FEATURE-BASED
TOOL CONDITION MONITORING FOR MILLING

DONG JIANFEI

NATIONAL UNIVERSITY OF SINGAPORE

2004

FEATURE-BASED
TOOL CONDITION MONITORING FOR MILLING

DONG JIANFEI
(B.Eng., NWPU)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2004

ACKNOWLEDGEMENTS

I express my deep sense of gratitude to my supervisors, Associate Professor G. S. Hong and Associate Professor Y. S. Wong, for their valuable supervision, constructive guidance, inspiration and friendly approaches throughout my research work.

I sincerely thank the National University of Singapore for sponsoring my study and providing excellent research environment.

I would also like to thank Mr. Lee, Mr. Lim, Mr. Wong, and all the technicians in Workshop 2 for their kind and valuable help in the whole process of experiments.

I wish to convey my gratitude to all my colleagues and friends in Control and Mechatronics Lab, especially Mr. Wang Wenhui and Mr. Cheng Zhaolin, for their help, support, and friendship.

I would like to show my appreciation to my parents and my brother for their encouragement, love, and understanding.

Dong Jianfei

Table of Contents

SUMMARY

The main objective of this project is to investigate the effectiveness of various features for tool condition monitoring (TCM) during milling processes. Sixteen different features extracted from force signals are considered, which have all been shown to be effective for TCM. These include residual errors derived from autoregressive models, statistical quantities, and frequency characteristics of force signals. Cutting experiments have been conducted under various conditions. A five-step approach has been proposed to extract the 16 features from the force signals measured in the experiments. Two innovative methodologies for neural networks are introduced and adopted in TCM, which are Bayesian interpretations for support vector machines (BSVM) and automatic relevance determination (ARD). Based on these approaches, two relevant feature sets have been identified from the 16 features for two main tasks in TCM: tool wear estimation (TWE) and tool wear recognition (TWR). The generalization capabilities of the entire, selected, and rejected feature sets have been tested and compared. Good generalization results have been achieved for both TWE and TWR using the selected features, which are superior to those using either the entire or the rejected feature set. The results prove that the selected features are relatively more relevant to tool wear processes, and draw attention to using the BSVM methodologies in TCM.

NOMENCLATURE

AAEE	averaged absolute estimation error
$\overline{A}_c(i,t)$	chip load of insert *i* at time *t*
A/D	analog to digital
ADC	analog-to-digital converter
AE	acoustic emission
AR	autoregressive model
ARD	automatic relevance determination
ART	adaptive resonance theory
a(t)	residual error or disturbance at time *t*
BSVC	Bayesian support vector classification
BSVM	Bayesian support vector machine
BSVR	Bayesian support vector regression
C	parameter controlling the distribution of noise
Cov	covariance function
D	training data set
DAQ	data acquisition
df	variable force
doc	depth of cut
f(i,j)	the *j-th* force sample in the *i-th* tool rotation
$f_a(i,j)$	total amplitude of cutting force
$f_d(t)$	different cutting force at time *t*
f_m	maximum force level

$\mathbf{f}_{MP}$	MAP estimation of function *f*
fm	maximum force level
fod	first order differencing
fr	feed rate
fstd	standard deviation of the force components in tool breakage zone
$\bar{f}_t$	feed per tooth
Δf	combined incremental force changes
F'	estimated force using high order AR model
F_a	average force
F_{med}	median cutting force
$F_p(i)$	peak value of the cutting force during the *i-th* tooth period
F_R	radial cutting force
F_T	tangential cutting force
F_x	cutting force along x-direction
F_y	cutting force along y-direction
$\Delta F_a(i)$	first order differencing of the average force during the *i-th* tooth period
$\Delta^2 F_a(i)$	second order differencing of the average force during the *i-th* tooth period
I	identity matrix
k_0	average power of the latent function
k_b	variance of the offset to the latent function
k_l	ARD parameter
K_{pr}	peak rate of cutting forces
K_R	radial cutting force coefficient

K_T	tangential cutting force coefficient
K(t)	estimation gain in AR^1 model
kts	kurtosis
$K(\mathbf{x}, \mathbf{x}_i)$	kernel function
G	harmonics of cutting force
GW	Gradual Wear
l	loss function
LDF	linear discriminant function
lw	length of the workpiece
MAP	maximum a posteriori
MLP	multi-layer perceptron
N_m	maximum rotation number within one pass
N_{max}	maximum number of samples
*P(**D**)*	prior probability of training data
$P(\mathbf{D}\|\mathbf{f})$	likelihood
$P(\mathbf{f})$	prior probability of latent function
$P(\mathbf{D}\|\boldsymbol{\theta})$	evidence
P_{TH}	total harmonic power
$\boldsymbol{R}$	real number space
r_a	amplitude ratio
RBF	radial basis function
RCE	restricted *Coulomb* energy
$\boldsymbol{R}^d$	d-dimensional real function space
re	residual error
rt	effective radius of the tool holder

sod	second order differencing
SOM	self-organizing map
sr	sampling rate
sre	sum of the squares of residual errors
S_s	spindle speed
std	standard deviation
SV	support vectors
SVC	support vector classification
SVM	support vector machine
SVR	support vector regression
TBD	tool breakage detection
TCM	tool condition monitoring
T_{CR}	time for the cutter to move by a distance of its radius
T_d	delay time of a timer routine
T_{ij}	the i^{th} tooth period during the j^{th} spindle rotation
thp	total harmonic power
T_p	Processing Time
TWD	tool wear detection
TWE	tool wear estimation
TWR	tool wear recognition
VB_{max}	maximum width flank wear
v_d	digitalized voltage level
$W(i,t)$	interruption function
α	a large constant in AR^1 model
α_i	Lagrange multiplier

δ	disturbance
$\delta\varepsilon$	incremental radial run-out
ε	run-out, parameter of Huber's function
η	normalization factor
$\bar{\theta}_i(t)$	angular position of insert *i* at time instance *t*
λ	forgetting factor in AR^1 model
$\mathbf{\Lambda}$	diagonal matrix
$\mathbf{\Sigma}$	covariance matrix
Φ	parameters in AR model
ψ	feature extraction function
$\boldsymbol{\omega}$	weight vector of neural networks

List of Figures

List of Tables

CHAPTER 1

INTROUDUCTION

1.1 BACKGROUND

In recent years, significant advances have been achieved in the manufacturing environment. Manufacturing systems are fast converting into fully automated environments such as computer integrated manufacturing systems (CIMS) and flexible manufacturing systems (FMS). However, in order to meet the need of industries for saving cost, improving quality, and reducing production time, robust and practical process monitoring systems have to be further developed and introduced [Byrne, 1995]. Tool condition monitoring (TCM) systems are among such kind of systems, which are considered to be the most crucial and determining factor to successful maximization of the metal cutting process [Dimla, 1996].

Tool condition monitoring is primarily for tool wear monitoring [Lange, 1992]. Tool failure resulted from wear represents about 20% of machine tool down-time and negatively impacts the work quality in the context of dimensions, finish, and surface integrity [Liang, 2002]. As a result, considerable research has been carried out in this area, including turning [Emel, 1988; Abu-Zahra, 1997; Niu, 1998], milling [Altintas, 1989; Elbestawi, 1991; Tarng, 1994], and drilling [Tansel, 1992; Elwardany, 1996; Huseyin, 2001]. No matter for which kind of processes the tool condition monitoring system is developed, it can be viewed as an information flow and processing system.

The information flow in the tool condition monitoring systems starts at the data acquisition stage, when signals are measured from the process using sensors. The sensor systems can be categorized into direct and indirect measurement systems. Direct measurement techniques measure the tool geometry directly, such as optical scanning of tool tips [Yamazaki, 1974], laser displacement and intensity measurement of tool geometric failures [Ryabov, 1996], and optical measurement of the flank wear land [Kurada and Bradley, 1997]. These systems possess a high degree of accuracy. However, they are unsuitable for practical deployment due to installation problems and the harsh environment of the practical cutting processes [Byrne, 1995]. Indirect measurement systems measure some process-borne quantities, from which the actual tool wear can be deduced. These include measurement of cutting forces [Altintas, 1988; Elbestawi, 1990; Tansel, 1994], acoustic emissions (AE) [Sampath, 1987; Wilcox, 1997; Jemielniak, 1998], vibrations [Lee, 1987; Coker, 1996; Li, 2000], and feed drive current [Rangwala, 1987; Altintas, 1992]. These systems are less complex and more suitable for practical application [Byrne, 1995]. The sensor systems can also be categorized into multiple-sensor and single-sensor systems, according to the types of the sensors deployed. Multiple-sensor systems [Silva, 1997; Choi, 1999] provide richer information about the process by various kinds of signals, and thus ensure a better performance. Single-sensor systems [Yao, 1993; Purushothaman, 1994] are easier to implement and more suitable for real-time applications due to the smaller amount of information to process.

The information processing in the tool condition monitoring system is responsible for extracting meaningful features from raw signals and making decisions on tool conditions. For the direct measurement systems, tool wear can be directly obtained from the acquired data. For example, the flank wear land can be directly extracted

from the captured tool images [Kurada and Bradley, 1997]. For the indirect measurement systems, the acquired data have to be mapped to tool wear in quite different approaches. Multiple features are usually extracted to replace the raw data. Then they are fed into an empirical model to deduce tool wear, such as a stochastic-process model [Altintas, 1988] and a neural network [Tansel, 1994].

Some commercial tool condition monitoring systems are now in the market and are used in industry. However, the systems have narrow range of performance or require substantial training or setup time to function correctly [Byrne, 1995; Liang, 2002]. The current research activities in TCM aim to develop systems with higher reliability and flexibility.

1.2 LITERATURE REVIEW

This study focuses on milling process monitoring using force signals, due to its high sensitivity to tool wear [Altintas, 1989], robustness in harsh working environments and convenience in installation [Byrne et. al., 1995]. The review of the literature concentrates on some of the relevant studies. These can be generally categorized into three methodologies, including model-based method, statistical-stochastic analysis, and artificial intelligence approaches.

1.2.1 MODEL-BASED METHODS

The research on tool life can be traced back to Taylor's work around 1906. He built a model, in which the tool life was related to the cutting speed by a power function relationship [Taylor, 1906]. This model is based on empirical results rather than on a physical model of the wear process, and therefore does not always work in tool life prediction.

Rabinowica [1977] developed a quantitative description of the abrasive wear process over the entire range of abrasive hardnesses. It was a model of abrasive force and was dependent on the hardness of the tool and the inclusions in the workpiece. This model adequately explained the relationship between the wear and mechanical activation.

Kramer [1986] suggested that there were other causes of tool wear, and separated the mechanisms controlling the wear rate of a tool materials into three regimes, depending on the cutting temperature and the properties of the tool and workpiece materials. The first is the low-temperature regime, where the wear of the tool material is determined primarily by its hardness. Rabinowica's abrasive model works well in this regime. The other two regimes are under higher cutting temperatures, with the solid solubility and the chemical dissolution of the tool material determining the wear resistance. Based on this understanding, Kramer came up with a chemical dissolution wear model, and combined it with the abrasive model, which resulted in a composite wear rate model.

Koren [1978] developed a flank wear model using a linear control theory. He assumed two principal mechanisms as wear causes: a thermally activated one and a mechanically activated one. The wear process is mathematically treated as a positive feedback process, whereby the wear raises the cutting forces and temperature and it thereby raises the wear growth rate.

The model-based methods mentioned above contribute to the understanding of the physical mechanisms of tool wear process, the determination of optimal cutting

conditions, and the design of tool materials. However, they are the functions of cutting conditions and dependent on the properties of the tool and workpiece materials. To implement, a large database must be established through numerous experiments to furnish the constants in the models.

1.2.2 STATISTICAL-STOCHASTIC ANALYSIS

In 1980's and the early 1990's, the trend of the research on tool condition monitoring is based on statistical and stochastic analysis. These methodologies are employed to evaluate the relationships between tool wear processes and the characteristics of the signals in both the time domain and the frequency domain. Thresholds are commonly imposed on the results from the analysis to make a judgment on tool state.

Time series analysis has been successfully adopted by many researchers to sense tool breakages. Lan [1986] monitored the feed forces in milling using a very high-order autoregressive time series filter (AR^{15}) to detect tool breakages. Altintas [1988] suggested that high-order time-series filters are not practical for real time applications due to the large computation time and the inefficiency in distinguishing the transient cutting from the tool breakage event. He thus proposed an AR^{1} model to predict the cutting force and calculate the difference between the actual measurement and predicted value, which was called as the residual error of the cutting force. He found that when the process suddenly and sharply deviates from its normal course, which means a breakage occurs, the model becomes unable to track the process for several intervals. He used this force variation phenomenon to detect tool breakages in milling. A similar approach can be found in Yan's work [1995]. Also by using AR models

(20^{th}~24^{th} order), Tansel [1993a] further evaluated the estimation error by calculating the sum of the squared residual errors in each tooth period.

Without the prediction steps in time-series analysis methodologies, some statistical quantities of cutting force signals can be calculated and used to monitor tool status. Altintas [1989] used the first and second order differencing of a time averaged resultant force to detect tool failures in milling. Tarn [1989] calculated four quantities from each tooth period to monitor tool and cutting conditions in milling, which included maximum force level, total amplitude of the cutting force, combined incremental force changes, and amplitude ratio. Zhang [1995] used the peak rate of cutting forces, and the relative eccentricity rate of the cutter to detect tool breakages. The force peak rate of the adjacent tooth periods was defined as the ratio between the difference and the sum of force peaks in adjacent tooth periods, which was claimed to be independent of the cutting conditions such as cutting depth, cutting thickness and feed, etc.

Signal processing techniques have also been successfully used in monitoring tool failures. Tarng [1990] defined a tool breakage zone, which is located within the frequency range between the d.c. component and the tooth passing frequency. And he found that the force components within this zone correlate to the tool breakage very well. He extracted the tool breakage zone components using a band-pass filter. Then the standard deviation of the force data was calculated. Elbestawi [1991] et al performed FFT on the cutting force signal to obtain the spectrum of the cutting force. Then the ratio between the harmonics which are most and least sensitive to wear was

calculated. However, a database has to be established and used for searching of the harmonics which are most and least sensitive to wear.

The major difficulty of the statistical-stochastic analysis methodologies lies in the determination of the threshold, which could be quite sensitive to various cutting conditions and tool-workpiece properties.

1.2.3 ARTIFICIAL INTELLIGENCE APPROACHES

Recently, it has been widely acknowledged that a better solution for TCM systems lies in artificial intelligence approaches [Monostori, 1993]. These approaches include pattern recognition, expert system, neural network, and fuzzy logic. Like the statistical-stochastic analysis methodologies, it is also necessary to extract meaningful features from raw signals in using these approaches. However, tool failure detection using artificial intelligence approaches is more sophisticated than just using thresholds, because of the complicated procedure in making a decision.

Elbestawi [1989] designed a linear discriminant function (LDF) classifier to partition the feature space into signal classes. He found that the harmonic contents of cutting forces and spindle vibrations are sensitive to tool flank wear. So he summed up the powers at the fundamental tooth frequency and its harmonics and derived a total harmonic power. Then the total harmonic powers of cutting forces and spindle vibrations were mapped into one of the partitions through the classifier. And then a decision could be made on tool status.

Unlike LDF operators, neural networks have the advantages of realizing complicated nonlinear mappings. They have been widely used in TCM systems, both for tool failure detection and for tool wear estimation. Leem [1995] used a customized neural network in online monitoring of cutting tool wear. Power spectrum and four statistics (mean, standard deviation, skew, and kurtosis) were extracted from cutting force and AE signals. Tool wear levels were first topologically ordered by Kohonen's self organizing map (SOM). Then the input features were transformed via input feature scaling to make the decision boundaries of the neural network approximate those of error-minimizing Bayes classifier. Tansel [1992] compared two types of neural networks, the restricted Coulomb energy (RCE) and the adaptive resonance theory (ART2), in tool breakage detection. 10 normalized averages within one full tool revolution were used as input features. RCE-type neural networks were found to be convenient and beneficial for detection of tool breakage in processes with constant cutting conditions. ART2 was found to be better in varying cutting conditions and heavy tool wear, due to the continuous learning capability. Tarng [1992] applied a multi-layer perceptron (MLP)-type neural network in sensing tool breakage. The average force and the variable force, derived by subtracting the median force from the average force, were used as input features. In the later work of Tansel [1995], wavelet transformations were used in compressing the force signals and eliminating the high-frequency components. Then the estimated parameters of the wavelet transformations were classified by using ART2-type neural networks. Better performances were achieved than using the 10 averages in one revolution in his earlier work.

Neural networks have also been widely used in tool wear estimation. Using neural networks to model complex data can be considered as performing a curve fitting

operation in multidimensional space. Elanayar [1995] used radial basis function neural networks (RBF) to map feed rate and spindle speed to flank and crater wear. Good results were reported for flank wear estimation. However, the performance for estimating crater wear was not reliable. Santanu [1996] mapped average force and cutting conditions to flank wear using MLP-type neural networks. Reasonably close assessment of target flank wear values was achieved. A similar approach can be found in Lin's work [1996]. Besides the neural network approach, Lin also established and evaluated two regression models. The 6-24-12-1 network model was finally proven to be more accurate in tool wear prediction.

According to these prior studies, the advantages of neural networks in TCM applications can be summarized as follows:

- fault tolerance and adaptability;
- data-driven nature;
- noise suppression; and
- parallel processing capabilities.

1.3 OBJECTIVES AND SCOPE OF THIS STUDY

The TCM methodologies based on the statistical-stochastic analysis and artificial intelligence approaches are listed in Table 1.1 in chronological order. It can be clearly seen that there are many different kinds of features. Although all of these features have been shown to be effective for TCM, it is only until recently that few studies have been done to compare them [Goebol, 2000; Sun, 2002]. The necessity to do the comparison is two-fold. First of all, in the implementation of online systems, a compact feature set means less computation time and therefore better real-time performance. Besides, the

Table 1.1 TCM Methodologies

No	Objective	Features	Decision Making	Reference
1	TBD[1]	Residual Error	Thresholding	Lan, 1986
2	TBD	Residual Error	Thresholding	Altintas, 1988
3	TWD[2]	1st & 2nd order differencing	Thresholding	Altintas, 1989
4	TWD	Maximum Force Level, Total Amplitude of Cutting Force, Combined Incremental Force Changes, Amplitude Ratio	Thresholding	Tarn, 1989
5	TWD	Power Spectral Density of Force and Spindle Vibration	LDF-Classifier	Elbestawi, 1989
6	TBD	Force Components in Tool Breakage Zone	Thresholding	Tarng, 1990
7	TWD	Ratio between Force Harmonics	Thresholding	Elbestawi, 1991
8	TBD	10 Normalized Averages in One Tool Revolution	RCE, ART2	Tansel, 1992
9	TBD	Average Force and The Variable Force	MLP	Tarng, 1992
10	TBD	Sum of the Squares of Residual Errors	Thresholding	Tansel, 1993
11	TWD	Power Spectral Density and Mean, Standard Deviation, Skew, Kurtosis of Force and AE	SOM	Leem, 1995
12	TBD	Peak Rate of Cutting Forces, Relative Eccentricity Rate of Cutter	Thresholding	Zhang, 1995
13	TWD	Wavelet Transformations Coefficient	ART2	Tansel, 1995
14	TWE[3]	Feed Rate, Spindle Speed	RBF	Elanayar, 1995
15	TWE	Average Force, Cutting Conditions	MLP	Santanu, 1996
16	TWE	Average Force, Cutting Conditions	MLP, Regression Models	Lin, 1996

1. TBD: Tool Breakage Detection; 2. TWD: Tool Wear Detection; 3. TWE: Tool Wear Estimation.

proper selection of features is a vital issue in using neural networks. Including irrelevant features can ultimately lead to poor performance, because it is inevitable that the irrelevant features can be more closely associated with the targets by chance than are the truly relevant ones (Neal, 1996). Based on these two considerations, a small but efficient feature set is a key factor for the implementation of practical TCM systems. As a result, the main focus of this study is to select more relevant features from the known features.

In this study, force signal is used as the sensor information for monitoring face milling processes, because of its high sensitivity to wear and low noise. 16 well known features based on the force signal are extracted. The automatic relevance determination (ARD) algorithm, originated by MacKay [1992] and Neal [1996], is used to select a subset of the features with higher relevance to tool wear processes. The feature selection procedures are conducted for both tool wear recognition (TWR) using Bayesian support vector classification (BSVC) algorithm and tool wear estimation (TWE) using Bayesian support vector regression (BSVR) algorithm. The generalization capabilities using the entire feature set, the selected feature set, and the rejected feature set are compared for both TWR and TWE to verify the relevance of the selected features to tool wear processes.

1.4 ORGANIZATION OF THE THESIS

Chapter 1 gives a brief introduction on tool condition monitoring and its methodologies. 16 different feature extraction algorithms are discussed in details in Chapter 2. Chapter 3 introduces the Bayesian support vector classification and regression algorithms, as well as the automatic relevance determination approach. The

experimental setup for data acquisition and a software structure for online tool condition monitoring are described in Chapter 4. The feature selection results and the comparisons of the generalization capabilities using the entire, selected, and rejected feature sets are given in Chapter 5. Conclusions are given in the last chapter together with a recommendation for future work.

CHAPTER 2

FEATURE EXTRACTION METHODOLOGIES

The challenge in developing a TCM system is in choosing suitable sensing techniques and robust decision making strategies. For monitoring milling processes, force signal is widely used due to its high sensitivity to tool wear, low noise to signal ratio, and satisfactorily accurate force models [Altintas, 1989]. As mentioned in Chapter 1, considerable research has been undertaken for the development of feature extraction methodologies based on force signals. In this chapter, a mechanistic force model of milling processes is first given as a theoretical background. Then 16 different feature extraction methodologies are introduced.

2.1 MECHANISTIC FORCE MODEL OF MILLING PROCESSES

Force mechanisms of milling processes have been well understood. And satisfactorily accurate models have been established (Fu, 1984; Zheng, 1999).

Figure 2.1 shows the cut geometry used in this study. If there is no run-out, the expression for the chip area cut by insert *i* at time *t* is given by:

$$\overline{A}_c(i,t) = \bar{f}_t \sin\left(\overline{\theta}_i(t)\right) \cdot W(i,t) \cdot doc \tag{2.1}$$

where $\bar{f}_t$ is the feed per tooth, *doc* is the depth of cut, and $\overline{\theta}_i(t)$ is the angular position of insert *i* from the negative *Y* axis in the clockwise direction. $W(i,t)$ is the interruption

function that assumes values 1 or 0 depending on whether or not insert *i* is cutting at time *t*.

The tangential and radial cutting forces, F_T and F_R, acting on an insert *i*, are expressed as the product of the chip area $\overline{A}_c(i,t)$ and the cutting force coefficients K_T and K_R, respectively:

$$F_T(i,t) = K_T \cdot \overline{A}_c(i,t) \tag{2.2}$$

$$F_R(i,t) = K_R \cdot \overline{A}_c(i,t) \tag{2.3}$$

The radial and tangential forces acting on insert *i* can be transformed to the global *X*, *Y* coordinate frame and summed over all the *N* inserts to express the forces acting on the cutter as:

$$\begin{bmatrix} F_x(t) \\ F_y(t) \end{bmatrix} = \sum_{i=0}^{N-1} \begin{bmatrix} \cos\theta_i(t) & \sin\theta_i(t) \\ -\sin\theta_i(t) & \cos\theta_i(t) \end{bmatrix} \cdot \begin{bmatrix} F_T(i,t) \\ F_R(i,t) \end{bmatrix} \tag{2.4}$$

where $\theta_i(t)$ is the angular position of insert *i* from the negative *Y* axis in the clockwise direction.

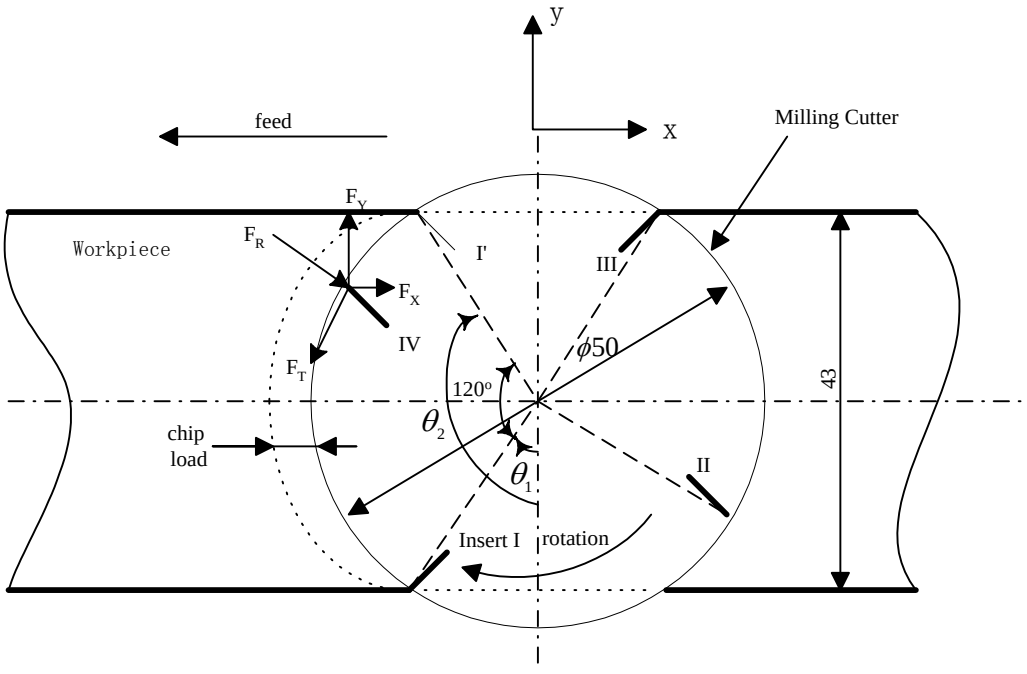


Figure 2.1 Face Milling Geometry

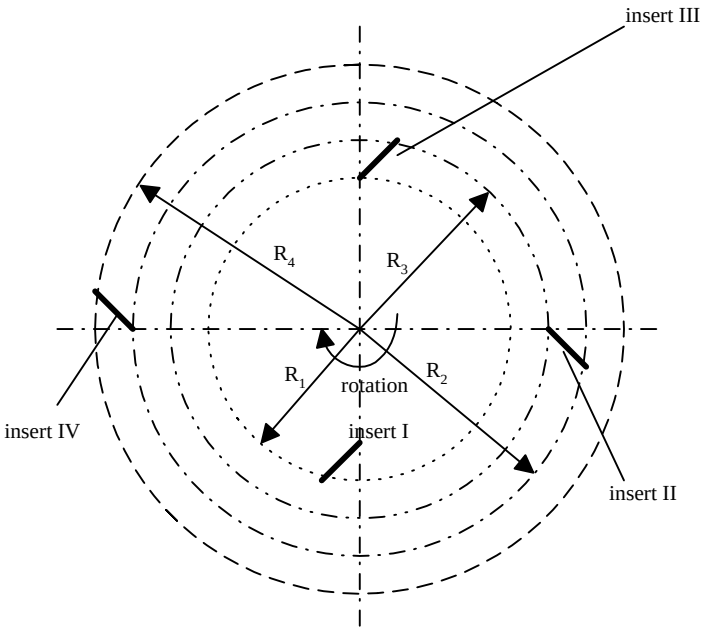


Figure 2.2 Cutter Geometry with Runout

In the presence of radial runout, the chip load equation and the subsequent force models must be modified. Figure 2.2 shows the radial position of the teeth on a cutter with radial runout. The radial runout of insert *i* can be expressed as:

$$\varepsilon_i = \left| R_i - R \right| \tag{2.5}$$

where *R* is the true cutting radius. Then the chip load equation can be modified to:

$$A_c(i,t) = \left(\bar{f}_t \sin\left(\bar{\theta}_i(t)\right) + \delta\varepsilon_i \right) \cdot W(i,t) \cdot doc \tag{2.6}$$

where $\delta\varepsilon_i$ is the incremental radial runout faced by insert *i*. For clockwise rotation of the cutter, the incremental radial runout for insert *i* is given by:

$$\delta\varepsilon_i = \min\left\{ \varepsilon_i - \varepsilon_{i-1}, \varepsilon_i - \varepsilon_{i-2} + \bar{f}_t \sin\bar{\theta}_i(t), \varepsilon_i - \varepsilon_{i-3} + 2\bar{f}_t \sin\bar{\theta}_i(t), \cdots \right\} \tag{2.7}$$

Force models can be then modified by substituting $A_c(i,t)$ for $\bar{A}_c(i,t)$ in Equ. (2.2) and Equ. (2.3). Both the models with and without runout are used in the subsequent analysis of cutting force.

2.2 FEATURE EXTRACTION METHODOLOGIES

(i) Residual Errors

Altintas [1988] built a first order autoregressive (AR^1) model to predict the cutting force and evaluated the difference between the actual measurement and predicted value, which was called the residual error of the cutting force. He found that when tool breakages occur, the model becomes unable to track the process, and therefore produces a large residual error.

An autoregressive model with order *p* can be written as:

$$F(t) = \Phi_1 F(t-1) + \Phi_2 F(t-2) + \cdots + \Phi_p F(t-p) + a(t) \tag{2.8}$$

where *F(t)* and *a(t)* are respectively the measured signal and the disturbance at time *t*, and $\Phi_1, \Phi_2, \cdots, \Phi_p$ are the filter parameters. The first order AR model is the one step ahead estimation of *F(t)* at time *(t-1)*:

$$F(t) = \Phi_1 F(t-1) + a(t) \tag{2.9}$$

Based on Equation 2.9, Altintas' AR^1 model can be expressed as:

$$f_d(t) = \Phi f_d(t-1) + a(t) \tag{2.10}$$

$$\& \quad f_d(t) = F_a(t) - F_a(t-1) \tag{2.11}$$

where $F_a(t)$ is the average force over the *t*-th tooth period. The residual error can therefore be calculated by:

$$a(t) = f_d(t) - f_d(t-1) \cdot \hat{\Phi}(t-1) \tag{2.12}$$

where $\hat{\Phi}(t-1)$ is the estimated value of Φ, which can be evaluated by:

$$\hat{\Phi}(t) = \hat{\Phi}(t-1) + K(t-1) \cdot a(t) \tag{2.13}$$

where *K(t)* is the estimation gain:

$$K(t) = \frac{P(t) \cdot f_d(t)}{\lambda + P(t) \cdot f_d^2(t)} \tag{2.14}$$

The λ in the above equation is the forgetting factor with a value between 0.9 and 1. And *P(t)* can be updated by:

$$P(t+1) = \frac{P(t)}{\lambda}[1 - K(t) \cdot f_d(t)] \tag{2.15}$$

Figure 2.3 illustrates the procedures for using this model, with the following initial conditions: $\Phi(0) = 0$, and $P(0) = \alpha$ where α is a large number. An example of the residual errors during the whole process of an experiment is shown in Figure 2.4. The force samples from the same experiment are used for extracting all the other features introduced in this chapter.

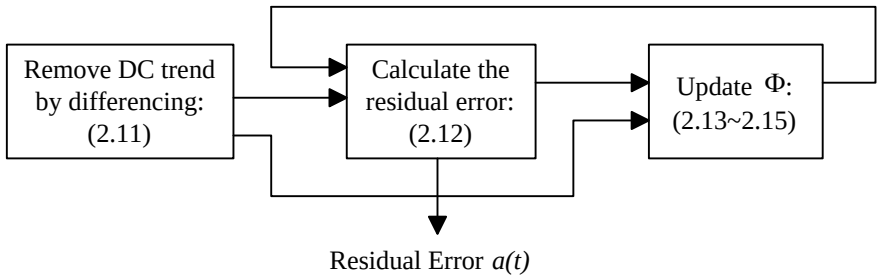


Figure 2.3 Procedures for Calculating Residual Errors

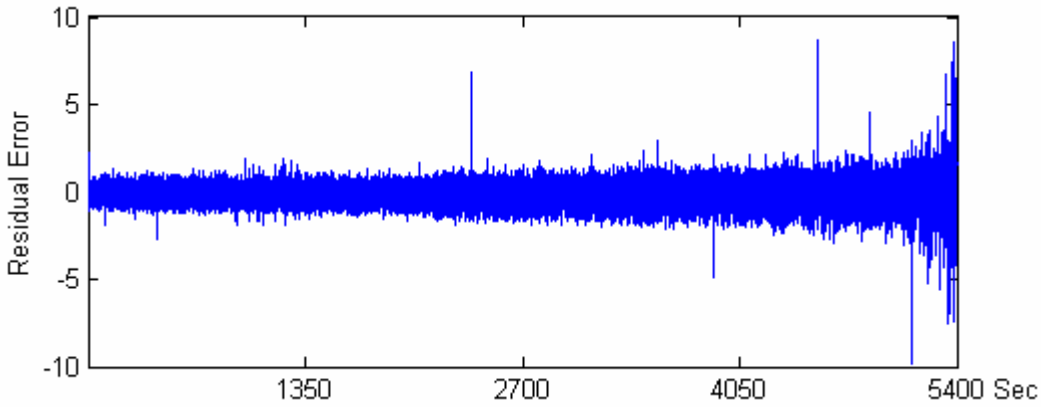


Figure 2.4 Residual Errors
Spindle Speed: 1000rpm, Feed Rate: 100mm/min,
Depth of Cut: 1 mm, Insert Number = 2, Insert Type: AC325.

(ii, iii) First & Second Order Differencing

In another attempt by Altintas [1989], the first and second order differencing of a time averaged resultant force were found effective in recognition of tool breakages in milling. The first order differencing of the average cutting forces compares the cutting performances of the adjacent teeth:

$$\Delta F_a(i) = F_a(i) - F_a(i-1) \tag{2.16}$$

where $F_a(i)$ is the average force during the *i-th* tooth period. The second order differencing can be evaluated from $\Delta F_a(i)$:

$$\Delta^2 F_a(i) = \Delta F_a(i) - \Delta F_a(i-1) = F_a(i) - 2F_a(i-1) + F_a(i-2) \tag{2.17}$$

These two features are shown in Figures 2.5 and 2.6, respectively.

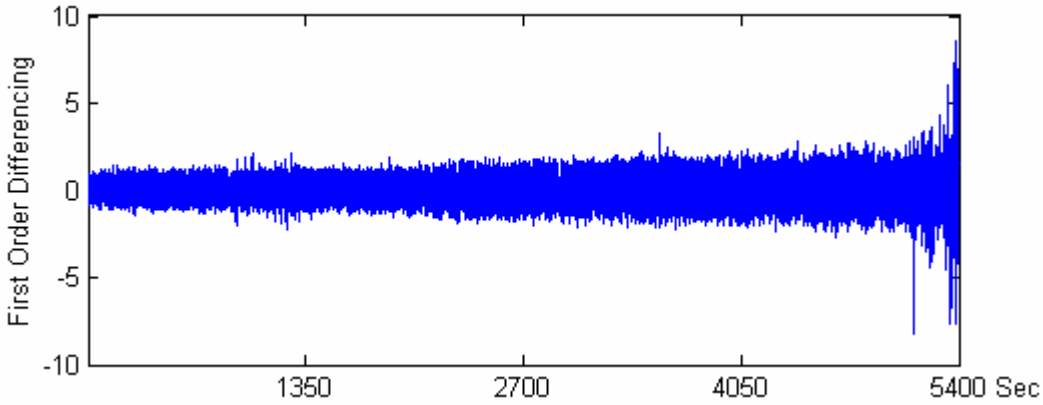


Figure 2.5 First Order Differencing of Cutting Force

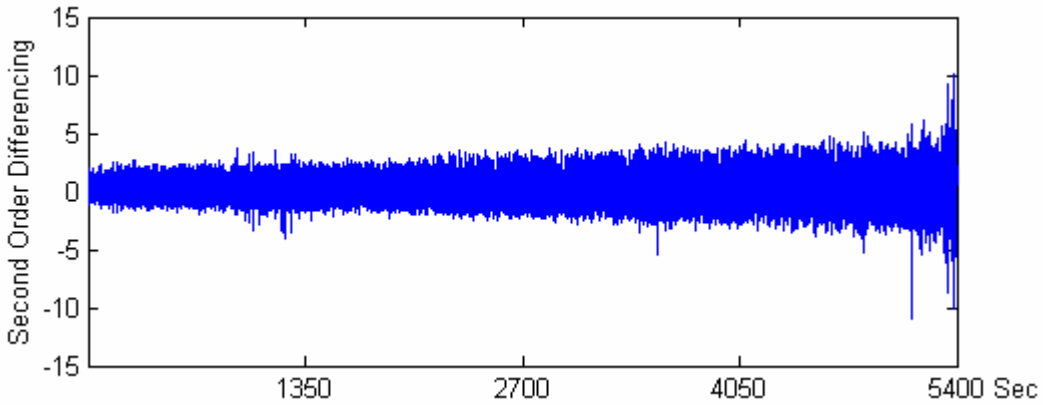


Figure 2.6 Second Order Differencing of Cutting Force

(iv~vii) Maximum Force Level, Total Amplitude of the Cutting Force, Combined Incremental Force Changes, and Amplitude Ratio

Tarn [10] calculated four quantities from each tooth period to monitor tool and cutting conditions in milling. The first two features, maximum force level (f_m) and total amplitude of cutting force (f_a), represent the steady-state and variational portion of the instantaneous cutting force. They can be derived from Equations (2.18) and (2.19), respectively:

$$f_m(i,j) = \max_{t \in T_{ij}} \left| f(i,j,t) \right| \tag{2.18}$$

$$f_a(i,j) = \max_{t \in T_{ij}} \left[f(i,j,t) \right] - \min_{t \in T_{ij}} \left[f(i,j,t) \right] \tag{2.19}$$

where i denotes the i^{th} cutting edge, j denotes the j^{th} spindle rotation, and $f(i,j,t)$ denotes the cutting force where t varies over the tooth period, T_{ij}. From Equation 2.19, it can be seen that $f_a(i,j)$ is actually the peak-to-peak value of the force waveform during the i^{th} tooth period.

The third feature, combined incremental force changes, indicates the changes in cutting conditions. It combines the incremental changes in the first two features between the j^{th} and $(j+1)^{th}$ spindle rotation:

$$\Delta f(i,j) = \left(f_m(i,j+1) - f_m(i,j) \right) + \left(f_a(i,j+1) - f_a(i,j) \right) \tag{2.20}$$

The fourth feature is called as the amplitude ratio. It can be evaluated from $f_a(i,j)$ and $f_a(i+1,j)$:

$$r_a(i,j) = \frac{\max\left[f_a(i,j),\, f_a(i+1,j) \right]}{\min\left[f_a(i,j),\, f_a(i+1,j) \right]} \tag{2.21}$$

When cutting geometry changes in the $(i+1)^{th}$ tooth period, $f_a(i+1,j)$ will be different from $f_a(i,j)$ (chip area changes). So the deviation of this quantity from unity indicates the changes in cutting edge geometries. Figures 2.7 to 2.10 illustrate the four features extracted from the force data of the same experiment as described in Figure 2.4.

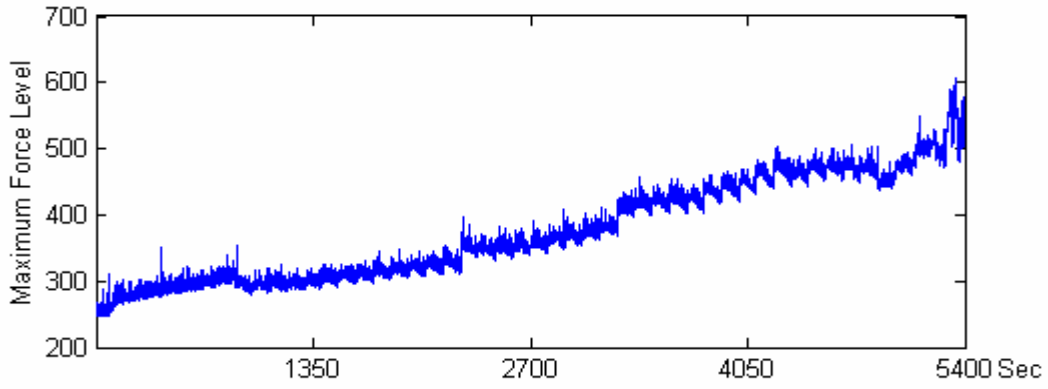


Figure 2.7 Maximum Force Level

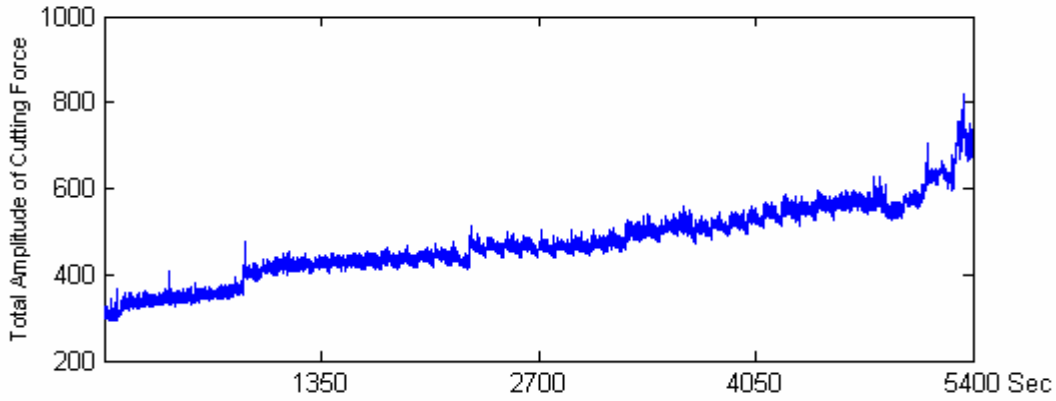


Figure 2.8 Total Amplitude of the Cutting Force

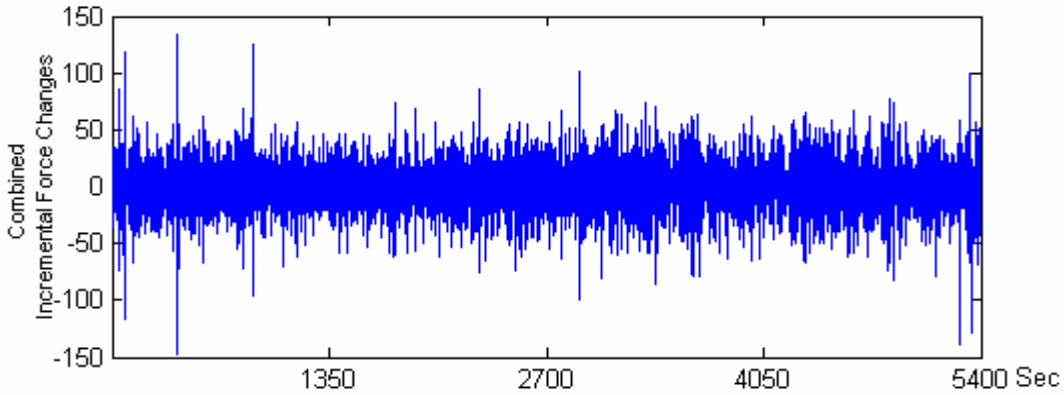


Figure 2.9 Combined Incremental Force Changes

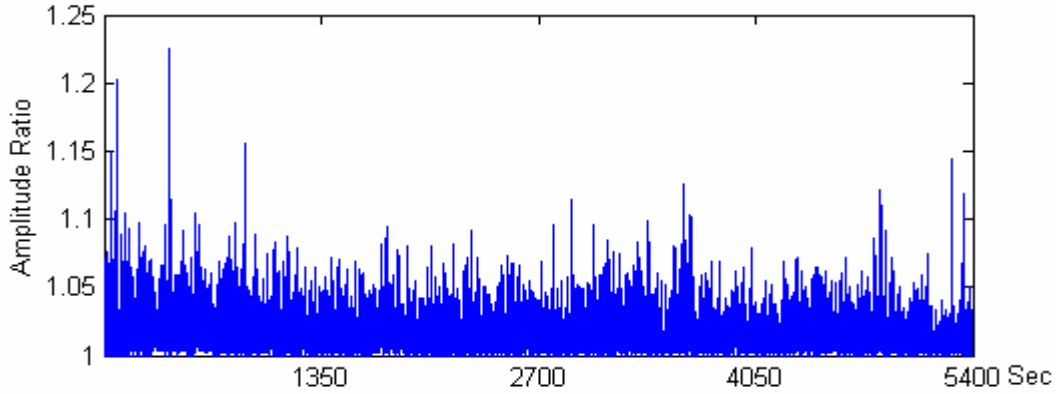


Figure 2.10 Amplitude Ratio

(viii) Standard Deviation of the Force Components in Tool Breakage Zone

Tarng [1990] defined a tool breakage zone, which is located within the frequency range between the DC component and the tooth frequency. He found that the force components within this zone correlate to tool breakage very well. The tool breakage zone components were extracted using a band-pass filter. Then the standard deviation of the filtered force data was calculated.

Figure 2.11 shows the cutting force signal and its spectrum at both the fresh and the highly worn stage. When the tool was still fresh, the two peaks in a single rotation were not equal due to the large runout of the cutter; and therefore the component at the spindle rotation frequency was very large. During the cutting process, the runout was gradually compensated by the uneven wear of the two teeth. As the tool became highly worn, the two peaks became similar; and therefore the tooth passing frequency dominated in the spectrum.

If the runout is negligible compared to the feed per tooth, the opposite situation will become true. During the fresh stage, the force peaks appear to be identical; and the spectral power concentrates on the tooth passing frequency. When the tool gets highly worn or broken, there will be a severe fluctuation in the force waveform; and therefore the component at the spindle rotation frequency will in turn dominate. This is shown in Figure 2.12 by simulated force data (in most of the real cases, runout is comparable to the feed per tooth and cannot be ignored).

No matter whether runout is negligible or not, the spectral components in the tool breakage zone change a lot from the fresh stage to the worn stage. So they can be used to distinguish failed tools from fresh ones.

