluminium turning)

It was found that wavelet analysis sometimes yielded an improvement in the trend, but the improvement is small and the same result can be achieved with digital filtering. Furthermore, wavelet analysis is computationally much slower than regular digital filtering, which hampers it as a possibility for on-line implementation. It should be kept in mind that this conclusion was only reached for TCM applications and not for other applications of wavelet analysis. Consequently, the features generated by wavelet analysis were discarded from the feature selection step because the improvement was too small to make it worthwhile for on-line implementation. It should be mentioned here that there exists a TCMS for ball-end milling which is using wavelet packet analysis as a filter, and the method is successfully running as a demonstration system [239]. It would nevertheless be interesting to investigate the necessity of wavelet packet analysis to make the system run effectively.

6.2.3 Feature selection and feature space reduction

In Chapters 4 and 5 the methods for feature selection were described. A combination of engineering judgment and automated feature selection methods were used. The automated feature selection methods were based on the agreement between a feature's trend and tool wear. The methods were employed to quantify this agreement, namely the correlation function and difference minimisation by means of GAs. Both methods usually yielded more or less the same results and as a last step engineering judgement was used to select particular features. Normally, only four signal features were chosen for monitoring purposes.

Another method of verifying that the best features have indeed been selected is by calculating the Statistical Overlap Factor (SOF). The SOF determines the degree of separation of a feature between the new and worn tool conditions. Ideally, a feature should show a high degree of separation due to the worn condition and a low degree of variance due to noise. The SOF is one method that can assist in investigating the behaviour of a feature. The SOF is defined by:

$$SOF = \left| \frac{\bar{x}_1 - \bar{x}_2}{(\sigma_1 + \sigma_2)/2} \right| \tag{6.2}$$

where $\bar{x}_1$ is the mean and σ_1 is the standard deviation of vector x_1. Vector x_1 should be data collected from new tools and x_2 should be data collected from worn tools. As a consequence the SOF will yield a value that is an indication of a feature's ability to separate between new and worn conditions. Comparisons between the correlation coefficient approach and the SOF were made for several experiments (correlation coefficient approach described in earlier chapters: The correlation is calculated between tool wear and the feature vectors). One result is shown as a bar graph comparison in Figure 6.11 for 30 different features calculated from the Aluminium turning data. Ideally, features should be chosen that exhibit a high degree of SOF and correlation.

One problem with the SOF method is that it will not separate between linearly dependant features. Thus, there is the risk of selecting linearly dependant features when only the SOF method is employed. Another method that can assist to remove linear dependency is Principal Component Analysis (PCA). However, it is known that TCM problems are often one-dimensional (all the signal features increase when tool wear increase), and thus features that are linearly dependant to some degree will have to be

selected. Consequently, as a last step, engineering judgement must be used to select features that are not completely linearly dependant and also have high SOF and correlation values. Referring to Figure 6.11, features 7, 14, 17 and 25 were selected.

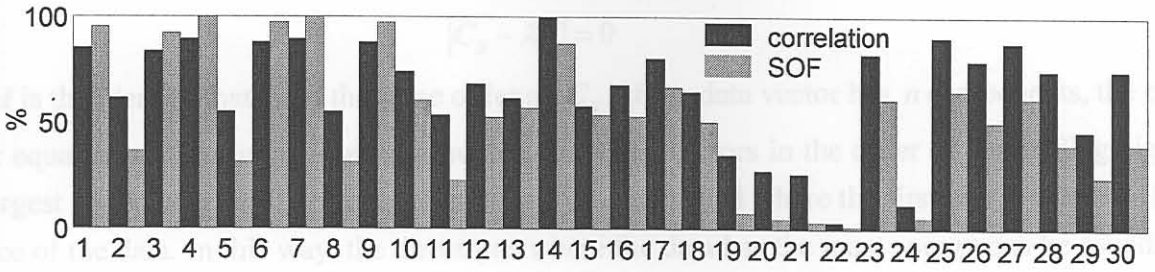

Figure 6.11: Comparison of correlation coefficient and SOF

In many cases, choosing many features and then using a feature space reduction method such as PCA to reduce the feature space can achieve a better result. Fortunately in data sets with many features, groups of features often move together. One reason for this is that more than one variable may be measuring the same driving principle governing the behaviour of the system. In many systems there are only a few such driving forces. But an abundance of instrumentation allows us to measure dozens of system variables. When this happens, the problem can be simplified by replacing a group of variables with a single new variable. PCA is a one method for achieving this simplification. The method generates a new set of variables, called principal components. 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 component as a whole forms an orthogonal basis for the space of the data, and is essentially based on the statistical representation of a random variable. Suppose there is a random vector population $\boldsymbol{x}$ where:

$$\boldsymbol{x} = (x_1, \ldots, x_n)^T \tag{6.3}$$

and the mean of that population is denoted by

$$\mu_x = E\{\boldsymbol{x}\} \tag{6.4}$$

and the covariance of the same data set is

$$\boldsymbol{C}_x = E\left\{(\boldsymbol{x} - \mu_x)(\boldsymbol{x} - \mu_x)^T\right\} \tag{6.5}$$