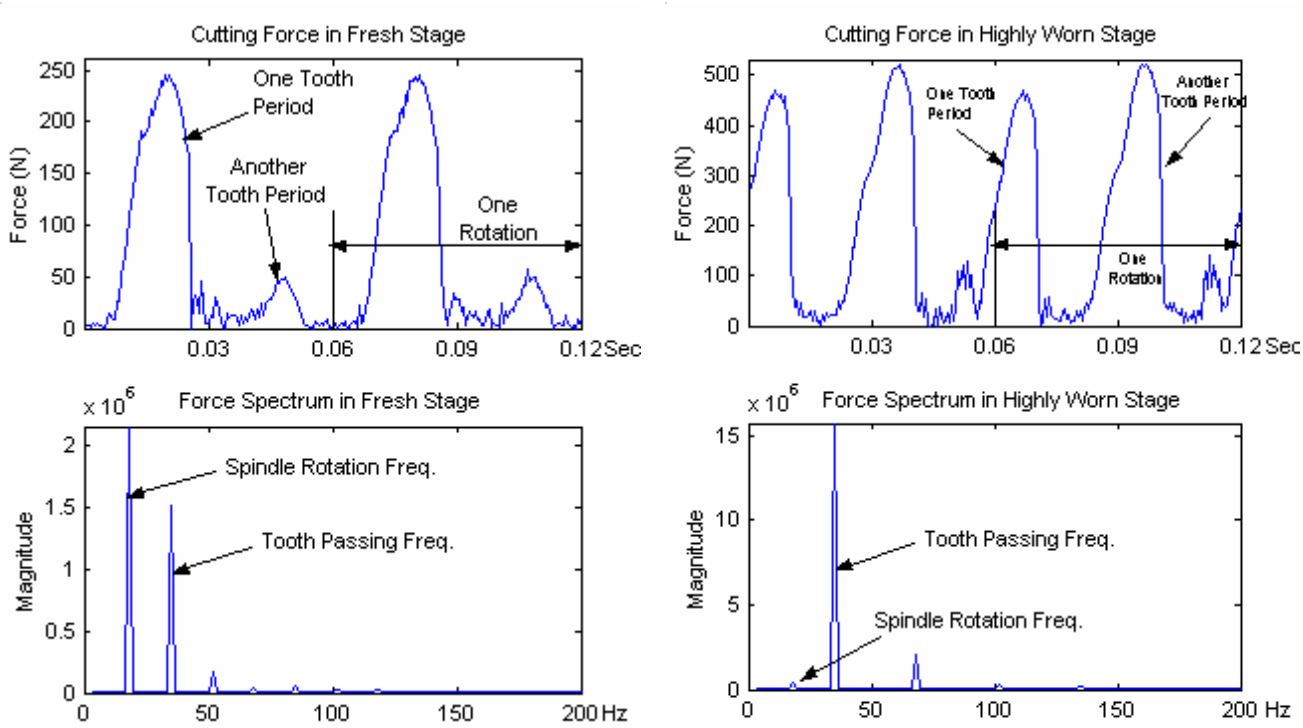


Figure 2.11 Cutting Force and Its Spectrum in Two Rotations

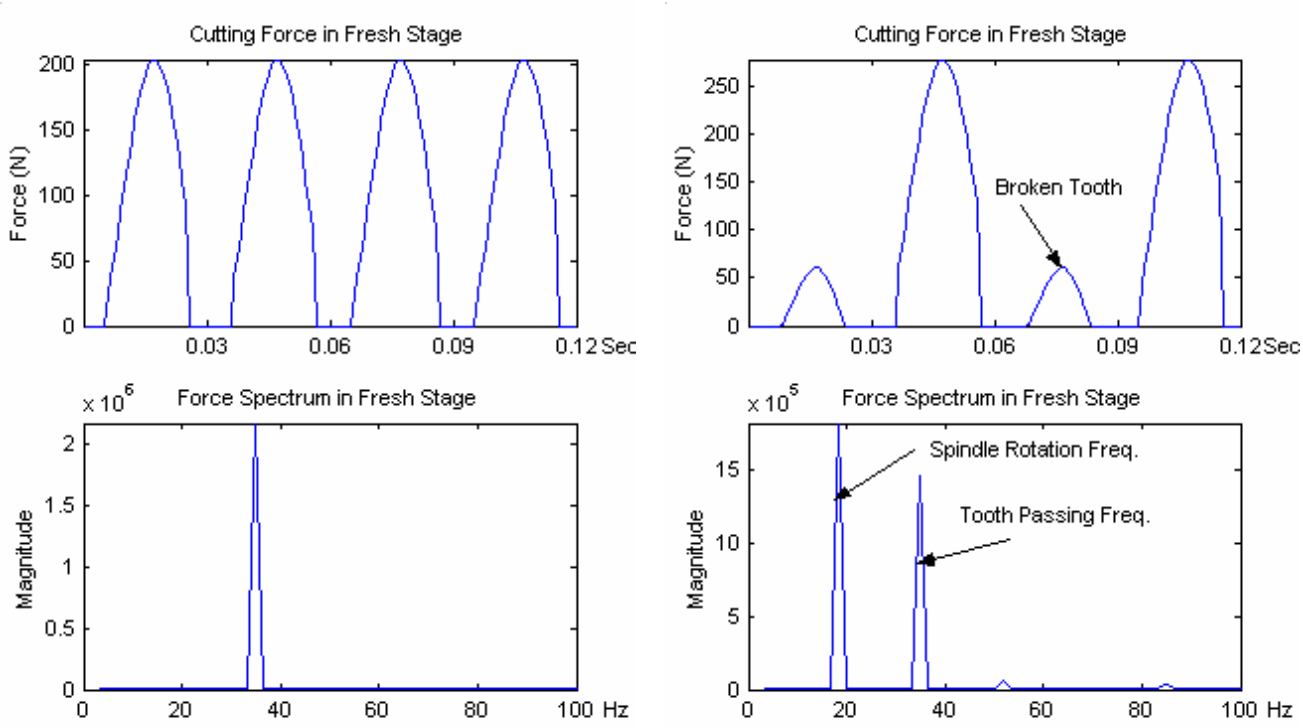


Figure 2.12 Simulated Cutting Force and Its Spectrum

The procedures for calculating the standard deviation of the force components in tool breakage zone are shown in Figure 2.13. And Figure 2.14 shows an example of this feature.

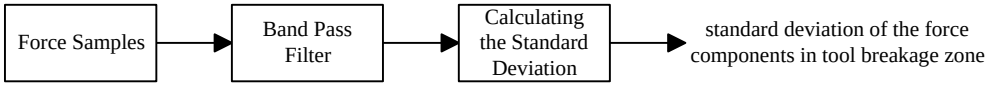


Figure 2.13 Procedures for Calculating the Feature

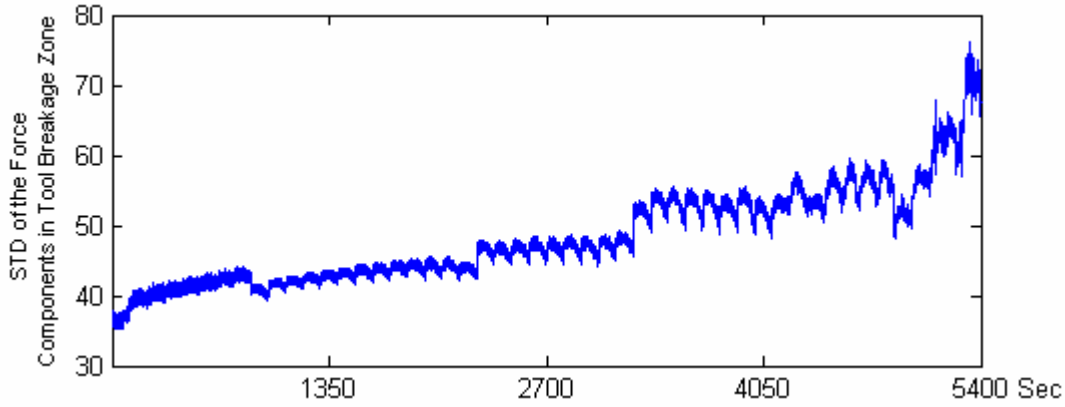


Figure 2.14 Standard Deviation of the Force Components in Tool Breakage Zone

(ix) Sum of the Squares of Residual Errors

By using high-order AR models (20th order), Tansel [1993] also derived the residual errors of the cutting force. But he further evaluated the estimation error by summing up the squares of the residual errors in each tooth period. The force estimation at the time instance *i* is calculated by the *n-th* (*n*=20) order AR model:

$$F'(i) = \sum_{k=1}^{n} f(i-k) \times \Phi_k(j-1) \tag{2.22}$$

where $f(i-k)$ is the *(i-k)*th measured force during the j^{th} tooth period, $\Phi_k(j-1)$ are the parameters of the model estimated at the end of the previous tooth period. The residual error of the model at the time instance *i* can be calculated by:

$$E(i) = f(i) - F'(i) \tag{2.23}$$

The amount of the error for each tooth period *j* can then be calculated by the sum of squares of the residual errors *E(i)*:

$$S(j) = \sum_{k=1}^{l} E^2(j \times l + k) \tag{2.24}$$

where *l* is the number of force samples per tooth period. Figure 2.15 shows an example of this feature.

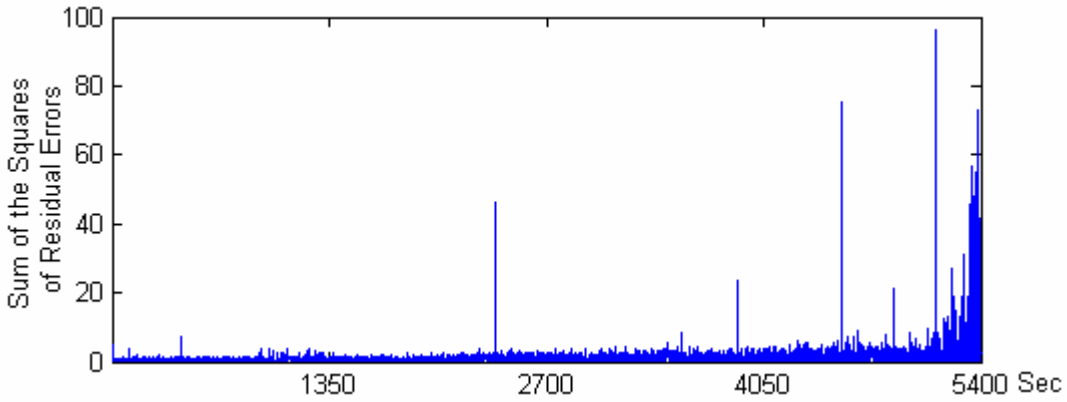


Figure 2.15 Sum of the Squares of Residual Errors

(x) Peak Rate of Cutting Forces

Zhang [1995] used the peak rate of cutting forces to detect tool breakages. It was defined as the ratio between the difference and the sum of force peaks in adjacent tooth periods:

$$K_{pr}(i, j) = \frac{F_p(i, j) - F_p(i-1, j)}{|F_p(i, j)| + |F_p(i-1, j)|} \tag{2.25}$$

where $F_p(i,j)$ is the peak value of the cutting force in the *i-th* tooth period during the *j-th* tool rotation. Introduce Equation (2.18) into (2.25):

$$K_{pr}(i, j) = \frac{f_m(i, j) - f_m(i-1, j)}{|f_m(i, j)| + |f_m(i-1, j)|} \tag{2.26}$$

The force peak rate is dimensionless and independent of the cutting conditions such as cutting depth, cutting thickness and feed. Large values of K_{pr} indicate tool breakage,

because of the large difference between the adjacent periods. An example of this feature is given in Figure 2.16.

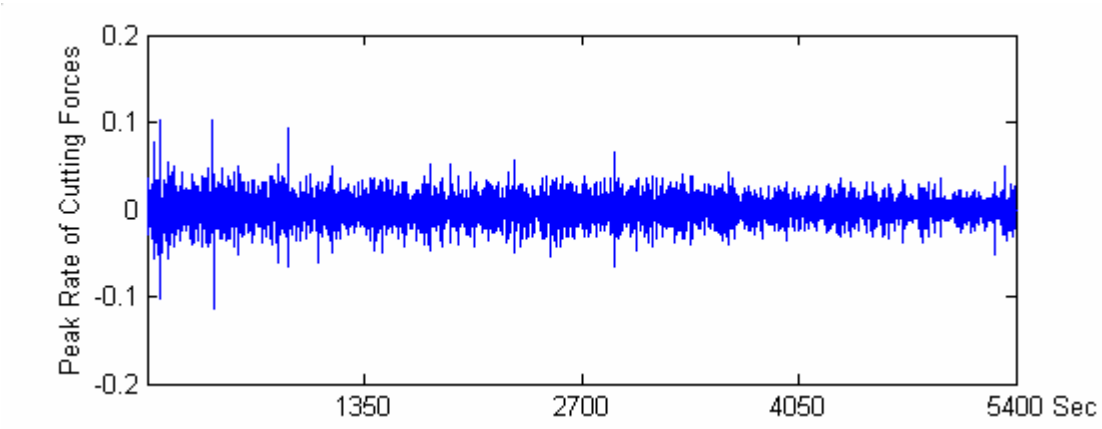


Figure 2.16 Peak Rate of Cutting Force

(xi) Total Harmonic Power of Cutting Force

Elbestawi [1989] found that the harmonic contents of cutting forces are sensitive to tool flank wear. This is because when the tool gets worn, there is an obvious increasing trend in the magnitudes of the fundamental tooth frequency and its harmonics. This phenomenon can be seen in Figure 2.11 and 2.12. Thus, the total harmonic power of the force spectrum can be used as an indicator of tool failures:

$$P_{TH}(i) = \sum_{m=1}^{N} G(m) \tag{2.27}$$

$G(m)$ is the power at the fundamental tooth frequency and its harmonics. N is the desired order which defines the frequency range of interest. The features are then mapped to tool status through a linear discriminant function classifier. An example of this feature is given in Figure 2.17.

(xii) Average Force

The average force is widely used in both tool failure detection [Elbestawi, 1989; Tarng, 1994; Leem, 1995; Kim, 1995] and wear estimation [Lin, 1996; Santanu, 1996],

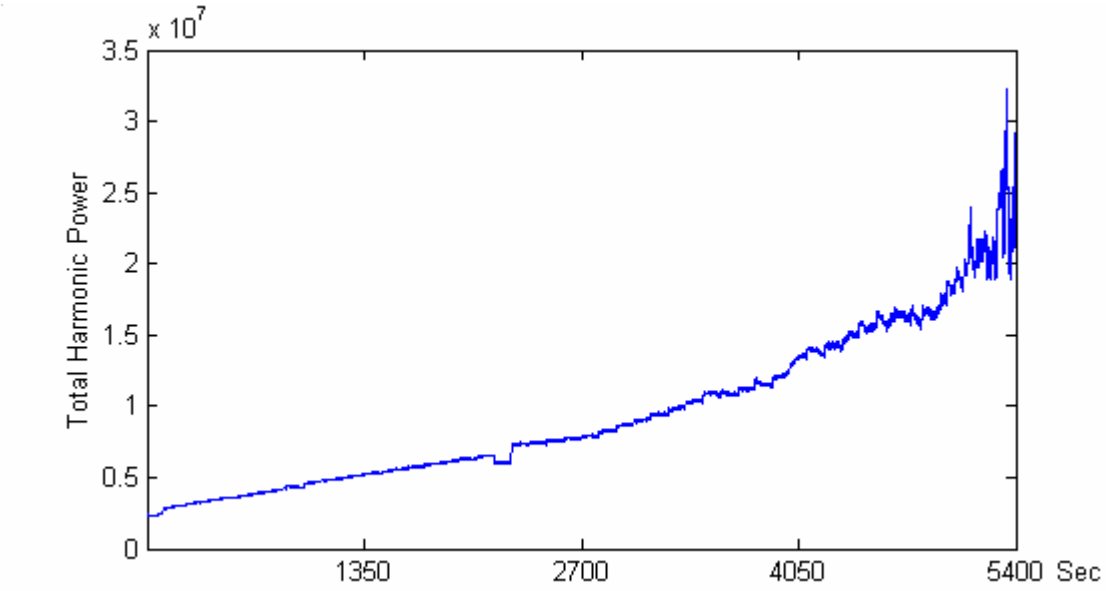


Figure 2.17 Total Harmonic Power

because its trend correlates very well to the growth of flank wear, which can be clearly seen from the figures in Appendix A. The average force (F_a) within a spindle rotation can be calculated by:

$$F_a(i) = \frac{1}{N}\sum_{j=1}^{N} f(i, j) \tag{2.28}$$

where *f(i,j)* denotes the *j-th* force sample in the *i-th* tool rotation, *N* is total number of the force samples in a rotation. The average force is illustrated in Figure 2.18.

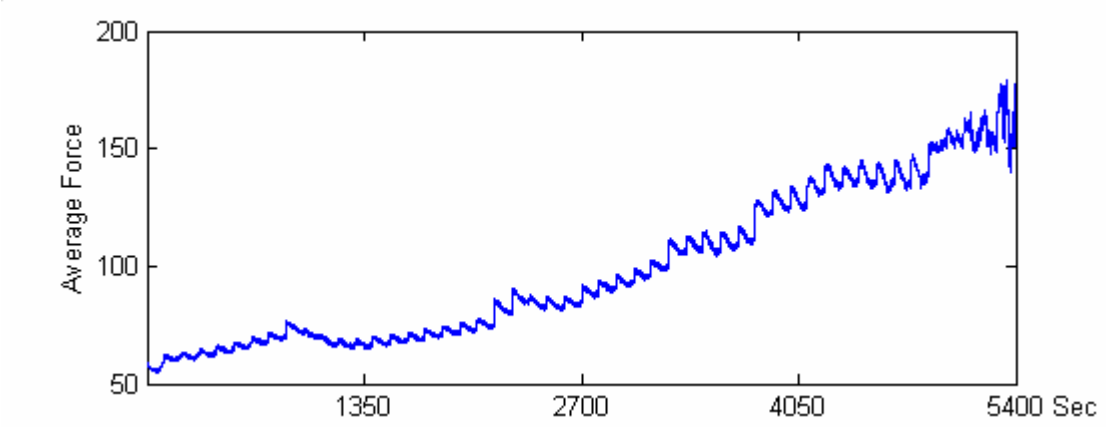


Figure 2.18 Average Force

(xiii) Variable Force

Tarng [1994] defined a variable force to evaluate the variation of cutting force due to tool failures. First, the average cutting force data are passed through a nonlinear median filter to obtain the deterministic component:

$$F_{med}(i) = median\left(F_a(i+j), j = -k, \cdots, k; k = {(m+1)}/{2} \right) \tag{2.29}$$

where F_{med} is called median cutting force, and m is the order of the median filter. Then the variable cutting force can be obtained by subtracting the median cutting force from the average force:

$$\Delta F_a(i) = F_a(i) - F_{med}(i) \tag{2.30}$$

This process can be illustrated in Figure 2.19. And an example of the variable force is shown in Figure 2.20.

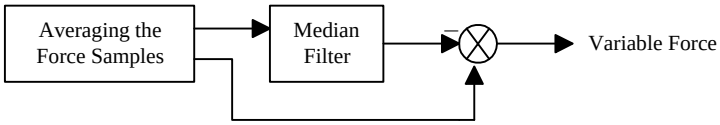


Figure 2.19 Calculation of Variable Force

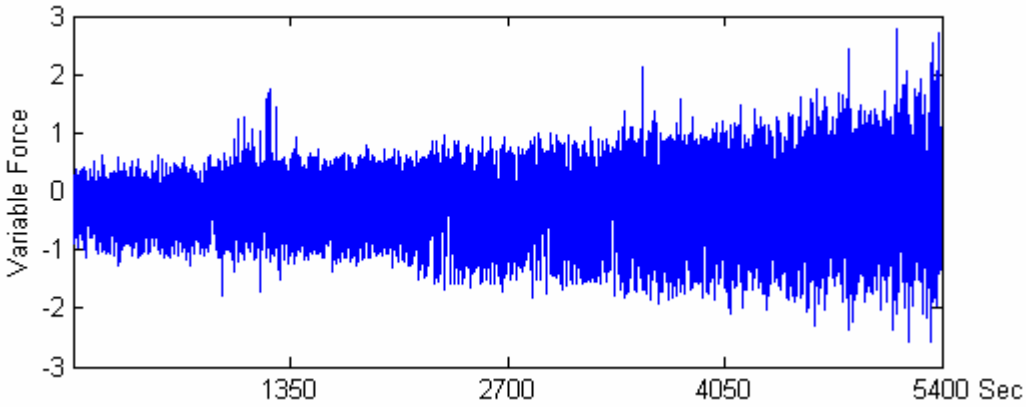


Figure 2.20 Variable Force

(xiv~xvi) Standard Deviation, Skewness, and Kurtosis

Leem [1995] extracted four statistics from the cutting force for monitoring tool wear. These include mean, standard deviation, skewness, and kurtosis. The mean can be derived by Equation (2.28). And the other three features can be calculated by the following three equations, respectively:

$$\sigma(i)=\sqrt{\frac{1}{N-1}\cdot\sum_{j=1}^{N}\left[f(i,j)-F_a(i)\right]} \tag{2.31}$$

$$s(i)=\frac{N}{(N-1)(N-2)}\sum_{j=1}^{N}\left(\frac{f(i,j)-F_a(i)}{\sigma(i)}\right)^3 \tag{2.32}$$

$$k(i)=\frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum_{j=1}^{N}\left(\frac{f(i,j)-F_a(i)}{\sigma(i)}\right)^4-\frac{3(N-1)^2}{(N-2)(N-3)} \tag{2.33}$$

where *f(i,j)* is the *j-th* force sample in the *i-th* rotation, and *N* is total number of the force samples in a rotation. The examples of these three features are shown in Figures 2.21 to 2.23.

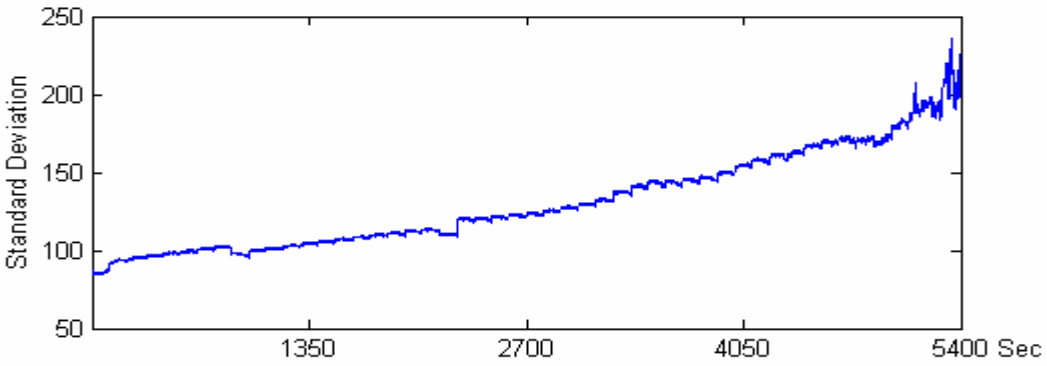


Figure 2.21 Standard Deviation

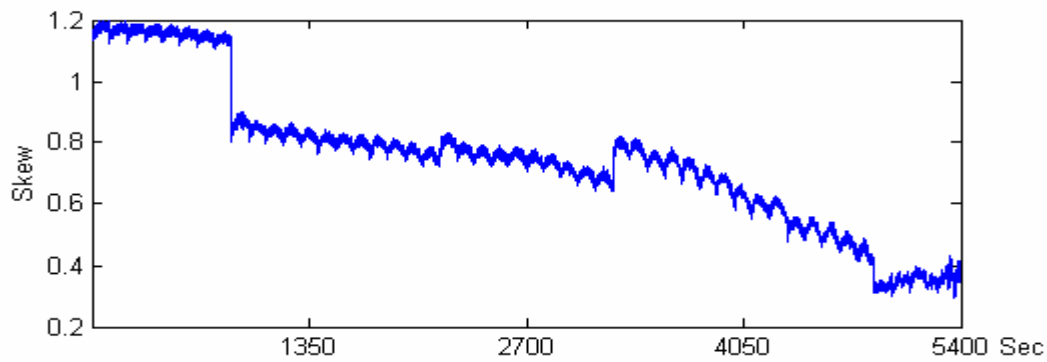


Figure 2.22 Skewness

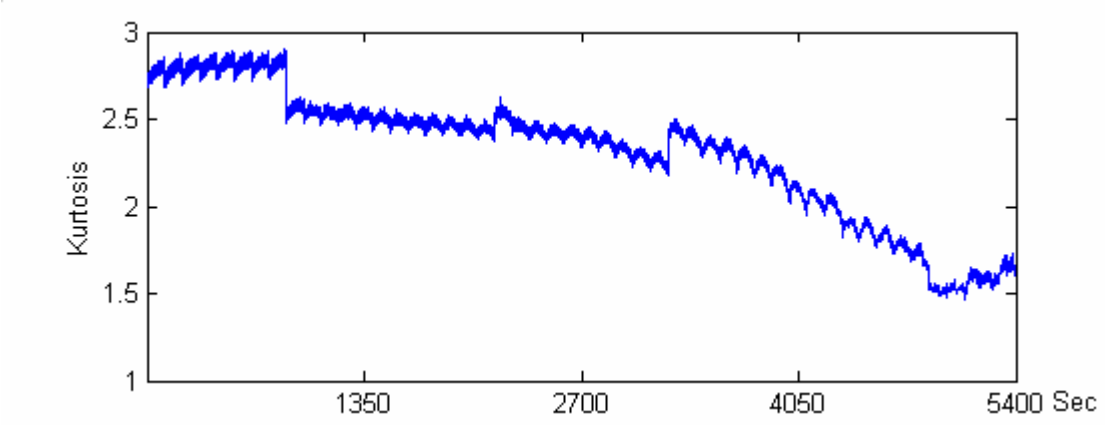


Figure 2.23 Kurtosis

2.3 SUMMARY OF THE FEATURE EXTRACTION METHODS

16 different feature extraction methodologies are introduced and discussed in this chapter, which have all been shown to be indicative of tool condition monitoring. These methodologies are summarized in Table 2.1. The relationships among them are illustrated in Figure 2.24. These 16 features form the scope of the feature selection in this study. However, they cannot be effectively used due to the large noise, as shown in the feature graphs. Further processing must be made to make them suitable for subsequent applications, which is described in Chapter 4.

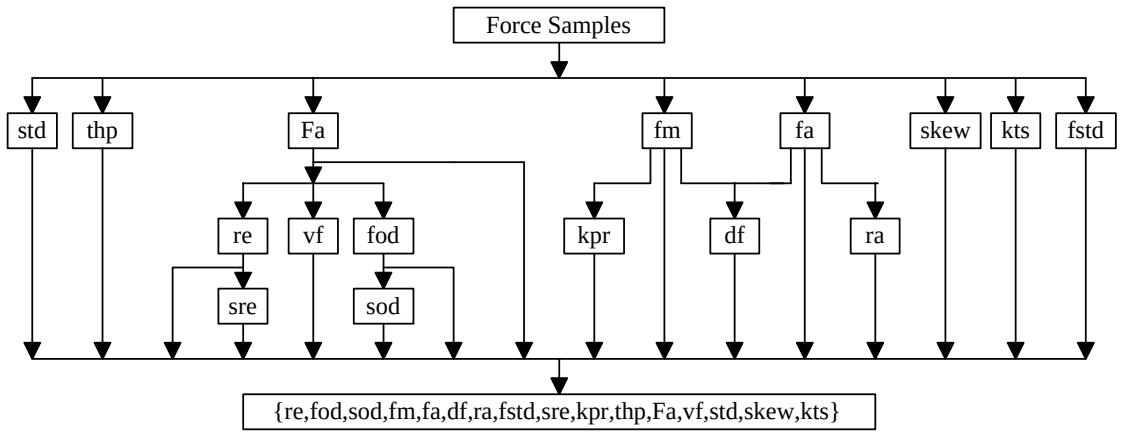


Figure 2.24 Relationships among the Features

Table 2.1. Feature Extraction Methodologies

No	Feature	Notation	Objective*	Decision Making Strategy	Reference
1	Residual Error	re	TBD	Thresholding	Altintas, 1988
2	First Order Differencing	fod	TWD	Thresholding	Altintas, 1989
3	Second Order Differencing	sod	TWD	Thresholding	Altintas, 1989
4	Maximum Force Level	fm	TWD	Thresholding	Tarn, 1989
5	Total Amplitude of Cutting Force	fa	TWD	Thresholding	Tarn, 1989
6	Combined Incremental Force Changes	df	TWD	Thresholding	Tarn, 1989
7	Amplitude Ratio	ra	TWD	Thresholding	Tarn, 1989
8	Standard Deviation of the Force Components in Tool Breakage Zone	fstd	TBD	Thresholding	Tarng, 1990
9	Sum of the Squares of Residual Errors	sre	TBD	Thresholding	Tansel, 1993
10	Peak Rate of Cutting Forces	kpr	TBD	Thresholding	Zhang, 1995
11	Total Harmonic Power	thp	TWD	LDF-Classifier	Elbestawi, 1989
12	Average Force	Fa	TBD, TWD, TWE	MLP	Tarng, 1994; Leem, 1995; Lin, 1996; Santanu, 1996
13	Variable Force	vf	TBD	MLP	Tarng, 1994
14	Standard Deviation	std	TWD	SOM	Leem, 1995
15	Skew	skew	TWD	SOM	Leem, 1995
16	Kurtosis	kts	TWD	SOM	Leem, 1995

* TBD: Tool Breakage Detection; TWD: Tool Wear Detection; TWE: Tool Wear Estimation.

CHAPTER 3

BAYESIAN SUPPORT VECTOR MACHINES AND AUTOMATIC RELEVANCE DETERMINATION

16 different features based on force signals are introduced in Chapter 2. As mentioned in Chapter 1, the main objective of this study is to compare these features and select a subset of the features with higher relevance. For this purpose, the automatic relevance determination (ARD) algorithm, originated by MacKay [1992] and Neal [1996], is used. The feature selection procedures are implemented for both tool wear recognition (TWR) using Bayesian support vector classification (BSVC) algorithm and tool wear estimation (TWE) using Bayesian support vector regression (BSVR) algorithm. In this chapter, a literature review of the algorithms is given first, followed by a theoretical background of the ARD, BSVC, and BSVR.

3.1 INTRODUCTION

Support vector machines (SVM), as described by Vapnik [1995], exploit the idea of mapping input data into a high dimensional (often infinite) reproducing kernel Hilbert space. The SVM methods have many advantages, including a global minimum solution as the minimization of a convex programming problem, relatively fast training speed, and sparseness in solution representation (i.e. only a proportion of training

points are relevant). However, as pointed out by Tipping [1999], the traditional SVM methodology also exhibits significant disadvantages. For example, it cannot produce probabilistic predictions. The application of Bayesian approaches to neural networks, originated by Buntine and Weigend [1991], MacKay [1992] and Neal [1996], can solve this problem effectively. Bayesian probability theory provides a unifying framework for data modeling which offers several benefits, such as optimizing the model parameters and handling uncertainty in a natural manner [Mackay, 1992].

Bayesian interpretations of support vector machines are based on MacKay's evidence framework [MacKay, 1992]. These include support vector classification (SVC) [Seeger, 1999; Kwok, 2000; Chu, 2003] and support vector regression (SVR) [Law, 2001; Chu, 2001]. In this study, the Bayesian SVC algorithm and the Bayesian SVR algorithm proposed by Chu [2001, 2003] are used, because of the good generalization capabilities.

Based on the Bayesian approaches, MacKay and Neal proposed a new method, called automatic relevance determination (ARD). The aim of ARD is to automatically determine which of many inputs to a neural network are relevant to prediction of the targets. This is done by making the weights on the connections out of each input unit have a distribution that is controlled by a hyperparameter associated with that input, allowing the relevance of each input to be determined automatically as the values of these hyperparameters adapt to the data [Neal, 1996]. The ARD method can be directly implemented in Bayesian support vector machines.

3.2 BAYESIAN SUPPORT VECTOR REGRESSION

3.2.1 BAYESIAN LEARNING

Bayesian learning theory provides a unifying framework for data modeling. The result of it is a probability distribution over model parameters that expresses the beliefs regarding how likely the different parameter values are. It can be realized through evaluating the posterior probabilities of the model parameters when training data are given [Neal, 1996]:

$$P\left(\theta \middle| x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right) \propto \underbrace{L\left(\theta \middle| x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right)}_{likelihood} \underbrace{P(\theta)}_{prior} \tag{3.1}$$

where θ is a parameter vector, and $x^{(1)}, x^{(2)}, \cdots, x^{(n)}$ are the training data. The posterior distribution combines the likelihood function, which contains the information about θ derived from observation, with the prior, which contains the information about θ derived from our background knowledge. The application of Bayesian learning in neural networks aims to infer the network parameters, which effectively solves the overfitting problem by controlling model complexity [Mackay, 1995].

3.2.2 BAYESIAN SUPPORT VECTOR REGRESSION

In regression problems, a set of training data $\mathbf{D} = \left\{ (\mathbf{x}_i, y_i) \middle| i = 1, \ldots, n, \mathbf{x}_i \in \mathbf{R}^d, y_i \in \mathbf{R} \right\}$ is collected by randomly sampling a function *f*, defined on $\boldsymbol{R}^d$. As the measurements are usually corrupted by noise, training samples can be represented as

$$y_i = f(\mathbf{x}_i) + \delta_i \qquad i = 1,2,\ldots,n \tag{3.2}$$

δ_i are independent, identically distributed random variables, whose probability distribution can be assumed to be:

$$p(\delta_i) = \frac{1}{Z_s} \exp\left(-C \cdot l(\delta_i)\right) \tag{3.3}$$

where $Z_s = \int \exp\left(-C \cdot l(\delta_i)\right) d\delta_i$, C is a parameter greater than zero, and $l(\delta_i)$ is the loss function, which is in Huber's form:

$$l(\delta)=\begin{cases}-\delta-\varepsilon, & if \quad \delta\in\Delta_{C^*}=(-\infty,-2\varepsilon)\\ \dfrac{\delta^2}{4\varepsilon}, & if \quad \delta\in\Delta_M=[-2\varepsilon,2\varepsilon]\\ \delta-\varepsilon, & if \quad \delta\in\Delta_C=(2\varepsilon,+\infty)\end{cases} \tag{3.4}$$

where ε>0. Huber's loss function is non-quadratic with low sensitivity to the outliers and differentiable allowing appropriate approximations to be used in the Bayesian approach. This function is illustrated in Figure 3.1. Therefore, Z_s can be written as:

$$Z_s=\int \exp(-C\cdot l(\delta_i))d\delta_i=\frac{2}{C}\exp(-C\varepsilon)+2\sqrt{\frac{\pi\varepsilon}{C}}\cdot erf(\sqrt{C\varepsilon}) \tag{3.5}$$

where $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt$.

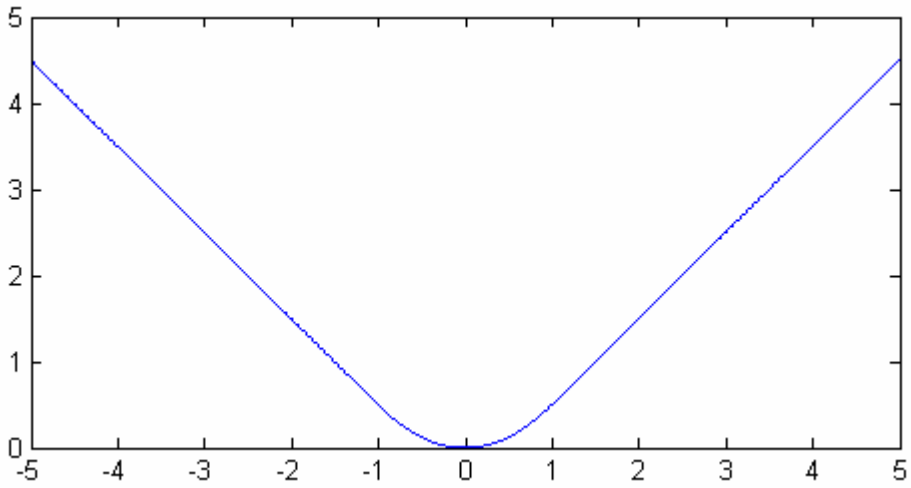


Figure 3.1 Huber's loss function (ε=0.5)

The regression aims to infer the function f in (3.2), or an estimate of it, from the finite data set ***D***. In the Bayesian approach, we regard the function f as the realization of a random field with a known prior probability. The posterior probability of f given the training data ***D*** can then be derived by Bayes' theorem:

$$P(\mathbf{f}|\mathbf{D})=\frac{P(\mathbf{D}|\mathbf{f})P(\mathbf{f})}{P(\mathbf{D})} \tag{3.6}$$

where $\mathbf{f} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n)]^T$. Since the training data are usually sampled arbitrarily, $P(\mathbf{D})$ is not meaningful [MacKay, 1995]. So the posterior is only determined by the prior $P(\mathbf{f})$ and the likelihood $P(\mathbf{D}|\mathbf{f})$.

The prior probability $P(\mathbf{f})$ can be modeled as a multivariate Gaussian with zero mean and a $n \times n$ covariance matrix $\mathbf{\Sigma}$, given by (3.7):

$$P(\mathbf{f}) = \frac{1}{Z_f} \exp\left(-\frac{1}{2}\mathbf{f}^T \mathbf{\Sigma}^{-1} \mathbf{f}\right) \tag{3.7}$$

where $Z_f = (2\pi)^{n/2} \sqrt{|\mathbf{\Sigma}|}$. The *ij*-th component of $\mathbf{\Sigma}$ is:

$$Cov\left[f(\mathbf{x}_i), f(\mathbf{x}_j)\right] = Cov(\mathbf{x}_i, \mathbf{x}_j) = k_0 \exp\left(-\frac{1}{2}\sum_{l=1}^{d} k_l (\mathbf{x}_i^l - \mathbf{x}_j^l)^2\right) + k_b \tag{3.8}$$

where k_0>0 denotes the average power of $f(\boldsymbol{x})$; k_l >0, $l = 1,2,\cdots,d$ is the ARD parameter that determines the relevance of the *l-th* input dimension to the prediction of the output variables; k_b>0 denotes the variance of the offset to the function $f(\boldsymbol{x})$; and $\boldsymbol{x}^l$ denotes the *l-th* element of the input vector $\boldsymbol{x}$.

The likelihood $P(\mathbf{D}|\mathbf{f})$ is a model of noise, which can be evaluated by

$$P(\mathbf{D}|\mathbf{f}) = \prod_{i=1}^{n} P(y_i - f(\mathbf{x}_i)) = \prod_{i=1}^{n} P(\delta_i) \tag{3.9}$$

Introducing Equation (3.3) into (3.9), the likelihood function can be expressed as:

$$P(\mathbf{D}|\mathbf{f}) \propto \exp\left(C \cdot \sum_{i=1}^{n} l(y_i - f(\mathbf{x}_i))\right) \tag{3.10}$$

Based on Bayes' theorem (3.6), prior probability (3.7), and the likelihood (3.10), the posterior probability of $\mathbf{f}$ can be written as:

$$P(\mathbf{f}|\mathbf{D}) = \frac{1}{Z}\exp(-S(\mathbf{f})) \tag{3.11}$$

where $S(\mathbf{f}) = C\sum_{i=1}^{n} l(y_i - f(\mathbf{x}_i)) + \frac{1}{2}\mathbf{f}^T\boldsymbol{\Sigma}^{-1}\mathbf{f}$ and $Z = \int \exp(-S(\mathbf{f}))d\mathbf{f}$. The maximum a posteriori (MAP) estimation of the function values is therefore the minimization of the following optimization problem:

$$\min_{\mathbf{f}} S(\mathbf{f}) = C\sum_{i=1}^{n} l(y_i - f(\mathbf{x}_i)) + \frac{1}{2}\mathbf{f}^T\boldsymbol{\Sigma}^{-1}\mathbf{f} \tag{3.12}$$

Define $\omega_i = -C\frac{\partial l(y_i - f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)}\Big|_{f(\mathbf{x}_i)=f(\mathbf{x}_{MP})} \forall i$ and let $\boldsymbol{\omega}$ be the column vector formed by ω_i. Then $\mathbf{f}_{MP}$ can be written as:

$$\mathbf{f}_{MP} = \boldsymbol{\Sigma}\cdot\boldsymbol{\omega} \tag{3.13}$$

It can also be decomposed to:

$$f_{MP} = k_0\sum_{i=1}^{n}\omega_i K(\mathbf{x},\mathbf{x}_i) + b(\mathbf{x}) = \sum_{i=1}^{n}\omega_i k_0 K(\mathbf{x},\mathbf{x}_i) + k_b\sum_{i=1}^{n}\omega_i \tag{3.14}$$

where $b = k_b\sum_{i=1}^{n}\omega_i$ and $K(\mathbf{x},\mathbf{x}_i) = \exp\left(-\frac{1}{2}\sum_{l=1}^{d}k_l\left(\mathbf{x}^l - \mathbf{x}_i^l\right)^2\right)$ is just the Gaussian kernel in classical support vector regression [Haykin, 1999].

The dual problem of Equation 3.12 can be finally developed into:

$$\min_{\alpha,\boldsymbol{\alpha}^*} S(\boldsymbol{\alpha},\boldsymbol{\alpha}^*) = \frac{1}{2}\sum_{i}^{n}\sum_{j}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)Cov(x_i,x_j) - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i + \frac{\varepsilon}{C}\sum_{i=1}^{n}\left(\alpha_i^2 + \alpha_i^{*2}\right) \tag{3.15}$$

subject to $0 \le \alpha_i \le C$, and $0 \le \alpha_i^* \le C$.

The optimal value of the primal variables $\mathbf{f}$ can be written as:

$$\mathbf{f}_{MP} = \boldsymbol{\Sigma} \cdot \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^* \right) \tag{3.16}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_n]^T$ and $\boldsymbol{\alpha}^* = [\alpha_1^*, \alpha_2^*, \cdots, \alpha_n^*]^T$. At the optimal solution, the training samples $(\mathbf{x}_i, y_i)$ with associated $\alpha_i - \alpha_i^*$ satisfying $0 < |\alpha_i - \alpha_i^*| < C$ are called off-bound support vectors; the samples with $|\alpha_i - \alpha_i^*| = C$ are on-bound support vectors; and the samples with $|\alpha_i - \alpha_i^*| = 0$ are non-support vectors.

3.2.3 MODEL ADAPTATION AND ARD

Let $\boldsymbol{\theta}$ be the hyperparameter vector containing the parameters in the prior distribution and the likelihood function, i.e. $\boldsymbol{\theta} = \{\varepsilon, C, k_0, k_b, k_1, k_2, \ldots, k_d\}$. The optimal values of hyperparameters $\boldsymbol{\theta}$ can be inferred by maximizing the posterior probability $P(\boldsymbol{\theta}|\mathbf{D}) = \dfrac{P(\mathbf{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{D})}$. As we typically have little idea of suitable values of $\boldsymbol{\theta}$ before training data are available, we assume a flat distribution for P($\boldsymbol{\theta}$), i.e., P($\boldsymbol{\theta}$) is greatly insensitive to the values of $\boldsymbol{\theta}$. Therefore, the evidence P($\mathbf{D}$|$\boldsymbol{\theta}$) can be used to assign a preference to alternative values of the hyperparameters $\boldsymbol{\theta}$ [MacKay, 1992]:

$$P(\mathbf{D}|\boldsymbol{\theta}) = \int p(\mathbf{D}|\mathbf{f}, \boldsymbol{\theta}) p(\mathbf{f}|\boldsymbol{\theta}) d\mathbf{f} = Z_f^{-1} Z_s^{-n} \int \exp(-S(\mathbf{f})) d\mathbf{f} \tag{3.17}$$

where Z_s and Z_f are given in (3.5) and (3.7), respectively.

An explicit expression of the evidence P($\mathbf{D}$|$\boldsymbol{\theta}$) can be obtained from an integral over the f-space with a Taylor expansion at $\mathbf{f}_{MP}$ (where $\dfrac{\partial S(\mathbf{f})}{\partial \mathbf{f}}\Big|_{\mathbf{f}=\mathbf{f}_{MP}} = 0$) and retaining terms up to the second order:

$$\begin{aligned} S(\mathbf{f}) &\approx S(\mathbf{f}_{MP}) + \frac{1}{2}(\mathbf{f} - \mathbf{f}_{MP})^T \frac{\partial^2 S(\mathbf{f})}{\partial \mathbf{f} \partial \mathbf{f}^T}|_{\mathbf{f}=\mathbf{f}_{MP}} (\mathbf{f} - \mathbf{f}_{MP}) \\ &= S(\mathbf{f}_{MP}) + \frac{1}{2} \cdot (\mathbf{f} - \mathbf{f}_{MP})^T \cdot (\boldsymbol{\Sigma}^{-1} + C \cdot \boldsymbol{\Lambda}) \cdot (\mathbf{f} - \mathbf{f}_{MP}) \end{aligned} \tag{3.18}$$

$$P(\mathbf{D}|\boldsymbol{\theta}) \approx \exp(-S(\mathbf{f}_{MP}))\cdot|\mathbf{I}+C\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{\Lambda}|^{-\frac{1}{2}}\cdot Z_s^{-n} \tag{3.19}$$

where $\mathbf{I}$ is an $n\times n$ identity matrix; $\boldsymbol{\Lambda}$ is a diagonal matrix with the *ii-th* entry being $\frac{1}{2\varepsilon}$ if the corresponding training sample is an off-bound support vector, otherwise the entry is zero. Therefore, only a sub-matrix of $\boldsymbol{\Sigma}$ plays a role in the determinant $|\boldsymbol{I}+C\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{\Lambda}|$, due to the sparseness of $\boldsymbol{\Lambda}$. Let $\boldsymbol{\Sigma}_M$ be the $m\times m$ sub-matrix of $\boldsymbol{\Sigma}$ obtained by deleting all the rows and columns associated with the on-bound support vectors and non-support vectors, i.e. keeping the m off-bound support vectors only. Then the negative log probability of data given hyperparameters is:

$$-\ln P(\mathbf{D}|\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^T\boldsymbol{\Sigma}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)+C\sum_{i=1}^{n}l(y_i-f_{MP}(x_i))+\frac{1}{2}\ln\left(\left|\mathbf{I}+\frac{C}{2\varepsilon}\boldsymbol{\Sigma}_M\right|\right)+n\ln Z_s \tag{3.20}$$

where $\mathbf{I}$ is an $m\times m$ identity matrix.