The components of $\boldsymbol{C}_x$, denoted by c_{ij}, represent the covariances between the random variable components x_i and x_j. The component c_{ii} is the variance of the component x_i. The variance of a component indicates the spread of the component around its mean. If two components x_i and x_j of the data are uncorrelated, their covariance is zero $\left(c_{ij} = c_{ji} = 0\right)$. The covariance matrix will always be symmetric. From a sample of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$, the mean and covariance matrices can be calculated. From the covariance matrix, an orthogonal basis can be calculated by determining eigenvalues and eigenvectors. The eigenvectors $\boldsymbol{e}_i$ and the corresponding eigenvalues λ_i are the solutions of the following equation:

$$C_x e_i = \lambda_i e_i \text{ with } i = 1,...,n \tag{6.6}$$

For simplicity it is assumed that λ_i is distinct. These can be determined by solving the characteristic equation:

$$|C_x - \lambda_i I| = 0 \tag{6.7}$$

where I is the identity matrix of the same order as C_x. If the data vector has n components, the characteristic equation will have order n. By ordering the eigenvectors in the order of descending eigenvalues (largest first), an ordered orthogonal basis can be established where the first eigenvector has largest variance of the data. In this way, the directions where the data has the most energy can be found. If the data is transformed with these principal components, it reduces the amount of data but retains the information containing the most energy.

Principal component analyses were performed on the hard turning and Aluminium turning experiments. The results of the 1st two principal components, together with the features chosen for the AI monitoring method in the previous chapters, are plotted in Figure 6.12 and Figure 6.13. It can be seen that the 1st principal component follows the increasing trend of features with less noise than features themselves. The 2nd principal component is unfortunately only noise, because the TCM problem is one-dimensional. Features generally increase (or sometimes decrease) with increasing tool wear. The tool wear is also a monotonically increasing variable. As a result, the 1st principal component contains the information along the increasing axis, and the 2nd principal component is generated by the inherent noise in the data. An advantage of PCA is that the 1st principal component is less noisy than the feature values themselves. Care must be taken not to contaminate the principal components with noise from redundant data.

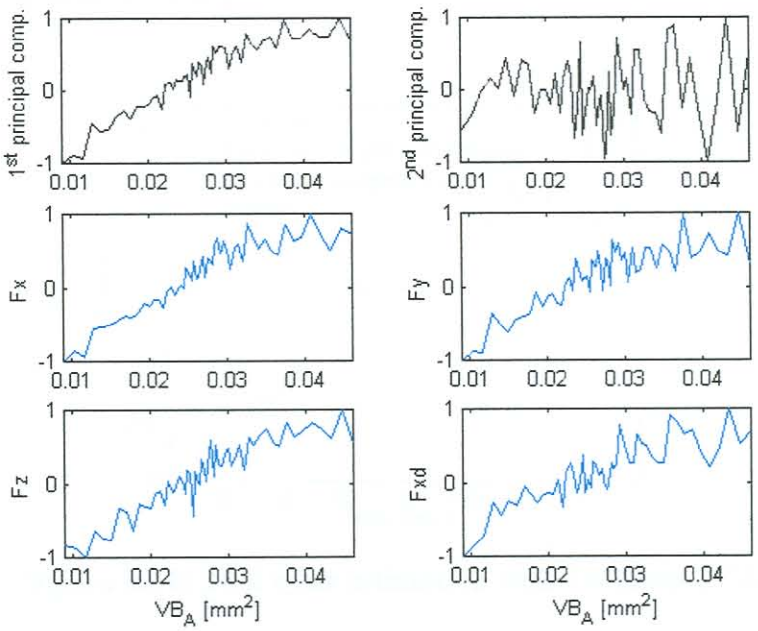

Figure 6.12: Principal component analysis Hard turning (data normalised)