Gradient based optimization can then be used for minimizing (3.20). We usually use $\{\ln\varepsilon,\ln C,\ln k_0,\ln k_b,\ln k_1,\ln k_2,\ldots,\ln k_d\}$ as the set of variables to tune, to eliminate the constraints (>0). And the gradients are given by:

$$\frac{\partial -\ln P(\boldsymbol{D}|\boldsymbol{\theta})}{\partial \ln C} = C\sum_{i=1}^{n}l(y_i-f_{MP}(x_i))+\frac{1}{2}trace\left[\left(\frac{2\varepsilon}{C}\boldsymbol{I}+\boldsymbol{\Sigma}_M\right)^{-1}\boldsymbol{\Sigma}_M\right]$$

$$-\frac{n}{Z}\left(\sqrt{\frac{\varepsilon\pi}{C}}\cdot erf\left(\sqrt{C\varepsilon}\right)+\frac{2}{C}\exp(-C\varepsilon)\right) \tag{3.21}$$

$$\frac{\partial -\ln P(\boldsymbol{D}|\boldsymbol{\theta})}{\partial \ln \varepsilon} = -C\left[\sum_{\delta_i\in\Delta_M}\frac{\delta_i^2}{4\varepsilon}+\sum_{\delta_i\in\Delta_C\cup\Delta_{C^*}}\varepsilon\right]-\frac{1}{2}trace\left[\left(\frac{2\varepsilon}{C}\boldsymbol{I}+\boldsymbol{\Sigma}_M\right)^{-1}\boldsymbol{\Sigma}_M\right]$$

$$+\frac{n}{Z}\sqrt{\frac{\varepsilon\pi}{C}}\cdot erf\left(\sqrt{C\varepsilon}\right) \tag{3.22}$$

$$\frac{\partial -\ln P(\boldsymbol{D}|\boldsymbol{\theta})}{\partial \ln k_j} = \frac{k_j}{2} trace\left[\left(\frac{2\varepsilon}{C}\boldsymbol{I} + \boldsymbol{\Sigma}_M\right)^{-1} \frac{\partial \boldsymbol{\Sigma}_M}{\partial k_j}\right] + \frac{k_j}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T \frac{\partial \boldsymbol{\Sigma}}{\partial k_j}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)$$

(3.23)

where $k_j \in \{k_0, k_b, k_1, k_2, \cdots, k_d\}$, $\delta_i = y_i - f_{MP}(\mathbf{x}_i)$, and $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ are the optimal solution of (3.15).

Note that $k_l > 0$, $l = 1,2,\cdots,d$ controls the contribution of the *l-th* input dimension to the prediction of the output variables. The larger the value of k_l, the more relevant the *l-th* input dimension to the prediction. The procedures for inferring these parameters are called automatic relevance determination.

Based on the Bayesian support vector regression and the automatic relevance determination algorithms, feature selection and training the regression network can be conducted through the following steps:

1) Assume an initial hyperparameter set $\boldsymbol{\theta}$.
2) Use the maximum a posteriori methods to get $\mathbf{f}_{MP}$.
3) Use the gradient-based optimization methods to infer the optimal values of the hyper-parameters.
4) If the sum square error given by $(\mathbf{Y} - \mathbf{f}_{MP})^T \cdot (\mathbf{Y} - \mathbf{f}_{MP})$ is smaller than the predetermined threshold, then end the iteration; else return to the Step (2).
5) Select the set of k_l greater than a threshold to be the relevant feature set.

The procedures for implementing the BSVR and ARD algorithms are illustrated in Figure 3.2.

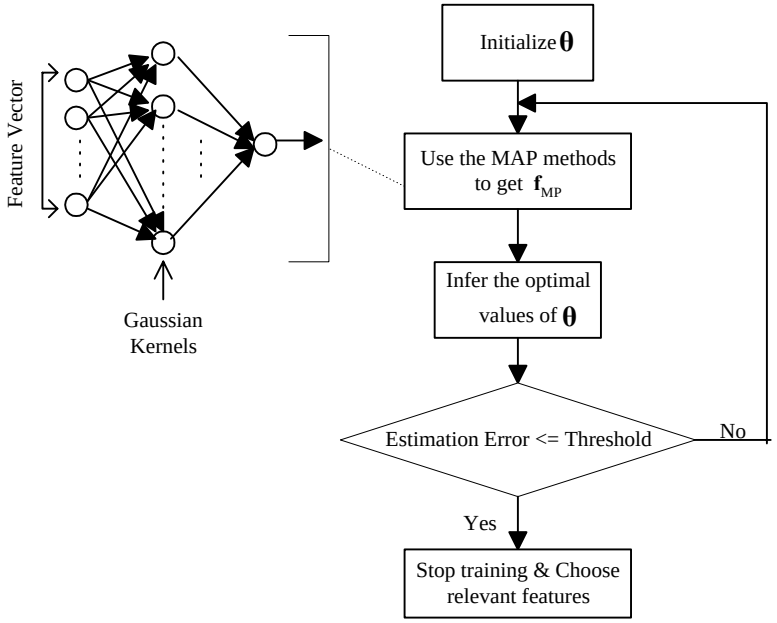


Figure 3.2 Procedures for Implementing BSVR and ARD

3.3 BAYESIAN SUPPORT VECTOR CLASSIFICATION

3.3.1 BAYESIAN SUPPORT VECTOR CLASSIFICATION

Unlike the curve fitting problem of regression, classification is defined as the process whereby a received pattern/signal is assigned to one of a prescribed number of classes (categories). A pattern is represented by a set of *m* observables, which may be viewed as a point **x** in an *m*-dimensional observation space. Feature extraction is usually used to maximize the information contained in **x**, which is defined as a transformation that maps the point **x** into an intermediated point **y** in a *q*-dimensional feature space. The classification is itself described as a transformation that maps the point **y** into one of the classes in an *r*-dimensional decision space, where *r* is the number of classes to be distinguished [Haykin, 1999]. This concept can be illustrated in Figure 3.3.

The Bayesian approach for classification can also be developed through the framework described in 3.2. The difference lies in the form of the loss function and the

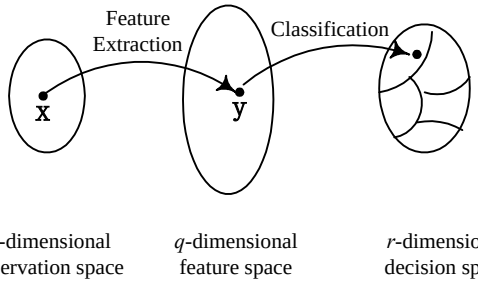


Figure 3.3 Concept of Classification

likelihood function resulted from it. For binary classification, the trigonometric loss function, proposed by Chu [2001], is a good solution because it satisfies the following characteristics:

1. naturally normalized in likelihood evaluation;
2. possessing a flat zero region that results in sparseness property;
3. smooth and continuous up to the first order derivative.

The trigonometric loss function takes the following form:

$$l_t(y_{\mathbf{x}} \cdot f_{\mathbf{x}}) = \begin{cases} +\infty & \text{if} \quad y_{\mathbf{x}} \cdot f_{\mathbf{x}} \in (-\infty, -1] \\ 2\ln\sec\left(\frac{\pi}{4}(1 - y_{\mathbf{x}} \cdot f_{\mathbf{x}})\right) & \text{if} \quad y_{\mathbf{x}} \cdot f_{\mathbf{x}} \in (-1, +1) \\ 0 & \text{if} \quad y_{\mathbf{x}} \cdot f_{\mathbf{x}} \in [+1, +\infty) \end{cases} \tag{3.24}$$

and its first order derivative is therefore:

$$\frac{\partial l_t(y_{\mathbf{x}} \cdot f_{\mathbf{x}})}{\partial f_{\mathbf{x}}} = \begin{cases} -y_{\mathbf{x}} \frac{\pi}{2} \tan\left(\frac{\pi}{4}(1 - y_{\mathbf{x}} \cdot f_{\mathbf{x}})\right) & \text{if} \quad y_{\mathbf{x}} \cdot f_{\mathbf{x}} \in (-1, +1) \\ 0 & \text{if} \quad y_{\mathbf{x}} \cdot f_{\mathbf{x}} \in [+1, +\infty) \end{cases} \tag{3.25}$$

where $\mathbf{x} \in \mathbf{R}^{\mathrm{d}}$ is the input vector, $y_{\mathbf{x}} \in \{+1, -1\}$ is the class label, and $f_{\mathbf{x}}$ denotes the latent function at **x**. The loss function is illustrated in Figure 3.4. The trigonometric likelihood function can then be written as:

$$P_t(y_{\mathbf{x}} | f_{\mathbf{x}}) = \begin{cases} 0 & \text{if} \quad y_{\mathbf{x}} \cdot f_{\mathbf{x}} \in (-\infty, -1] \\ \cos^2\left(\frac{\pi}{4}(1 - y_{\mathbf{x}} \cdot f_{\mathbf{x}})\right) & \text{if} \quad y_{\mathbf{x}} \cdot f_{\mathbf{x}} \in (-1, +1) \\ 1 & \text{if} \quad y_{\mathbf{x}} \cdot f_{\mathbf{x}} \in [+1, +\infty) \end{cases} \tag{3.26}$$

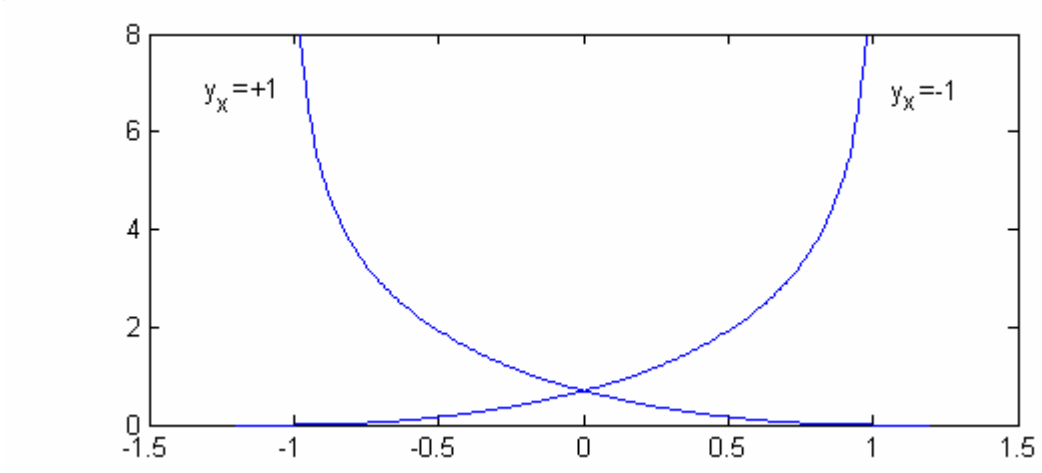


Figure 3.4 Trigonometric Loss Function

In the Bayesian approach, we can infer the latent function $f_{\mathbf{x}}$ by maximizing the posterior, which is defined by Bayes' theorem in Equation (3.6). The prior probability $P(\mathbf{f})$ is the same as defined in Equation (3.7) and (3.8). And with the likelihood function in (3.26), the likelihood $P(\mathbf{D}|\mathbf{f})$ can be written as:

$$P(\mathbf{D}|\mathbf{f}) = \prod_{i=1}^{n} P_t\left(y_{\mathbf{x}_i} \middle| f_{\mathbf{x}_i}\right) \tag{3.27}$$

Introducing (3.7) and (3.27) into (3.6), the posterior probability of f is:

$$P(\mathbf{f}|\mathbf{D}) = \frac{1}{Z}\exp(-S(\mathbf{f})) \tag{3.28}$$

where $S(\mathbf{f}) = \frac{1}{2}\mathbf{f}^T\boldsymbol{\Sigma}^{-1}\mathbf{f} + \sum_{i=1}^{n} l_t\left(y_{\mathbf{x}_i} \cdot f(\mathbf{x}_i)\right)$ and $Z = \int \exp(-S(\mathbf{f}))d\mathbf{f}$. The maximum a posteriori (MAP) estimation of the function values is therefore the minimization of the following optimization problem:

$$\min_{\mathbf{f}} S(\mathbf{f}) = \sum_{i=1}^{n} l_t\left(y_{\mathbf{x}_i} \cdot f(\mathbf{x}_i)\right) + \frac{1}{2}\mathbf{f}^T\boldsymbol{\Sigma}^{-1}\mathbf{f} \tag{3.29}$$

Define $\omega_i = -\left.\frac{\partial l_t\left(y_{\mathbf{x}_i} \cdot f(\mathbf{x}_i)\right)}{\partial f(\mathbf{x}_i)}\right|_{f(\mathbf{x}_i)=f(\mathbf{x}_{MP})} \forall i$ and let $\boldsymbol{\omega}$ be the column vector formed by ω_i. Then $\mathbf{f}_{MP}$ can be written as:

$$\mathbf{f}_{MP} = \boldsymbol{\Sigma} \cdot \boldsymbol{\omega} \tag{3.30}$$

It can also be decomposed into the form:

$$f_{MP} = \sum_{i=1}^{n} \omega_i k_0 K(\mathbf{x}, \mathbf{x}_i) + k_b \sum_{i=1}^{n} \omega_i \tag{3.31}$$

where $K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{2} \sum_{l=1}^{d} k_l \left(\mathbf{x}^l - \mathbf{x}_i^l \right)^2 \right)$ is just the Gaussian kernel in classical support vector classification [Haykin, 1999].

The dual problem of Equation 3.29 can be finally developed into:

$$\min_{\alpha} S(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{i}^{n} \sum_{j}^{n} \left(y_{\mathbf{x}_i} \alpha_i \right) \left(y_{\mathbf{x}_j} \alpha_j \right) Cov(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \left[\frac{4}{\pi} \alpha_i \arctan\left(\frac{2\alpha_i}{\pi} \right) - \ln\left(1 + \left(\frac{2\alpha_i}{\pi} \right)^2 \right) \right] \tag{3.32}$$

subject to $\alpha_i \geq 0, \forall i$. At the optimal solution, the training samples $\left(\mathbf{x}_i, y_{\mathbf{x}_i} \right)$ associated with non-zero Lagrange multiplier α_i are called support vectors (SVs); the samples with zero α_i are not support vectors, which do not involve in the solution representation.

3.3.2 MODEL ADAPTATION AND ARD

Let $\boldsymbol{\theta}$ be the hyperparameter vector containing the parameters in the prior distribution and the likelihood function, i.e. $\boldsymbol{\theta} = \{k_0, k_1, k_2, \cdots, k_d, k_b\}$. The optimal values of hyperparameters $\boldsymbol{\theta}$ can be inferred through the same approach described in 3.2.3. The evidence in this case can be written as:

$$p(\mathbf{D}|\boldsymbol{\theta}) = \exp\left[-S(\mathbf{f}_{MP}) \right] \cdot \left| \mathbf{I} + \boldsymbol{\Sigma} \cdot \boldsymbol{\Lambda} \right|^{-1/2} \tag{3.33}$$

where $\mathbf{I}$ is an $n \times n$ identity matrix; $\boldsymbol{\Lambda}$ is a diagonal matrix coming from the second order derivative of the trigonometric loss function, whose entry is non-zero only if the

corresponding training sample is a support vector. Let $\mathbf{\Sigma}_M$ and $\mathbf{\Lambda}_M$ be the $m \times m$ sub-matrix of $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$ by keeping their non-zero entries. The negative logarithm of P(**D**|**θ**) is therefore:

$$-\ln P(\mathbf{D}|\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\upsilon}_M^T \mathbf{\Sigma}_M \boldsymbol{\upsilon}_M + 2\sum_{m \in SVs} \ln \sec\left(\frac{\pi}{4}\xi_m\right) + \frac{1}{2}\ln\left(\left|\mathbf{I} + \mathbf{\Sigma}_M \cdot \mathbf{\Lambda}_M\right|\right) \qquad (3.34)$$

where $\boldsymbol{\upsilon}_M$ is the sub-vector of $\boldsymbol{\upsilon} = \left[y_{\mathbf{x}_1}\alpha_1, y_{\mathbf{x}_2}\alpha_2, \cdots, y_{\mathbf{x}_n}\alpha_n\right]^T$ by keeping the entries associated with support vectors; and $\xi_m = 1 - y_{\mathbf{x}_m} \cdot f_{MP}(\mathbf{x}_m), \forall m \in SVs$.

Gradient based optimization can then be used for minimizing (3.34). The derivatives of $-\ln P(\mathbf{D}|\boldsymbol{\theta})$ with respect to $\ln\theta$ are given by:

$$\frac{\partial -\ln P(\mathbf{D}|\boldsymbol{\theta})}{\partial \ln\theta} = \frac{\theta}{2} trace\left(\left(\mathbf{\Lambda}_M^{-1} + \mathbf{\Sigma}_M\right)^{-1}\frac{\partial \mathbf{\Sigma}_M}{\partial\theta}\right) - \frac{\theta}{2}\boldsymbol{\upsilon}_M^T \frac{\partial \mathbf{\Sigma}_M}{\partial\theta}\boldsymbol{\upsilon}_M$$
$$-\frac{\theta}{2}\sum_{m \in SVs} \boldsymbol{\upsilon}_M^m \left(\left(\mathbf{\Lambda}_M^{-1} + \mathbf{\Sigma}_M\right)^{-1} \cdot \mathbf{\Sigma}_M\right)_{mm} \left(\mathbf{\Lambda}_M^{-1}\left(\mathbf{\Lambda}_M^{-1} + \mathbf{\Sigma}_M\right)^{-1}\frac{\partial \mathbf{\Sigma}_M}{\partial\theta}\boldsymbol{\upsilon}_M\right)^m \qquad (3.35)$$

where the superscript *m* denotes the *m*-th entry of a vector. Feature selection and training the classification network can then be conducted through the same procedures mentioned in 3.2.3. The network structure is shown in Figure 3.5.

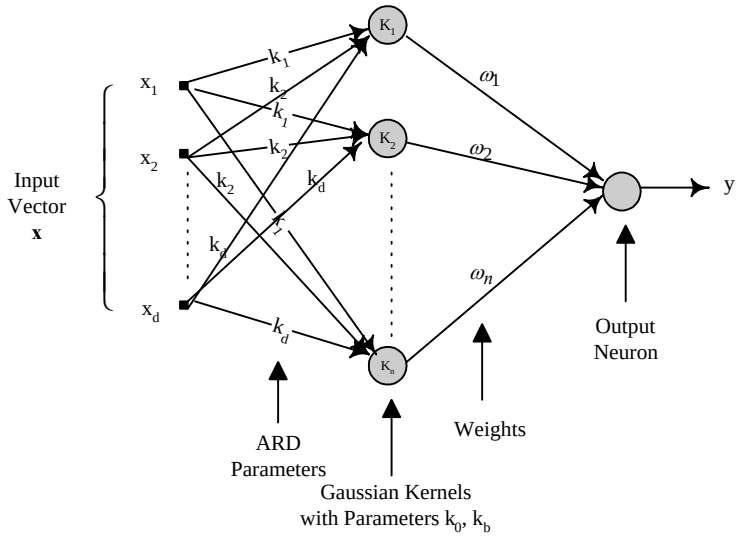


Figure 3.5 Structure of Bayesian Support Vector Machines

CHAPTER 4

EXPERIMENTAL SETUP AND DATA PROCESSING

In this chapter, the experimental setup for the data acquisition is first described, followed by the analysis of the signal and the implementation of the feature extraction methodologies mentioned in Chapter 2. An online tool condition monitoring strategy is given at the end of this chapter.

4.1 EXPERIMENTAL SETUP

Force signal is used in this study to monitor face milling processes, due to its high sensitivity to tool wear, low noise, and satisfactorily accurate force models (Altintas, 1989). Figure 4.1 shows the scheme of the experimental setup, whose components are listed in Table 4.1. The actual systems are shown in Figures F1 and F2 (Appendix F).

Table 4.1 Experimental Components

Components
Makino CNC milling machine with Funuc controller
EGD 4450R cutter with AC325 and A30N inserts
ASSAB718HH workpiece $(206mm \times 43mm \times 106mm)$
Kistler 9265B Quartz 3-Component Dynamometer
Kistler 5019A Multi-channel Charge Amplifier
NI-DAQ PCI 1200 Board
Olympus microscope and Panasonic digital camera
Computer with Pentium III 600MHz and 128M SDRAM

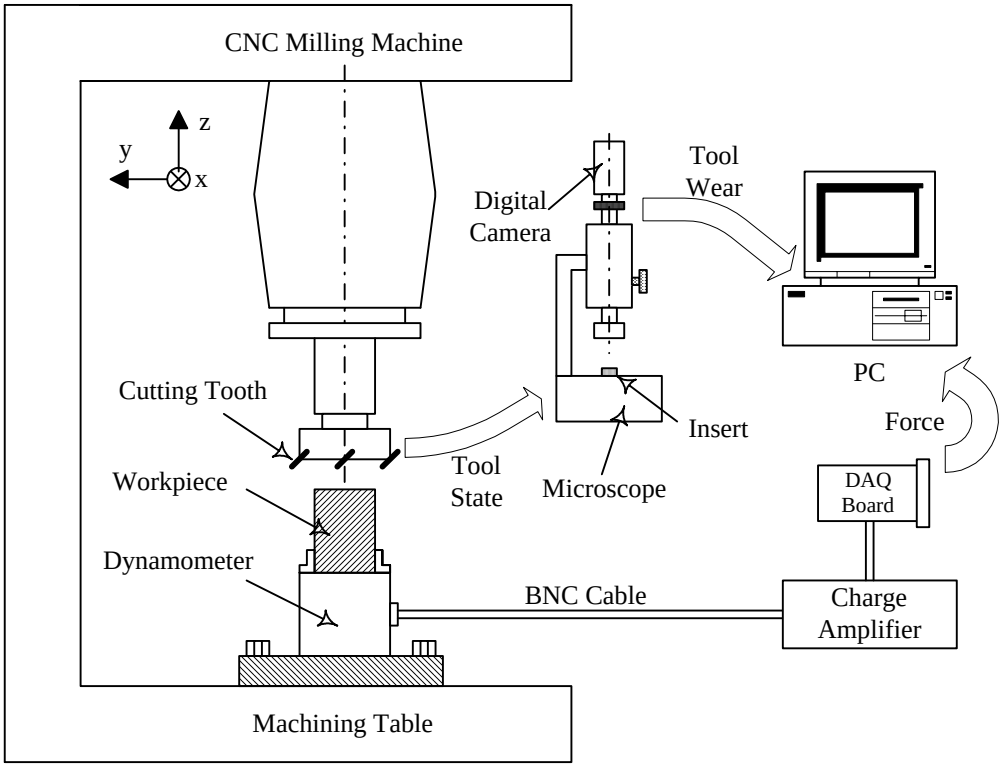


Figure 4.1 Experimental Setup

4.2 INSTRUMENTATION & DATA ACQUISITION

The cutting force along the y-direction (traverse cutting force) is monitored, due to its high sensitivity to tool wear and consistency in direction. The signal is captured by the Kistler dynamometer in the form of charges, and converted to voltages by the Kistler charge amplifier. The charge amplifier is grounded to provide a reference for the signal, with the parameters specified in Table 4.2. *TS* determines the sensitivity of the dynamometer to the cutting force. And *SC* determines the output magnitude of the charge amplifier.

Table 4.2 Specification of Parameters of the Charge Amplifier

Channel	1
Transducer Sensitivity/TS [pC/Mechanical Unit]	7.85
Scale/SC [Mechanical Units/volt]	600
Low-Pass Filter/LP	1KHz
Time Constant/TC (High-Pass Filter)	Long
Operation/OP	Enb.

The voltage signal from the charge amplifier is sampled by NI's PCI 1200 DAQ board at 2 kHz and directly streamed to the hard disk of the computer. In the board, there is a gain amplifier, a 12-bit *ADC* (analog-to-digital converter), and a 16-bit *FIFO* (first in first out) memory buffer. If necessary, the analog signal can first be amplified to increase resolution and accuracy through the gain amplifier. Then it is converted to digital format using the *ADC*. When an *A/D* conversion is completed, the *ADC* clocks the result into the *FIFO*, which serves as a buffer (4096 words deep) to the *ADC*.

The analog input polarity is selected as "bipolar $\pm 5V$" to accommodate negative signals. Therefore, the least significant bit (LSB) of the *ADC* is 10V/4095. Based on the scale of the charge amplifier listed in Table 4.2, the force can be calculated by:

$$F = v_d \cdot \frac{10}{4095} \cdot 600 \tag{4.1}$$

where v_d is the digitalized voltage level.

Because the charge amplifier is grounded, the Non-Referenced Single Ended (NRSE) mode is used to connect the signal. In this mode, all signals are referenced to the same common mode voltage, which is allowed to float with respect to the analog ground of the DAQ board. Any potential difference between the ground of the board and the signal ground appears as a common-mode signal at both the positive and negative inputs of the instrumentation amplifier and is therefore rejected by the amplifier. Figure 4.2 shows how to connect the voltage signal from the charge amplifier to the DAQ board in this mode. Channel 2 (ACH1) is used as the input terminal. AIGND and AISENSE stand for analog input ground and analog input sense, respectively.

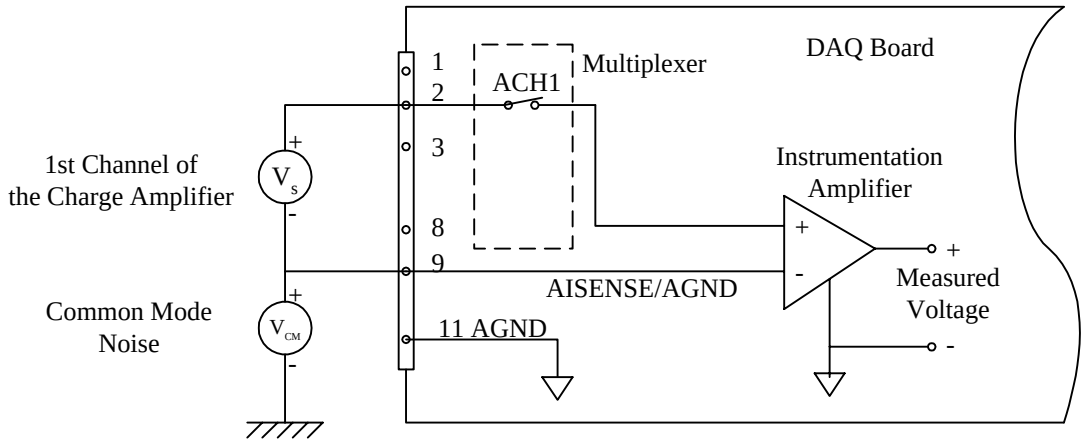
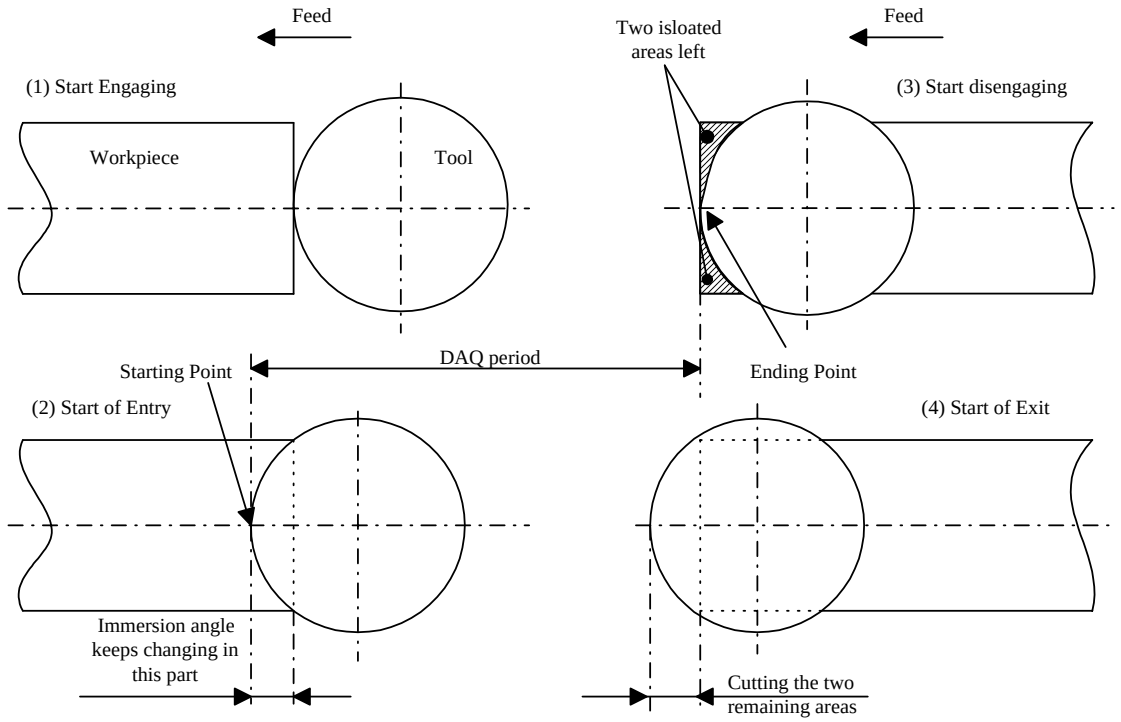


Figure 4.2 Connection of the Charge Amplifier to the DAQ Board

The DAQ board can be driven by the NI-DAQ driver software, which has a library of calling functions enabling configuration, initialization, and implementation of DAQ processes. With this library, a synchronous DAQ operation can be easily realized. However, because of the intermittent nature of milling processes, the starting and ending point of the DAQ operation must be carefully controlled.

When the tool is not engaged in cutting, the signal is useless and not necessary to measure. On the other hand, when the tool is engaged in cutting, the process can be partitioned into three sections. The first one is between position (1) and position (2) as shown in Figure 4.3. During this period, the immersion angle of the tool changes from 0 to 120 degrees (the geometry of the cutting is shown in Figure 2.1). The second one is between position (2) and position (3), during which the immersion angle remains 120 degrees and therefore the force waveform is almost consistent (except the fluctuation caused by chipping, breakage and large wear). The last procedure is between position (3) and position (4), during which there are only two isolated uncut areas left. So the chip load pattern of each tooth turns from one single pulse into two separate pulses, which is totally different from the pattern in the second procedure.



(a) Entry Point (b) Exit Point

Figure 4.3 Starting and Ending Point of DAQ

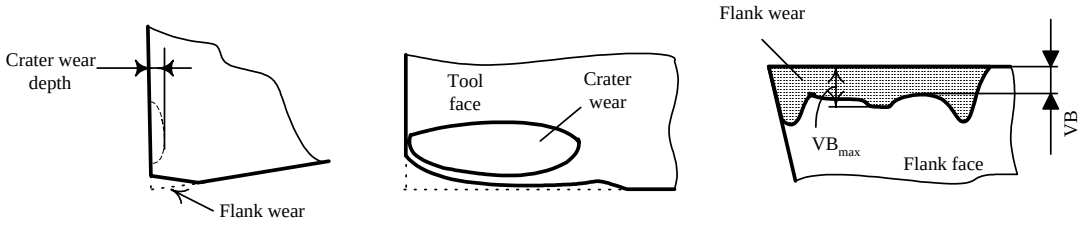
Based on these considerations, the useful signal can only be acquired after the tool is at start of entry in cutting and before the tool starts disengaging.

The starting point of DAQ can be controlled either manually or automatically. The method for automatically detecting the point is discussed in Section 4.5. The manual method is used to acquire the data for offline analysis. This means that the DAQ process is started when it is observed that nearly half of the tool holder is about to pass the front edge of the workpiece. The ending point of the DAQ process can be determined by counting the total number of samples. The maximum number of samples (N_{max}) can be calculated as follows:

$$N_{\max} = \frac{lw - rt}{fr} \times 60 \times sr \tag{4.2}$$

where *lw* denotes the length of the workpiece; *rt* is the effective radius of the tool holder; *fr* and *sr* stand for feed rate and sampling rate, respectively.

The flank wear of each individual tooth is measured at an interval of 5 tool passes by the Olympus microscope, and at each time an average is calculated from all the teeth mounted on the cutter. The tool wear patterns can be illustrated in Figure 4.4.



(a) End View (b) Top View (c) Side View

Figure 4.4 Illustration of Tool Wear Measurement

According to ISO 8688 (International Standard, 1989), the threshold in terms of flank wear for determining the tool life is 0.5mm. In practice, this threshold is a bit larger, because when the flank wear is above 0.4mm, it causes a severe vibration, which may damage the machine. Therefore, 0.4mm is used as the threshold. It means when the flank wear of any of the teeth inserted in the tool exceeds this value, the tool is regarded as worn. Although flank wear is the main factor determining the tool life under normal cutting conditions, some abrupt events may also occur, such as chipping and breakage, which could cause more catastrophic damage to the workpiece and the machine than the flank wear. So it is also necessary to check whether these events happen. This can be observed by the digital camera. According to ISO 8688, the measurement of tool chipping volume is illustrated in Figure 4.5. And the criterion for determining the type of chipping is listed in Table 4.3.

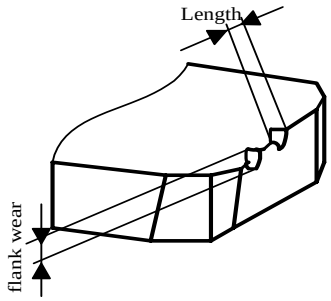


Figure 4.5 Measurement of Chipping Volume

Table 4.3 Types of Chipping

Type	Length (mm)
Micro-chipping	<0.3
Macro-chipping	0.3 to 1.0
Breakage	>1.0

Based on the methods for measuring the force signal and the tool wear mentioned before, 20 experiments are conducted on the Makino CNC milling machine. Two kinds of inserts, AC325 and A30N, are used in the experiments. The geometries of the inserts and the face mill are shown in Figures F3 and F4 (Appendix F).The cutting conditions are listed in Table 4.4.

Table 4.4 Cutting Conditions

Cutting Condition	**Value**
Spindle Speed (rpm)	600, 800, 1000, 1200
Feed Rate (mm/min)	100, 150, 200, 300
Depth of Cut (mm)	1, 2
Insert No.	2, 4
Immersion Rate	Full
Workpiece	ASSAB718HH
Milling Cutter	EGD4450R
Insert	AC325, A30N

4.3 EXPERIMENTAL DATA ANALYSIS

In this study, the influence of cutter run-out on the force signal is examined. After the first experiment, it was observed that the force waveform was severely distorted from the ideal pattern. Four inserts were used in the cutting. Instead of four identical peaks within the waveform of one single rotation, only three were found with large differences in their magnitudes at the fresh stage of the inserts. This phenomenon is illustrated in Figure 4.6. Because the inserts were still fresh, the fluctuation could not be due to tool failure, but due to the radial run-out of the cutter itself (the axial run-out of the cutter is usually negligible compared to the depth of cut). To interpret this, the

radial positions of the teeth were measured using a gauge. The difference in the radial positions of the four inserts is shown in Figure 4.7.

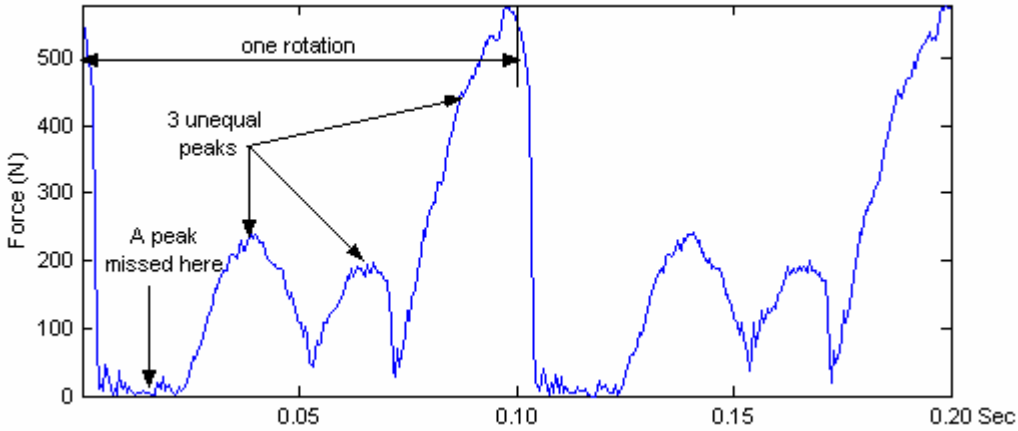


Figure 4.6 Experimental Force Waveform in Two Rotations
Spindle Speed: 600rpm, Feed Rate: 100mm/min,
Depth of Cut: 1 mm, Insert Number: 4, Insert Type: AC325.

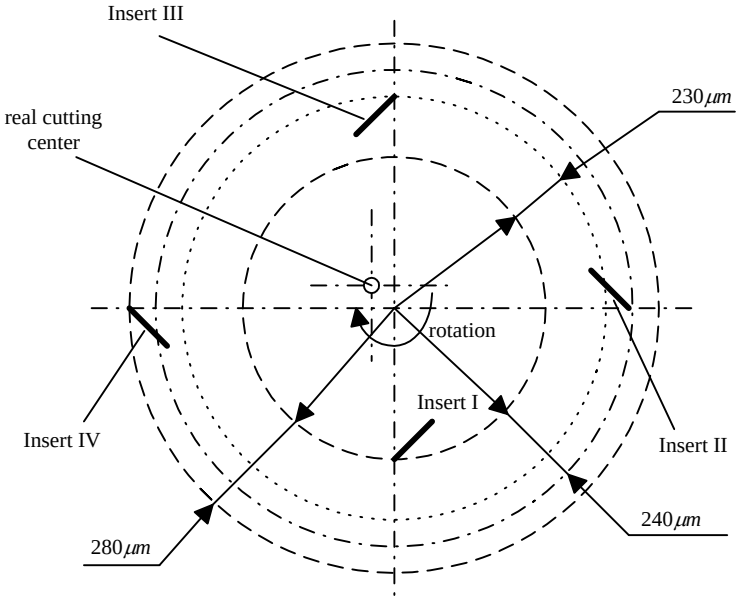


Figure 4.7 Radial Positions of Four Inserts

As can be seen from Figure 4.7, the radial run-out of the cutter is very big, because it has been used for almost ten years in the workshop. Based on these values and the force models described in Section 2.1, the chip load pattern and the transverse force (F_y) can be simulated, as shown in Figure 4.8 and 4.9. Since the tip of Insert I is more than 200 microns closer to the center of the tool holder than the other three inserts and

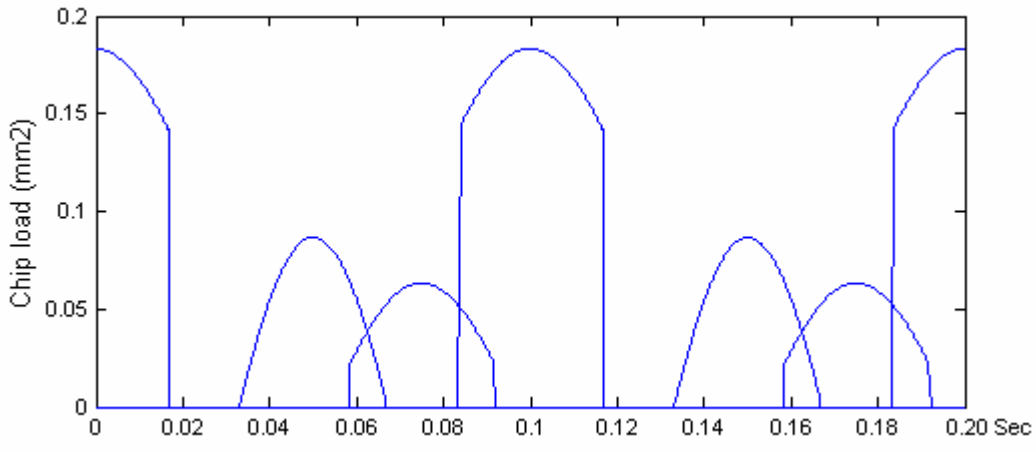


Figure 4.8 Simulated Chip Load Pattern with Run-out

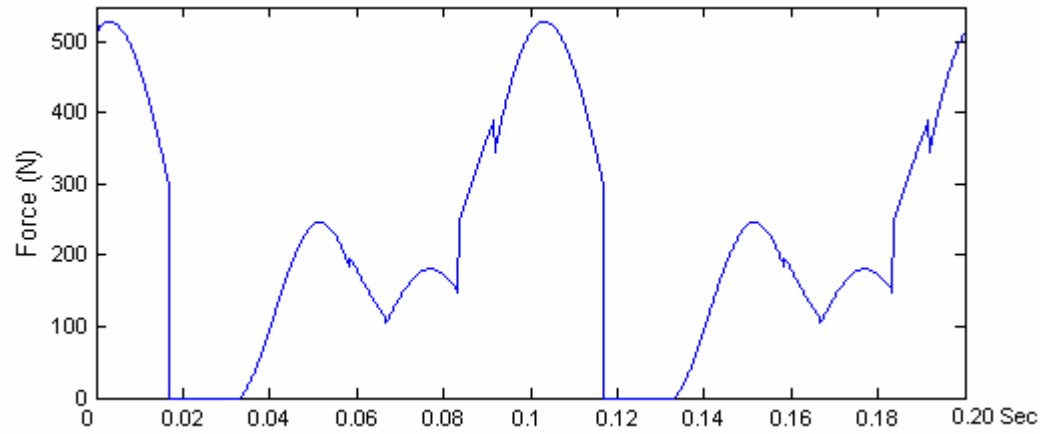


Figure 4.9 Simulated Transverse Force with Run-out

the feed per tooth is only 42 microns, Insert I can never engage in cutting, and it leaves uncut metal for the subsequent teeth. This results in uneven wear among the four teeth, which can not only shorten the tool life but also degrade the machined surface quality [Liang, 1994]. There are some methods for automatically compensating run-out, such as chip load manipulation [Liang, 1994] and spindle speed variation [Sastry, 1999]. These methods involve complex machine control elements, which are beyond the scope of this study. To avoid the serious run-out problem, a new face mill, with moderate run-out as shown in Figure 4.10, was used in the subsequent experiments. Figure 4.11 illustrates the simulated chip load pattern, the simulated force, and the sampled force with 2 inserts. As can be seen from the figures, the trouble caused by cutter run-out is greatly reduced.

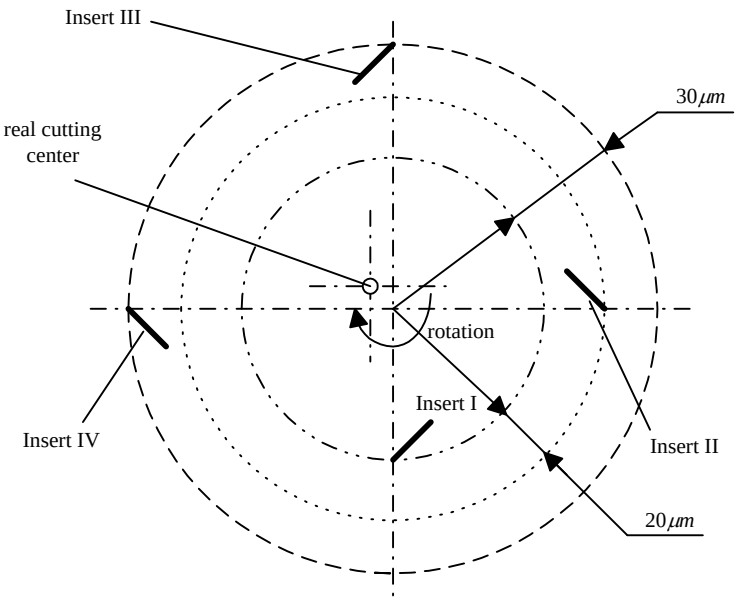


Figure 4.10 Run-out of the New Face Mill

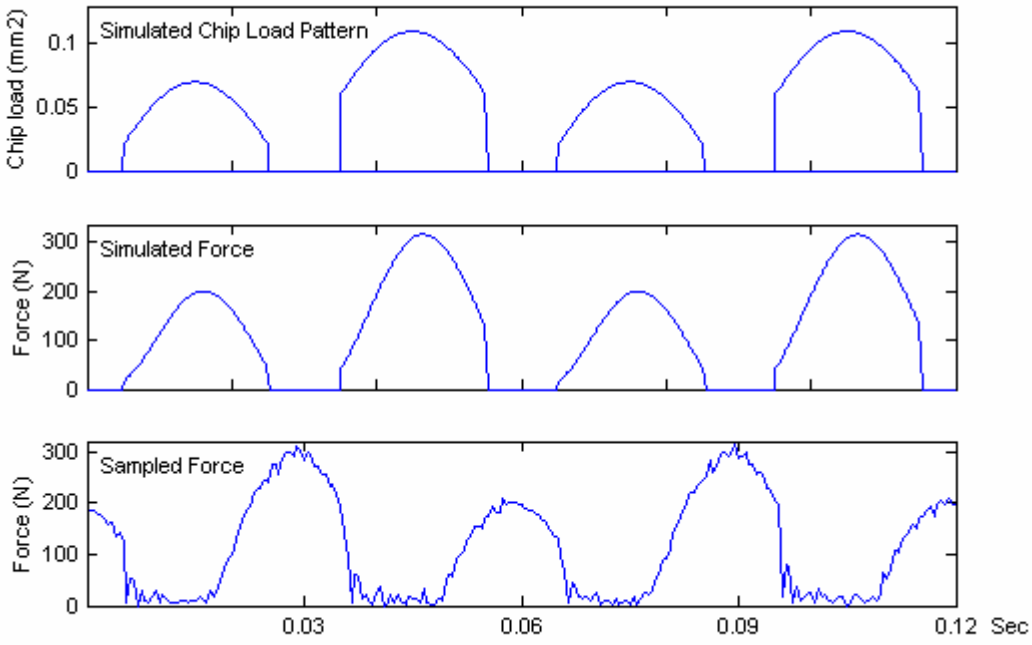


Figure 4.11 Simulated Chip Load Pattern/Force and Sampled Force
Spindle Speed: 1000rpm, Feed Rate: 200mm/min,
Depth of Cut: 1 mm, Insert Number: 2, Insert Type: AC325.