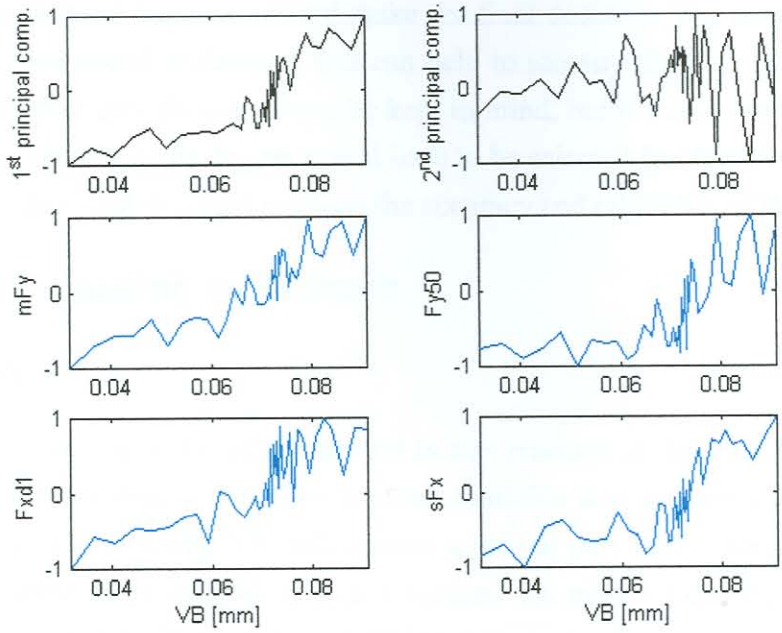

Figure 6.13: Principal component analysis Aluminium turning (data normalised)

It is clear from the examples from the two different turning processes that a PCA can actually increase the reliability of the TCMS, despite the fact that the problem is one-dimensional. One possibility would be to keep the existing four features and just add a fifth, namely the 1st principal component of the data. Another possibility would be to base the TCMS only on the 1st principal component. This was done with data from Aluminium turning and the results are compared in Figure 6.14. It can be seen from the figure that adding the 1st principal component achieved a slight increase in the reliability of the TCMS. However, using the 1st principal component only did not work very well. This is an important result, because it shows that using data that are linearly correlated to some degree yields better results than using only the principal component.

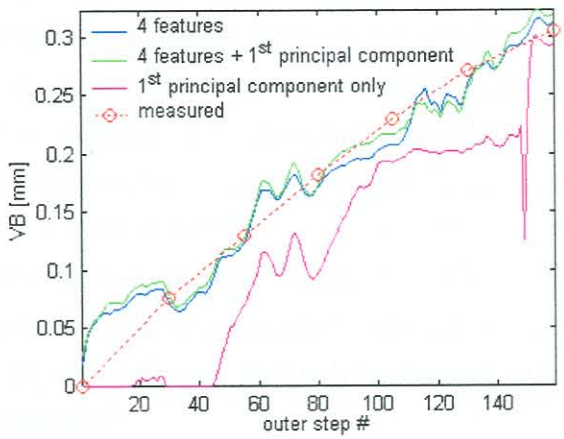

Figure 6.14: Tool wear estimation with / without PCA

The conclusion from the investigations into feature selection is that available techniques can be used to assist in the decision but the engineer should make the final decision. The correlation coefficient, the SOF and PCA are three useful techniques that can help to identify the best features. The problem of selecting linear dependant data should always be kept in mind, but it was shown that selecting features that are linearly dependant to some degree would have to be selected for TCM. Adding the 1st principal component as an addition feature could increase the accuracy and reliability of the TCMS.

6.3 Alternative modelling techniques

6.3.1 Introduction

One of the advantages of the proposed AI method in this research is that certain parts of it can be replaced by other methods - should a method become available that is more accurate and fit for shop floor implementation. The dynamic NN will remain a crucial part of the method, but the static NNs can be replaced by some other method, if such a method can model a chosen signal feature. Models that can calculate static cutting forces for sharp tools are common. However, due to the complex nature of tool wear, it is difficult to estimate the worn tool forces with most other methods. Features such as those derived from the frequency spectrum would be very difficult if not impossible to determine by means of theoretical models only.

One possibility would be to combine available theoretical models with the AI approach. An advantage of theoretical approaches is that they can handle changing machining conditions with more ease and accuracy than AI models. The reason for this is the *a priori* knowledge of the theoretical models of what the effect of changing conditions would be, whereas the AI model needs appropriate training samples to obtain this knowledge. In this section, some promising theoretical methods are discussed that may enhance the accuracy of a TCMS.

6.3.2 Finite Element Method (numerical models)

The use of the FEM to model machining operations was discussed in Section 2.5. Besides the FEM, various other numerical computer simulation methods are available or are under development. This approach, though still in the development stage for many machining operations, seems to be one of the most promising to assist is sensor-based TCM. A numerical model can be used instead of the static NN for TCM. This is what makes the new formulation proposed in this work particularly useful: Instead of modelling the tool wear as an output of the model it is actually used as input to the static NN. Hence, with tool wear as an input, a numerical method such as the FEM can estimate the static cutting forces for many combinations of tools, workpieces and machining parameters. The dynamic NN will still be included to follow the development of wear, but will be trained with error between the FEM model simulation and on-line measurements.

One problem is that this approach will be computationally very slow with available computing power. In future years computers might become fast enough to use the FEM model iteratively to train the dynamic NN. In fact, with available computing speed, it might not be extremely slow because the dynamic NN usually requires less than ten iterations. The use of response surfaces could also prove very

useful to lessen the number of FEM simulations. Unfortunately, the FEM approach could not be investigated for this work but is suggested for further research. FEM models could also be used to normalise data with respect to machining conditions. To achieve this will require intensive research and collaboration between various research groups.

6.3.3 Theoretical models

The basics of theoretical models to predict cutting forces were discussed in Chapter 1. A pure theoretical model that can accurately predict worn tool forces does not exist and establishing such a model would be virtually impossible. The only possible use of these models is to assist another method, such as the AI approach. For instance, an accurate theoretical model can be used to:

- Predict the sharp tool cutting forces
- Normalise the AI approach for cutting conditions
- Assist as a validation procedure

In this research, the analytical method described in Section 2.2.3 was evaluated as a possible method of calculating the sharp tool forces and then adding the worn tool forces by means of the AI approach. The analytical procedure involves determining the oblique cutting constants through orthogonal cutting tests. When this is found, an oblique cutting transformation is applied to two regions along the radius of the tool insert. The complete model relies on the accuracy in determining certain constants (some of which will be available from a database for some tool and workpiece combinations [240]) and the validity of a few assumptions. Because this and other theoretical models rely on cutting tests and underlying assumptions, it raises the question if it is worthwhile to implement the method if it can only assist with TCM in part, namely predicting sharp tool cutting forces. Also, the dynamic behaviour of the tool is not included in the model. After careful consideration it was decided that it is not worthwhile to attempt a shop floor implementation of the method because it will raise the complexity of a problem that the AI approach can already handle to satisfaction.

6.3.4 Mechanistic models

The underlying assumption of mechanistic cutting models is that the cutting forces are proportional to the uncut chip area. The constants of proportionality depend on the cutting conditions and geometry and material properties. Kapoor *et al.* [241,242] also described a method that can be incorporated into a basic mechanistic approach that can predict dynamic cutting forces. This could be achieved by impact tests on the machine structure. They also describe a worn tool force model [223,243] for turning operations. This is the first model based on theoretical foundations for predicting worn tool forces. Of course, other experimental models exist that rely on the accurate determination of certain empirical constants.

Because the mechanistic worn tool force model is a possible complete replacement of the static NNs, it was decided to attempt an implementation of the model. The complete mechanistic approach for predicting worn tool forces is described in Appendix G. The method was applied to the hard turning data after personal correspondence with Kapoor and DeVor from UIUC indicated that the method would

after personal correspondence with Kapoor and DeVor from UIUC indicated that the method would possibly be applicable to the type of processes investigated in this research. To determine the sharp tool forces is basically a matter of calculating empirical constants from cutting experiments. With this a calibration procedure is carried out that will yield the mechanistic constants. The part of model that deals with determining the worn tool forces relies on a number of assumptions, the most important being:

- There is a linear growth of the plastic flow region on the tool flank
- There exists a critical value of flank wear after which plastic flow will be observed
- The worn tool forces are governed only by flank wear
- An accurate calculation of the maximum effective stress in the workpiece is possible

It was found that for most of the hard turning experiments, the critical value of flank wear could not be determined. Even after extensive testing, it was not always observed. Attempts were made to calculate the maximum contact stress in the workpiece by the method described in Appendix G, and a result is shown in Figure 6.15. When this is accurately established and verified, the 3-D worn tool forces can be calculated by the equations described in Appendix G. However, the maximum contact stress could not be verified because the assumptions made for the mechanistic model does not seem to apply for hard turning. It was mentioned several times in Chapter 4 that the stability of hard turning is governed largely by crater wear. During hard turning, the crater wear also has an influence on the cutting forces, but crater wear is not included in the mechanistic approach. In fact, a theoretical inclusion of the crater wear is rather difficult because the mechanics of cutting with crater wear is not as clear as the case with flank wear.

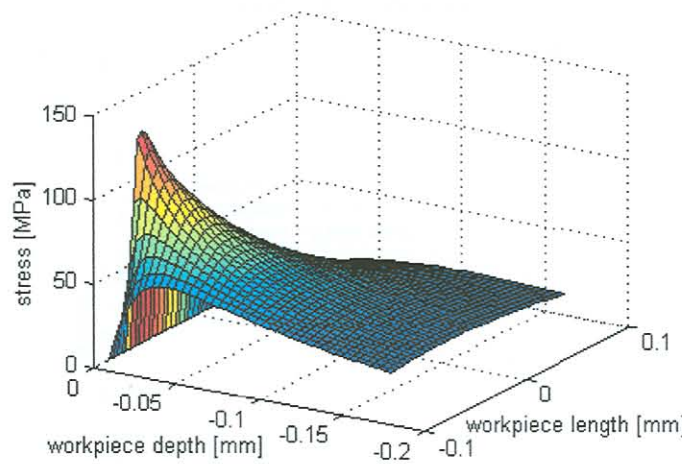

Figure 6.15: Calculating maximum contact stress in workpiece (hard turning)

The linear growth of the plastic flow and the critical flank wear was also not so clearly observed with the CBN tools, and as a result the maximum contact stress could not be determined. However, the plastic and elastic zones were better observed in the Aluminium turning experiments. Because the aim of this work is to determine a method of TCM that can be treated as a general methodology for turning, it was decided not to attempt an implementation of the model on Aluminium turning because the method did not apply to hard turning as well.