4.4 FEATURE EXTRACTION

The 16 different features discussed in Chapter 2 are extracted from the experimental force data. The feature extraction process follows a five-step procedure as shown in Figure 4.12. This procedure can reduce computational redundancy, becau-

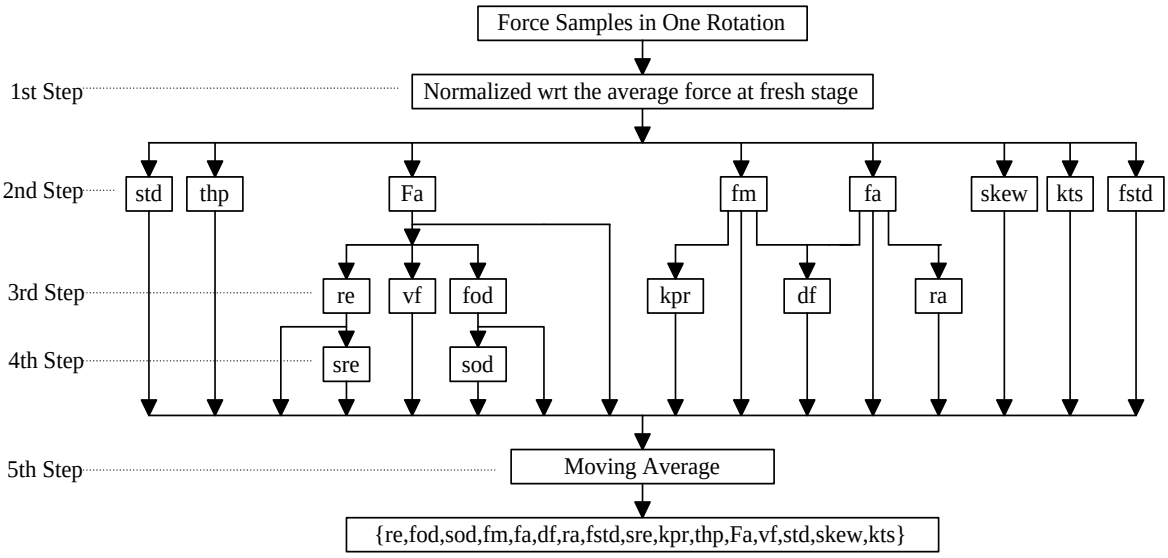


Figure 4.12 Feature Extraction Procedure

se of the dependency of the lower level features on the upper ones. Note that samples within one spindle rotation instead of one tooth period are used in feature extraction. The purpose is to avoid the influence of run-out. As can be seen from either Figure 4.6 or 4.11, the run-out causes fluctuation to the force magnitudes of the teeth even at the fresh stage. However, the fluctuating behavior of cutting force is just the sign to distinguish failed tools from fresh ones. Since in most cases the run-out of the milling cutter is not negligible, it is safer to extract the features over each rotation to avoid expressing the fluctuation caused only by run-out.

The force samples are first normalized with respect to the average force when the tool is still fresh:

$$\eta = \frac{\sum_{j=1}^{M} F(j)}{M} \tag{4.2}$$

$$F'(i) = F(i)/\eta, \forall i > M \tag{4.3}$$

where M is total number of samples during the first n rotations of the first pass; $\forall i > M$ means all the samples after the first M; and η will never be zero because the force values along y direction are always positive. This procedure can effectively reduce the sensitivity of the features to cutting conditions [Altintas, 1989; Kim, 1996].

At the second step, eight features are derived from the normalized force samples $F'(i)$ in one spindle rotation, including Fa, fm, fa, skew, kts, fstd, std, and thp. The feature extraction methodologies are described in details in Chapter 2. In the next step, six features based on Fa, fm, and fa are extracted, including re, vf, fod, kpr, df, and ra. In the fourth step, sre and sod are evaluated from re and fod, respectively.

Through the first four steps, all of the 16 primary features are obtained. However, these features are still quite noisy and not reliable for the subsequent processing. Therefore, a moving average step is performed for each of the features to make it smooth. Let *w* be the size of the moving window. Then the moving average can be calculated by:

$$\Psi'(m) = \frac{1}{w} \sum_{i=m}^{m+w-1} \Psi(i) \tag{4.4}$$

where $\Psi(i)$ can be any of the 16 features.

In addition, as the magnitudes of the features differ greatly, as shown in the feature examples of Chapter 2, scales should be used to make them comparable. In this study, all the magnitudes of the features are scaled to around 5 to make sure the outputs from the Gaussian kernels in the neural networks significant enough. Large magnitudes can make the covariance matrix singular (details in Chapter 3). The scales can easily be obtained by calculating the ratio between 5 and the magnitudes of the features. This step is necessary for the feature selection in Chapter 5 to make sense. The five steps result in a feature vector of $\{\text{re,fod,sod,fm,fa,df,ra,fstd,sre,kpr,thp,Fa,vf,std,skew,kts}\}$, which forms the candidate set of features for the feature selection. Figure 4.13 shows these features after completing all the five steps.

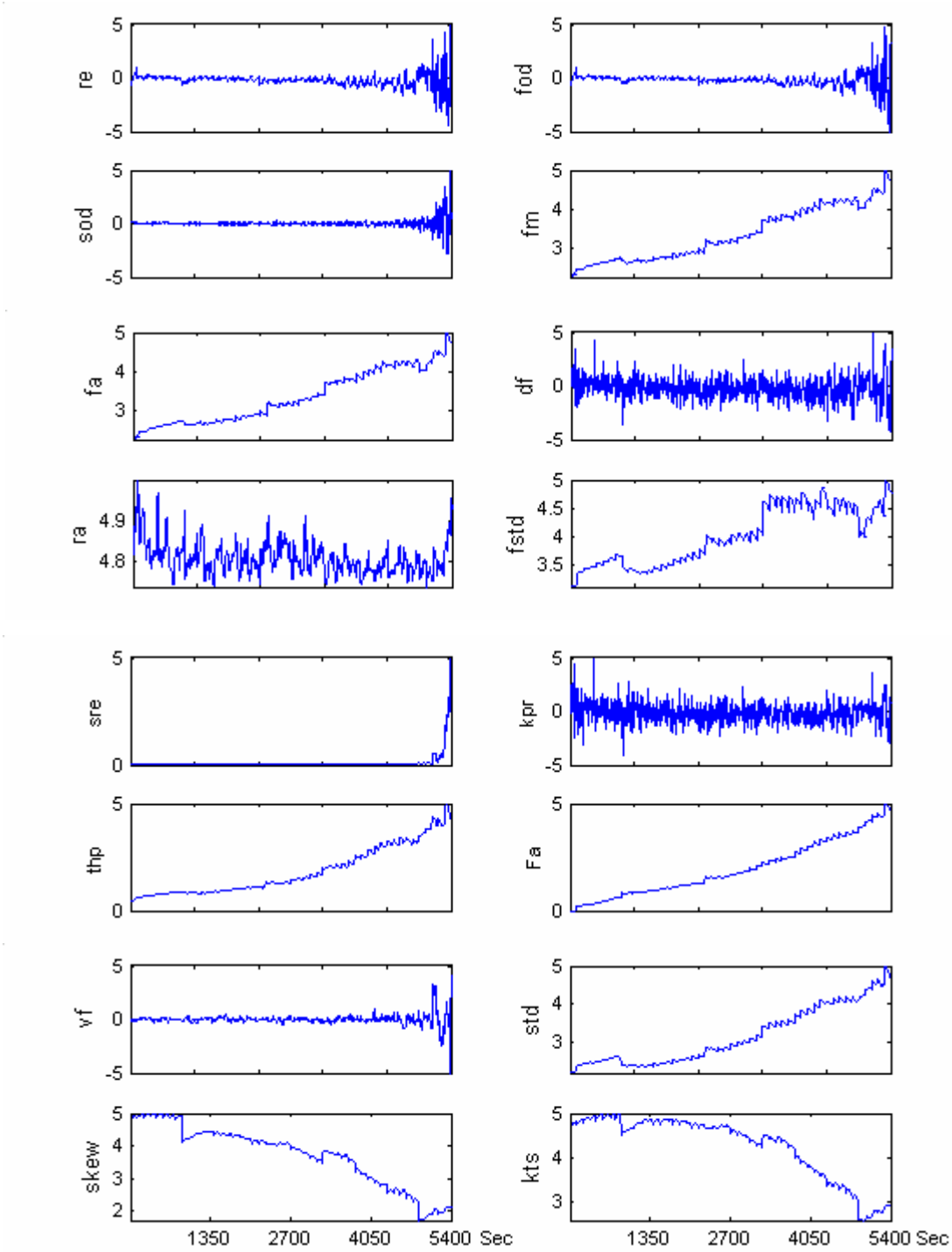


Figure 4.13 Feature Extraction Results
Spindle Speed: 1000rpm, Feed Rate: 100mm/min,
Depth of Cut: 1 mm, Insert Number = 2, Insert Type: AC325.

4.5 ONLINE TCM STRATEGY

The online tool condition monitoring system is an integrated system of testing hardware and monitoring software. The hardware setup and its configuration have

been described in Section 4.1 and 4.2. In this section, the software structure is outlined. The software fulfills two main tasks: control of the DAQ board and realization of tool wear recognition as well as tool wear estimation.

For the first task, the software calls the functions in the library of NI-DAQ divers, and therefore it is compatible with various kinds of DAQ boards. More specifically, "DAQ_Op", a synchronous DAQ operation, is called to sample force data to a buffer. Then features can be extracted from the data, which can be streamed to the hard disk of the computer from the buffer. Figure 4.14 shows the main structure of the software. Upon started, it initializes the parameters using the input values from the user interface. Then it calculates three quantities, T_d (delay time of a timer routine), N_m (maximum rotation number within one pass), and T_{CR} (the time for the cutter to move by the distance of its radius), for controlling the execution of data acquisition. The data acquisition and decision making procedures are implemented in a timer routine, which controls the starting and ending point of the effective data acquisition. The timer routine is illustrated in Figure 4.15.

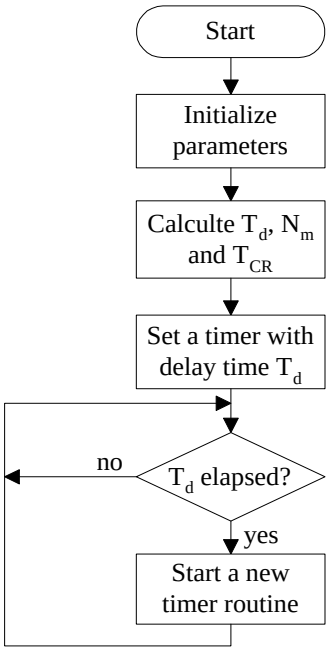


Figure 4.14 Main Routine of the Software

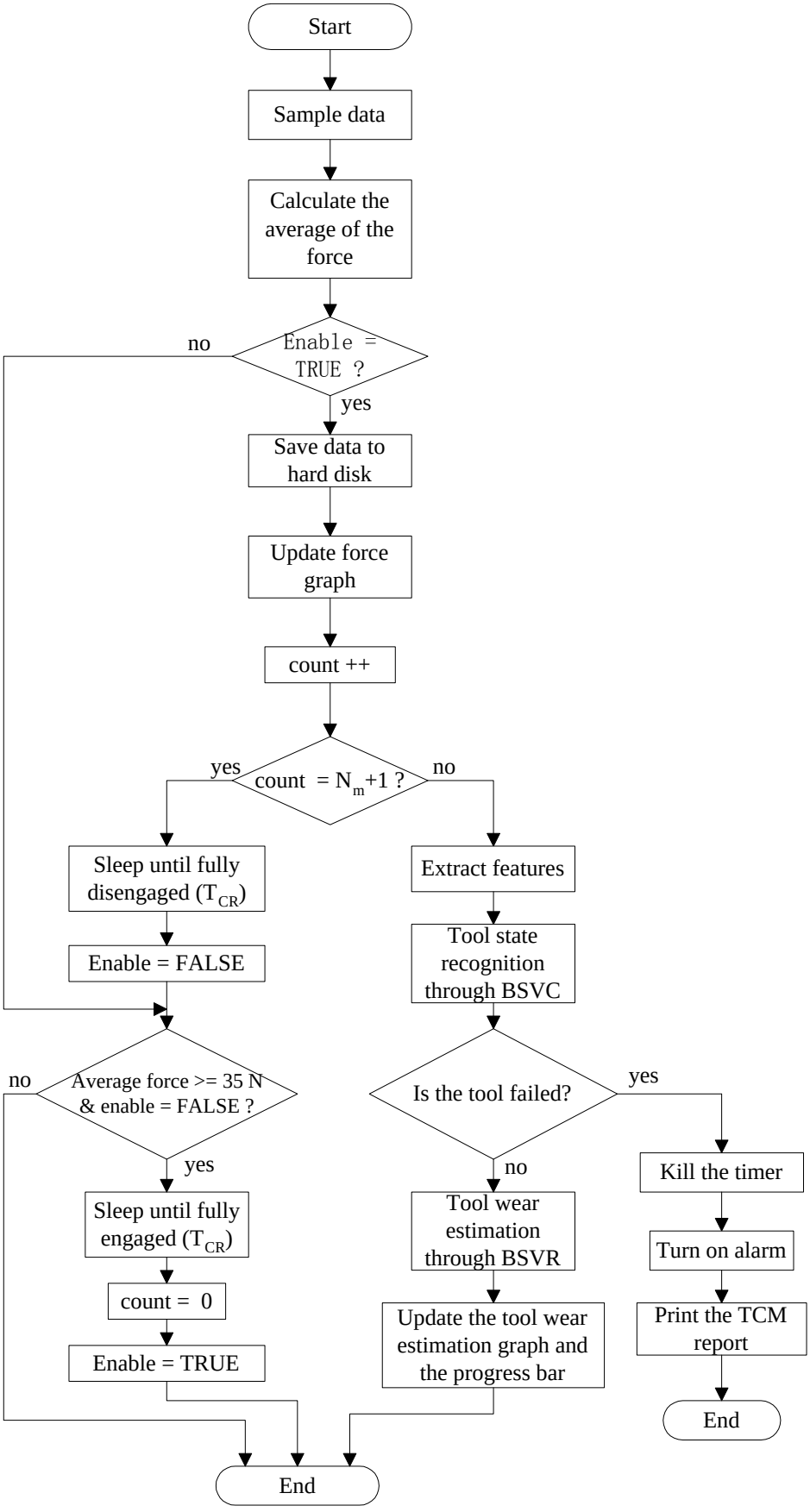


Figure 4.15 Timer Routine

T_d, N_m, and T_{CR} can be calculated as follows.

Notation:

Spindle Speed (rpm):	S_s	Length of Workpiece (mm):	L_w
Feed Rate (mm/min):	F_r	Sampling Rate (Hz):	S_r
Effective Radius of Tool Holder (mm):	R_t	Processing Time (ms):	T_p
Number of Rotations within a Timer Period:	N_T		

$$T_d = \left(\frac{60 \times N_T}{S_s} + T_p \right) \times 1000 \tag{4.5}$$

$$N_m = \frac{(L_w - R_t) \times 60 \times 1000}{F_r \times T_d} \times N_T \tag{4.6}$$

$$T_{CR} = \frac{R_t}{F_r} \times 60 \times 1000 \tag{4.7}$$

where T_d and T_{CR} are in milliseconds. N_T is a constant number, which is set to 10. The processing time T_p includes the time used in feature extraction and decision making, which can be estimated according to the computational burden and the processing speed of the computer.

The timer routine can be explained as follows. When started, the system sets a timer. When the first timer period elapses, it begins to collect data, no matter whether or not the cutter is engaged in cutting. The program then keeps tracking the average value of the data during the period of one rotation. If it is found that the average changes to a value greater than 35 Newton (empirical value), the system will know that the cutter begins engaging the workpiece. The program will wait until the center of the cutter reaches the front edge of the workpiece. Then tool wear recognition and tool wear estimation begins to function. When the rotation number in one pass reaches the limit N_m, the program will wait for the cutter to fully disengage and set “enable” to

FALSE to indicate the data become inefficient again. In the case that tool failure is detected by the system during a tool pass, the program will kill the timer to stop the monitoring process and sound warning beeps.

The GUIs of the software are shown from Figure F5 to Figure F8 in Appendix F. Figure F5 is the view window, on which the force signal measured in each pass and the estimated wear values during the entire monitoring process are plotted. The force graph indicates what is going on during milling processes. There is an item named “TCM” in the menu of the view window. It contains three dialogs, Milling Properties, DAQ Specifications, and Monitoring, which are illustrated in Figures F6, F7, and F8, respectively. Parameters can be fed into the system through the first two dialogs. The third dialog can only be enabled after the “OK” button in both of the first two dialogs is pressed. This dialog is responsible for indicating the growth of tool wear and raising alarm. A tool condition monitoring report can be printed at any point after the software is started, via “Print” item in “File” menu. An example is given in Figure F9. Figure F10 shows what the view window looks like when the software is working.

CHAPTER 5

RESULTS AND DISCUSSION

In this chapter, the feature selection results for tool wear estimation are given first, followed by the comparisons of the generalization capabilities using the entire, selected, and rejected feature sets. Then the feature selection results and similar discussions for tool wear recognition are presented. The results are summarized at the end of this chapter.

5.1 FEATURE SELECTION RESULTS FOR TWE

Twenty experiments are conducted using the setup described in Chapter 4. The cutting conditions for these experiments are listed in Table 5.1. Note that the item "Average Wear at Tool Failure Point" is the average wear value among all the teeth inserted in the tool when any of the teeth is found worn. The sixteen different features $\mathbf{x} = \{\text{re,fod,sod,fm,fa,df,ra,fstd,sre,kpr,thp,Fa,vf,std,skew,kts}\}$ introduced in Chapter 2 are extracted from all of the experimental data. Figure A1 to Figure A20 in Appendix A show the features of all the 20 experiments. The vector **x** forms the scope of the feature selection.

The feature selection algorithms for regression are described in Section 3.2 of Chapter 3. The regression aims to find a mapping function between the feature vectors and the tool wear values. The feature selection for the regression is to find the most relevant features to tool wear from the candidate feature vector **x**. Sixteen hyperpara-

Table 5.1. Cutting Experiments

Insert Type	Test No	Spindle Speed (rpm)	Feed Rate (mm/min)	Depth of Cut (mm)	Insert Number	Average Wear at Tool Failure Point (mm)	Phenomenon*
AC325	Test_a1	800	150	1	4	0.3718	GW
	Test_a2	1000	100	1	2	0.4595	GW
	Test_a3	1000	100	1	4	0.4102	GW
	Test_a4	1000	200	1	2	0.3960	BK
	Test_a5	1000	300	1	4	0.3043	BK
	Test_a6	1200	150	1	2	0.3621	CP
	Test_a7	1200	200	1	2	0.4238	BK
	Test_a8	1200	300	1	4	0.3234	BK
	Test_a9	600	100	2	4	0.3520	GW
	Test_a10	600	200	2	4	0.2736	GW
	Test_a11	800	100	2	2	0.2915	GW
	Test_a12	1000	100	1	4	0.2633	CP
A30N	Test_b1	800	200	1	4	0.4200	GW
	Test_b2	800	300	1	4	0.3626	GW
	Test_b3	1000	200	1	2	0.3803	GW
	Test_b4	1000	300	1	4	0.3938	GW
	Test_b5	1000	300	2	4	0.3920	GW
	Test_b6	1200	100	1	2	0.3780	GW
	Test_b7	1200	200	1	4	0.4094	GW
	Test_b8	800	300	1	4	0.3382	GW

* GW: Gradual Wear, CP: Chipping, BK: Breakage.

meters are assigned for them $\{k_1, k_2, \ldots, k_{16}\}$. The automatic relevance determination algorithm is used to infer the optimal values of these sixteen hyperparameters. During the computation, the less relevant feature dimensions are effectively suppressed as their controlling hyperparameters are automatically reduced to negligible values. The feature selection process of Test_a1 is shown in Figure 5.1 as an example. Those of the other 19 experiments are illustrated in Figure B1 to B19 in Appendix B. Note that in order to plot the graphs clearly, the hyperparameters after each iteration are all normalized with respect to the maximum value of the hyperparameters at that iteration:

$$g_i^{(j)} = \frac{k_i^{(j)}}{\max\left(k_1^{(j)}, k_2^{(j)}, \cdots, k_{16}^{(j)}\right)}, i = 1,2,\cdots,16 \tag{5.1}$$

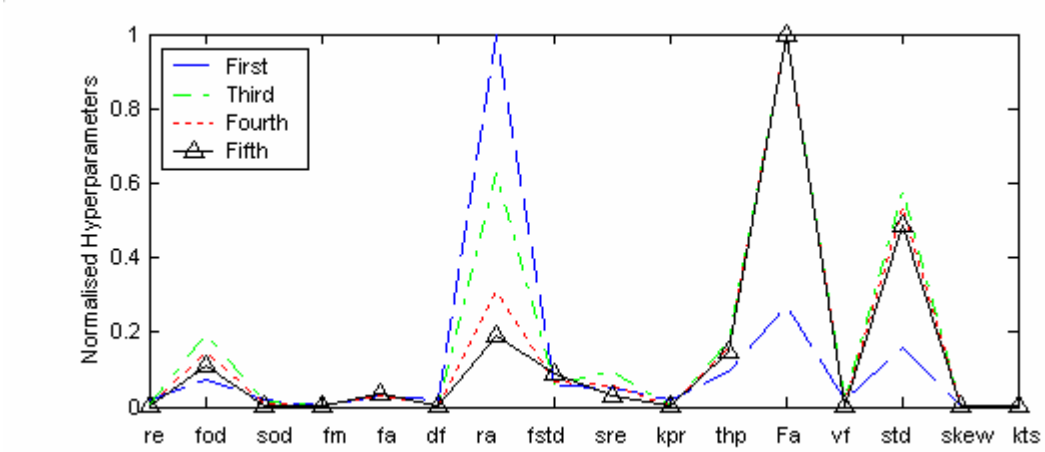


Figure 5.1. Illustration of the Feature Selection Processes of Test_a1 for TWE
Cutting conditions: spindle speed = 800rpm, feed rate = 150mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

where $k_i^{(j)}$ denotes the *i*-th hyperparameter after the *j*-th iteration. From the graphs, the convergence of the feature selection processes can be clearly seen.

The values of $k_1, k_2, \cdots, k_{16}$ at the last iteration of all the experiments are listed in Table 5.2. All the numbers are compared with "1". Those feature dimensions with the corresponding k_i's greater than 1 are selected as relevant features; otherwise, the features are rejected, because their contribution to the prediction is compromised by a factor smaller than 1. The feature selection results are listed in Table 5.3, with the "√" marks representing the selected feature set.

It can be seen from Table 5.3 that *Fa*, the average force, turns out to be the most relevant feature. This result coincides with the wide and successful use of the average force for tool wear estimation very well [Lin, 1996; Santanu, 1996]. The good correlation between the average force and the tool wear values can also be seen from Figure A1 (a) to A20 (a). And there are also 6 other features that appear to be relevant in some of the experiments: *std, ra, fstd, kts, fm,* and *fa*. As a result, a union of all these 7 features is taken to be the relevant feature set: {*fm, fa, ra, fstd, Fa, std, kts*}.

Table 5.2 Hyperparameter Values at the Last Iteration for TWE

Feature	*Test_A1*	*Test_A2*	*Test_A3*	*Test_A4*	*Test_A5*	*Test_A6*	*Test_A7*	*Test_A8*	*Test_A9*	*Test_A10*
re	0.0045	0.0060	0.2012	0.0056	0.0000	0.0013	0.0091	0.0000	0.0001	0.0001
fod	0.2836	0.0886	0.0473	0.0995	0.0723	0.2610	0.2249	0.0985	0.2102	0.1451
sod	0.0101	0.0375	0.0222	0.0178	0.0074	0.0255	0.0065	0.0191	0.0052	0.0159
fm	0.0078	0.0000	0.0000	0.0598	0.1218	0.0000	0.0847	0.0000	0.0000	0.0181
fa	0.0835	0.0326	0.0565	0.1915	0.0000	0.0535	1.4259	0.1269	0.0000	0.7220
df	0.0022	0.0129	0.0608	0.0181	0.0339	0.0307	0.0115	0.0100	0.0194	0.0122
ra	0.4926	1.8699	0.7049	7.0827	9.7756	5.5813	2.2112	8.6904	16.6580	122.1674
fstd	0.2273	0.0420	0.1119	0.9217	0.5760	0.6318	0.6086	0.1936	0.0000	1.2434
sre	0.0825	0.0568	0.0508	0.0138	0.0342	0.0952	0.0274	0.0035	0.0474	0.0480
kpr	0.0001	0.0111	0.0272	0.0075	0.0043	0.0201	0.0061	0.0084	0.0102	0.0074
thp	0.3771	0.1654	0.4950	0.0000	0.1985	0.0000	1.1116	0.9294	0.3233	0.9327
Fa	2.5578	3.3660	2.6926	2.2844	5.5087	3.3153	0.4292	3.1612	3.5657	3.6165
vf	0.0131	0.0347	0.0101	0.0148	0.0339	0.0225	0.0268	0.0320	0.0148	0.0270
std	1.2387	0.7704	0.8110	1.3942	2.1895	1.8808	4.7131	1.7750	0.6492	2.9707
skew	0.0017	0.0030	0.0109	0.0028	0.0000	0.2337	0.0032	0.0195	0.0000	0.0170
kts	0.0005	0.1502	0.0767	0.0035	0.0510	0.0000	0.0000	0.3732	0.0209	0.3707

Continued:

Feature	*Test_A11*	*Test_A12*	*Test_B1*	*Test_B2*	*Test_B3*	*Test_B4*	*Test_B5*	*Test_B6*	*Test_B7*	*Test_B8*
re	0.0091	0.0100	0.0000	0.0000	0.0021	0.0000	0.0000	0.0000	0.0169	0.0000
fod	0.0332	0.0721	0.0887	0.0462	0.1598	0.0586	0.1006	0.0347	0.1769	0.0943
sod	0.0021	0.0011	0.0041	0.0005	0.0045	0.0000	0.0210	0.0121	0.0000	0.0127
fm	0.0568	0.0012	0.0000	0.0000	0.0000	0.0000	0.0775	0.0000	0.0241	1.1234
fa	0.0000	0.0052	0.3823	0.2171	0.0000	0.3337	0.0094	0.2231	0.2835	0.0886
df	0.0061	0.0033	0.0000	0.1561	0.0000	0.1078	0.0221	0.0106	0.0009	0.0049
ra	31.1166	0.6721	0.3592	0.5525	7.2252	3.8678	3.7798	0.4125	0.3239	0.3929
fstd	0.0755	0.5198	0.0233	0.0113	0.2828	0.8648	0.8205	8.0873	0.0047	0.5429
sre	0.0547	0.1483	0.0219	0.0687	0.2660	0.2327	0.0501	0.0162	0.0225	0.0309
kpr	0.0029	0.0031	0.0000	0.0847	0.0218	0.0798	0.0038	0.0052	0.0007	0.0000
thp	0.6139	0.2867	0.5731	0.2050	0.2456	0.1488	0.1510	0.1509	0.1440	0.4488
Fa	9.6543	0.9099	4.8620	3.1382	6.6912	1.6410	3.3943	8.9221	4.0167	6.3221
vf	0.0138	0.0029	0.0218	0.1084	0.0226	0.1467	0.0646	0.0152	0.0071	0.0270
std	4.4762	1.0996	1.3115	0.9193	0.3197	0.4690	1.5296	6.4630	0.3368	2.3618
skew	0.8515	0.5174	0.0000	0.0704	0.0569	0.0358	0.1194	0.1634	0.0004	0.0737
kts	0.3929	1.1077	0.0002	0.0074	0.0320	0.2980	0.4034	2.9630	0.0015	0.1970

Table 5.3 Feature Selection Results for TWE

	re	*fod*	*sod*	*fm*	*fa*	*df*	*ra*	*fstd*	*sre*	*kpr*	*thp*	*Fa*	*vf*	*std*	*skew*	*kts*
Test_a1												√		√		
Test_a2							√					√				
Test_a3												√				
Test_a4							√					√		√		
Test_a5							√					√		√		
Test_a6							√					√		√		
Test_a7					√		√							√		
Test_a8							√					√		√		
Test_a9							√					√		√		
Test_a10							√	√				√		√		
Test_a11							√					√		√		
Test_a12														√		√
Test_b1												√		√		
Test_b2												√				
Test_b3							√					√				
Test_b4							√					√				
Test_b5							√					√		√		
Test_b6								√				√		√		√
Test_b7												√				
Test_b8				√								√		√		

5.2 VERIFICATION OF THE RELEVANCE OF THE SELECTED FEATURE SET FOR TWE

In order to verify the relevance of the selected features to tool wear, generalization tests are performed. The experimental data sets are divided into two groups, one for training and another one for testing. The training and testing data sets are listed in Table 5.4 and 5.5, respectively. The 7 selected features {*fm, fa, ra, fstd, Fa, std, kts*} are first used in training and testing. For comparison purpose, the entire 16 and the rejected 9 features are used to repeat the procedures. Figure 5.2 shows the results from the testing data set T1. Those obtained from the remaining 10 data sets are illustrated in Figures C1 to C10 in Appendix C. It can be clearly seen from these figures that the

Table 5.4 Training Data Sets for TWE

No.	Insert Type	Spindle Speed (rpm)	Feed Rate (mm/min)	Depth of Cut (mm)	Insert Number
1	AC325	1000	100	1	4
2	AC325	1000	200	1	2
3	AC325	1000	300	1	4
4	AC325	1200	300	1	4
5	AC325	600	200	2	4
6	AC325	1000	100	1	4
7	A30N	1000	300	2	4
8	A30N	1200	100	1	2
9	A30N	1200	200	1	4

Table 5.5 Testing Data Sets for TWE

No.	Insert Type	Spindle Speed (rpm)	Feed Rate (mm/min)	Depth of Cut (mm)	Insert Number
T1	AC325	800	150	1	4
T2	AC325	1000	100	1	2
T3	AC325	1200	150	1	2
T4	AC325	1200	200	1	2
T5	AC325	600	100	2	4
T6	AC325	800	100	2	2
T7	A30N	800	200	1	4
T8	A30N	800	300	1	4
T9	A30N	1000	200	1	2
T10	A30N	1000	300	1	4
T11	A30N	800	300	1	4

estimated wear values using the selected feature set closely follow the measured wear curve, whereas those derived from the rejected feature set scatter randomly on the graph. The estimated values using the entire feature set also appear to be not as good as

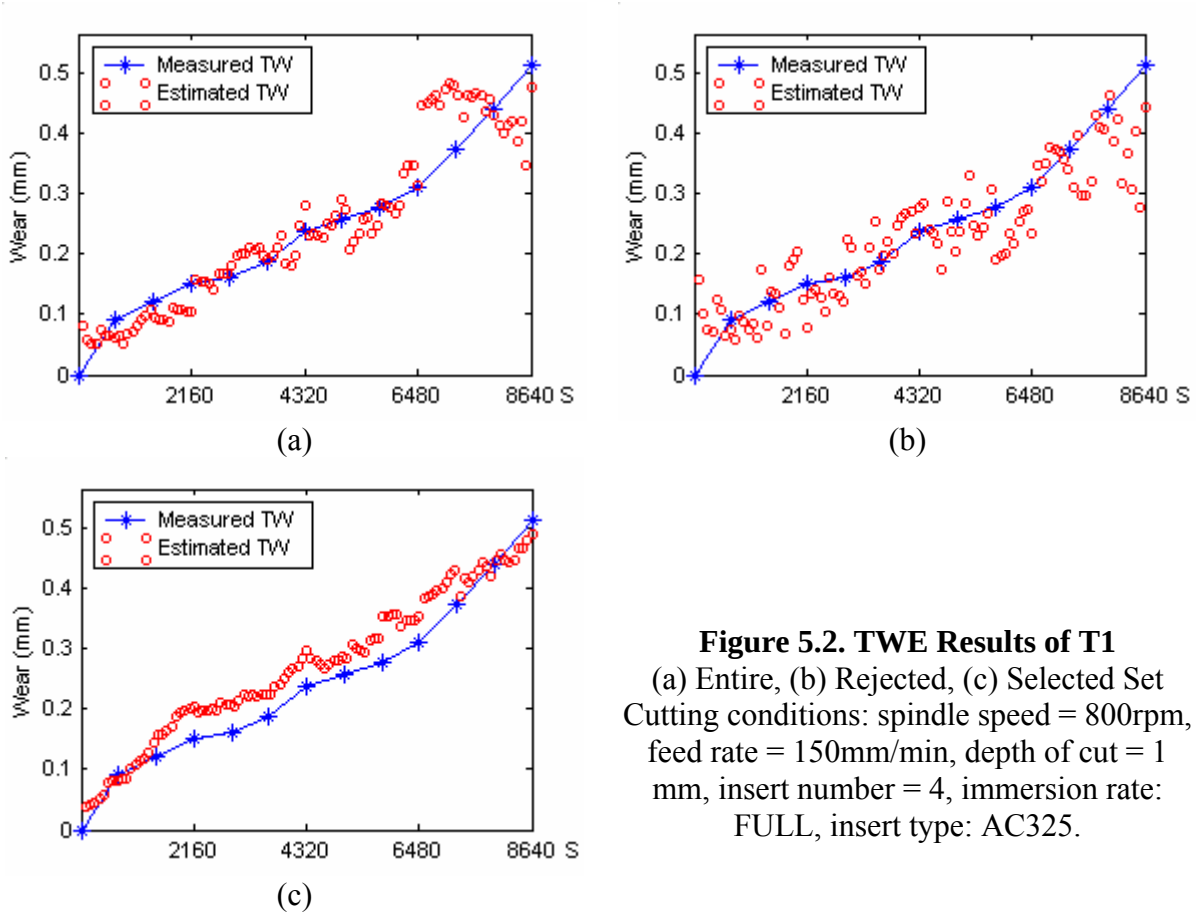


Figure 5.2. TWE Results of T1
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 800rpm, feed rate = 150mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

those using the selected feature set, because the input space is corrupted by the noisy rejected features.

It can also be observed that in T1, T4, and T11, the best estimates using the selected features are slightly biased. These results can be explained by analyzing the tool wear growth rate in terms of the amount of wear growth per Newton increase in cutting force (μm/N).

The average wear growth rate of the training data sets is 2.5 (μm/N). Due to the averaging effect of the neural estimator, the growth rate of the estimated wear is also around 2.5. If the true wear growth rate of a testing data set exceeds this number, the estimated wear values will be smaller than the measured ones, because the growth of

the former lags behind that of the latter; otherwise the estimated values will be larger. The true wear growth rate of T1, T4, and T11 is 2.2, 3.5, and 1.7, respectively. Therefore, the estimates of T1 and T11 are slightly larger than the measurements; while those of T4 are slightly smaller.

The generalization capabilities of the entire, selected, and rejected feature sets are compared in terms of the averaged absolute estimation error (AAEE), which is defined as:

$$AAEE = \frac{\sum_{i=1}^{N} \left| f_{MP}(\mathbf{z}_i) - y_i \right|}{N} \tag{5.2}$$

where $\mathbf{z}_i$ represents the input vector comprising the entire, selected, and rejected feature set, respectively; $f_{MP}(\mathbf{z}_i)$ is the MAP estimation output of the BSVR networks; y_i is the measured wear value; and N is the total number of the points in one testing data set. Then the ratios between the AAEEs and 400 microns are calculated to indicate the relative estimation errors. 400 microns is used here, because it is the threshold to determine whether the tool is failed or not (Chapter 4). The comparisons are listed in Table 5.6 and visualized in Figure 5.3.

Table 5.6 and Figure 5.3 indicate the same outcomes. It is clear that the generalization capabilities of the selected feature set are the best and much better than those of the rejected feature set. This proves that the selected features are relatively more relevant to tool wear processes.

5.3 FEATURE SELECTION RESULTS FOR TWR

The feature selection algorithms for classification are described in Section 3.3 of Chapter 3. Unlike the regression, the classification aims to map the feature vectors into

Table 5.6 Tool Wear Estimation Results

Test No.	Entire Features		Rejected Features		Selected Features	
	AAEE (micron)	$\frac{AAEE}{400}$ (%)	AAEE (micron)	$\frac{AAEE}{400}$ (%)	AAEE (micron)	$\frac{AAEE}{400}$ (%)
T1	36.1	9.0	43.7	10.9	34.9	8.7
T2	67.7	16.9	91.7	22.9	32.5	8.1
T3	54.9	13.7	69.6	17.4	25.2	6.3
T4	61.0	15.3	90.6	22.7	39.0	9.8
T5	30.9	7.7	60.5	15.1	19.1	4.8
T6	38.3	9.6	50.6	12.7	25.9	6.5
T7	34.2	8.6	108.3	27.1	33.3	8.3
T8	53.0	13.3	88.4	22.1	21.8	5.5
T9	28.4	7.1	51.7	12.9	18.1	4.5
T10	54.8	13.7	72.3	18.1	36.6	9.2
T11	75.8	19.0	86.1	21.5	35.5	8.9

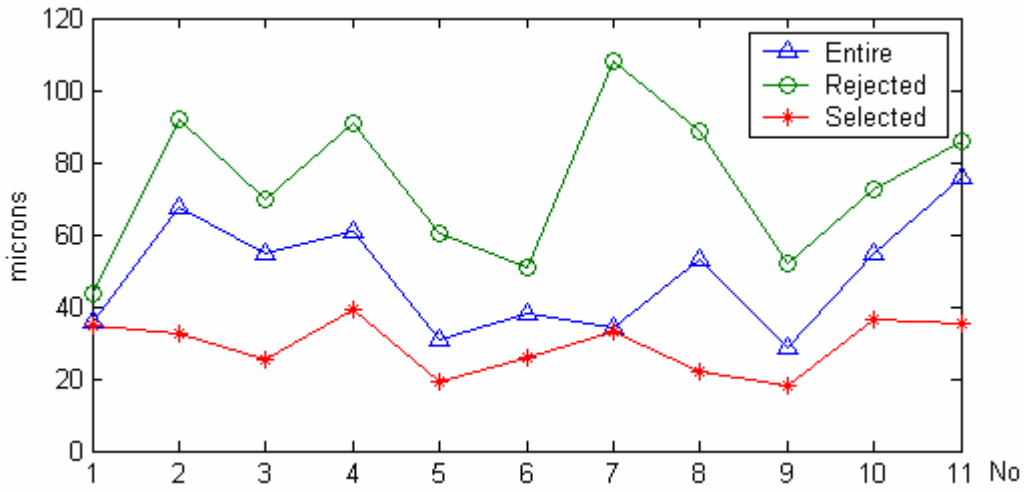


Figure 5.3 Comparisons of the Estimation Errors

tool conditions. The feature selection for the classification is to find the most relevant features to tool conditions from the same candidate feature vector **x** as described in Section 5.2. The automatic relevance determination algorithm is also used to infer the optimal values of the 16 hyperparameters $\{k_1, k_2, \ldots, k_{16}\}$ assigned for the 16 feature dimensions. The feature selection process of Test_a1 is shown in Figure 5.4 as an example. Those of the other 19 experiments are illustrated in Figures D1 to D19 in Appendix D. In these illustrations, the hyperparameters are also normalized with respect to the maximum values as defined in Equation 5.1.

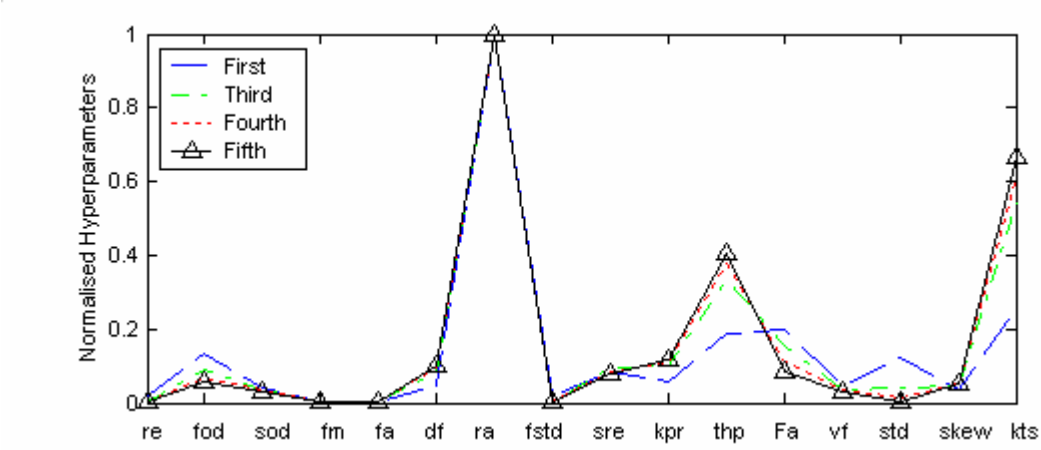


Figure 5.4. Illustration of the Feature Selection Processes of Test_a1 for TWR
Cutting conditions: spindle speed = 800rpm, feed rate = 150mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

The values of $k_1, k_2, \cdots, k_{16}$ at the last iteration of all the experiments are listed in Table 5.7. "1" is also used as the threshold. Those feature dimensions with the corresponding k_i's greater than 1 are selected as relevant features; otherwise, the features are rejected. The feature selection results are listed in Table 5.8, with the "√" marks representing the selected features.

It can be seen from Table 5.8 that besides the features chosen out for TWE, three more features turn out to be relevant to tool wear recognition, *thp*, *sre*, and *skew*. The selected feature set becomes {*fm, fa, ra, fstd, sre, thp, Fa, std, skew, kts*}, which spans a more complex feature space than that of TWE. The addition of the three relevant features makes the feature space more easily separable. Another point to note in the results is that there is not a single feature that stands out to be most relevant for tool condition recognition, unlike that in the case of TWE, where the average force is found to be most relevant. This is probably because no single feature among the candidate features is representative enough for reliable recognition of tool state. To make the feature space separable, multiple features have to work together.

Table 5.7 Hyperparameter Values at the Last Iteration for TWR

Feature	*Test_A1*	*Test_A2*	*Test_A3*	*Test_A4*	*Test_A5*	*Test_A6*	*Test_A7*	*Test_A8*	*Test_A9*	*Test_A10*
re	0.0000	0.2479	0.0258	0.0353	0.0000	0.1446	0.0120	0.0000	0.0000	0.0000
fod	0.1162	0.0249	0.0000	0.2366	0.1781	0.1342	0.0768	0.2221	0.1231	0.1410
sod	0.0654	0.1578	0.0326	0.0818	0.0525	0.1311	0.0514	0.0899	0.0217	0.0465
fm	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	7.8241
fa	0.0000	0.0000	0.0134	0.0000	1.2276	0.0000	0.4027	0.0000	0.0000	4.6805
df	0.2096	0.0950	0.1401	0.0931	0.1449	0.1925	0.0425	0.0802	0.0199	0.0581
ra	2.0530	3.3845	0.1858	6.2861	3.4543	9.7906	0.3993	5.0866	0.5338	46.7660
fstd	0.0000	0.0000	0.1060	3.3348	0.7533	0.0000	7.8723	0.0477	0.0000	0.0000
sre	0.1562	0.0000	0.0000	0.0006	0.0973	0.0196	0.0002	0.0205	0.1847	0.1071
kpr	0.2449	0.0732	0.0590	0.0595	0.1186	0.1015	0.0466	0.0707	0.0211	0.0699
thp	0.8351	0.0341	0.4521	0.0000	0.4484	1.0695	0.0000	1.7741	0.0815	2.3264
Fa	0.1723	0.4877	1.3042	0.7465	2.6544	1.3127	3.2243	0.6005	1.5639	3.7184
vf	0.0598	0.0626	0.0490	0.0896	0.0657	0.0832	0.0356	0.1137	0.0565	0.0752
std	0.0016	0.7829	0.0000	0.5916	3.4709	0.0000	0.1452	1.2748	0.0000	1.9404
skew	0.1101	0.3934	0.0241	0.0000	0.0000	0.0000	0.0000	0.0000	1.5876	0.2550
kts	1.3640	3.1751	0.0000	0.0000	0.0000	0.2779	0.0000	3.0119	0.0000	0.0000

Continued:

Feature	*Test_A11*	*Test_A12*	*Test_B1*	*Test_B2*	*Test_B3*	*Test_B4*	*Test_B5*	*Test_B6*	*Test_B7*	*Test_B8*
re	0.0176	0.2041	0.0000	0.2706	0.0000	0.0823	0.0000	0.1800	0.0000	0.0000
fod	0.0507	0.0818	0.0872	0.4298	0.0668	0.1118	0.3919	0.1407	0.0942	0.2033
sod	0.0263	0.1343	0.0382	0.4415	0.0274	0.1678	0.2210	0.2749	0.0698	0.0634
fm	0.0000	0.0000	1.7449	0.0000	0.0000	0.0000	9.8589	0.0000	0.0000	0.6970
fa	0.0000	0.6699	0.3162	0.0000	0.0000	0.6761	0.0006	0.0000	0.0000	2.4602
df	0.0252	0.3096	0.0000	0.3964	0.0026	0.1605	0.1110	0.1594	0.0188	0.0609
ra	101.2571	0.4799	1.6809	0.1980	1.2720	0.7564	4.3998	2.4523	0.0001	4.0414
fstd	0.0000	0.2614	0.0000	0.0000	2.5916	0.1744	0.2875	0.0000	1.6024	7.4351
sre	0.2333	0.9705	0.1981	27.7763	0.1435	2.9469	0.0205	1.7964	0.0478	0.0964
kpr	0.0331	0.2171	0.1819	0.3213	0.1111	0.2159	0.0491	0.1212	0.0124	0.0774
thp	0.0000	0.2289	0.2961	0.0819	0.0000	0.1119	0.5398	0.0000	3.2492	0.0000
Fa	15.5439	0.1676	0.9517	0.4316	6.7068	0.3767	1.1849	1.0308	0.6165	4.2867
vf	0.0877	0.1204	0.0368	0.4884	0.0262	0.1612	0.2088	0.3566	0.0392	0.0739
std	1.3626	4.9664	0.1880	0.0598	0.0000	0.4557	0.2795	0.3674	0.0000	0.0000
skew	0.0000	0.0286	0.0000	0.1622	1.1629	0.0519	0.0515	8.5447	0.0000	0.1369
kts	0.0000	0.5445	0.3297	0.5404	177.1931	0.6323	0.0000	4.9205	0.3759	0.2641

Table 5.8 Feature Selection Results for TWR

	re	*fod*	*sod*	*fm*	*fa*	*df*	*ra*	*fstd*	*sre*	*kpr*	*thp*	*Fa*	*vf*	*std*	*skew*	*kts*
Test_a1							√									√
Test_a2							√									√
Test_a3												√				
Test_a4							√	√								
Test_a5					√		√					√		√		
Test_a6							√				√	√				
Test_a7								√				√				
Test_a8							√				√			√		√
Test_a9												√			√	
Test_a10				√	√		√				√	√		√		
Test_a11							√					√		√		
Test_a12														√		
Test_b1				√			√									
Test_b2									√							
Test_b3							√	√				√			√	√
Test_b4									√							
Test_b5				√			√					√				
Test_b6							√		√			√			√	√
Test_b7								√			√					
Test_b8					√		√	√				√				

5.4 VERIFICATION OF THE RELEVANCE OF THE SELECTED FEATURE SET FOR TWR

In order to verify the relevance of the selected feature set to tool conditions, generalization tests are also performed. The training and testing data sets are listed in Table 5.9 and 5.10, respectively. The point where the tool is found worn in each data set is marked by the item “Average Wear at Tool Failure Point”, which is used to distinguish the failure stage from the fresh one. Note that these values vary from one data set to another. This is because each of these values is taken as the average of the wear value of every insert in the cutter when any insert is found worn; and the wear growth is usually uneven among different inserts.