6.3 AI approach improvement

6.3.1 Introduction

In this section, some aspects regarding the improvement of the AI methodology are investigated. In the opinion of the author, the improvement that can be obtained by optimising the network type, structure and activation functions are relatively small compared to selecting the correct measurement procedure, signal processing and feature selection steps. In this case, same aspects of the NN structure were investigated in order determine if it yields a worthwhile improvement or not.

6.3.2 Type of network

Many different types of network were compared for the best results. This included FF networks, FF networks with time delays, radial basis function networks, perceptrons, recurrent networks and unsupervised networks. In the case of the FF backpropagation networks, different activation functions and networks sizes were also compared. A comparison of the FF network as formulated in Chapter 5 and FF networks with time delays are shown Figure 6.16. The time delay networks required more training steps and when the convergence criteria were kept the same, the time delay networks did not yield very good results. Investigations only apply to the static NNs. If more training steps were taken, the time delay networks improved but did not yield better results than the initial formulation of the FF networks. Sick [171] has shown that time delay NNs should be used for TCM, because the TCM problem requires temporal information for accurate estimation. In the case of a the AI implemented in this research, the temporal information is already built into the time delay of the dynamic networks, and is not required for the static NNs. It was decided not to attempt to optimise time delays in the static NNs because it requires more training and also slows the training of the dynamic NNs down.

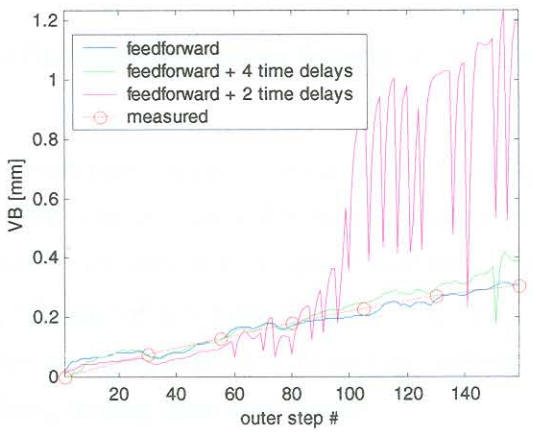

Figure 6.16: Time delay comparison (Aluminium turning)

In Figure 6.17 the result of using different network types is shown. The FF and the cascade forward networks yielded similar results. The Elman network requires longer training but the smooth response of the Elman network is a very nice attribute. The Elman network utilise feedback connections, and thus doubles the use of temporal information in the AI approach. The last investigation into network

formulation was to compare different activation functions. The results of using three different activation functions are shown in Figure 6.18. The tansigmoid and radial basis function yielded the best result.

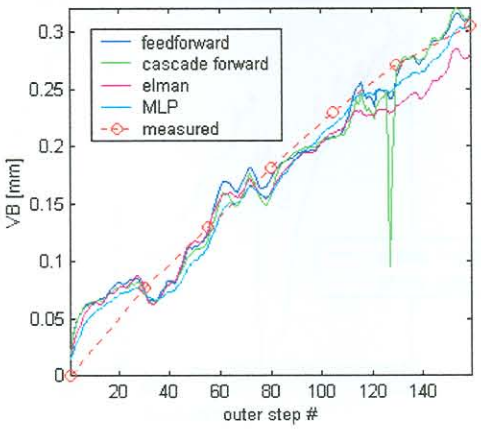

Figure 6.17: Different network types (Aluminium turning)

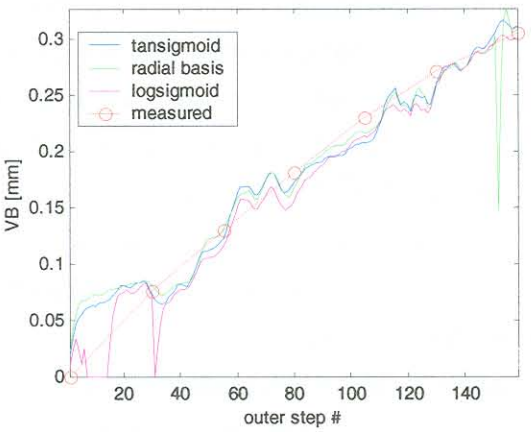

Figure 6.18: Different activation functions (Aluminium turning)

The results presented in this sub-section are only a fraction of the many investigations into formulating the best method for TCM. Unfortunately, only a few results can be shown and discussed here. In conclusion it can be stated that the FF network with a tansigmoid and linear layers yielded the best results and requires the least amount of time for adequate training. The Multilayer Perceptron (MLP) also exhibits good results, but requires a larger network. Elman type networks can also be considered due to their smooth response but training is very time consuming.