Table 5.9 Training Data Sets for TWR

No.	Insert Type	Spindle Speed (rpm)	Feed Rate (mm/min)	Depth of Cut (mm)	Insert Number	Average Wear at Tool Failure Point (mm)
1	AC325	800	150	1	4	0.3718
2	AC325	1000	100	1	4	0.4102
3	AC325	1200	150	1	2	0.3621
4	AC325	600	100	2	4	0.3520
5	AC325	1000	100	1	4	0.2633
6	A30N	800	200	1	4	0.4200
7	A30N	800	300	1	4	0.3626
8	A30N	1000	300	1	4	0.3938
9	A30N	800	300	1	4	0.3382

Table 5.10 Testing Data Sets for TWR

No.	Insert Type	Spindle Speed (rpm)	Feed Rate (mm/min)	Depth of Cut (mm)	Insert Number	Average Wear at Tool Failure Point (mm)
T1	AC325	1000	100	1	2	0.4595
T2	AC325	1000	200	1	2	0.3960
T3	AC325	1000	300	1	4	0.3043
T4	AC325	1200	200	1	2	0.4238
T5	AC325	1200	300	1	4	0.3234
T6	AC325	600	200	2	4	0.2736
T7	AC325	800	100	2	2	0.2915
T8	A30N	1000	200	1	2	0.3803
T9	A30N	1000	300	2	4	0.3920
T10	A30N	1200	100	1	2	0.3780
T11	A30N	1200	200	1	4	0.4094

The 10 selected features are first used in training and testing. Then the entire 16 and the rejected 6 features are used to repeat the procedures. Figure 5.5 shows the results from the testing data set T1. Those obtained from the remaining 10 data sets are illustrated in Figures E1 to E10 in Appendix E. Note that the first alarms given by the classifiers are marked with a vertical line jumping from the bottom to the top.

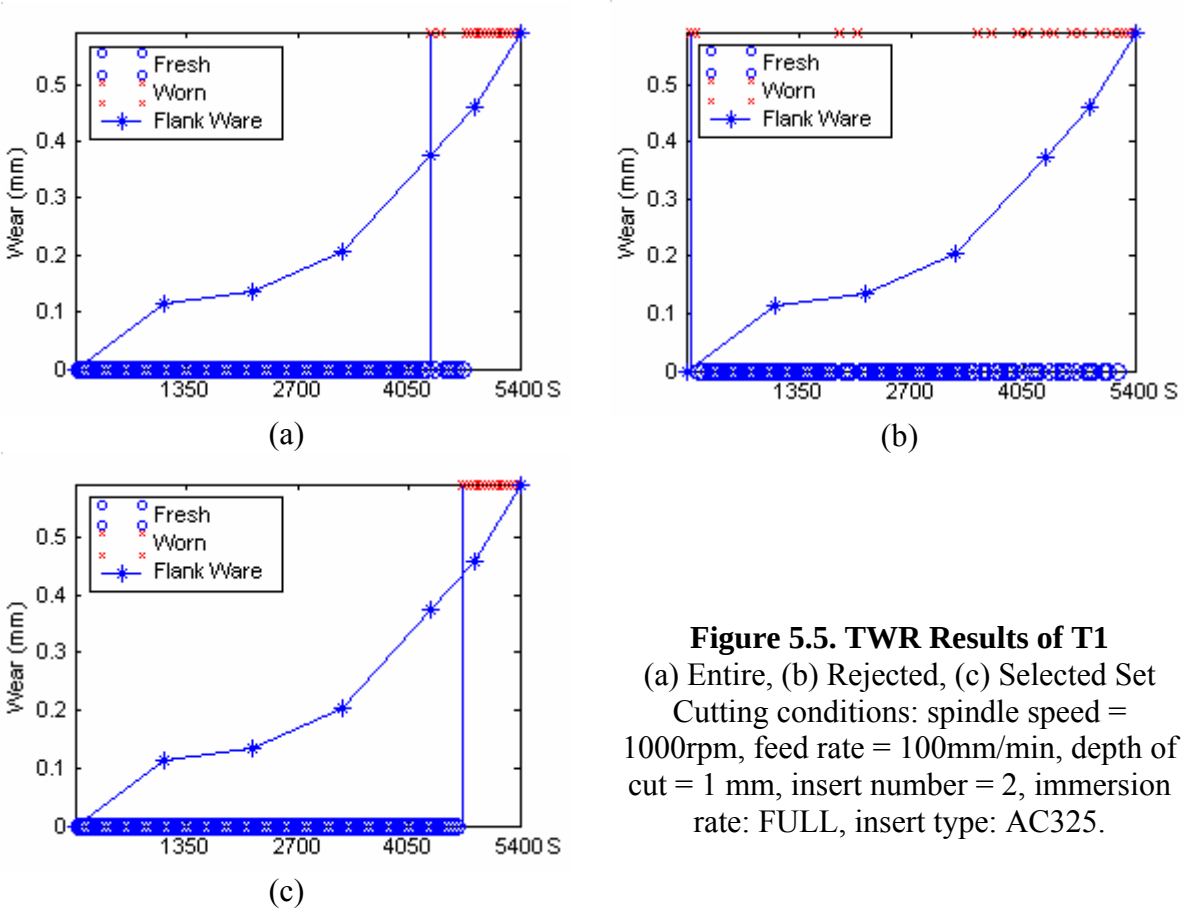


Figure 5.5. TWR Results of T1
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

It can be clearly seen from these figures that the classification results using the rejected feature set are quite noisy for all of the tests, with alarms given out even at the early stage of the tools. The results using the selected and the entire feature set are much better.

The generalization capabilities of the entire, selected, and rejected feature sets are compared in terms of the actual wear at the first alarm, and the successful classification rate. The actual wear at the first alarm is defined as the actual wear value when the classifier gives the first alarm. The successful classification rate is the ratio between the number of the successfully classified points and the total number of points. The comparisons are listed in Table 5.11.

Table 5.11 Tool wear recognition Results

Test No.	Successful Classification Rate			Actual Wear at the First Alarm			Average Wear at Tool Failure Point (mm)
	All Features	Rejected Features	Selected Features	All Features (mm)	Rejected Features (mm)	Selected Features (mm)	
T1	96.00%	83.00%	97.00%	0.3745	0.0056	0.4340	0.4595
T2	97.22%	88.89%	97.22%	0.3675	0.0205	0.3675	0.3960
T3	94.94%	67.09%	94.94%	0.3297	0.0300	0.3297	0.3043
T4	95.71%	70.00%	100.00%	0.3926	0.0125	0.4238	0.4238
T5	95.38%	72.31%	98.46%	0.3682	0.0105	0.3383	0.3234
T6	96.00%	67.00%	98.00%	0.1760	0.0941	0.2830	0.2736
T7	97.00%	33.65%	97.00%	0.2022	0.0132	0.2579	0.2915
T8	97.06%	72.06%	97.06%	0.3718	0.0213	0.3718	0.3803
T9	98.75%	88.75%	97.50%	0.4011	0.2006	0.4103	0.3920
T10	99.44%	87.78%	95.56%	0.3810	0.0029	0.3810	0.3780
T11	90.86%	41.71%	96.57%	0.2361	0.0116	0.4301	0.4094

By their definitions, the successful classification rate indicates the stability of the classifier (frequently varying classification results are regarded as unstable); while the actual wear at the first alarm represents the accuracy (the closeness of the detected failure point to the actual one). It is obvious that both the stability and the accuracy of the classifier using the selected feature set is much better than that using the rejected feature set. For most of the cases (except T9 and T10), the stability of the classifier using the selected feature set is not worse than that using the entire feature set. The reason why the successful rates using the entire feature set are comparable to those using the selected feature set is that the hyperparameters associated with the rejected feature dimensions are automatically reduced to negligible values in training. And therefore those features don't have much influence on the classification result.

From practical viewpoint, the identification error can be viewed as the absolute difference between the actual wear at the first alarm and the average wear at the tool failure point, which indicates the capability of the classifier in detecting tool failures in

time. The comparisons of the identification errors can be visualized in Figure 5.6. It can be clearly seen that the accuracy of the classifier using the selected feature set is the best, because the first alarms are given most punctually.

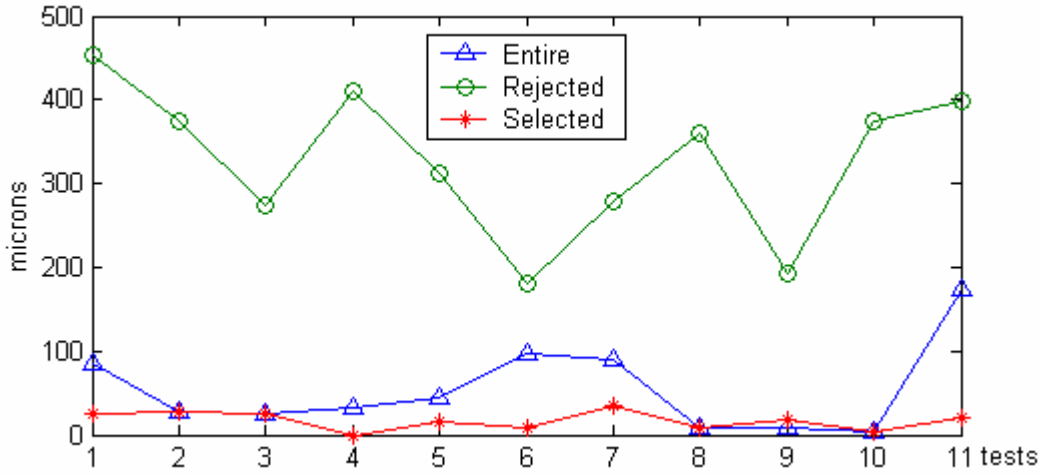


Figure 5.6 Comparisons of the Classification Errors

5.5 SUMMARY OF THE RESULTS

7 features {*fm, fa, ra, fstd, Fa, std, kts*} are identified to be relevant to tool wear estimation. And three more features, *thp*, *Fa*, and *skew*, are added to the selected feature set for tool wear recognition. The comparisons of the generalization capabilities using the entire, selected, and rejected feature sets indicate the successful identification of the relevant features in both the regression and the classification.

In addition, the performance of the estimator and the classifier designed using the Bayesian framework is good. With the selected feature set, an accuracy of more than 90% is achieved for tool wear estimation. On the other hand, a stable and accurate classifier is implemented for tool wear recognition, with a stability of more than 95% and an accuracy of recognizing a failed tool in the close vicinity of the actual failure point (identification error less than 35 microns).

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

In response to the latest development in using artificial intelligence approaches in tool condition monitoring, this work is devoted to study the effectiveness of various features for two main tasks in TCM, tool wear estimation and tool wear recognition. This is because the proper selection of features is a vital issue in using neural networks, which has also been widely acknowledged in the field of TCM [Leem, 1995].

Force signal is used, due to its high sensitivity to tool wear and low noise. Cutting experiments have been conducted under various conditions. The influence of the radial run-out on the cutting force and force features is examined through analysis of the force signals. It has been found that when the run-out is greater than the feed per tooth, the force pulses within one rotation are quite different from each other even at the fresh stage of the tool. Theoretically, this phenomenon should occur only when the tool is highly worn or some of the teeth are broken. In order to avoid the influence of the force fluctuating behavior caused by the run-out, all the features are extracted from the force signals within one rotation instead of one tooth period, so that the run-out can be eliminated as a “common mode noise” when comparing the force features between adjacent tool rotations.

16 different feature extraction methodologies are introduced and discussed, which have all been shown to be effective for tool condition monitoring. These include time-series analysis, statistical analysis, and signal processing approaches. According to the intrinsic relationships among these features, the feature extraction is performed through a five-step procedure. The 16 features form the scope for the feature selection in this study.

The feature selection is realized through the automatic relevance determination (ARD) approach, which by itself does not make sense and has to be implemented in specific neural networks. In this study, the ARD approach is implemented in Bayesian support vector machines, which is the combination of the Bayesian probability theory and the classic support vector machines.

To select features for tool wear estimation, the Bayesian support vector regression algorithm is used. The average force within one tool rotation has been proven to be the most relevant feature for tool wear estimation, because of its good correlation to tool wear processes. It forms the relevant feature set together with 6 other selected features, including amplitude ratio, standard deviation, maximum force level, kurtosis, total amplitude of cutting force, and standard deviation of the force components in tool breakage zone. The generalization capabilities of the entire, selected, and rejected feature sets are tested and compared. The results using the selected features turn out to be the best, proving that they are relatively more relevant to tool wear processes.

To select features for tool wear recognition, the Bayesian support vector classification algorithm is used. Besides the features chosen out for TWE, three more features turn out to be relevant to TWR. They are skew, total harmonic power, and

total amplitude of cutting force. The performance of the entire, selected, and rejected features is compared in terms of the successful classification rates and the identification errors. Once again, the performance of the selected features turns out to be the best, proving that they are more relevant to tool wear.

Besides the successful identification of the relevant features, good generalization capabilities have also been achieved for both TWE and TWR. An accurate tool wear estimator has been implemented using the Bayesian support vector regression algorithm, with an accuracy of more than 90%. And by using the Bayesian support vector classification algorithm, a stable and accurate classifier has been achieved for tool wear recognition, with a stability of more than 95% and an identification error less than 35 microns.

6.2 FUTURE WORK

According to Liang [2002], the future development of tool condition monitoring technology is likely to take on the following paths:

- Embedding sensors into the machine tool structure. For example, force sensing elements may be directly mounted in the tool holder to monitor flank wear.
- Miniaturizing system components. For example, a MEMS thermometer may be positioned right next to the cutter tip to measure the temperature at the tool-workpiece interface directly.
- Telecommunication-based and wireless process monitoring. This technology may enable the remote monitoring of machining processes.

These are the long-term goals for TCM. There are also some challenges to be faced in the nearest future, for example, the realization of robust TCM systems for industrial

application and the integration of multiple monitoring systems for comprehensive machining process monitoring.

Although considerable research has been done in the area of tool condition monitoring, industrial realization and commercial availability of TCM systems are still quite limited at present. Take milling process monitoring as an example. Most of the current systems are not robust enough to deal with various milling processes in industrial environment. One limitation is that they can only deal with some regular shapes of workpiece. Cubes are the most popular shape used in theoretical studies, because of the uniform cutting geometry. However, when it comes to shapes as complex as shown in Figure 6.1, most of the systems designed for cubic workpiece may not function properly. The problem lies in the feature extraction methodologies. For example, all the 16 features considered in this study are based on the waveform of the force signal. However, force waveforms can vary greatly from rotation to rotation just due to the change in cutting geometry, which will make the features too noisy to indicate tool states. Based on this consideration, alternative feature extraction methodologies insensitive to cutting geometry or alternative sensing techniques have

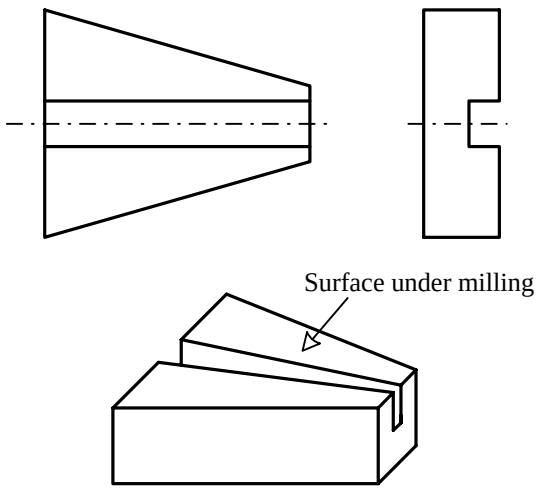


Figure 6.1 Illustration of a Complex Shape

to be developed in future work. A possible solution, for example, is to use computer vision systems, which can directly measure the wear volume rather than infer it from force signals. However, reliable vision systems, which can be fit into practical cutting environments, have to be developed in future work.

Besides tool condition monitoring, there are some other types of monitoring systems, e.g. chatter detection systems, which are also very important to machining processes. According to Du [1989], chatter is one of the most destructive malfunctions in machining processes. If uncontrolled, it can result in poor surface finish, premature wear, chipping and breakage of the cutting tool. So it could be interesting and necessary to combine the individual monitoring systems together. Approaches to chatter detection are commonly to investigate the spectral density of a process signal, such as cutting force [Du, 1989]. Therefore, it is possible to combine tool condition monitoring and chatter detection systems together by just using one dynamometer. Integration of the two systems could take the following form:

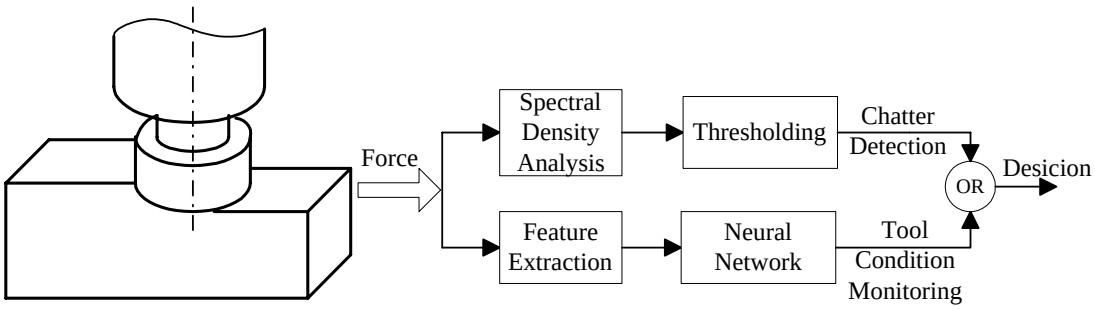


Figure 6.2 Combination of TCM and Chatter Detection

Not limited to this, the future machining process monitoring systems should be the integration of multiple subsystems for the monitoring of tool condition, chatter onset, part dimensions, surface roughness, spindle bearing failure, etc. Such systems are called supervisory systems, which are drawing more and more attention from the researchers in the field [Teltz, 1993; Landers, 1998].

REFERENCES:

[1] **Abu-Zahra, N.H.** and **Nayfeh, T.H.** Calibrated Method for Ultrasonic On-line Monitoring of Gradual Wear during Turing Operations, Int. J. Mach. Tools Manufact., Vol. 37, No. 10, pp. 1475-1484. 1997.

[2] **Altintas, Y.** In-Process Detection of Tool Breakage Using Time Series Monitoring of Cutting Forces, Int. J. Mach. Tools Manufact., Vol. 28, No. 2, pp. 157-172. 1988.

[3] **Altintas, Y.** and **Yellowley, I.** In-Process Detection of Tool Failure in Milling Using Cutting Force Models, J. Engng Ind., Vol. 111, pp. 149-157. 1989.

[4] **Altintas, Y.** Prediction of Cutting Forces and Tool Breakage in Milling from Feed Drive Current Measurement, ASME J. Eng. Ind., Vol. 114, pp. 386-392. 1992.

[5] **Buntine W. L.,** and **Weigend A. S.** Bayesian back-propagation, Complex Systems, Vol. 5, No. 6, pp. 603–643. 1991.

[6] **Byrne, G.** et al. Tool Condition Monitoring (TCM) – The Status of Research and Industrial Application, Ann. CIRP, Vol. 44, No.2, pp. 541-567. 1995.

[7] **Choi, D., Kwon, W. T.,** and **Chu, C. N.** Real-time Monitoring of Tool Fracture in Turning Using Sensor Fusion, Adv. Manuf. Eng., Vol. 15, No. 5, pp. 305-310. 1999.

[8] **Chu W., Keerthi S. S.,** and **Ong C. J.** A unified loss function in Bayesian framework for support vector regression, Proceeding of the 18th International Conference on Machine Learning, pp. 51–58. 2001.

[9] **Chu W., Keerthi S. S.,** and **Ong C. J.** Bayesian Trigonometric support vector classifier, to appear in Neural Computation. 2003.

[10] **Coker, S. A.** and **Shin, Y. C.** In Process Control of Surface Roughness due to Tool Wear Using Ultrasonic System, Int. J. Mach. Tools Manufact., Vol. 36, No. 3, pp. 411-422. 1996.

[11] **Dimla, D. E.**, **Lister, P. M.**, and **Leighton, N. J.** Neural Network Solution to the Tool Condition Monitoring Problem in Metal Cutting – A Critical Review of Methods, Int. J. Mach. Tools Manufact., Vol. 37, No. 9, pp. 1219-1241. 1997.

[12] **Du, R. X., Elbestawi, M. A.,** and **Ullagaddi, B. C.** Chatter Detection in Milling Based on the Probability Distribution of Cutting Force Signal, Mechanical Systems and Signal Processing, Vol. 6, No. 4, pp. 345-362. 1992.

[13] **Elanayar S.** and **Shin Y. C.**, Robust Tool Wear Estimation with Radial Basis Function Neural Networks, Journal of Dynamic Systems, Measurement, and Control, Vol. 117, pp. 459-467Dec. 1995.

[14] **Elbestawi M. A.**, **Marks J.** and **Papazafiriou T. A.** Process Monitoring in Milling by Pattern Recognition, Mechanical Systems and Signal Processing, Vol. 3, No. 3, pp. 305-315. 1989.

[15] **Elbestawi, M. A.**, **Papazafiriou, T. A.** and **Du, R. X.** In-process Monitoring of Tool Wear in Milling Using Cutting Force Signature, Int. J. Mach. Tools Manufact., Vol. 31, No. 1, pp. 55-73. 1991.

[16] **Elwardany, T. I., Gao, D.** and **Elbestawi, M. A.** Tool Condition Monitoring in Drilling Using Vibration Signal Analysis, Int. J. Mach. Tools Manufact., Vol. 36, No. 6, pp. 687-711. 1996.

[17] **Emel, E.**, and **Kannatey-Asibu, E.** Tool Failure Monitoring in Turning by Pattern Recognition Analysis of AE Signals, J. Engng Ind., Vol. 110, pp. 137-145. 1988.

[18] **Fu, H. J., DeVor R. E.,** and **Kapoor S. G.** A Mechanistic Model for the Prediction of the Force System in Face Milling Operations, J. Engng Ind., Vol. 106, pp. 81-88. 1984.

[19] **Goebol K.** and **Yan W.** Feature Selection for Tool Wear Diagnosis Using Soft Computing Techniques, Proceedings of IMECE2000. 2000.

[20] **Haykin, S.** Neural Networks-a comprehensive foundation. Prentice Hall, New Jersey, 1999. Proceedings of IMECE2000. 2000.

[21] **Huseyin, M. Errtunc, Kenneth, A. Looparo** and **Hasan, Ocak.** Tool Wear Condition Monitoring in Drilling Operations Using Hidden Marko Modes (HMMs), Int. J. Mach. Tools Manufact., Vol. 41, pp. 1363-1384. 2001.

[22] International Standard (ISO 8688-1, first edition): Tool Life Testing in Milling – part 1: Face milling. pp.14. 1989.

[23] **Kim, J. D.** and **Choi, I. H.** Development of a Tool Failure Detection System Using Multi-Sensors, Int. J. Mach. Tools Manufact., Vol. 36, No. 8, pp. 861-870. 1996.

[24] **Koren, Y.** Flank Wear Model of Cutting Tools Using Control Theory, J. of Eng. Ind., Vol. 100, pp. 103-109. 1978.

[25] **Kowk, J. T.** The Evidence Framework Applied to Support Vector Machines, Neural Networks, Vol. 14, pp. 257-274. 2001.

[26] **Kramer, B. M.** A Comprehensive Tool Wear Model, Ann. CIRP, Vol. 35, No.1, pp. 67-70. 1986.

[27] **Kurada, S.** and **Bradley, C.** A Machine Vision System for Tool Wear Assessment, Tribology International, Vol. 30, No. 4, pp. 295-304. 1997.

[28] **Lan, M. S.** and **Naerheim, Y.** In-Process Detection of Tool Breakage in Milling, J. of Eng. Ind., Vol. 108, pp. 191-197. 1986.

[29] **Landers, R. G.** and **Ulsoy, A. G.** Supervisory Machining control: Design approach and Experiments, Ann. CIRP, Vol. 47, No. 1, pp. 301-306. 1998.

[30] **Lange, k.** et al. Tool Life and Tool Quality in Bulk Metal Forming, Ann. CIRP, Vol. 41, No. 2, pp. 557-675. 1992.

[31] **Law, M. H.** and **Kowk, J. T.** Bayesian Support Vector Regression, Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics (AISTATS), PP. 239-244. 2001.

[32] **Lee, M. Y., Thomas, C. E.** and **Wildes, G.** Review – prospects for in-process diagnosis of metal cutting by monitoring vibration signals, J. Mat. Sci., Vol. 22, pp. 3821-3890. 1987.

[33] **Leem C. S., Dornfeld D. A.,** and **Dreyfus S. E.** A Customized Neural Network for Sensor Fusion in On-Line Monitoring of Cutting Tool Wear, J. Eng. Ind., Vol. 117, pp. 152-159. 1995.

[34] **Li, X., Dong, S.,** and **Venuvinod, P. K.** Hybrid Learning for Tool Wear Monitoring, Int. J. Adv. Manuf. Technol., Vol. 16, pp. 303-307. 2000.

[35] **Liang, S. Y.** and **Perry S. A.** In-Process Compensation for Milling Cutter Runout via Chip Load Manipulation, J. Engng Ind., Vol. 116, pp. 153-159. 1994.

[36] **Liang, S. Y, Hecker, R. L.,** and **Landers R. G.** Machining Process Monitoring and Control: the State-of-the-Art, Proceedings of IMECE2002. 2002.

[37] **Lin S. C.** and **Lin R. J.** Tool Wear Monitoring in Face Milling Using Force Signals, Wear, Vol. 198, pp. 136-142. 1996.

[38] **MacKay D. J. C.,** A practical Bayesian framework for back propagation networks, Neural Computation, Vol. 4, No. 3, pp. 448–472. 1992.

[39] **Mackay David J. C.** Bayesian Methods for Neural Networks: Theory and Application, Lecture Notes for Neural Network Summer School. 1995.

[40] **Monostori, L.** A Step towards Intelligent Manufacturing: Modelling and Monitoring of Manufacturing Processes through Artificial Neural Networks, Ann. CIRP, Vol. 42, pp. 485. 1993.

[41] **Neal, R. M.** "Bayesian Learning for Neural Networks", Lecture Notes in Statistics, Springer, 1996.

[42] **Niu, Y. M.**, **Wong, Y. S.**, **Hong, G. S.** and **Liu, T. I.** Multi-category Classification of Tool Conditions Using Wavelet Packets and ART2 Network, J. of Manufacturing Science and Engineering, Vol. 120, pp. 807-815. 1998.

[43] PCI-1200 User Manual: Multifunctional I / O Board for PCI Bus Computers, National Instrument Corporation, 1998.

[44] **Purushothaman, S.** and **Srinivasa, Y. G.** A Back-Propagation Algorithm Applied to Tool Wear Monitoring, Int. J. Mach. Tools Manufact., Vol. 34, No. 5, pp. 625-631. 1994.

[45] **Rabinowicz, E.** Abrasive Wear Resistance as a Materials Test, Lubrication Engineering, Vol. 33, No. 7, pp.378. 1977.

[46] **Rangwala, S.** and **Dornfeld, D.** Integration of Sensors via Neural Networks for Detection of Tool Wear States, Winter Annual Meetings of the ASME, PED, Vol. 25, pp. 109-120. 1987.

[47] **Ryabov, O., Mori, K., Kasashima, N.,** and **Uehara, K.** An In-Process Direct Monitoring Method for Milling Tool Failures Using A Laser Sensor, Ann. CIRP, Vol. 45, No. 1, pp. 97-100. 1996.

[48] **Sampath, A.** and **Vajpayee, S.** Tool Health Monitoring Using Acoustic Emission, Int. J. Prod. Res., Vol. 25, No. 5, pp. 703-719. 1987.

[49] **Santanu Das, Chattopadhyay, A. B.,** and **Murthy, A. S. R.** Force Parameters for On-line Tool Wear Estimation: A Neural Network Approach. Neural Networks, Vol. 9, No. 9, pp. 1639-1645. 1996.

[50] **Sastry S., Kapoor S. G.,** and **DeVor R. E.** Compensation of Progressive Radial Run-out in Face-Milling by Spindle Speed Variation, Int. J. Mach. Tools Manufact., Vol. 40, pp. 1121-1139. 2002.

[51] **Seeger, M.** Bayesian Model Selection for Support Vector Machines, Gaussian Processes and Other Kernel Classifiers, Advances in Neural Information Processing Systems, Vol. 12. 1999.

[52] **Silva, R. G., Reuben, R. L., Baker, K. L.** and **Wilcox, S. J.** Tool Wear Monitoring of Turning Operations by Neural Network and Expert System Classification of a Feature Set Generated from Multiple Sensors, Mechanical System and Signal Processing, Vol. 12, No. 2, pp. 319-332. 1998.

[53] **Sun J., Hong G. S., Rahman M.** and **Wong Y. S.** Feature Extraction and Selection in Tool Condition Monitoring System, Lecture Notes in Computer Science, Volume 2557, Springer. 2002.

[54] **Tansel, I. N.**, **Mekdeci, C.**, **Rodriguez, O.**, and **Uragun B.** Monitoring Drill Conditions With Wavelet Based Encoding and Neural Networks, Int. J. Mach. Tools Manufact., Vol. 33, No. 4, pp. 559-575. 1992.

[55] **Tansel, I.N.** and **McLaughlin, C.** Detection of Tool Breakage in Milling Operations – I. The Time Series Analysis Approach, Int. J. Mach. Tools Manufact., Vol. 33, No. 4, pp. 531-544. 1993a.

[56] **Tansel I. N.** and **McLaughlin C.** Detection of Tool Breakage in Milling Operations – II. the Neural Network approach, Int. J. Mach. Tools Manufact., Vol. 33, No. 4 pp. 545-558. 1993b.

[57] **Tansel, N. L.** et al. Detection of Tool Failure in End Milling with Wavelet Transformation and Neural Networks, Int. J. Mach. Tools Manufact., Vol. 35, No. 8, pp. 1137-1147. 1995.

[58] **Tarn J. H.** and **Tomizuka M.** On-Line Monitoring of Tool and Cutting Conditions in Milling, J. Eng. Ind., Vol. 111, pp. 206-212. 1989.

[59] **Tarng Y. S.** Study of Milling Cutting Force Pulsation Applied to the Detection of Tool Breakage, Int. J. Mach. Tools Manufact., Vol. 30, No. 4, pp. 651-660. 1990.

[60] **Tarng, Y. S., Hseih, Y. W.** and **Hwang, S. T.** Sensing tool breakage in face milling with a neural network, Int. J. Mach. Tools Manufact., Vol. 34, No. 3, pp. 341-350. 1994.

[61] **Taylor, F. W.** On the Art of Cutting Metals, Trans. ASME, 28, pp. 310–350. 1906.

[62] **Teltz, R.** and **Elbestawi, M. A.** Hierarchical Knowledge-Based Control in Turning, Journal of Dynamic Systems, Measurement, and Control, Vol. 115, No. 2A, pp. 122-132. 1993.

[63] **Tipping M. E.** The relevance vector machine, Neural Information Processing Systems, Vol. 12, pp. 652–658. MIT Press. 1999.

[64] **Vapnik** The Nature of Statistical Learning Theory, New York: Springer-Verlag, 1995.

[65] **Wilcox, S. J., Reuben, R. L.** and **Souquet, P.** The Use of Cutting Force and Acoustic Emission Signals for the Monitoring of Tool Insert Geometry During Rough Face Milling, Int. J. Mach. Tools Manufact., Vol. 37, No. 4, pp. 481-494. 1997.

[66] **Yamazaki, K., Yamada, A., Sawai, S.** and **Takeyama, H.** Ann. CIRP, Vol. 23, pp. 153-154. 1974.

[67] **Yan, D., El-Wardany, T. I.,** and **Elbstawi, M. A.** A Multi-Sensor Strategy for Tool Failure Detection in Milling, Int. J. Mach. Tools Manufact., Vol. 35, No. 3, pp. 383-398. 1995.

[68] **Yao, Y. L.,** and **Fang, X.** Assessment of Chip Forming Patterns with Tool Wear Progression in Machining via Neural Networks, Int. J. Mach. Tools Manufact., Vol. 33, No. 1, pp. 89-102. 1993.

[69] **Zhang D. Y.**, **Han Y. T.**, and **Chen D. C.** On-line detection of tool breakages using telemetering of cutting forces in milling, Int. J. Mach. Tools Manufact., Vol. 35, No. 1, pp. 19-27. 1995.

[70] **Zheng H. Q., Li, X. P., Nee A. Y. C.,** and **Wong Y. S.** Theoretical Modelling and Simulation of Cutting Forces in Face Milling with Cutter Runout, Int. J. Mach. Tools Manufact., Vol. 39, pp. 2003-2018. 1999.

Appendix A
Illustration of Cutting Force, Tool Wear, and Features

(a)

(b)

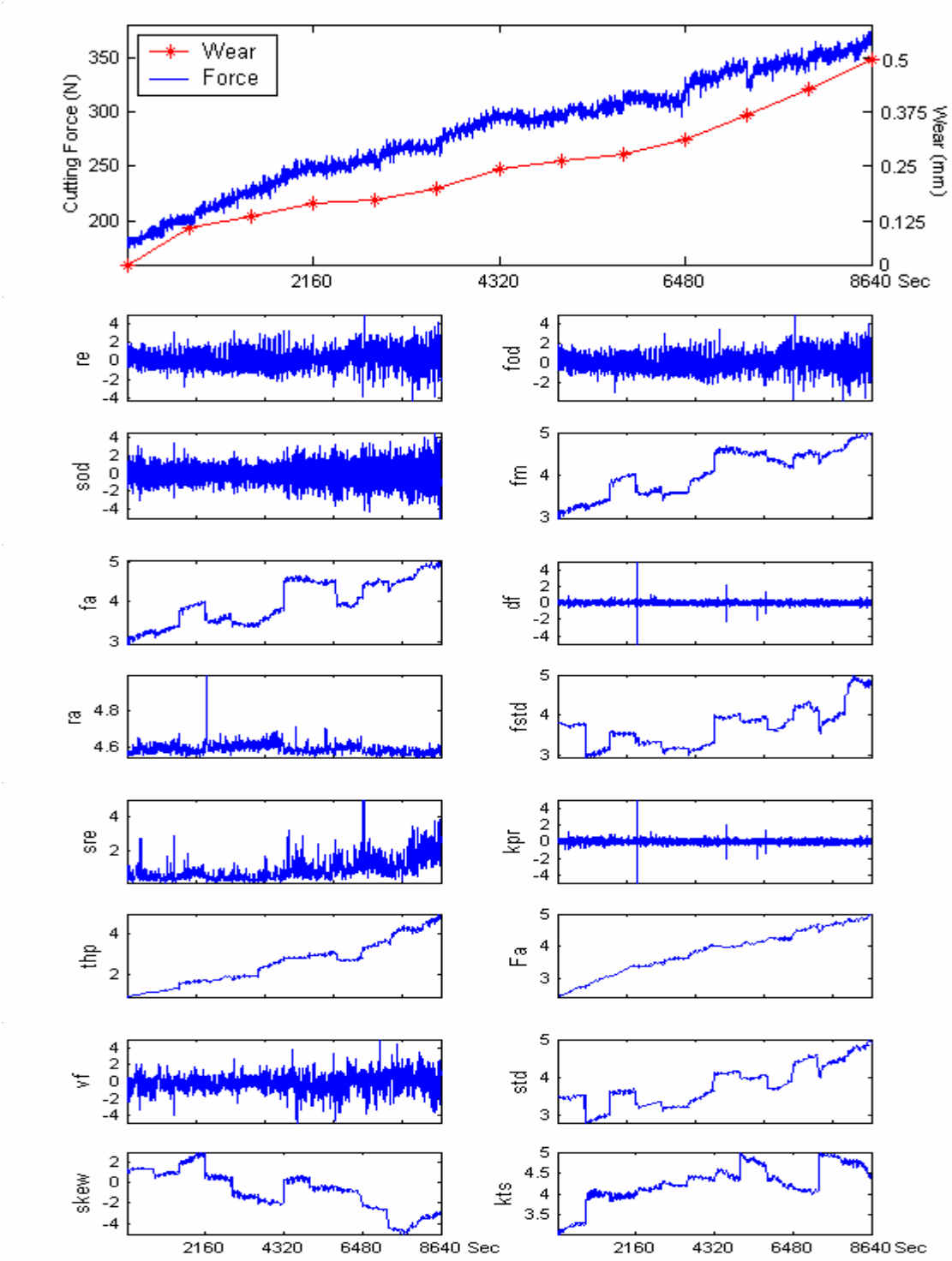


Figure A1. Illustration of the Cutting Force, Tool Wear, and Features of Test_a1
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 800rpm, feed rate = 150mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

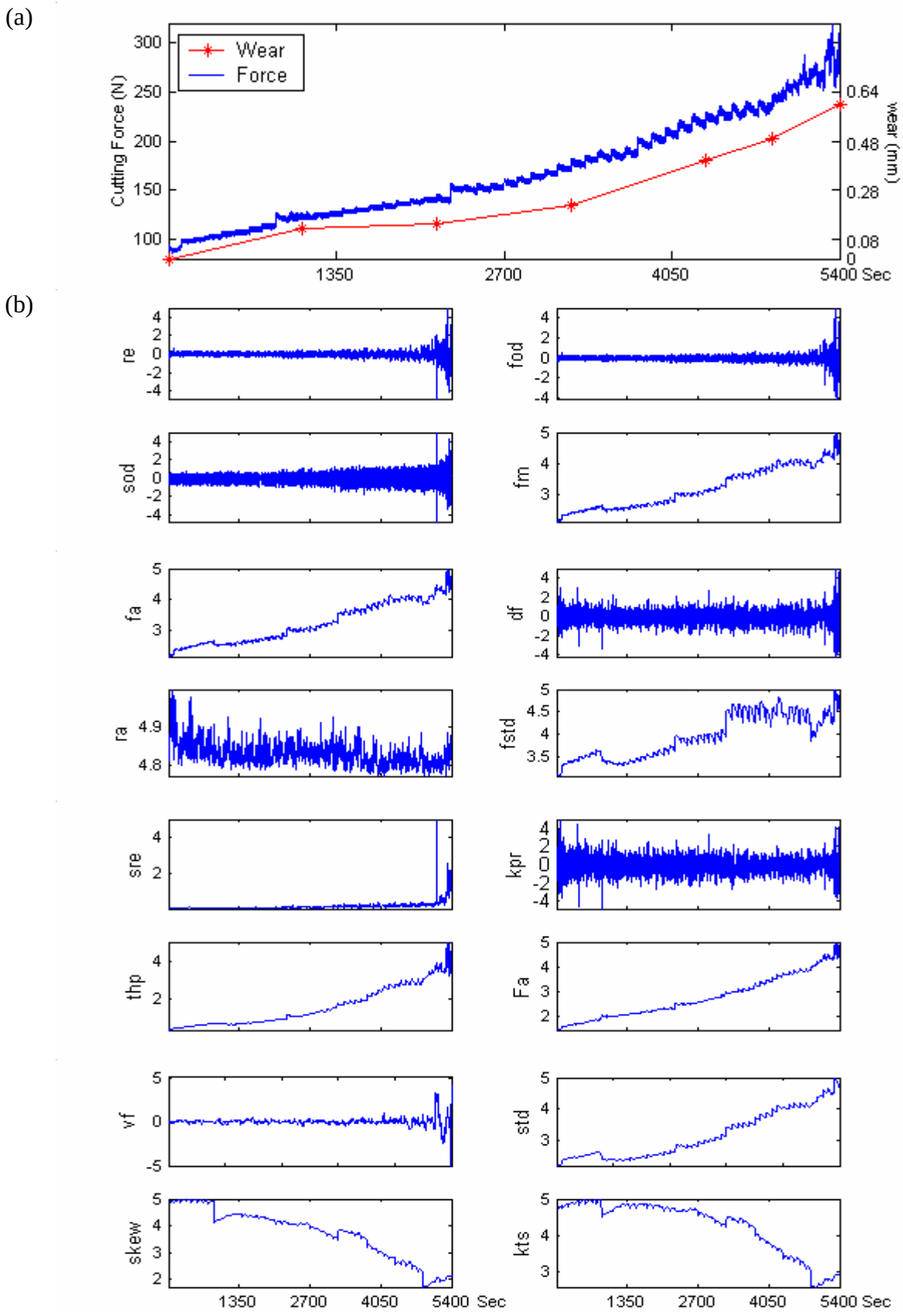


Figure A2. Illustration of the Cutting Force, Tool Wear, and Features of Test_a2
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

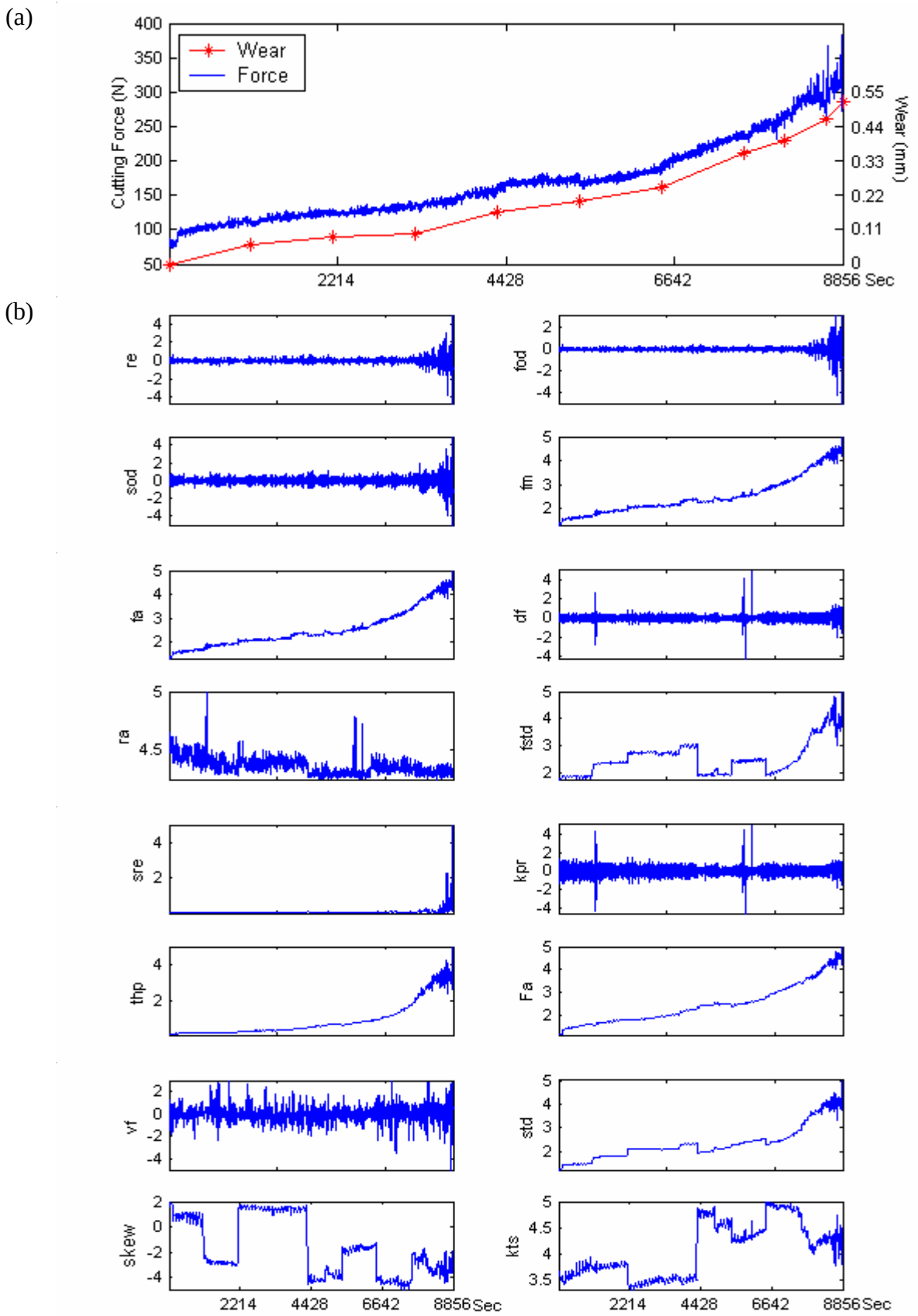


Figure A3. Illustration of the Cutting Force, Tool Wear, and Features of Test_a3
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

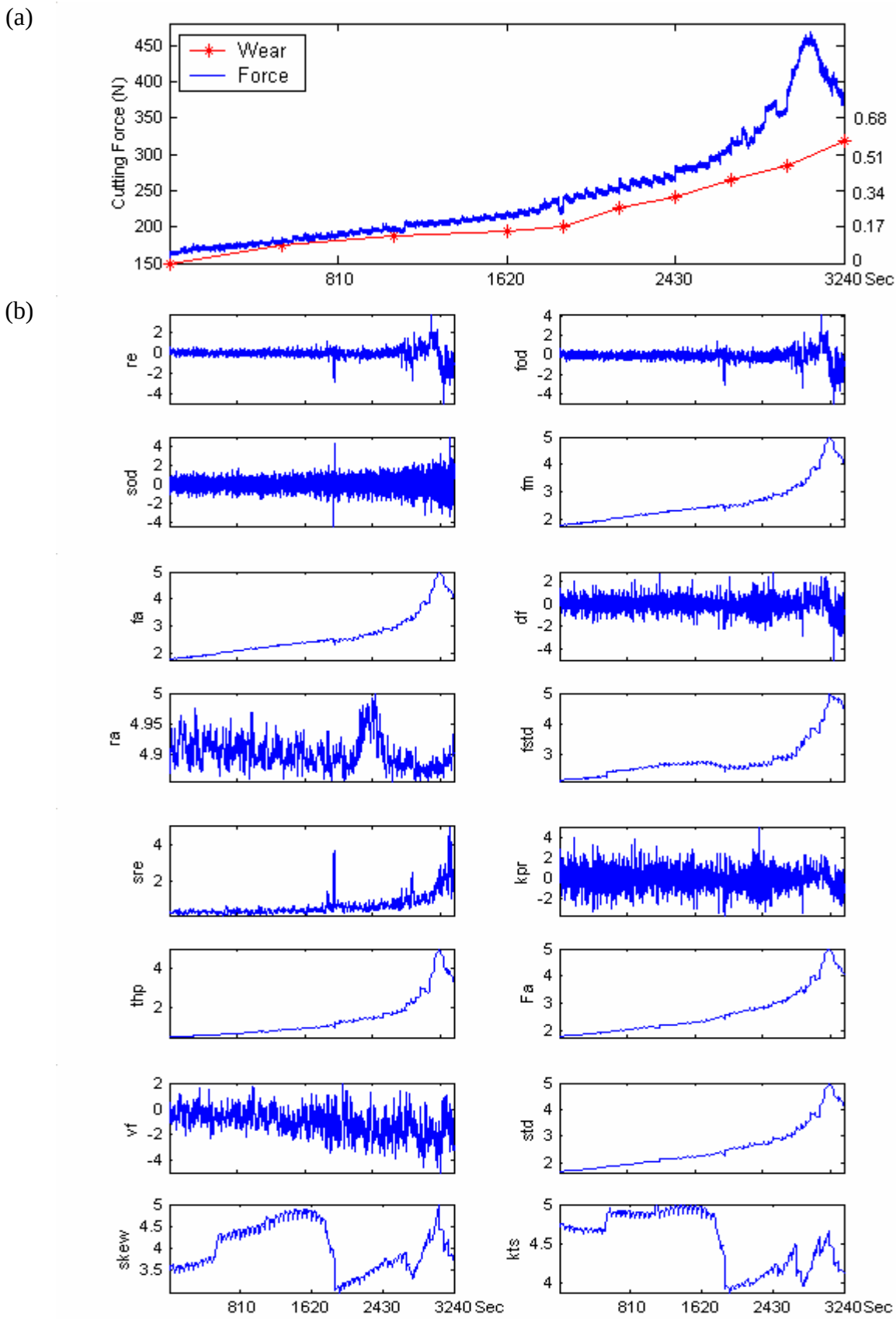


Figure A4. Illustration of the Cutting Force, Tool Wear, and Features of Test_a4
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

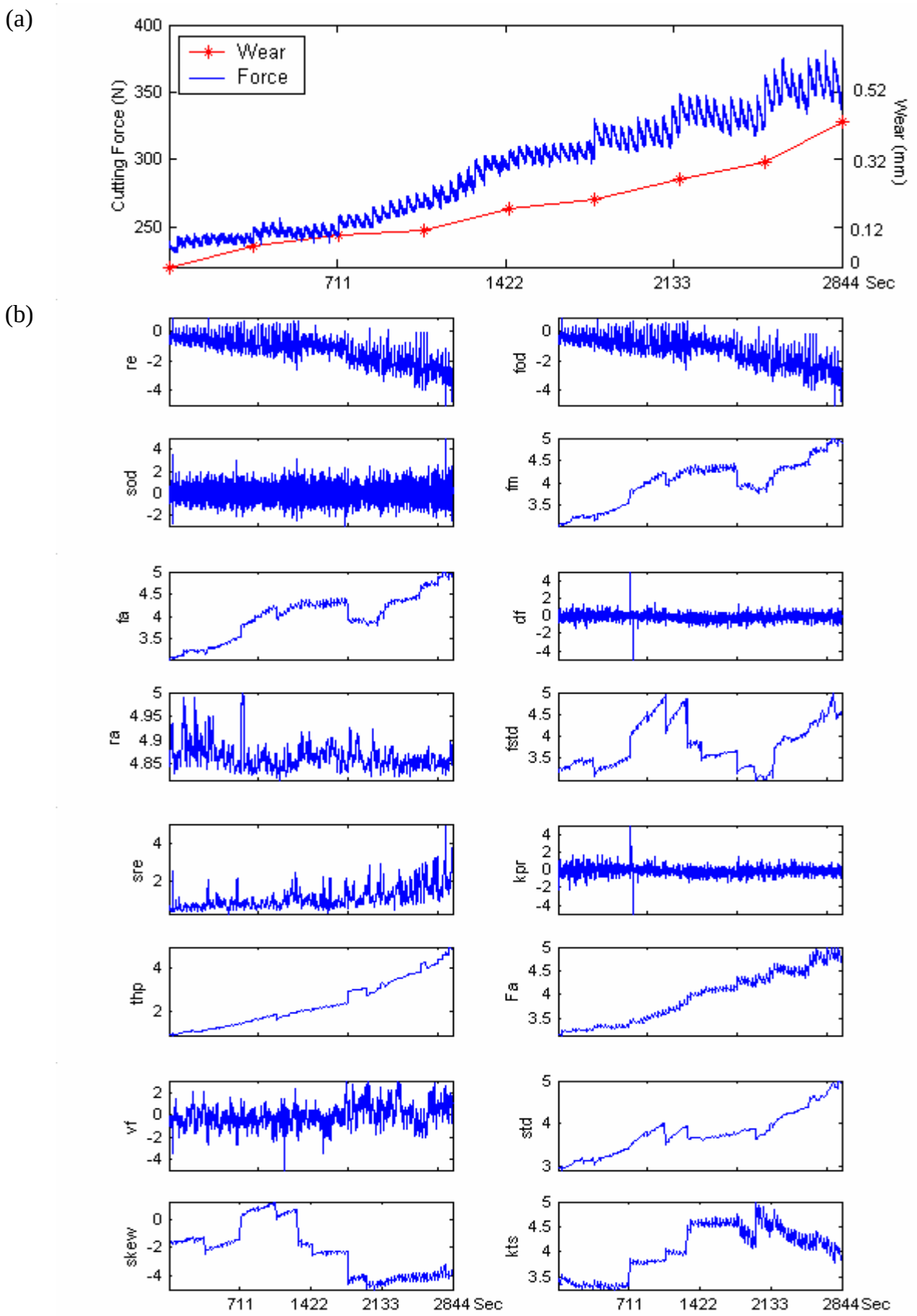


Figure A5. Illustration of the Cutting Force, Tool Wear, and Features of Test_a5
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

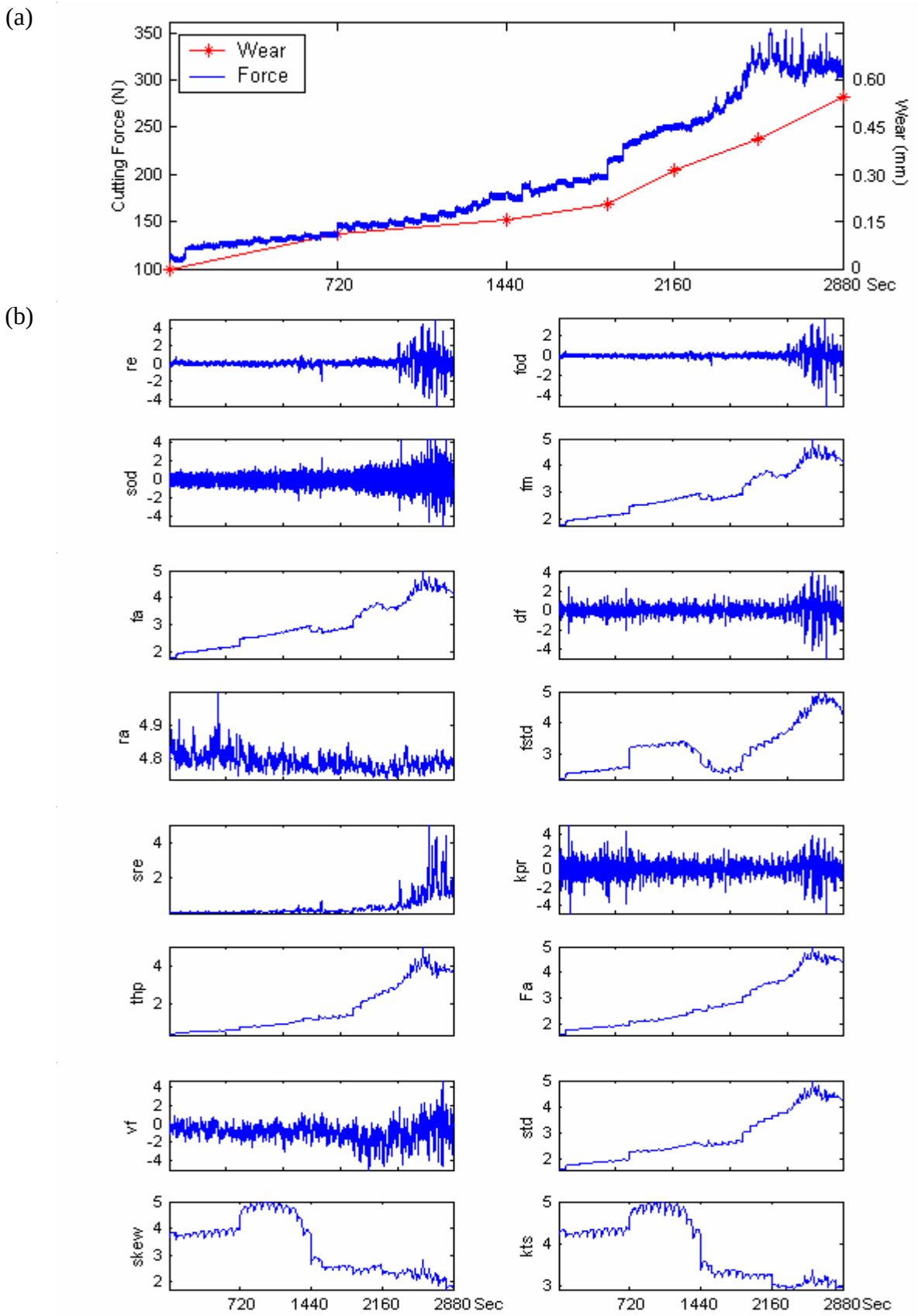


Figure A6. Illustration of the Cutting Force, Tool Wear, and Features of Test_a6
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1200rpm, feed rate = 150mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

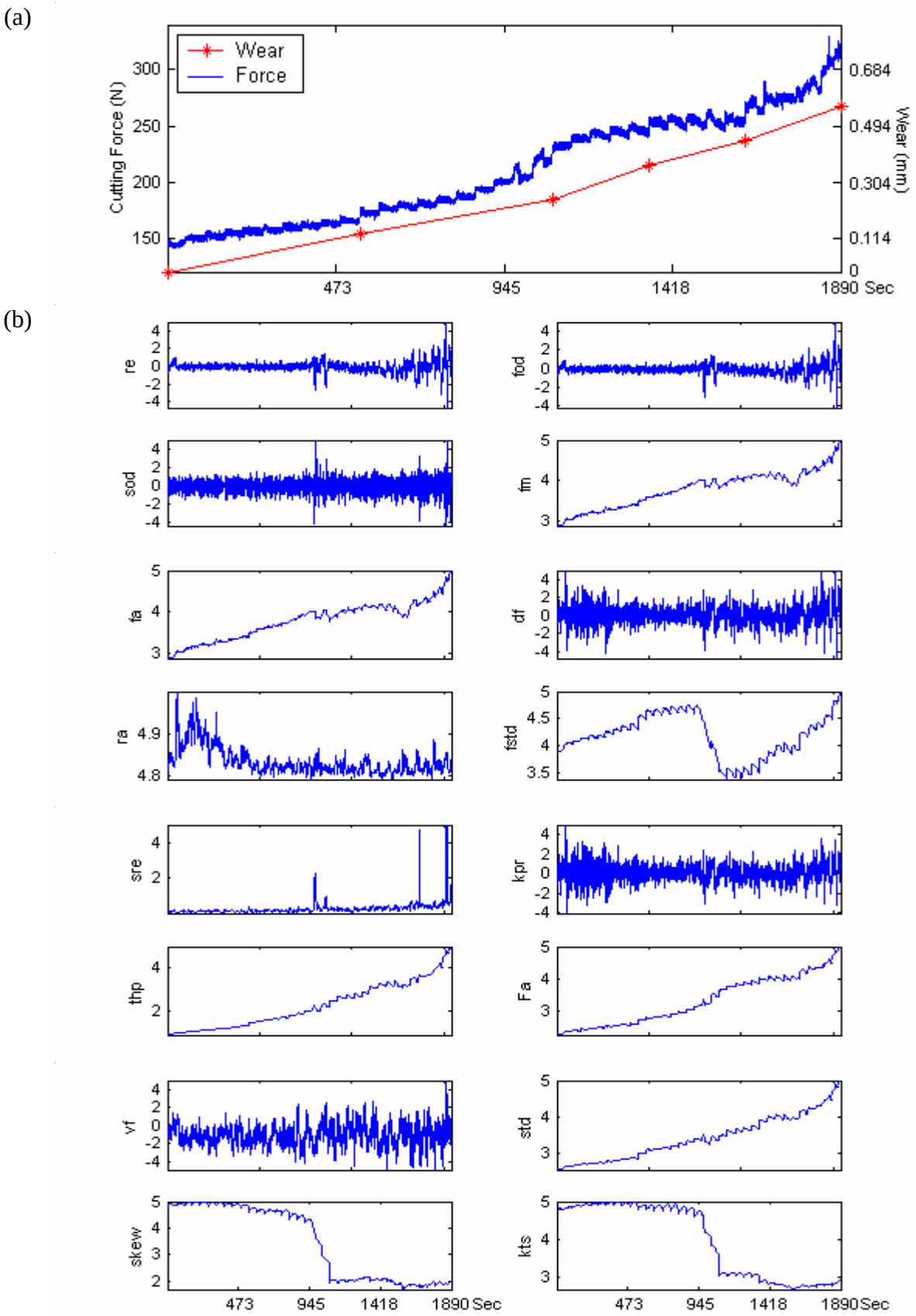


Figure A7. Illustration of the Cutting Force, Tool Wear, and Features of Test_a7
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

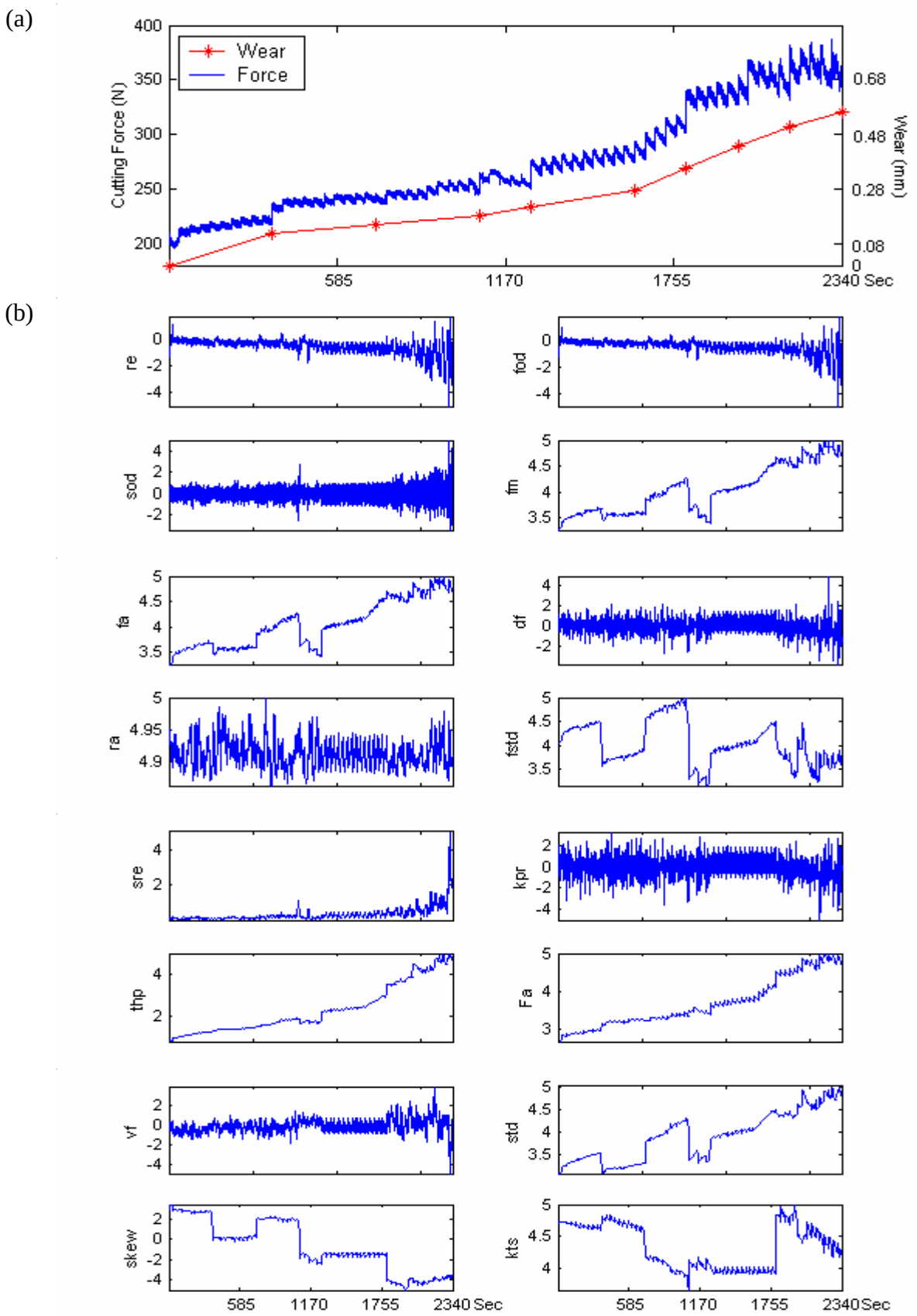


Figure A8. Illustration of the Cutting Force, Tool Wear, and Features of Test_a8
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1200rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

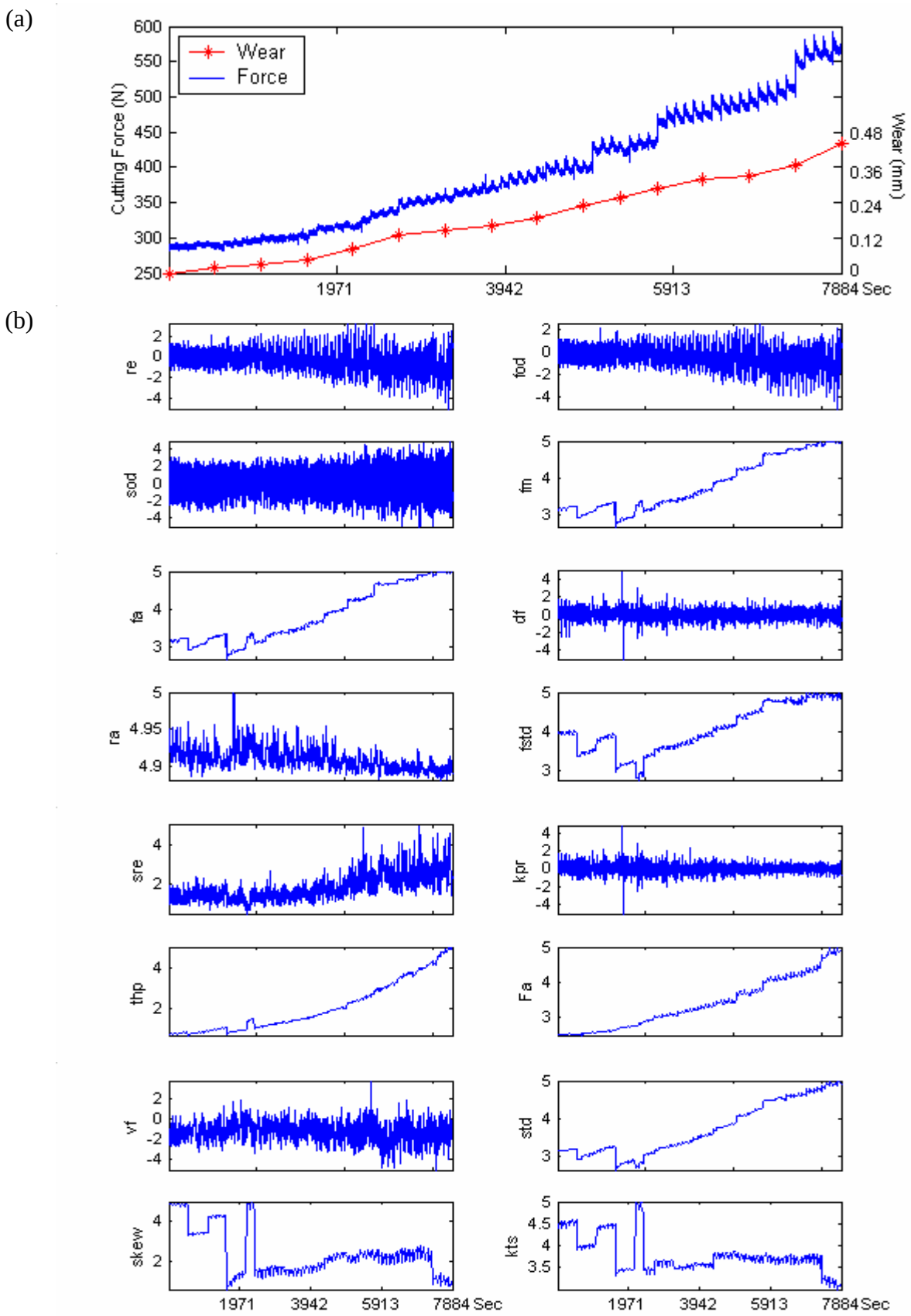


Figure A9. Illustration of the Cutting Force, Tool Wear, and Features of Test_a9
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 600rpm, feed rate = 100mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

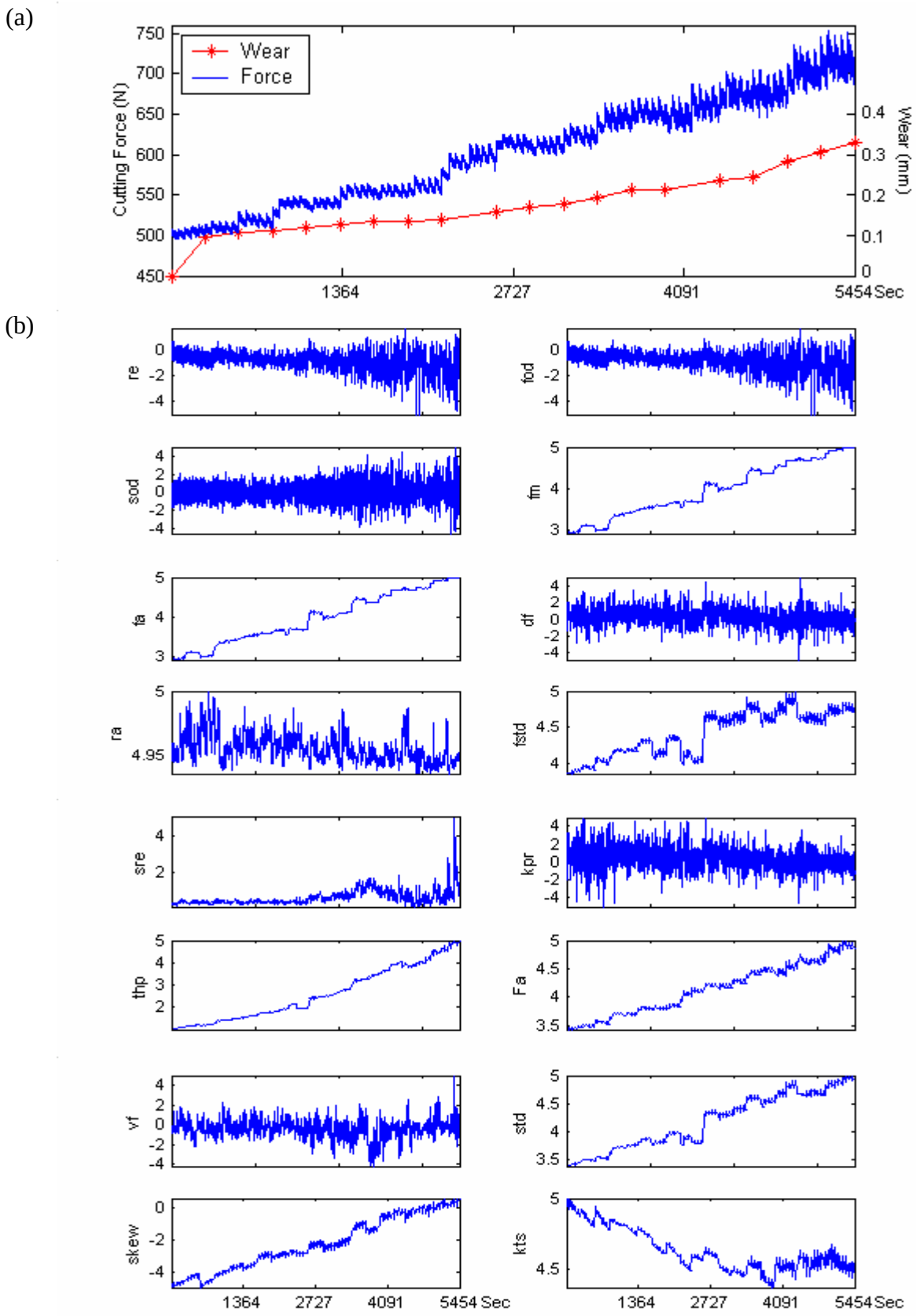


Figure A10. Illustration of Cutting Force, Tool Wear, and Features of Test_a10
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 600rpm, feed rate = 200mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

(a)

(b)

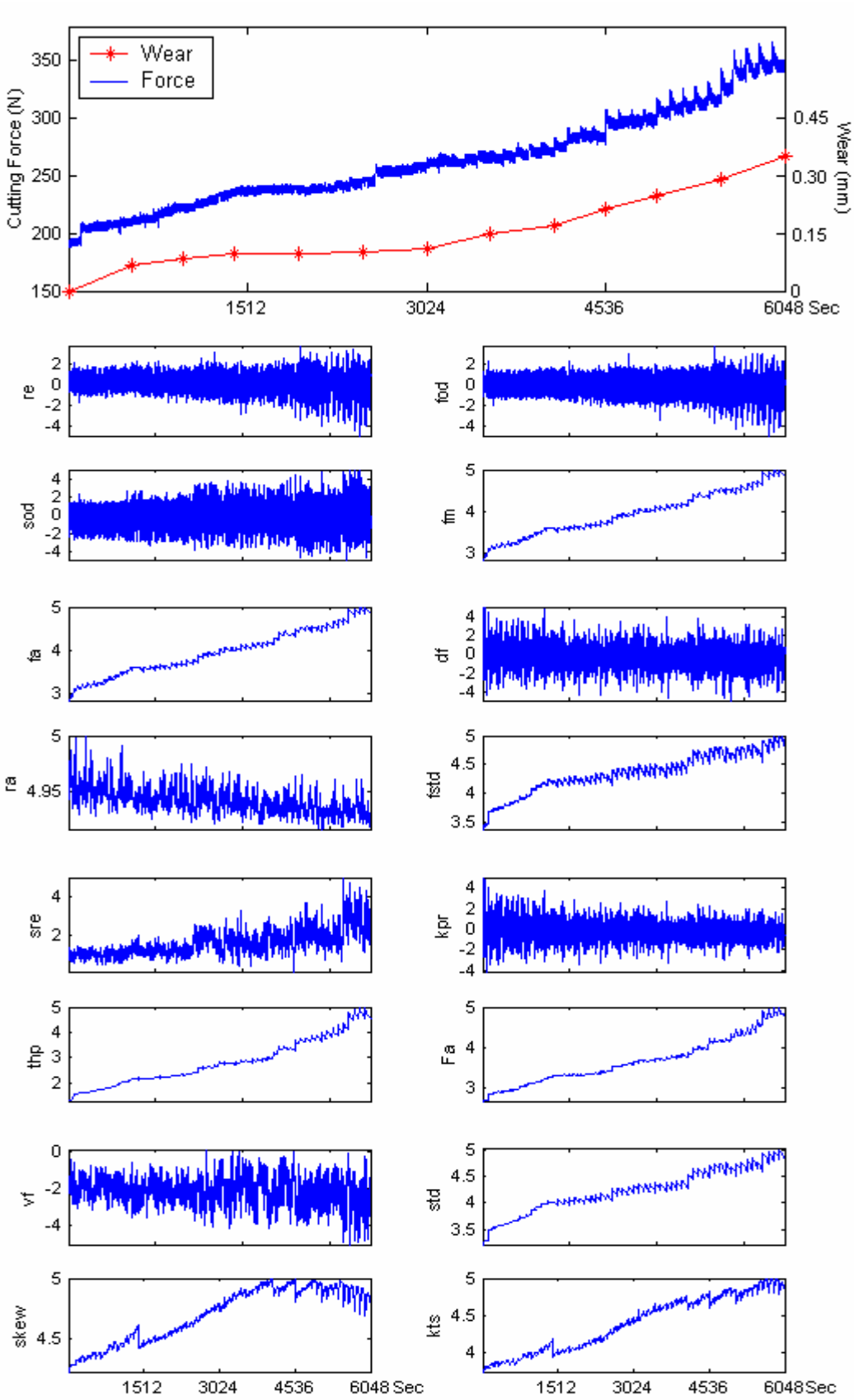


Figure A11. Illustration of Cutting Force, Tool Wear, and Features of Test_a11
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 800rpm, feed rate = 100mm/min, depth of cut = 2 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

(a)

(b)

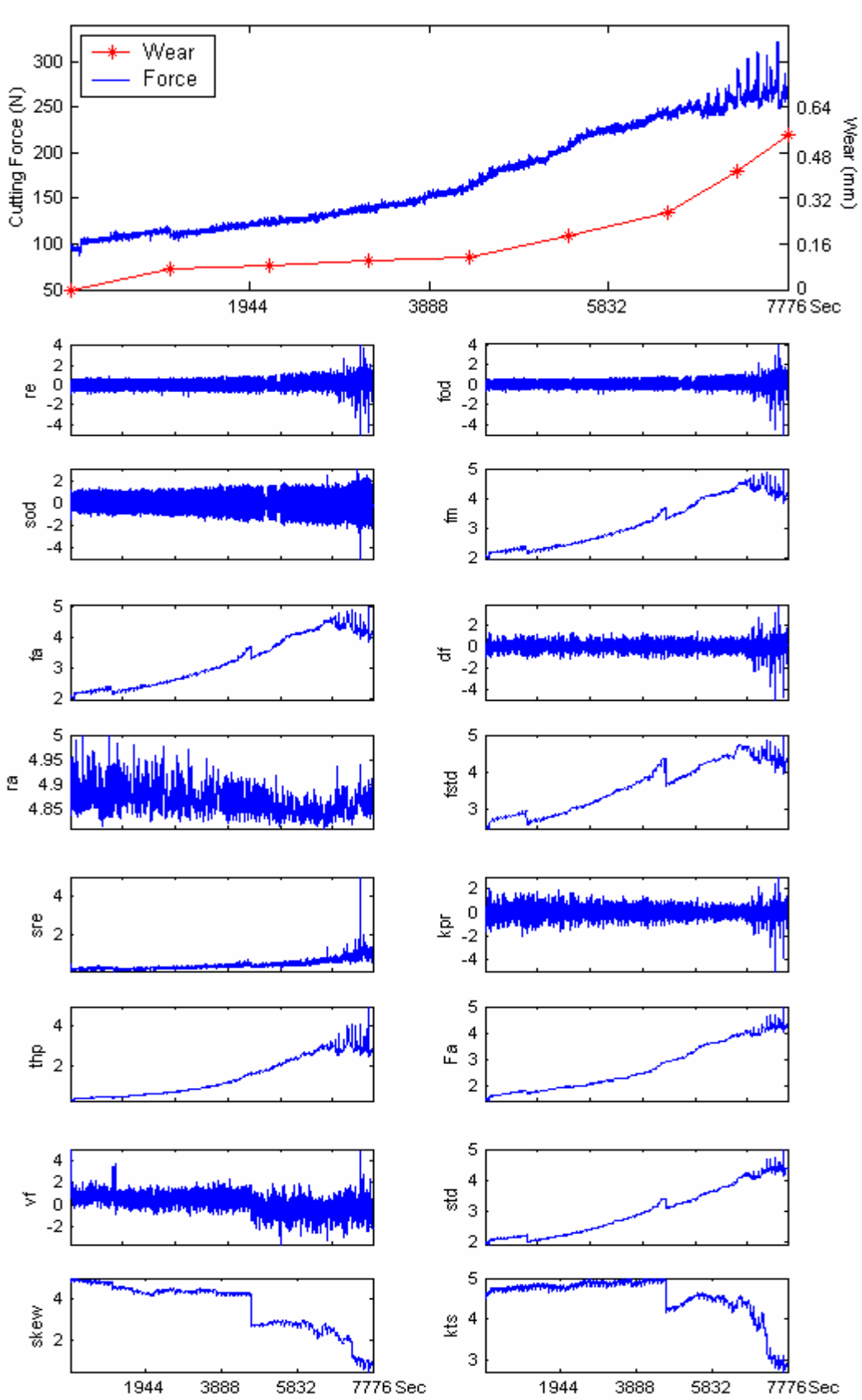


Figure A12. Illustration of Cutting Force, Tool Wear, and Features of Test_a12
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

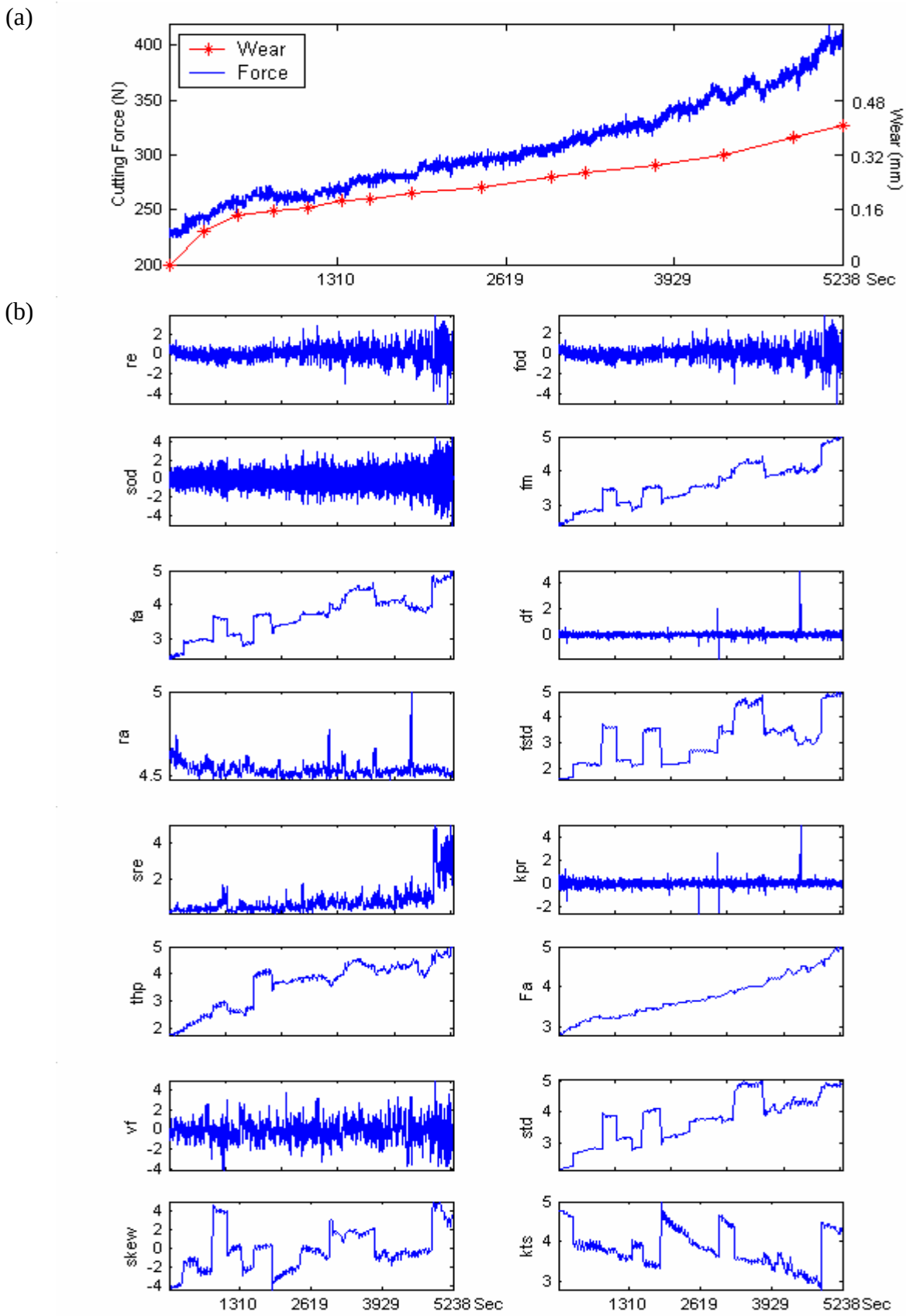


Figure A13. Illustration of Cutting Force, Tool Wear, and Features of Test_b1
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 800rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

(a)

(b)

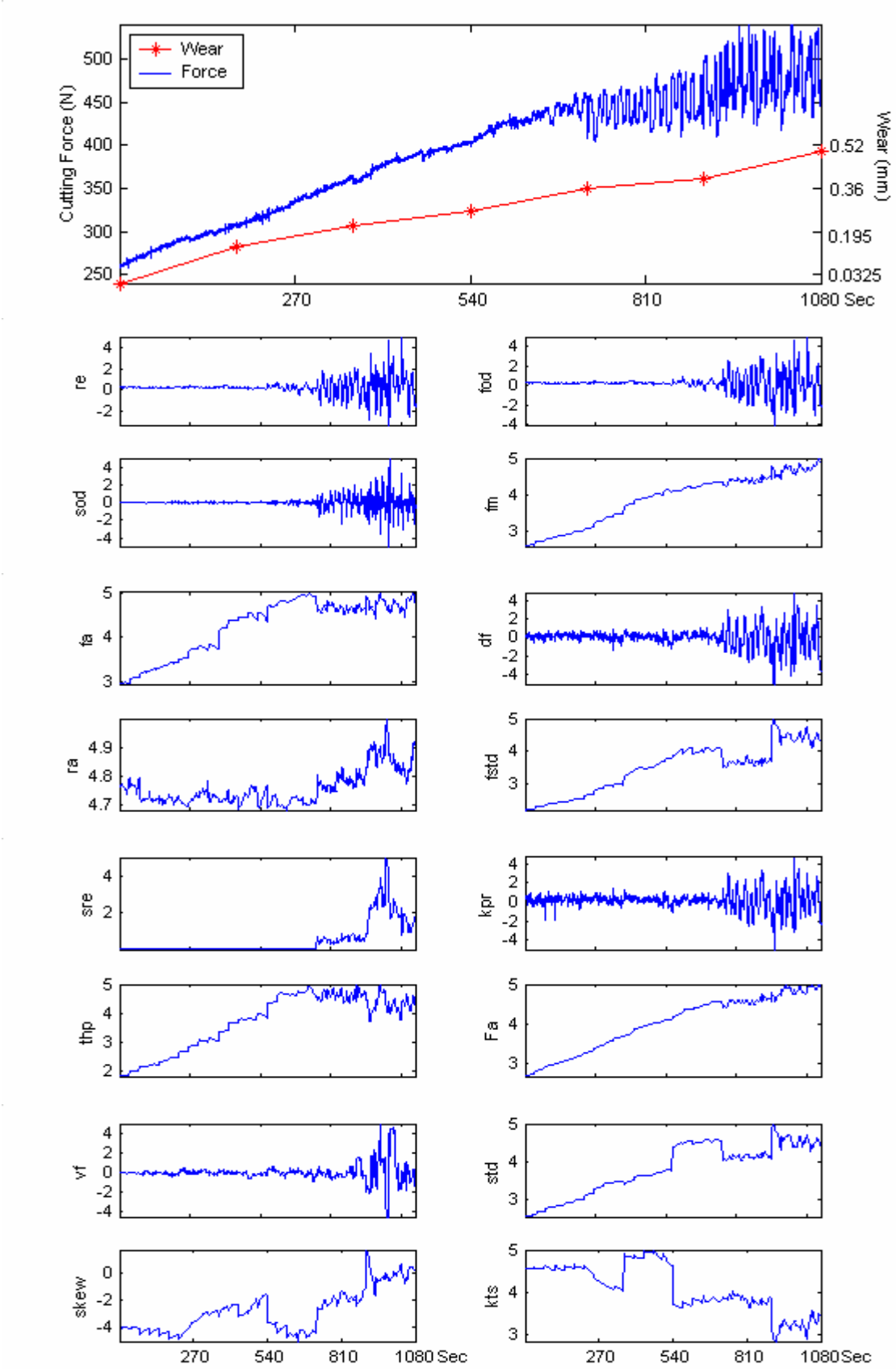


Figure A14. Illustration of Cutting Force, Tool Wear, and Features of Test_b2
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 800rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

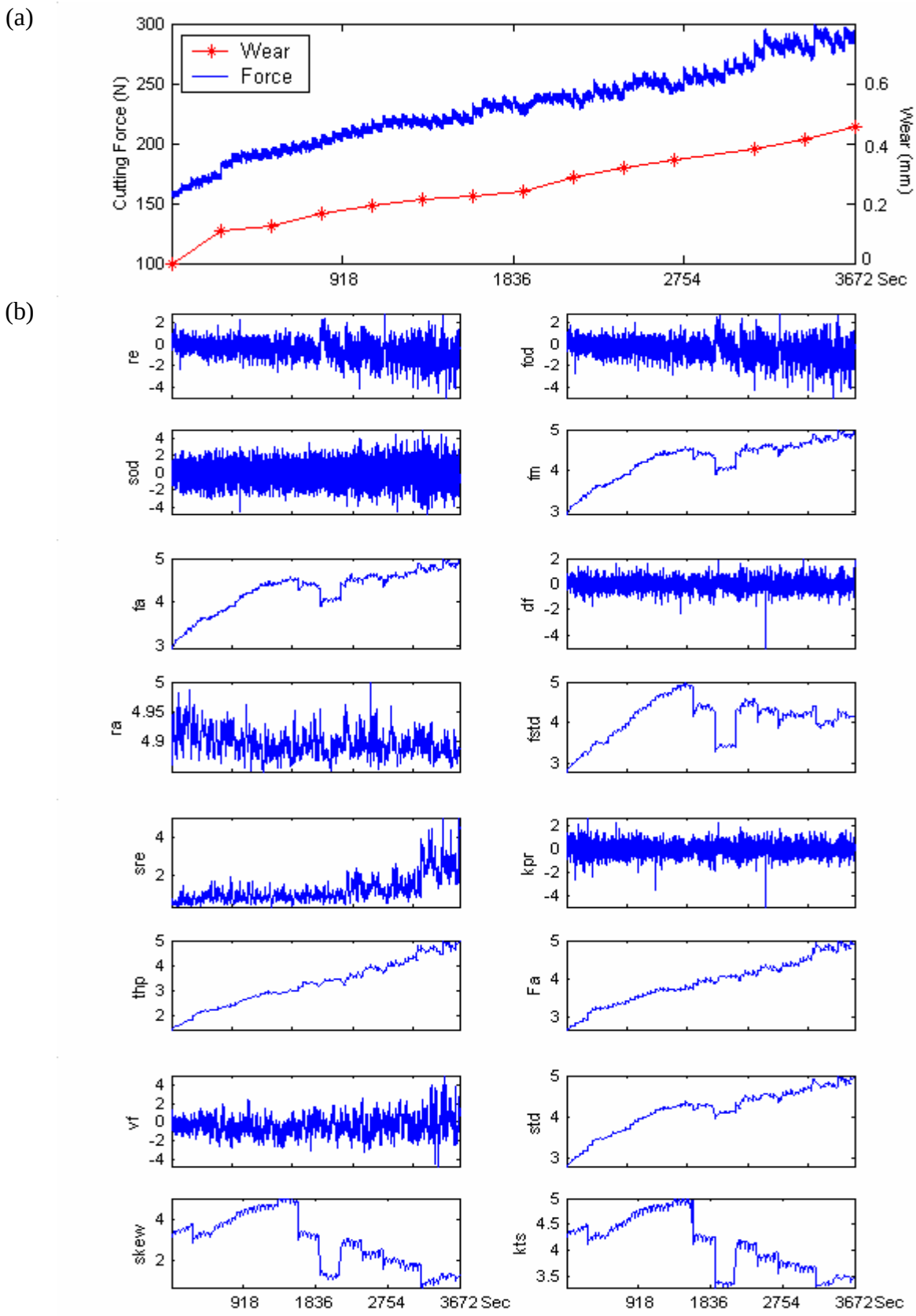


Figure A15. Illustration of Cutting Force, Tool Wear, and Features of Test_b3
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: A30N.