An important contribution of this research lies in the use of the dynamic NN for on-line monitoring. However, using only static NNs can also model the tool wear with the chosen features. For this reason a comparison of the two methods are included here to show that the proposed formulation of using static and dynamic networks is indeed the better one for practical applications. An example is shown in Figure 6.19, where simulations of a regular FF network is compared with that of the new formulation proposed in this study (without a sliding window output). The data is from three tools that wear from

new to approximately 0.1mm flank wear. It can be seen that the performance of the combined static and dynamic formulation is better than the FF network (static only). A much smoother network response is noted and the network has no difficulty to return to zero (automatic re-initialisation was used).

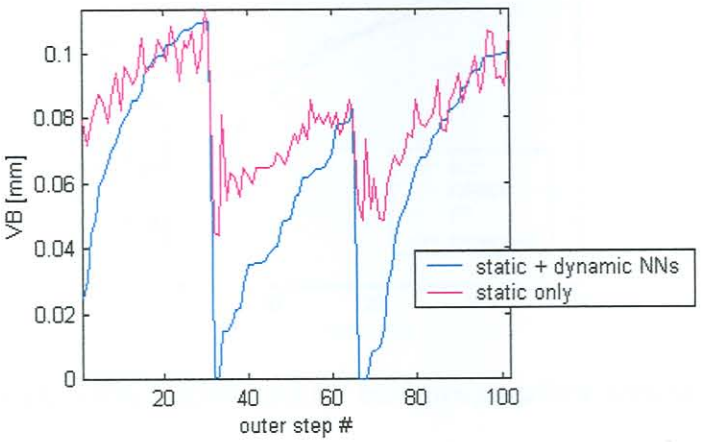

Figure 6.19: Comparison of formulations (Aluminium turning)

The new formulation was also compared to other conventional formulations, some of which performed well but do not hold the advantages of the combined formulation. No other NN formulation that was investigated outperformed the proposed combined formulation on the noisy shop floor data. Some of the results are reported in the figures that follow. The results are based on networks trained on the same data and tested on a previously unseen set of data. Where possible, the same network training tolerances were used. As a result, the sizes of the networks are not exactly the same for all the cases. The results of using of a Single Layer Perceptron (SLP), Adaptive Linear Neuron Networks (ADELINE) and adaptive linear filter network (linear layer with input delays) are plotted in Figure 6.20. The SLP can only be used for classification and is not recommend for continuous estimation. The ADELINE and linear adaptive filter display more or less the same result.

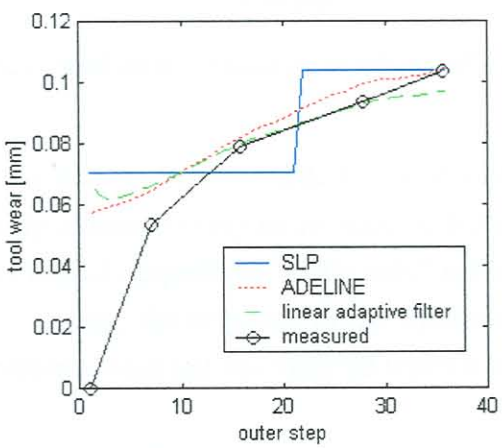

Figure 6.20: SLP, ADELINE and linear adaptive filter result

A Radial Basis Function (RBF) network is compared with a Generalised Regression Neural Network (GRNN), which is a RBF with an added linear layer at the output), and a two-layer FF network trained

with backpropagation in Figure 6.21. Again, the results are very similar.

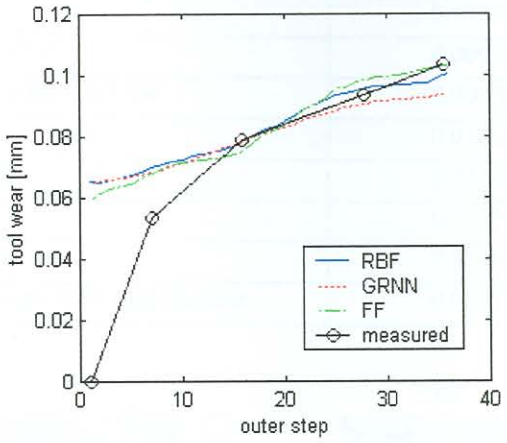

Figure 6.21: RBF, GRNN and FF backpropagation network result

A recurrent network (Elman formulation) is compared with the new formulation of combining the static and dynamic networks in Figure 6.22. The Elman network performs well but the new combined network still outperforms the Elman network.

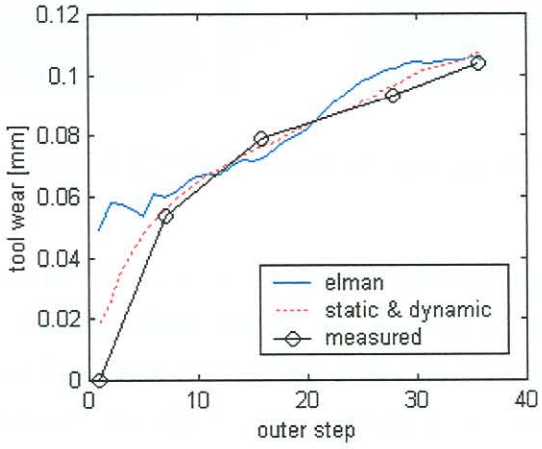

Figure 6.22: Recurrent (Elman) and new formulation with static and dynamic networks result

The results for this case study are summarised in Table 6.1. It should be kept in mind that the result reported here is for one case study although many more were performed to ensure that the newly proposed dynamic formulation does indeed outperform "static only" networks. From the graphs it can also be seen that besides increased accuracy, the response of the dynamic network is smoother and more stable. Training and stability problems were not encountered when using the PSOA.

As a last step the use of an unsupervised NN were investigated, one again in the form of the SOM (refer to Appendix H). The normalised training features were subjected to SOM training, and the result is shown in Figure 6.23. From the figure it can be seen that the features are automatically arranged in low and high regions.

Table 6.1: rms errors on training data

network type	*rms error [mm]*
SLP	0.0167
ADELINE	0.0097
linear adaptive filter	0.0092
RBF	0.0106
GRNN	0.0106
FF	0.0109
Elman	0.0106
combined static and dynamic	0.0057

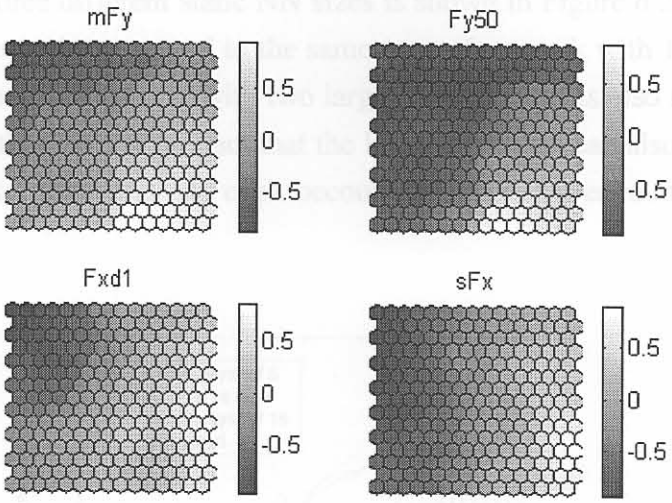

Figure 6.23: SOM result of training data

The labels of the training data, (in this case the flank wear values in mm) are plotted on the left hand side on Figure 6.24. A separate independent data set was labelled as "brand new", "new", "medium", "worn" and "replace". The Best Matching Units (BMUs) for this data were calculated the trajectory is plotted in Figure 6.25, together with the classification labels. From this it can be established that the SOM is very useful for interpreting the multi-dimensional data, but should rather be used for discrete classification (*e.g.* "new" or "worn") than continuous estimation. In this example, the SOM yielded very good results in terms of classification.

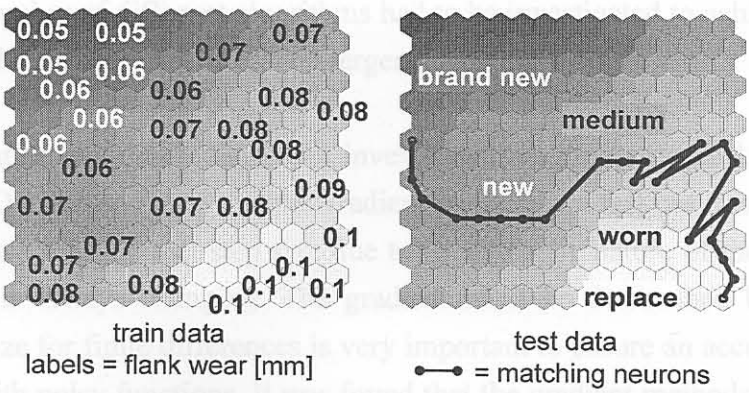

Figure 6.24: Training labels and testing classification

6.3.3 Size of networks

Although no mathematical optimisation procedure was performed, the size of the static and dynamic networks was optimised with manual iterations. To improve network generalisation, the networks were kept as small as possible to avoid overtraining. This also helps to improve training speed. If larger network sizes were used, the training was slow and the network could not generalise properly. If feed rate or cutting speed was included in the networks, a larger network size had to be used. In general, a middle layer of five neurons sufficed for a network with no machining parameters. If machining parameters were included, a middle layer of approximately ten neurons had to be used.