(a)

(b)

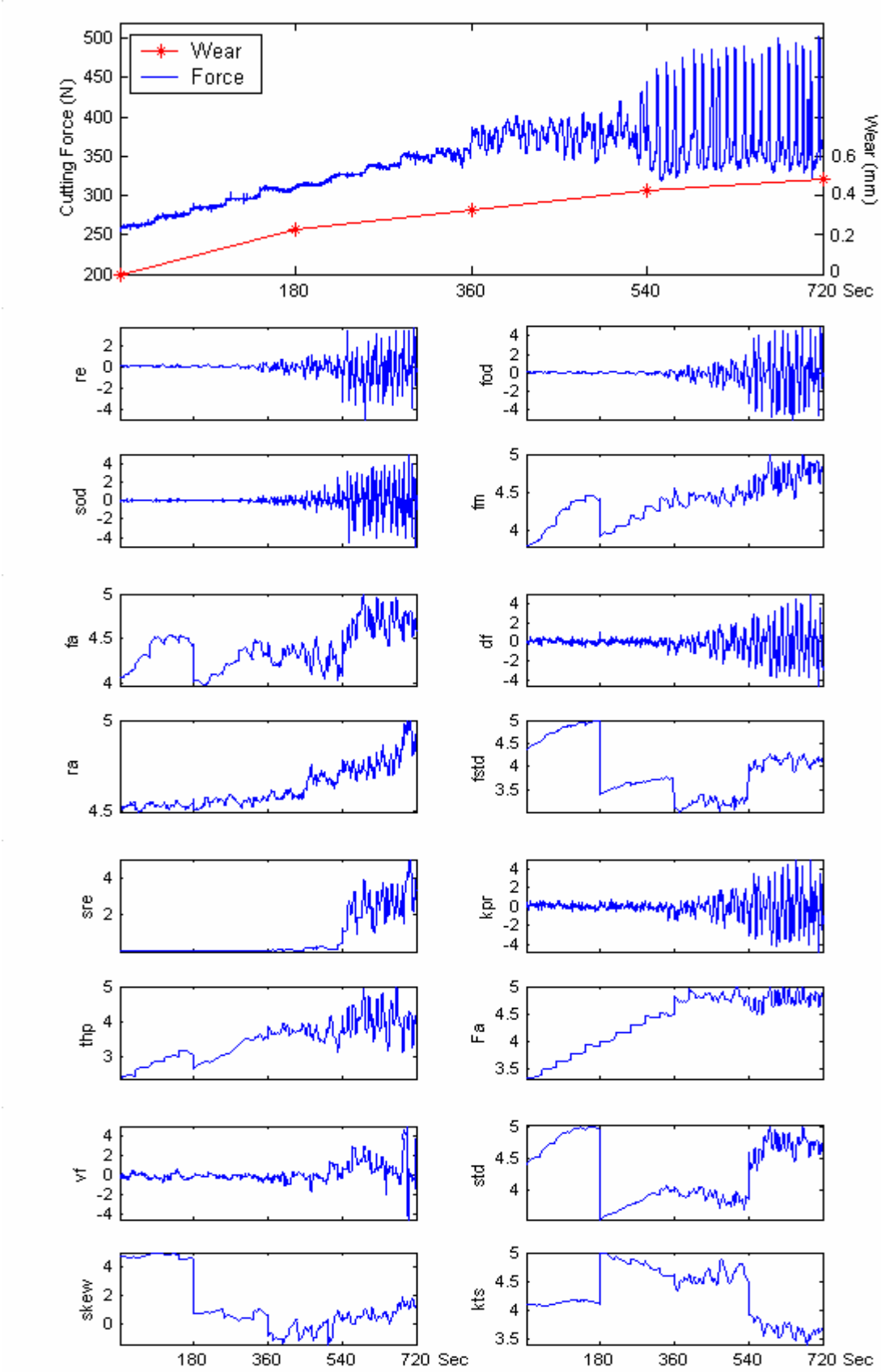


Figure A16. Illustration of Cutting Force, Tool Wear, and Features of Test_b4
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

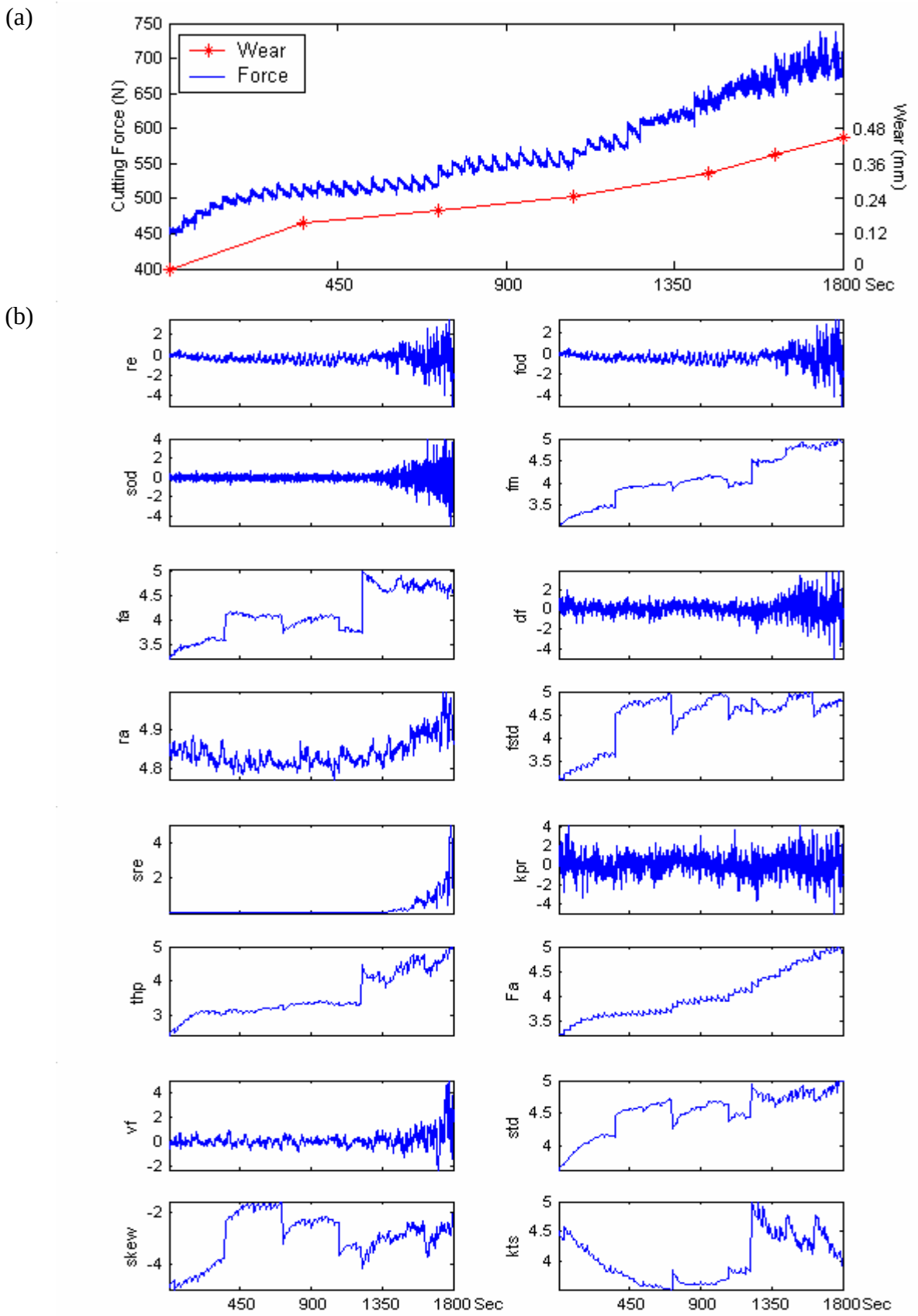


Figure A17. Illustration of Cutting Force, Tool Wear, and Features of Test_b5
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

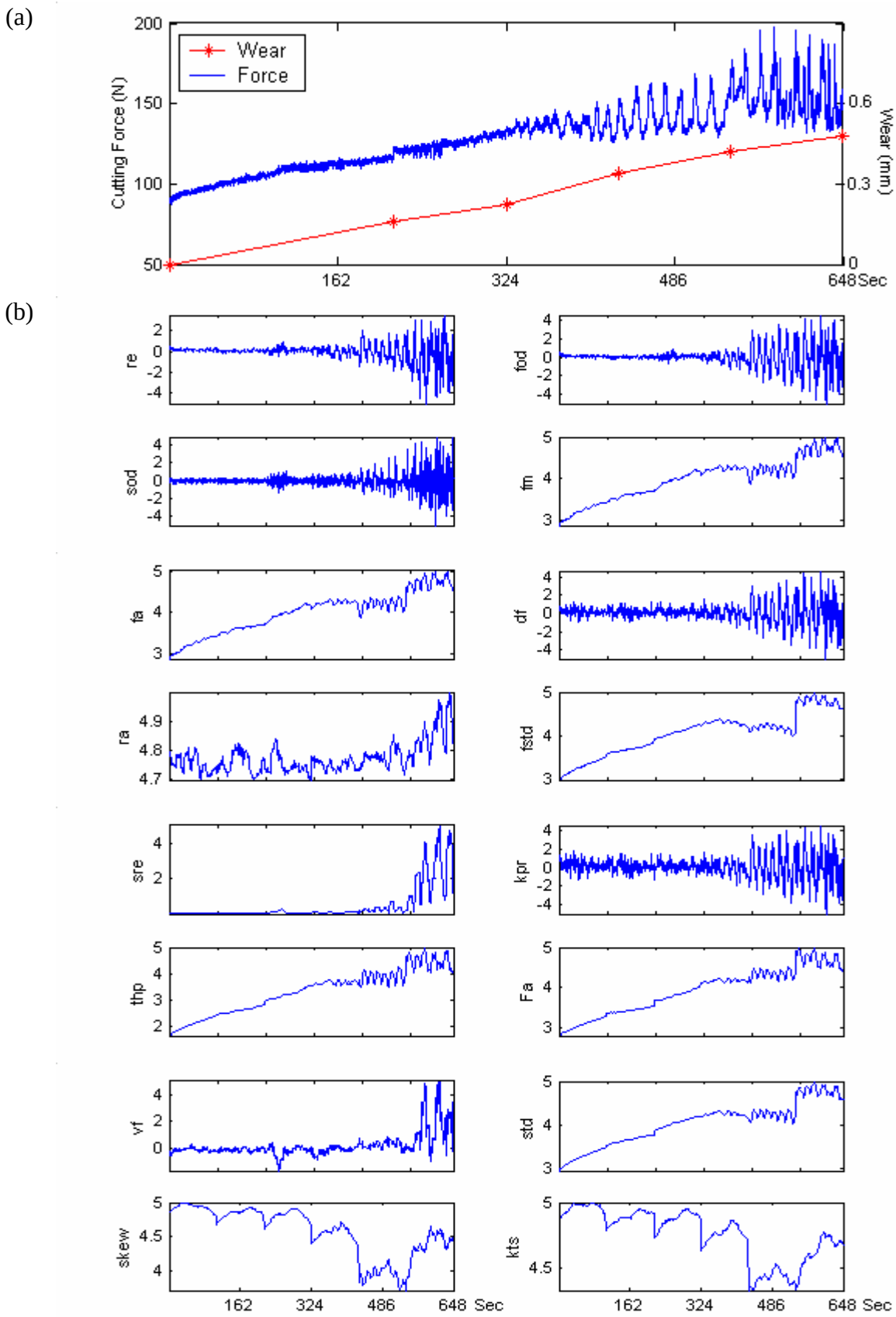


Figure A18. Illustration of Cutting Force, Tool Wear, and Features of Test_b6
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1200rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: A30N.

(a)

(b)

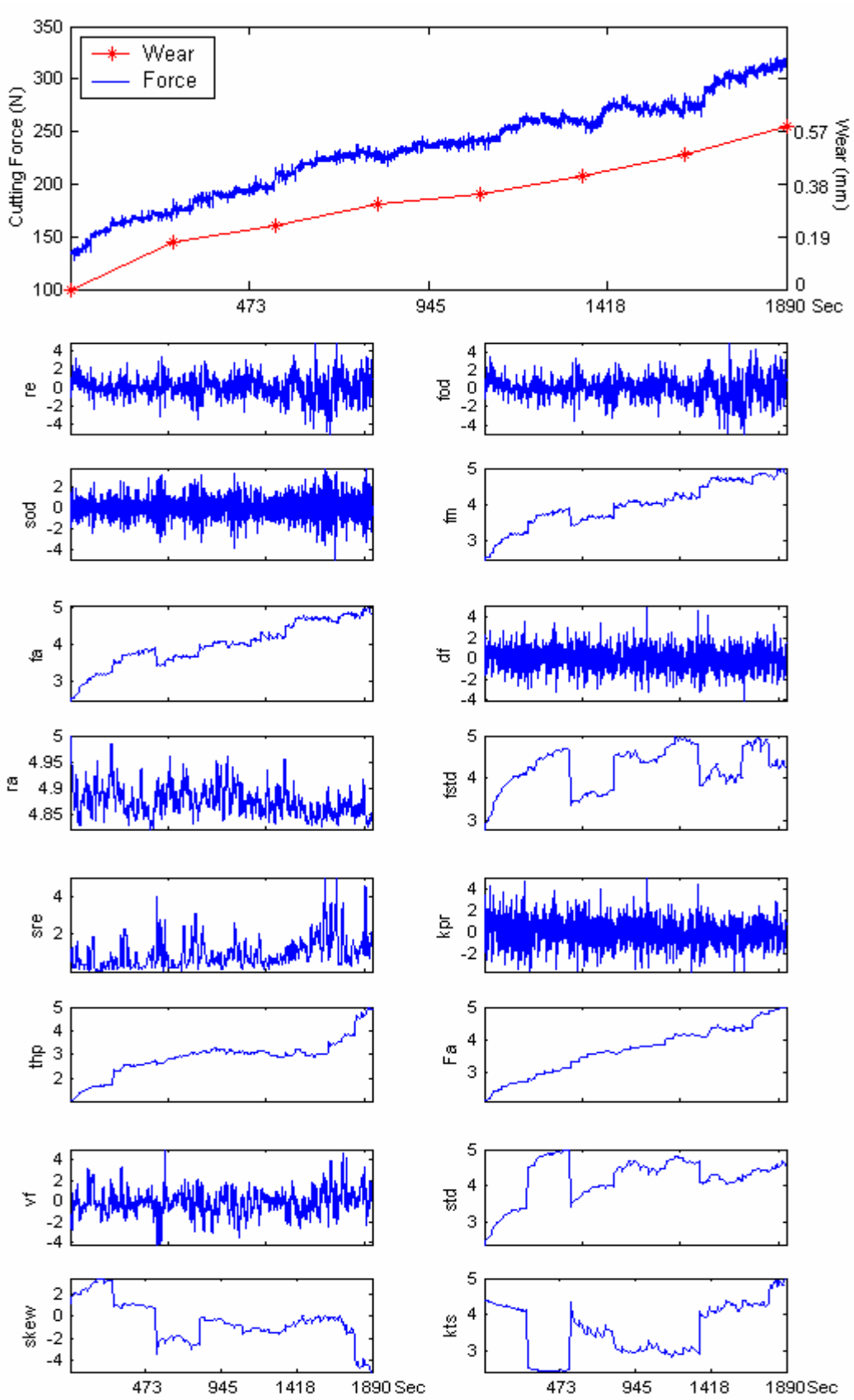


Figure A19. Illustration of Cutting Force, Tool Wear, and Features of Test_b7
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

(a)

(b)

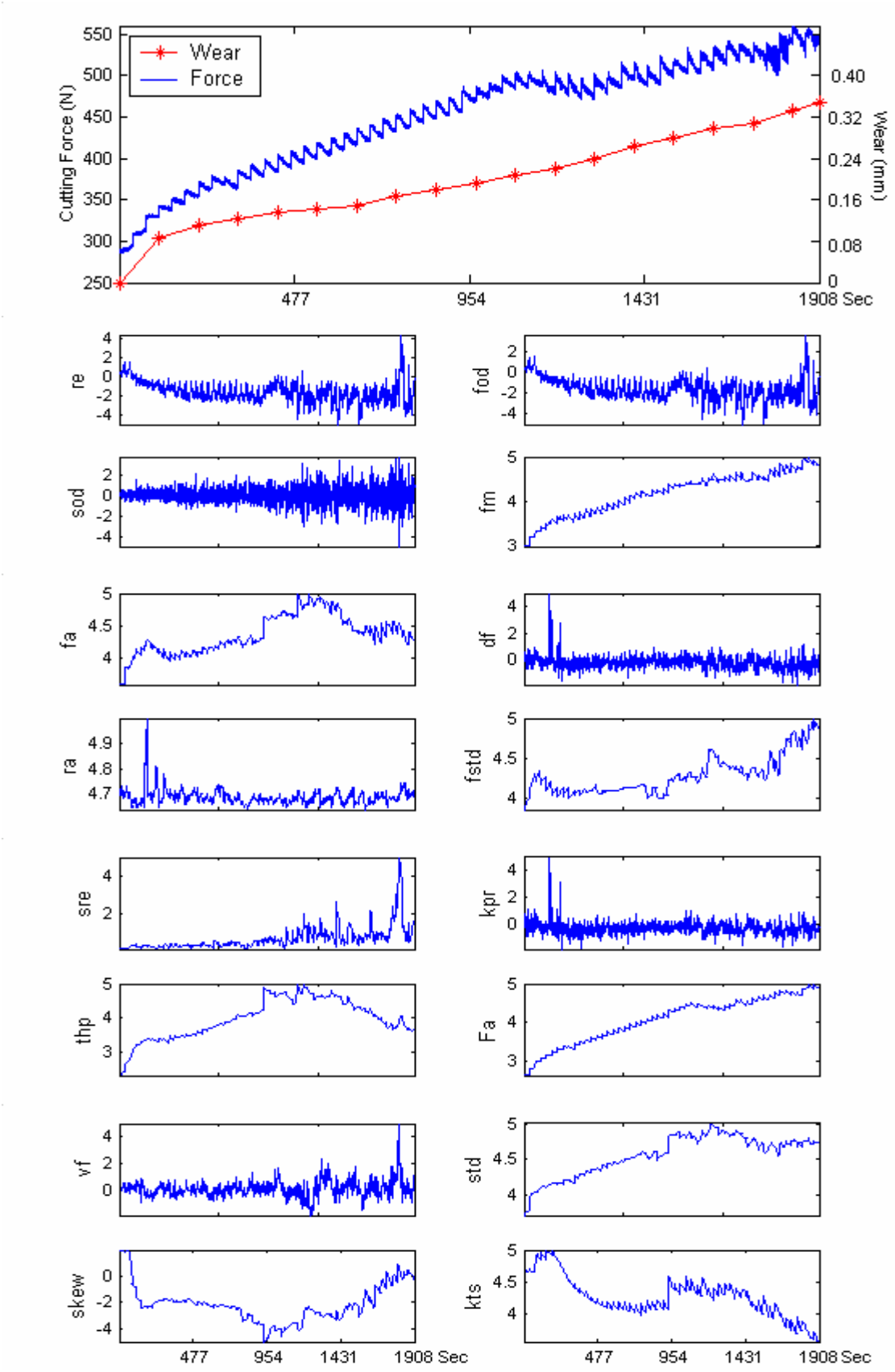


Figure A20. Illustration of Cutting Force, Tool Wear, and Features of Test_b8
(a) average cutting force & tool wear, (b) features
Cutting conditions: spindle speed = 800rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

Appendix B
Illustration of Feature Selection Processes for TWE

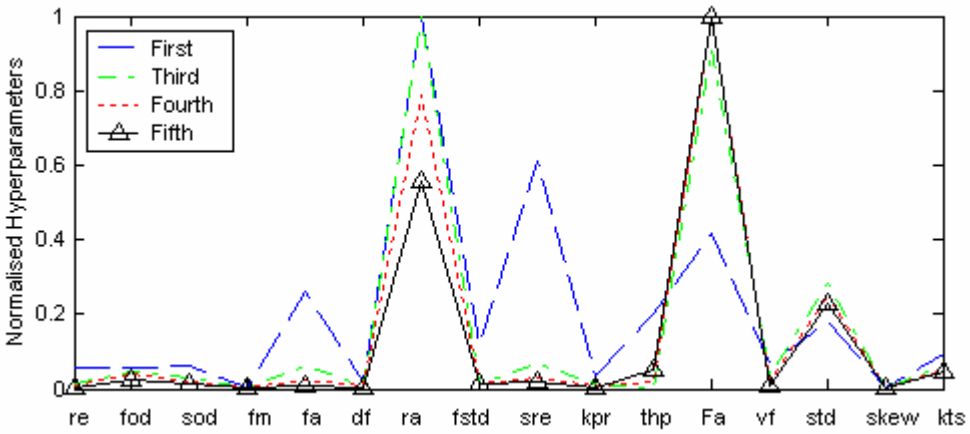


Figure B1. Illustration of the Feature Selection Processes of Test_a2
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

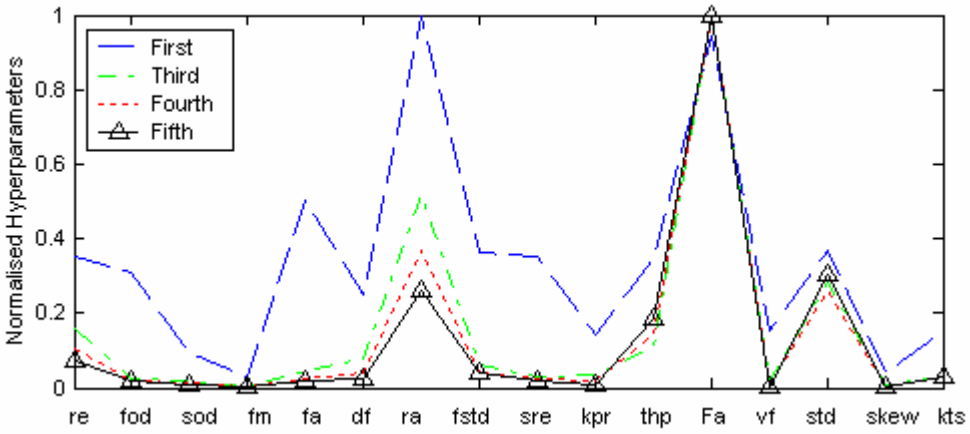


Figure B2. Illustration of the Feature Selection Processes of Test_a3
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

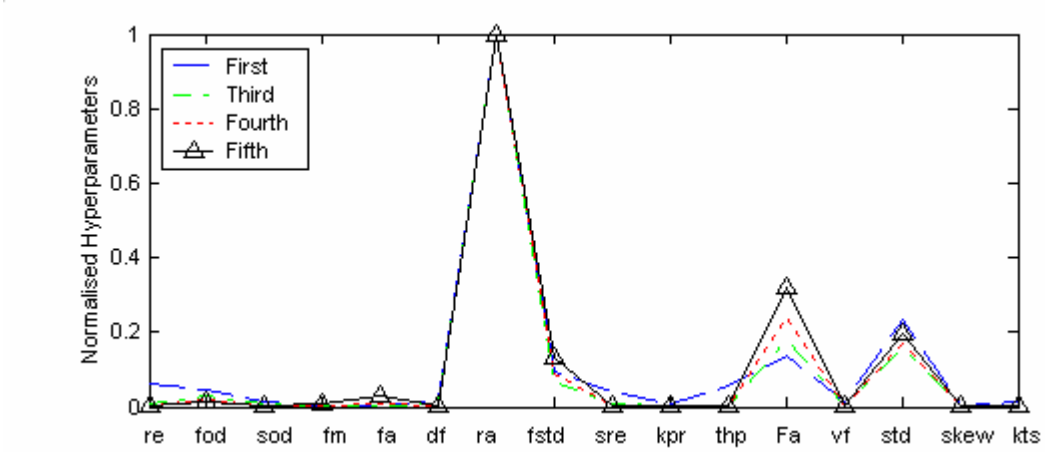


Figure B3. Illustration of the Feature Selection Processes of Test_a4
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

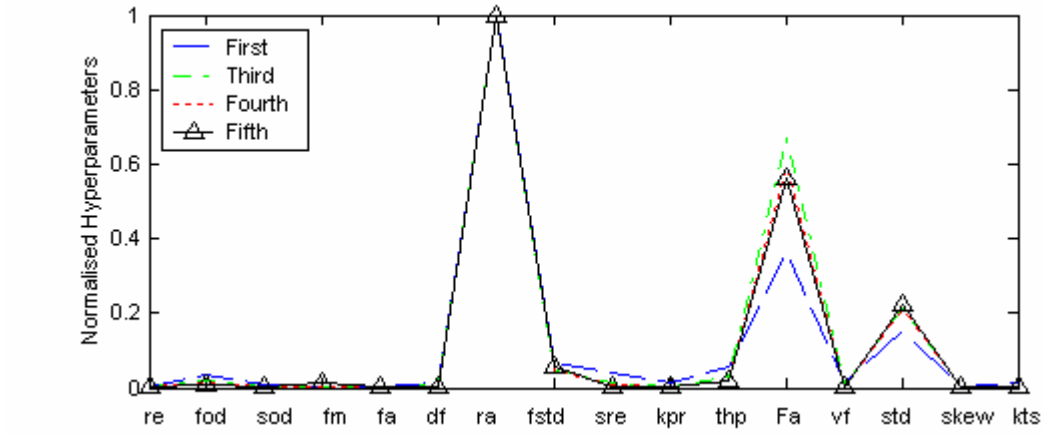


Figure B4. Illustration of the Feature Selection Processes of Test_a5
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

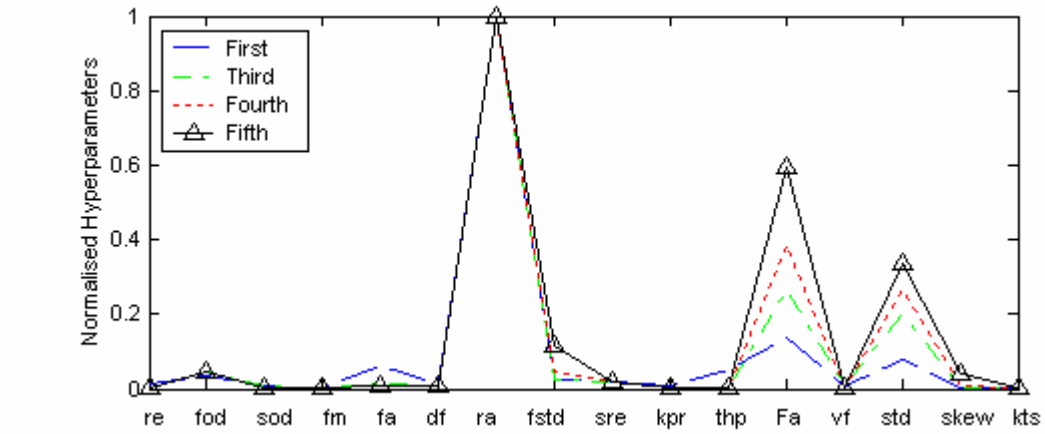


Figure B5. Illustration of the Feature Selection Processes of Test_a6
Cutting conditions: spindle speed = 1200rpm, feed rate = 150mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

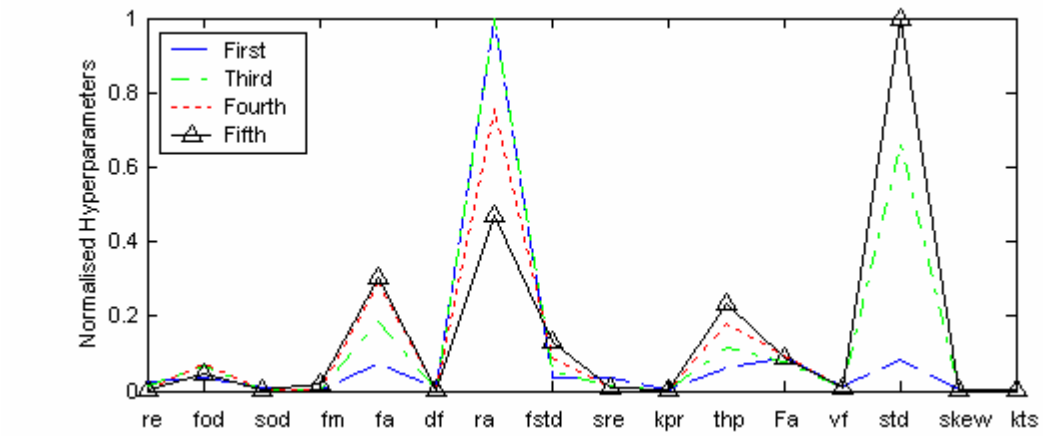


Figure B6. Illustration of the Feature Selection Processes of Test_a7
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

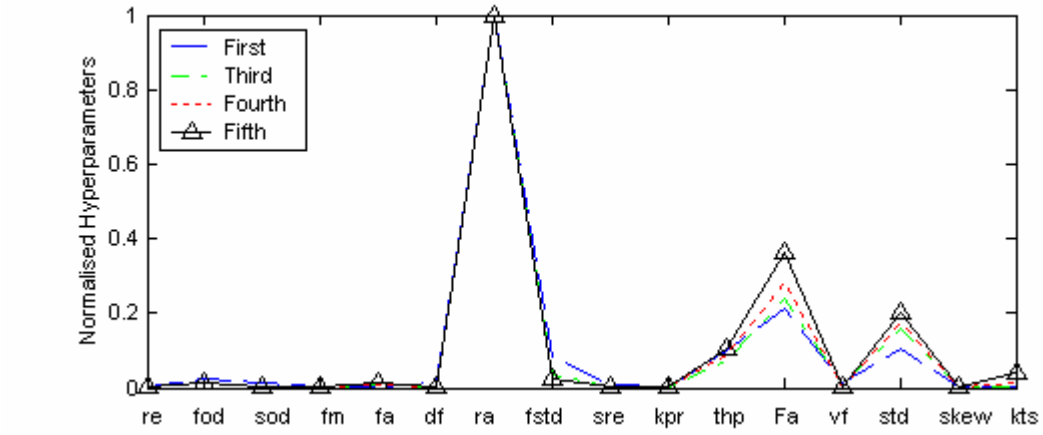


Figure B7. Illustration of the Feature Selection Processes of Test_a8
Cutting conditions: spindle speed = 1200rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

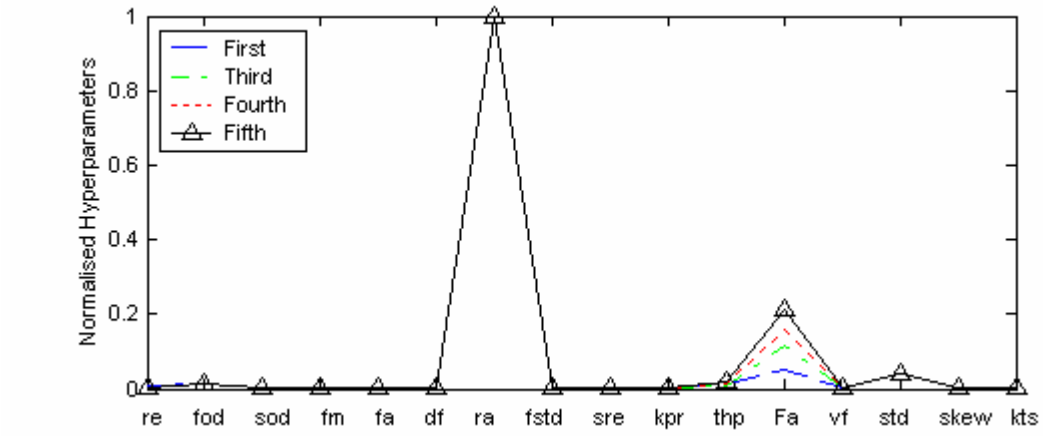


Figure B8. Illustration of the Feature Selection Processes of Test_a9
Cutting conditions: spindle speed = 600rpm, feed rate = 100mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

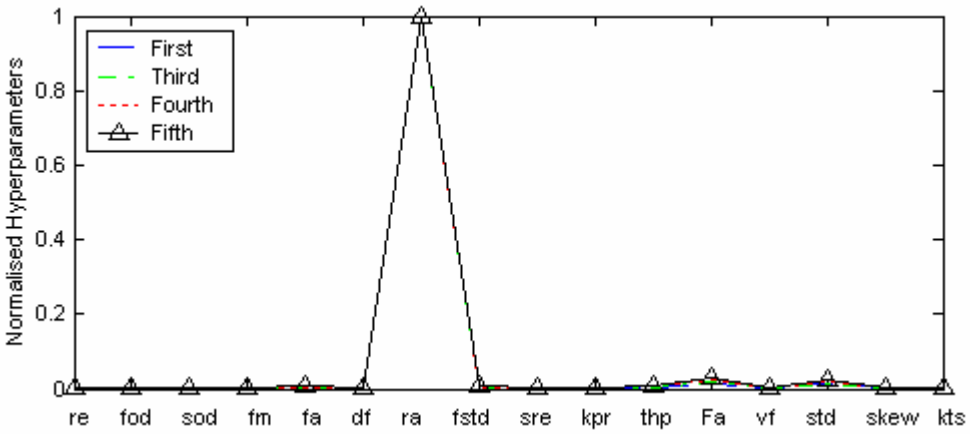


Figure B9. Illustration of the Feature Selection Processes of Test_a10
Cutting conditions: spindle speed = 600rpm, feed rate = 200mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

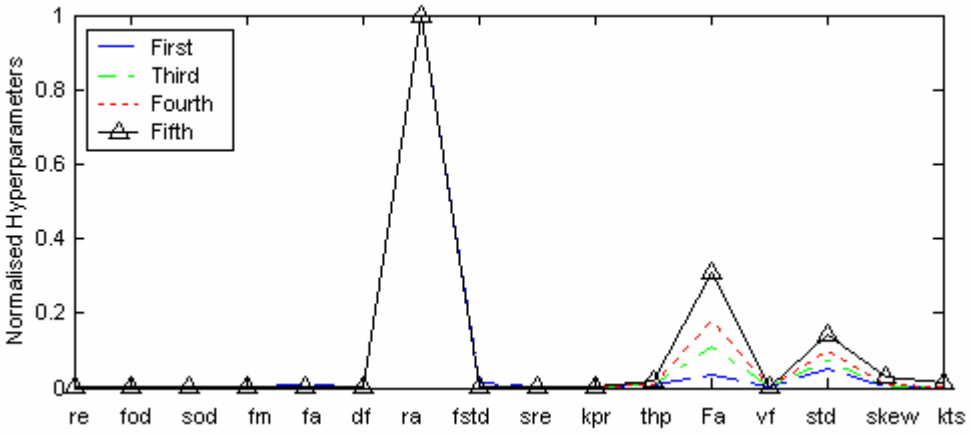


Figure B10. Illustration of the Feature Selection Processes of Test_a11
Cutting conditions: spindle speed = 800rpm, feed rate = 100mm/min, depth of cut = 2 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

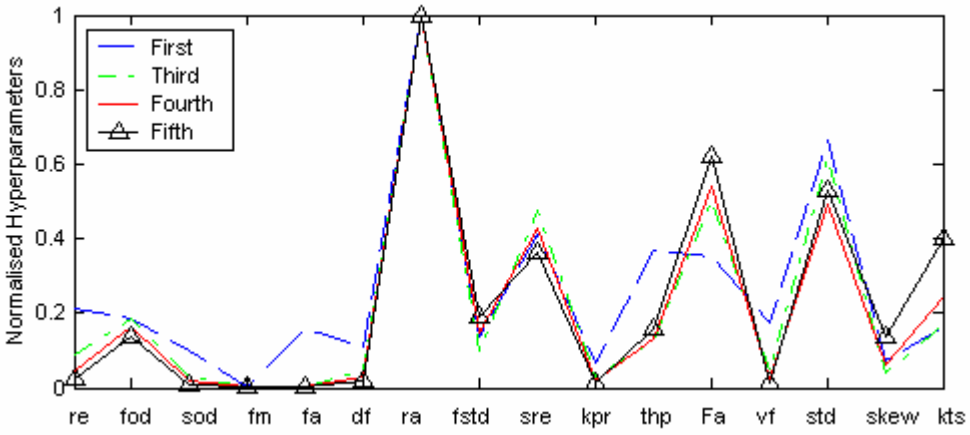


Figure B11. Illustration of the Feature Selection Processes of Test_a12
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

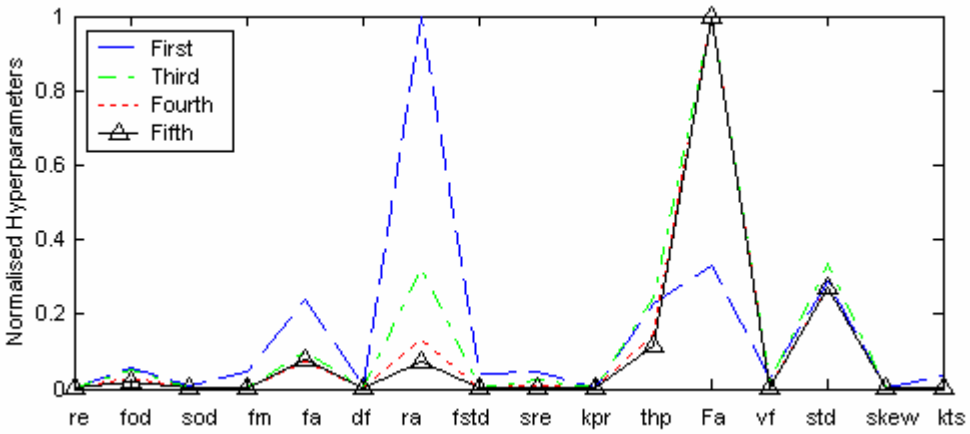


Figure B12. Illustration of the Feature Selection Processes of Test_b1
Cutting conditions: spindle speed = 800rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

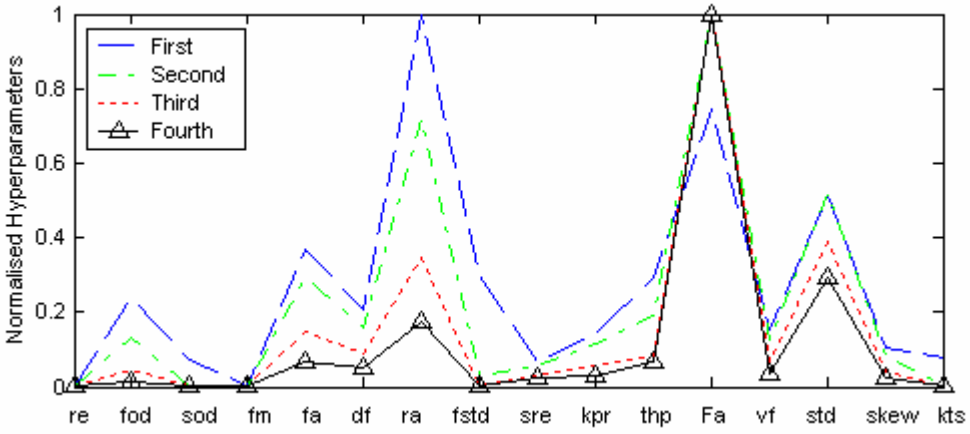


Figure B13. Illustration of the Feature Selection Processes of Test_b2
Cutting conditions: spindle speed = 800rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

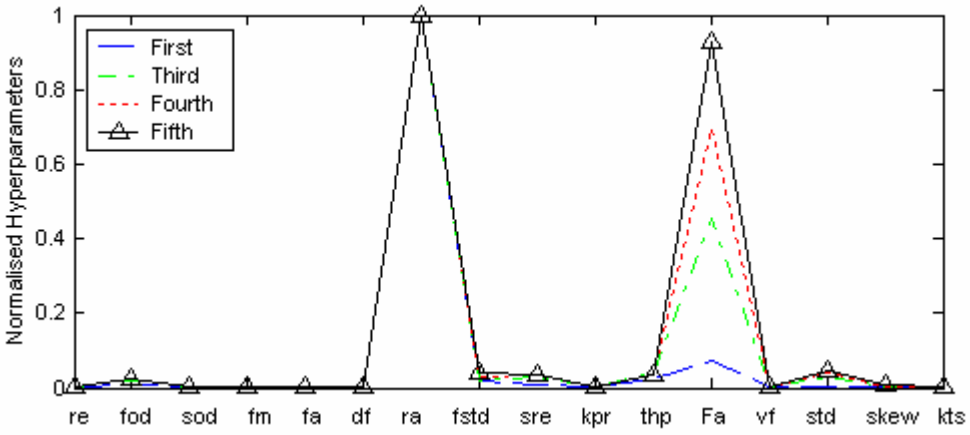


Figure B14. Illustration of the Feature Selection Processes of Test_b3
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: A30N.

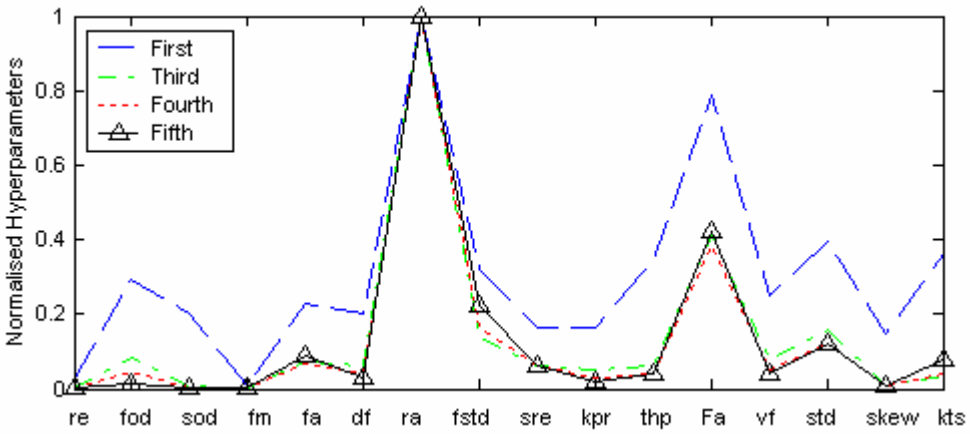


Figure B15. Illustration of the Feature Selection Processes of Test_b4
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

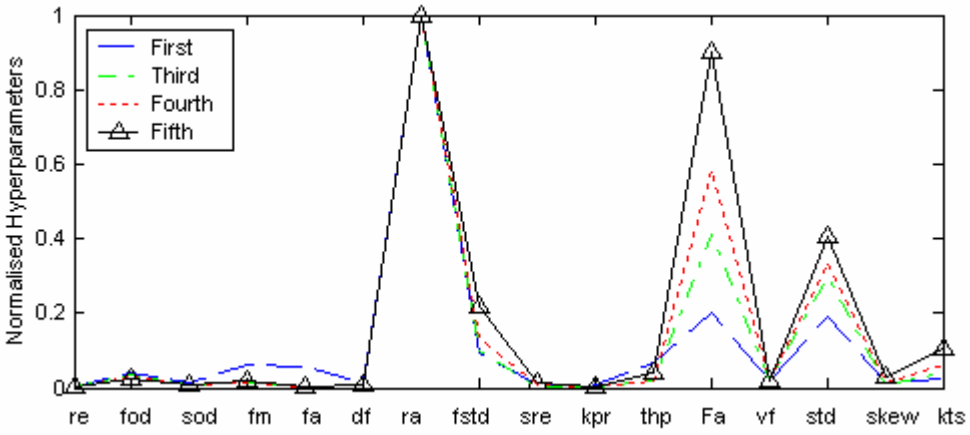


Figure B16. Illustration of the Feature Selection Processes of Test_b5
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

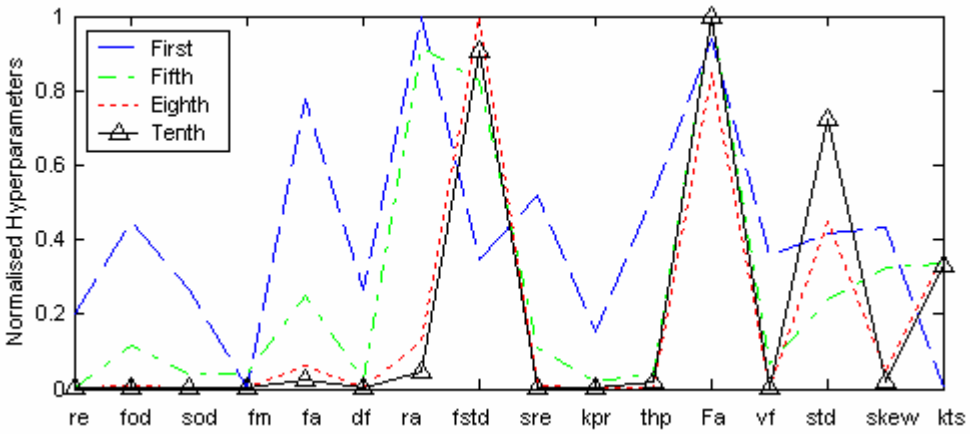


Figure B17. Illustration of the Feature Selection Processes of Test_b6
Cutting conditions: spindle speed = 1200rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: A30N.