As an example, a result of three different static NN sizes is shown in Figure 6.25. A FF network with five neurons in the middle layer is compared to the same type of network with 15 neurons in the middle layer. The result from another network, with two large hidden layers is also shown in the figure. It can be seen from the figure that despite the fact that the larger networks can also follow the tool wear, they are much more prone to noise and can even become unstable if it encounters a noisy measurement.

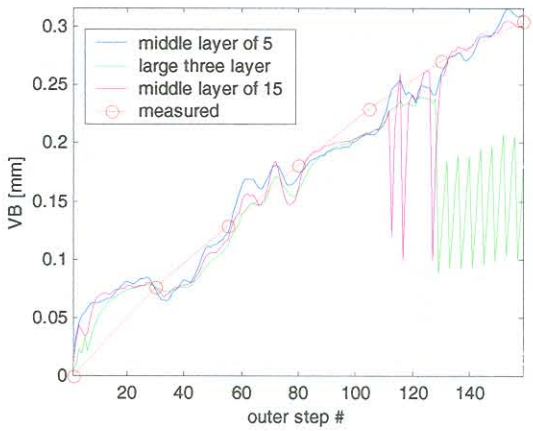

Figure 6.25: Different network sizes (Aluminium turning)

6.3.4 Training algorithm

The static networks were FF networks with backpropagation as a training procedure. The method was fast and accurate enough for the application without any convergence problems. In the case of the dynamic algorithms, a number of different algorithms had to be investigated to achieve fast on-line training. With conventional training procedures, convergence was slow or not at all.

The different optimisation algorithms that were investigated are discussed in Appendix D. These are ETOP, SQSD and LFOP. It was found that the gradient methods are fast but it is difficult to determine the gradient function with the correct step size due to the dynamic nature of the network. This is because the input data is always changing. The gradient must be determined by a finite difference method and the step size for finite differences is very important to ensure an accurate estimation of the gradient, especially with noisy functions. It was found that the gradient methods improve for dynamic training when the step size is decreased linearly when the objective function approach zero.

It was then decided to investigate another method that does not utilise a gradient calculation, and this was found in the PSOA. It was found that the PSOA provided quick and accurate training and rarely fails to converge. Furthermore, the PSOA does not have the problem of calculating gradient functions, and the random nature of the algorithm is ideal for this application. It was concluded that the PSOA is the best choice for on-line training of dynamic NNs.

6.3.5 Repeatability of simulations

Another important test for any implementation of NNs is to determine the repeatability of simulation results. This involves re-initialising and re-training all the networks and repeating the simulation on the same data. This was repeated for several data sets, and an example is plotted in Figure 6.26. Due to the nature of the PSOA and various convergence criterions to enforce generalisation, the simulations are not expected to be exactly the same every time, but should at least be very similar. This can be observed from the various simulations in Figure 6.26. Each follows the same progression of tool wear but they are not numerically the same. It can thus be concluded that the simulations are repeatable with newly trained networks.

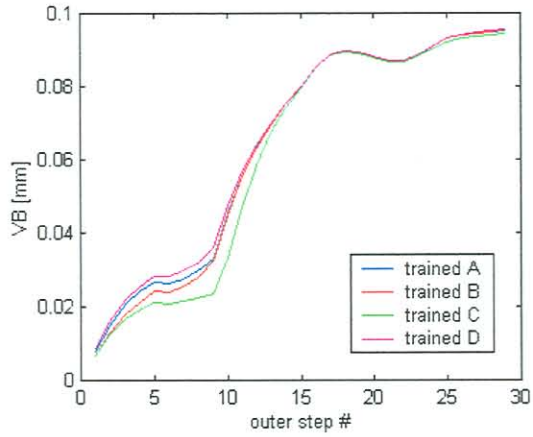

Figure 6.26: Repeated simulations with newly trained static NN

6.4 Conclusion

In this chapter, it was shown that:

- The coherence function is a method that can assist in TCM but is not universally applicable.
- Wavelet analysis can sometimes improve the accuracy of a TCMS but the same result (if not better) can be reached by simple digital filtering procedures.
- Feature selection could be done by the correlation function method and the SOF. PCA is another useful method and including the 1st principal component as another data feature can improve the accuracy of a TCMS.
- Engineering judgement remains the most important step of feature selection and care must be taken to avoid 100% linearly correlated data.
- Different modelling methods can assist in the AI method of TCM, but not all methods are universally applicable and often increase the complexity of required experiments and

mathematical formulation unnecessarily, making these methods insufficient for on-line implementation.

- Attempts were made to improve the formulation of the NNs in the AI approach for TCM by investigating network type, activation functions, network size and training algorithms.
- Comparisons between formulations were made and it was also shown that the result of the new formulation is repeatable.