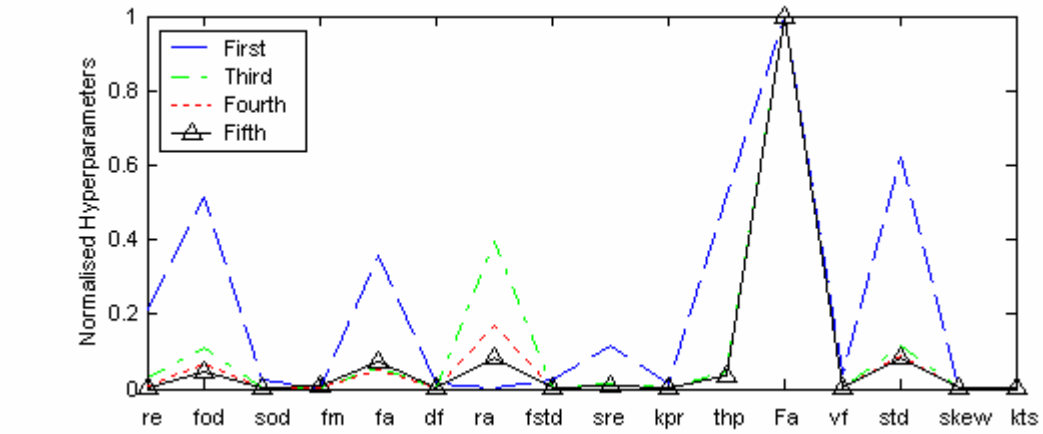


Figure B18. Illustration of the Feature Selection Processes of Test_b7
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

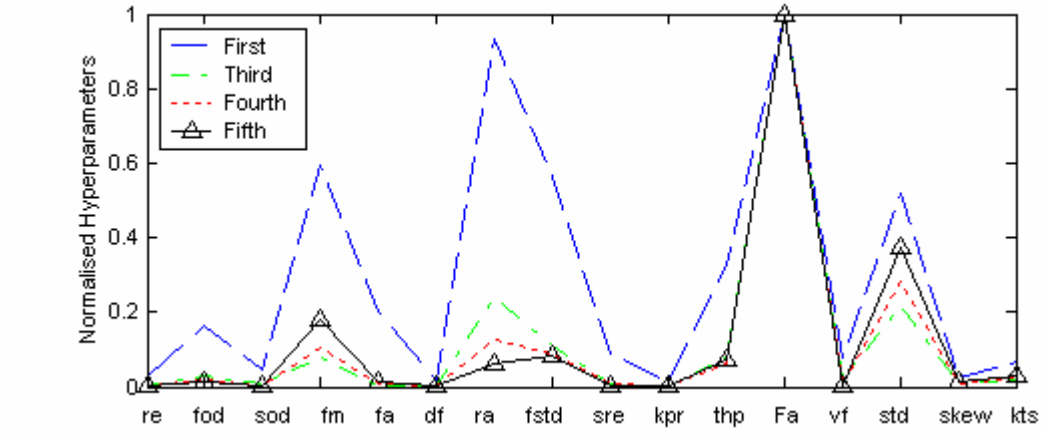
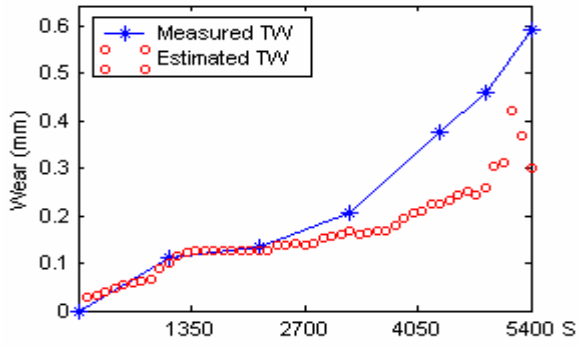
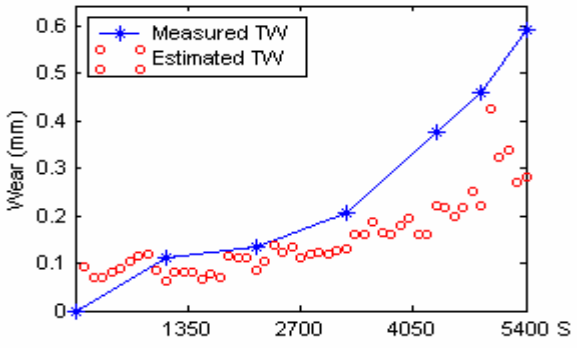


Figure B19. Illustration of the Feature Selection Processes of Test_b8
Cutting conditions: spindle speed = 800rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

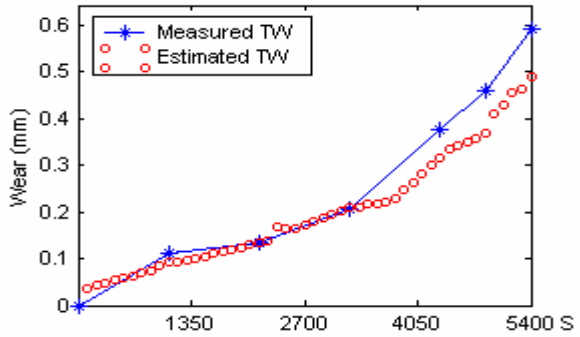
Appendix C
Tool Wear Estimation Results



(a)

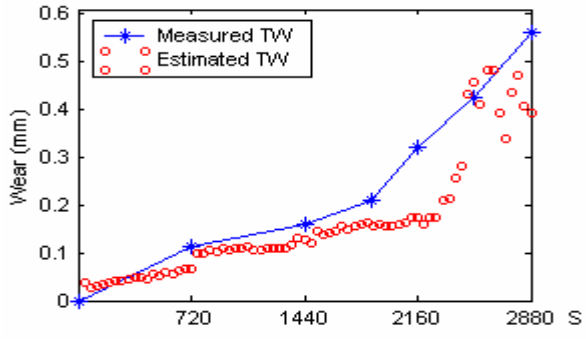


(b)

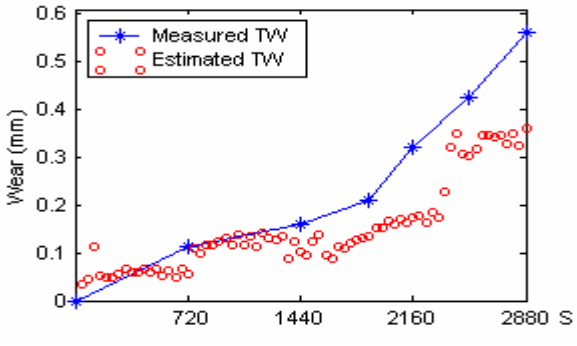


(c)

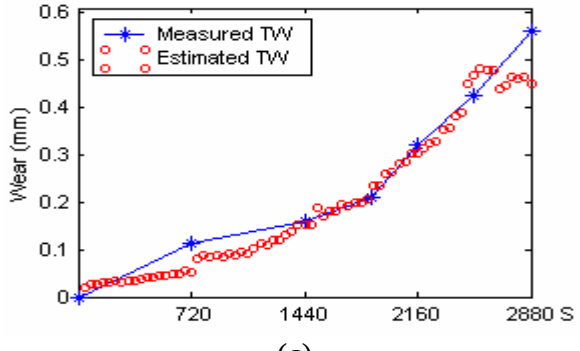
Figure C1. TWE Results of T2
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.



(a)

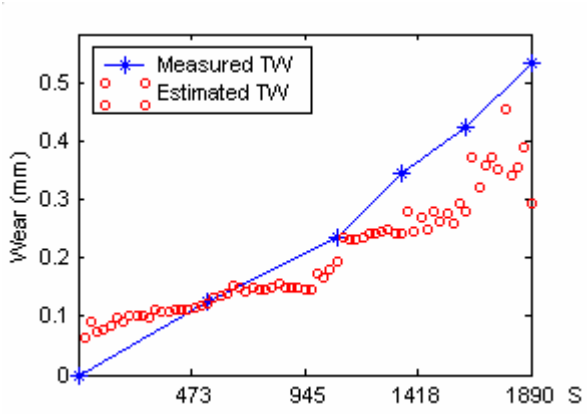


(b)

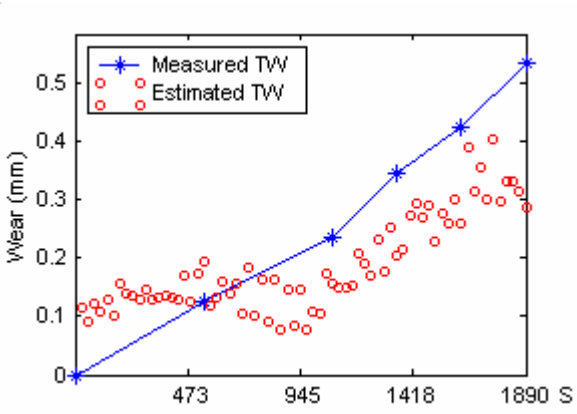


(c)

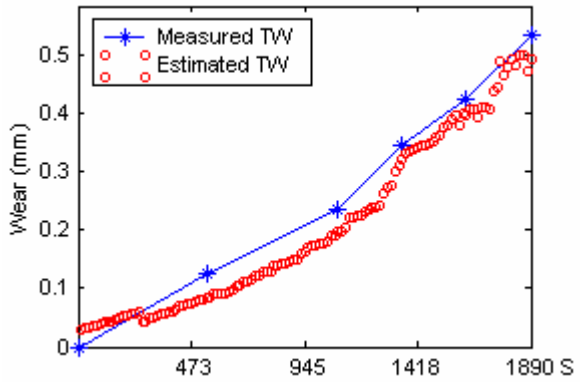
Figure C2. TWE Results of T3
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1200rpm, feed rate = 150mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.



(a)

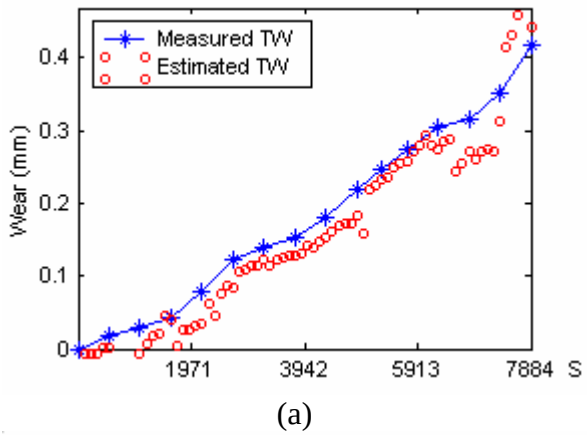


(b)

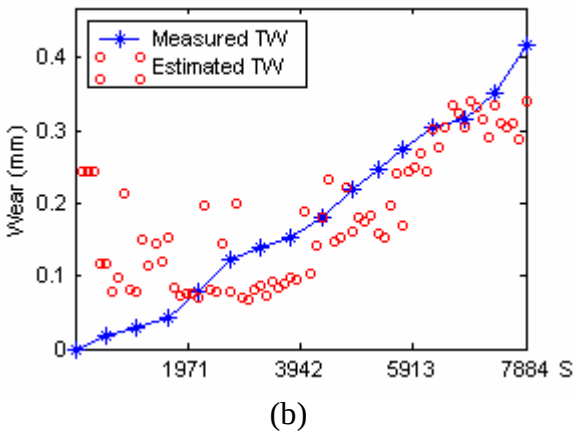


(c)

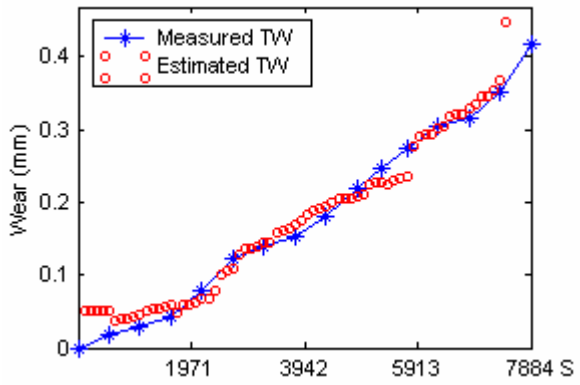
Figure C3. TWE Results of T4
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.



(a)

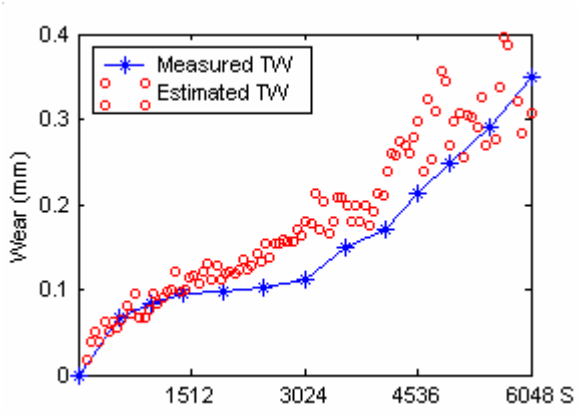


(b)

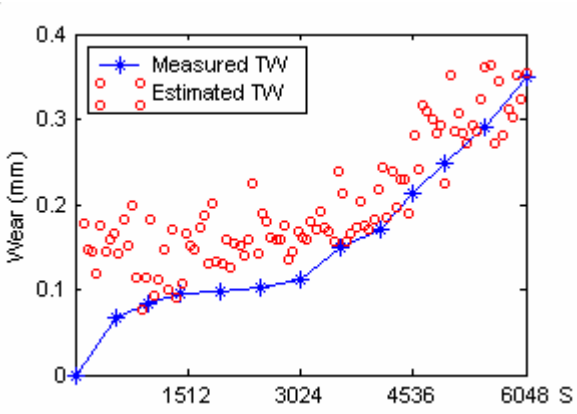


(c)

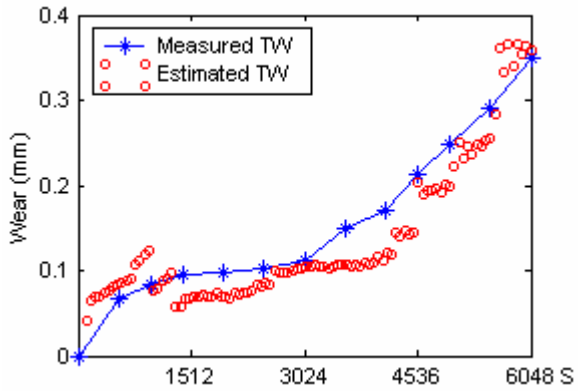
Figure C4. TWE Results of T5
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 800rpm, feed rate = 100mm/min, depth of cut = 2 mm, insert number = 2, immersion rate: FULL, insert type: AC325.



(a)

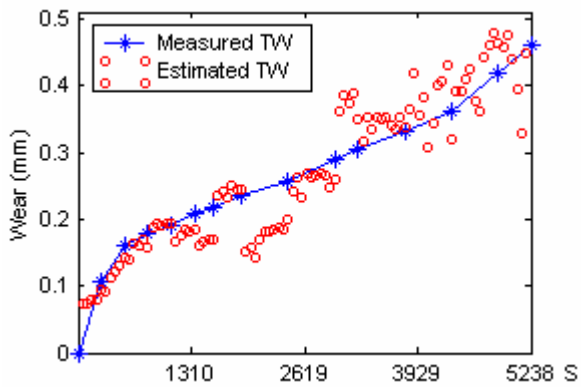


(b)

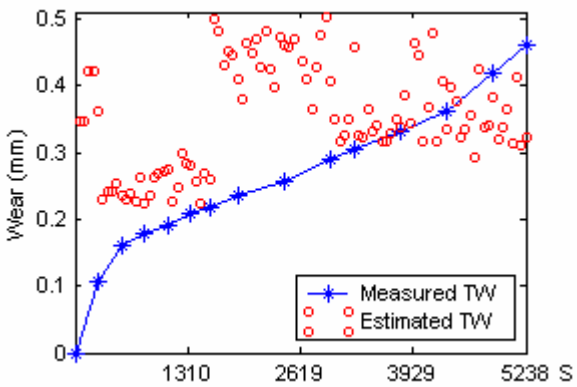


(c)

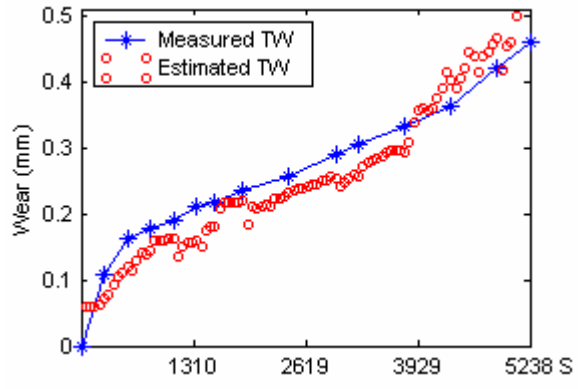
Figure C5. TWE Results of T6
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.



(a)

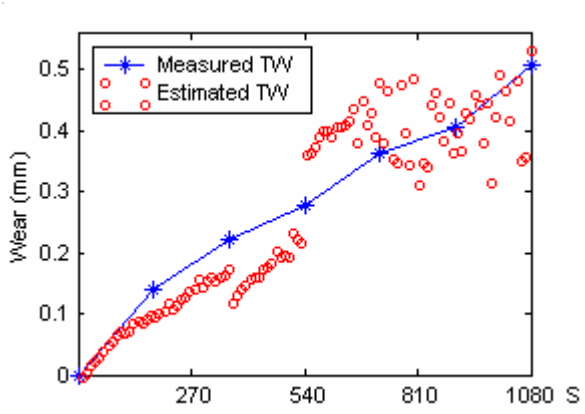


(b)

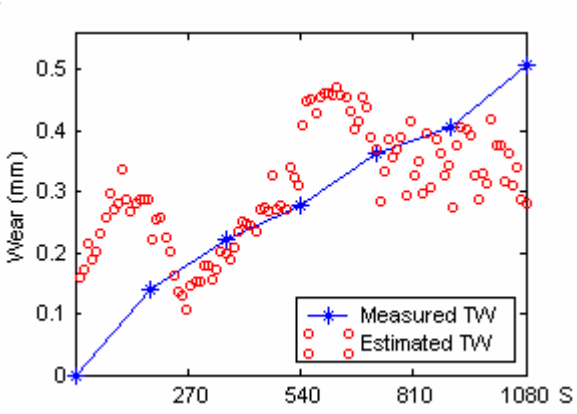


(c)

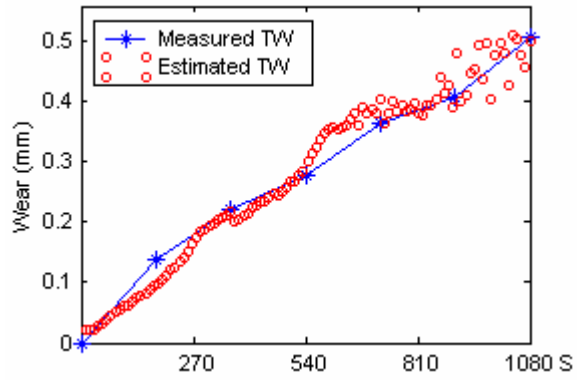
Figure C6. TWE Results of T7
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 800rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.



(a)

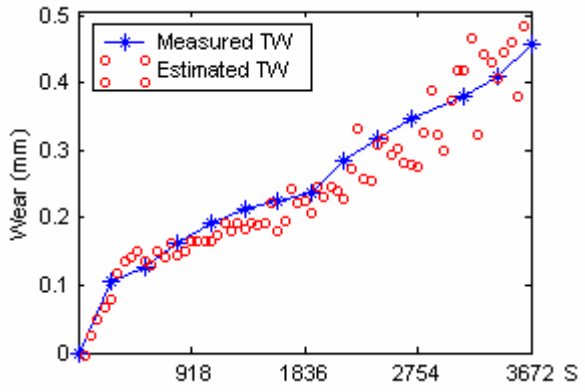


(b)

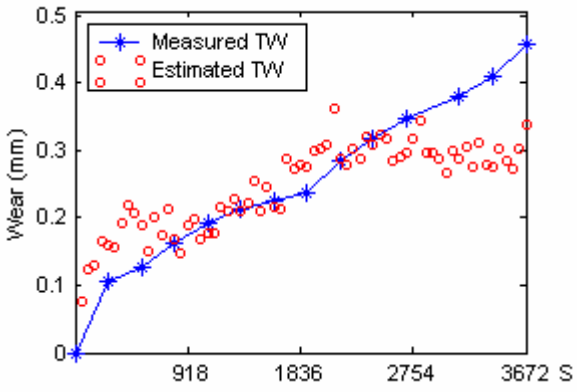


(c)

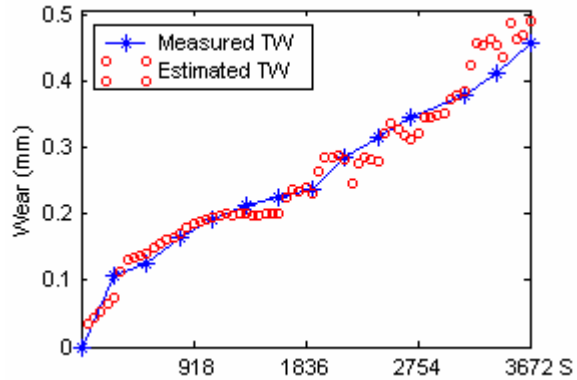
Figure C7. TWE Results of T8
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 800rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.



(a)

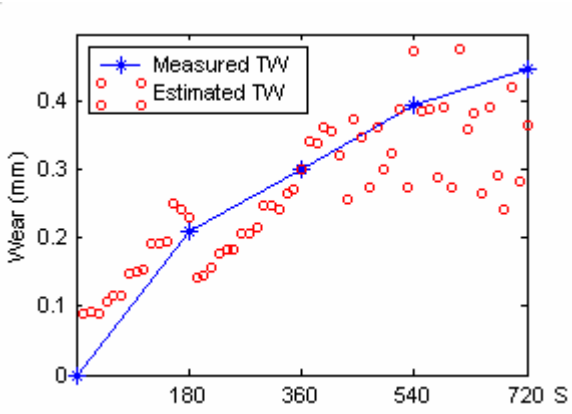


(b)

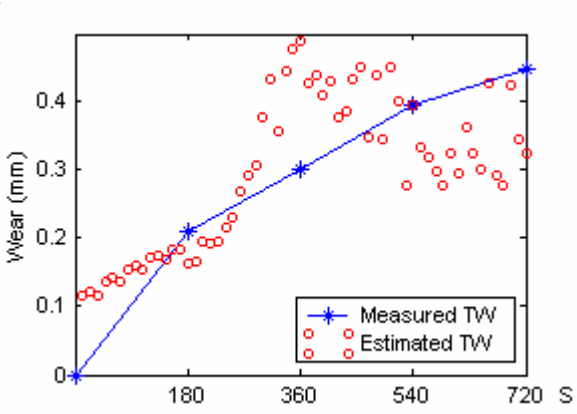


(c)

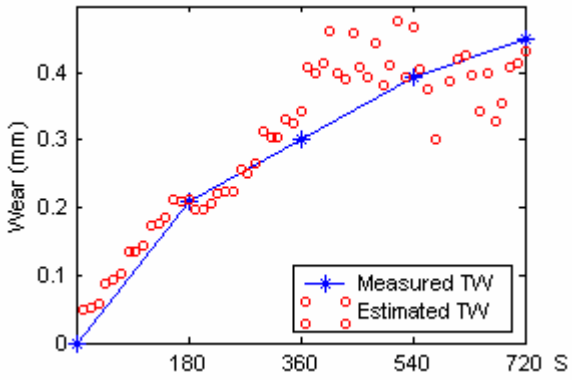
Figure C8. TWE Results of T9
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: A30N.



(a)

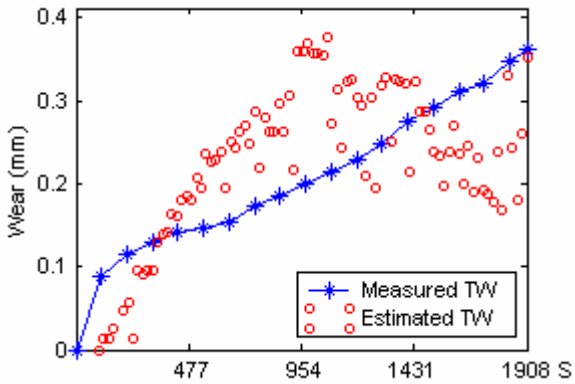


(b)

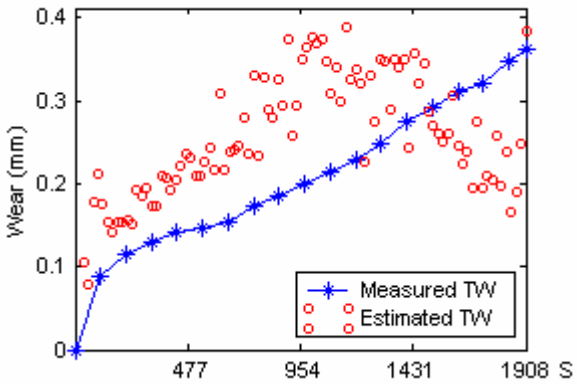


(c)

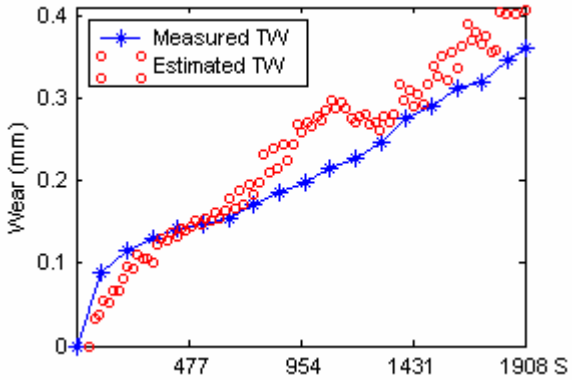
Figure C9. TWE Results of T10
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.



(a)



(b)



(c)

Figure C10. TWE Results of T11
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 800rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

Appendix D
Illustration of Feature Selection Processes for TWR

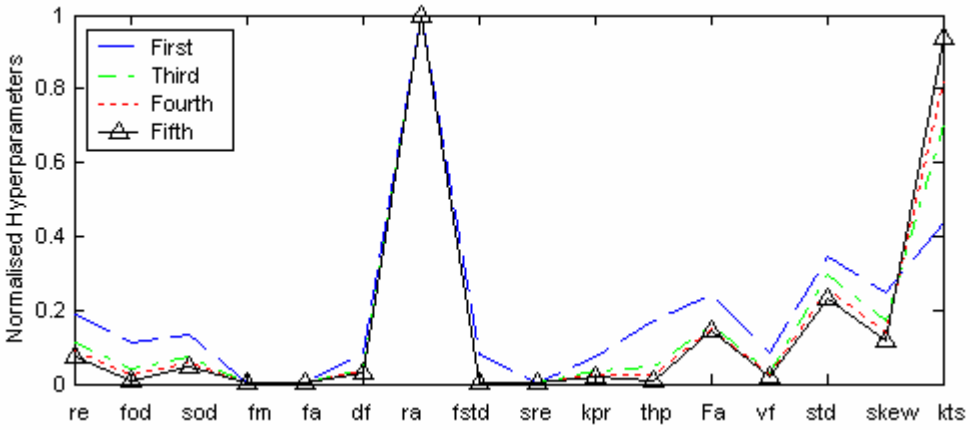


Figure D1. Illustration of the Feature Selection Processes of Test_a2
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

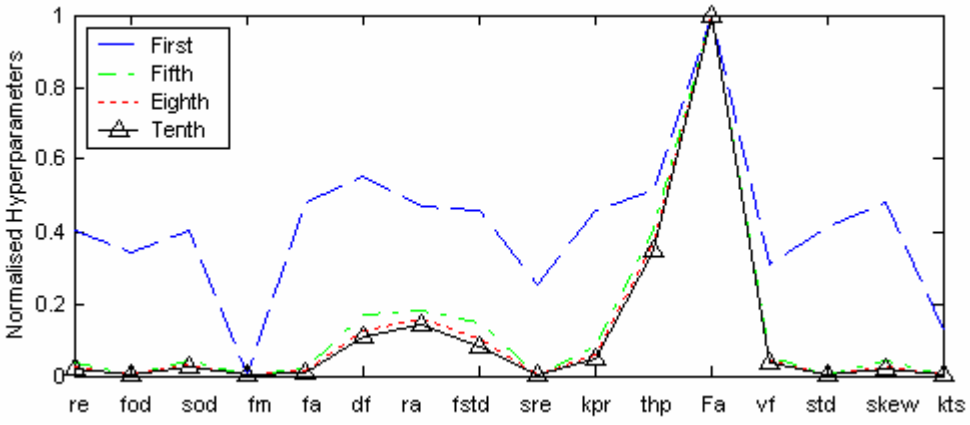


Figure D2. Illustration of the Feature Selection Processes of Test_a3
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

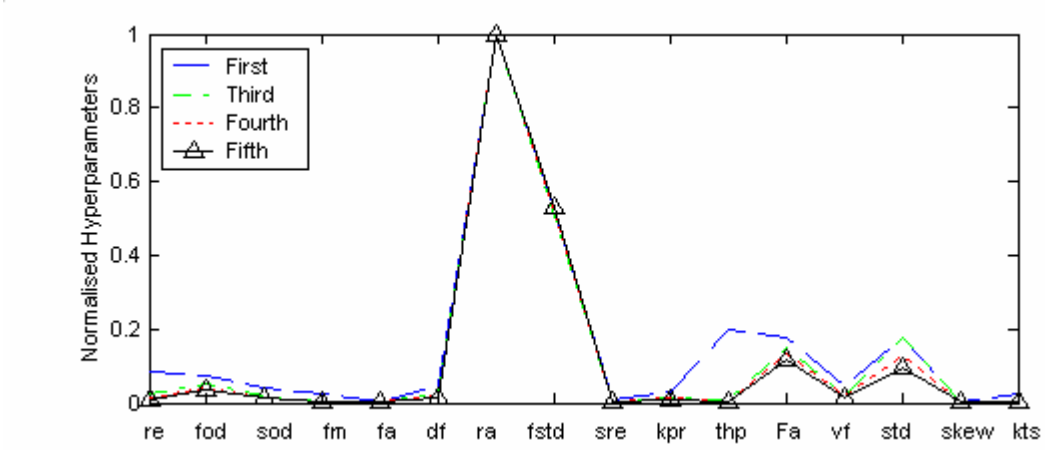


Figure D3. Illustration of the Feature Selection Processes of Test_a4
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

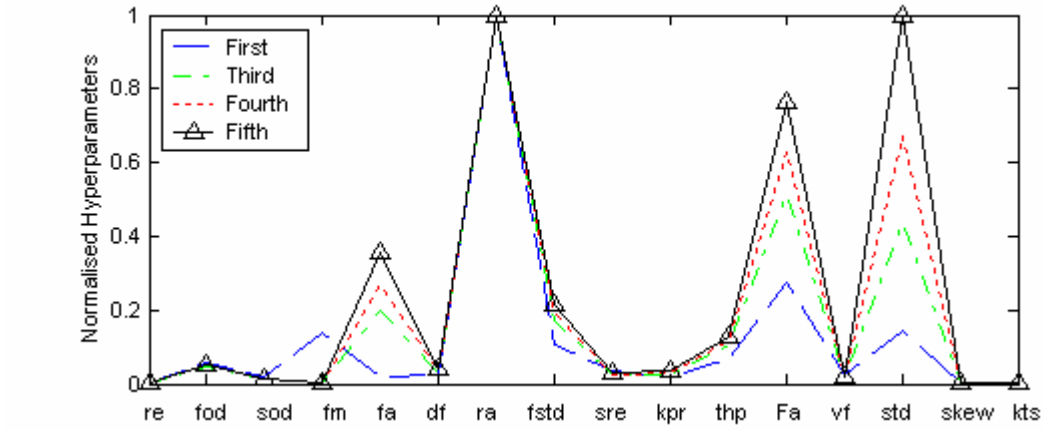


Figure D4. Illustration of the Feature Selection Processes of Test_a5
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

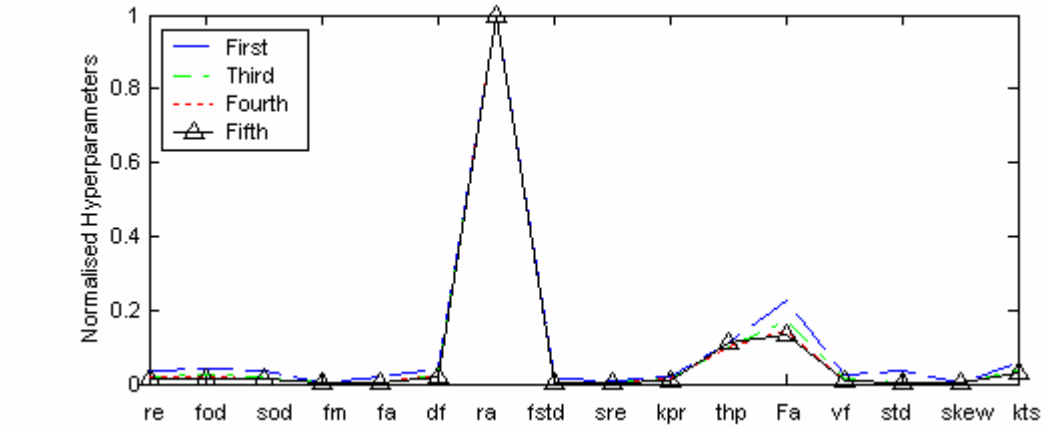


Figure D5. Illustration of the Feature Selection Processes of Test_a6
Cutting conditions: spindle speed = 1200rpm, feed rate = 150mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

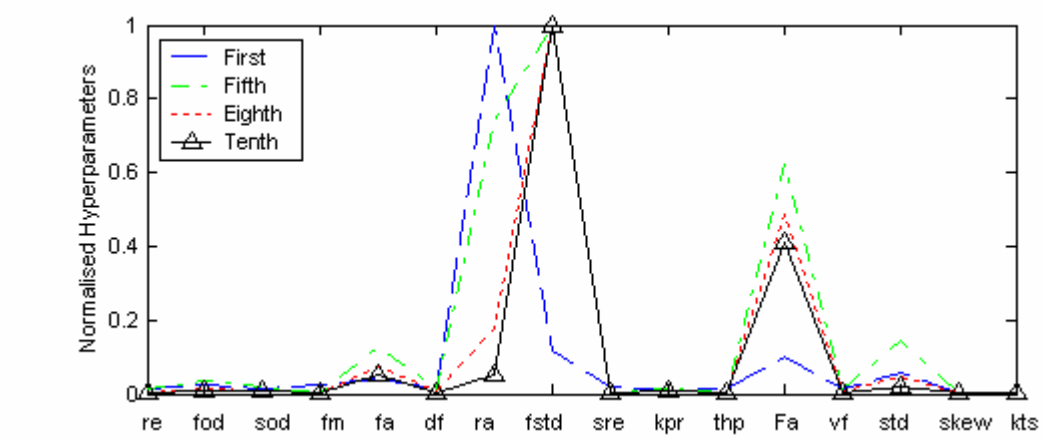


Figure D6. Illustration of the Feature Selection Processes of Test_a7
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

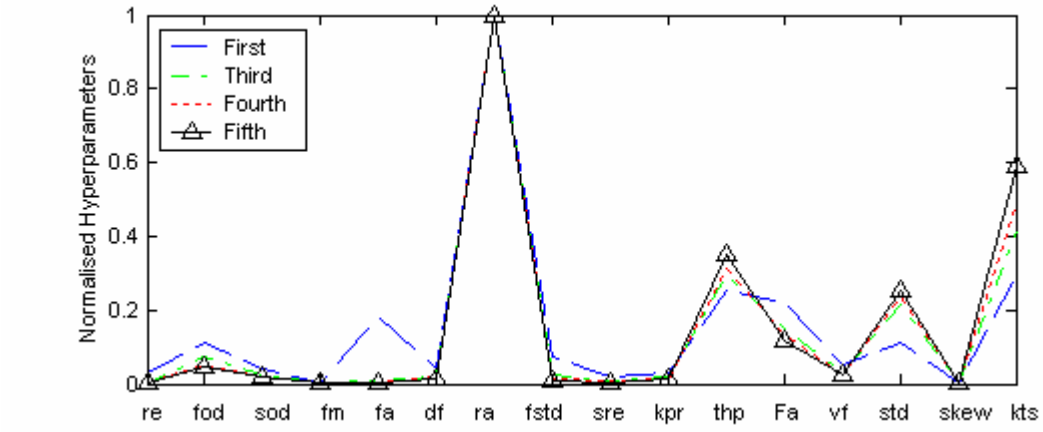


Figure D7. Illustration of the Feature Selection Processes of Test_a8
Cutting conditions: spindle speed = 1200rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

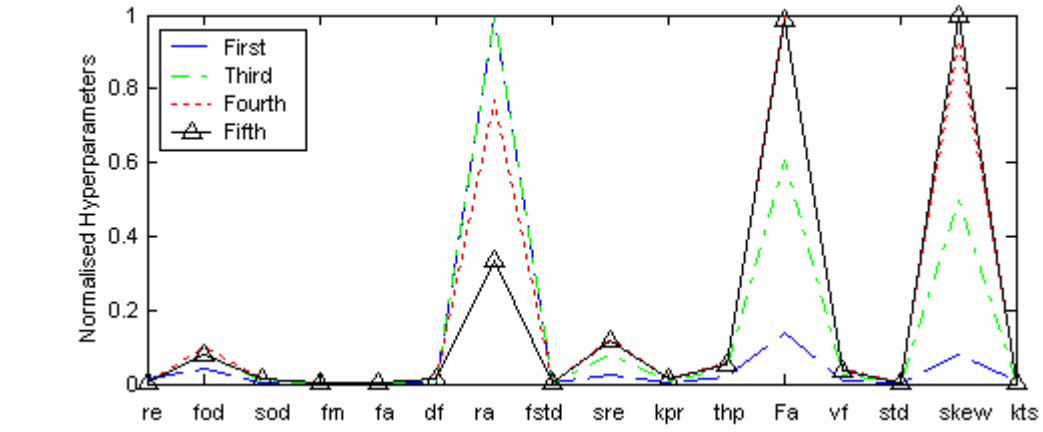


Figure D8. Illustration of the Feature Selection Processes of Test_a9
Cutting conditions: spindle speed = 600rpm, feed rate = 100mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

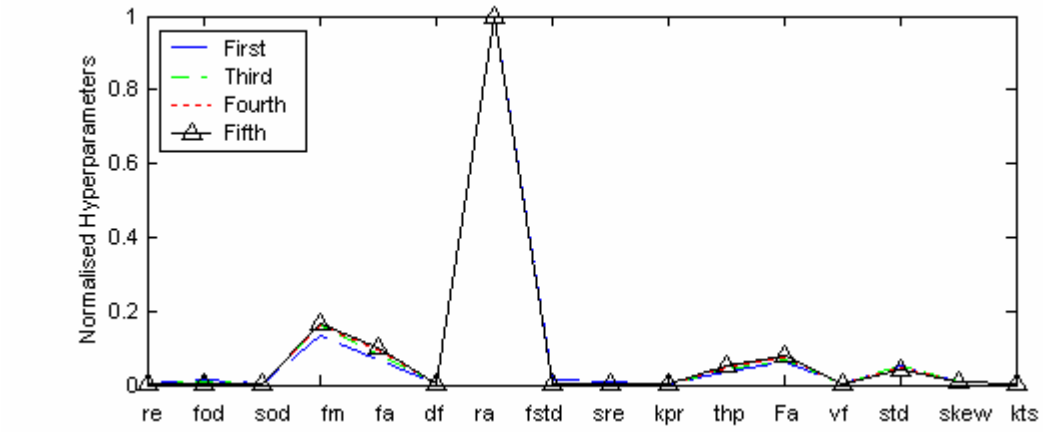


Figure D9. Illustration of the Feature Selection Processes of Test_a10
Cutting conditions: spindle speed = 600rpm, feed rate = 200mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

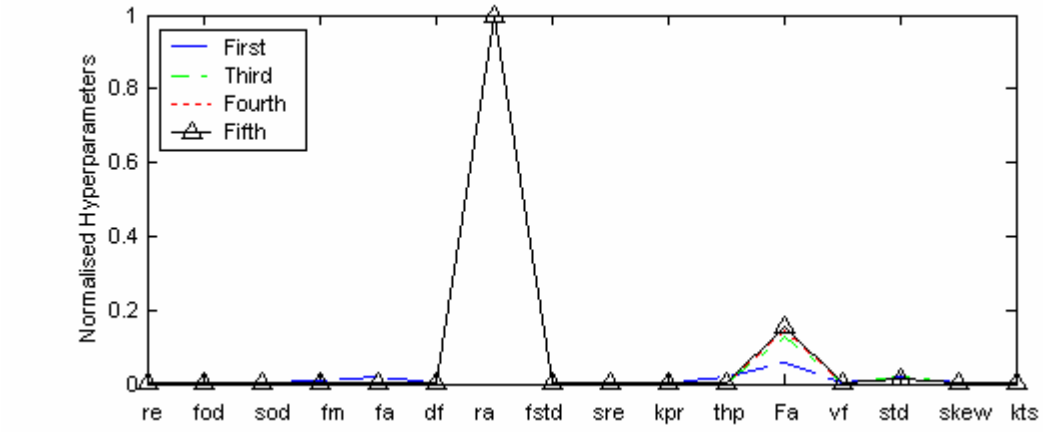


Figure D10. Illustration of the Feature Selection Processes of Test_a11
Cutting conditions: spindle speed = 800rpm, feed rate = 100mm/min, depth of cut = 2 mm, insert number = 2, immersion rate: FULL, insert type: AC325.

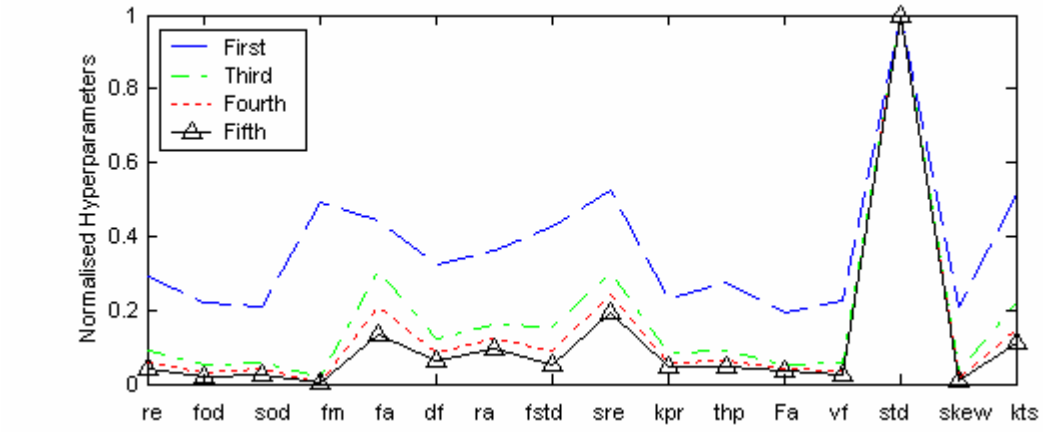


Figure D11. Illustration of the Feature Selection Processes of Test_a12
Cutting conditions: spindle speed = 1000rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.

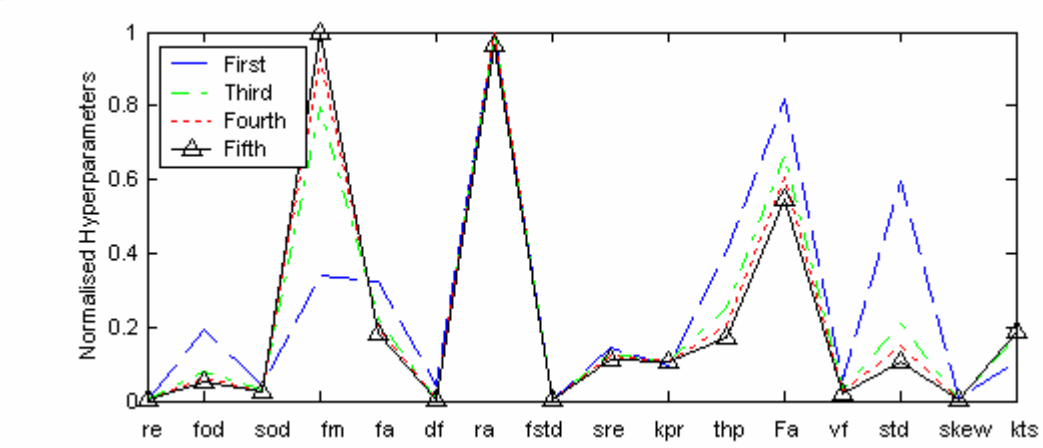


Figure D12. Illustration of the Feature Selection Processes of Test_b1
Cutting conditions: spindle speed = 800rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

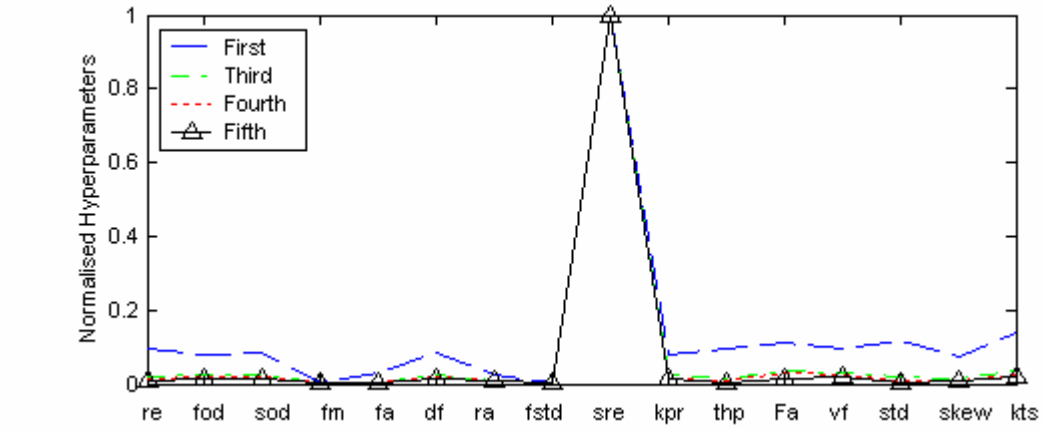


Figure D13. Illustration of the Feature Selection Processes of Test_b2
Cutting conditions: spindle speed = 800rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

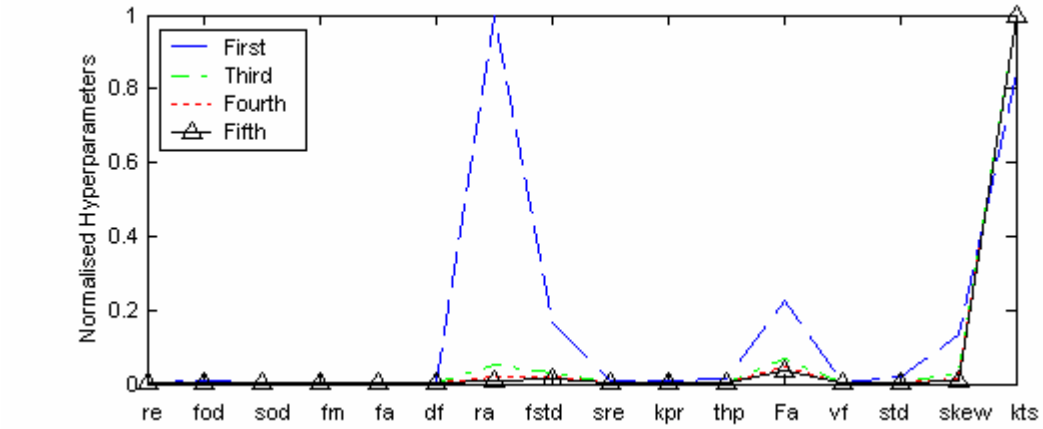


Figure D14. Illustration of the Feature Selection Processes of Test_b3
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: A30N.

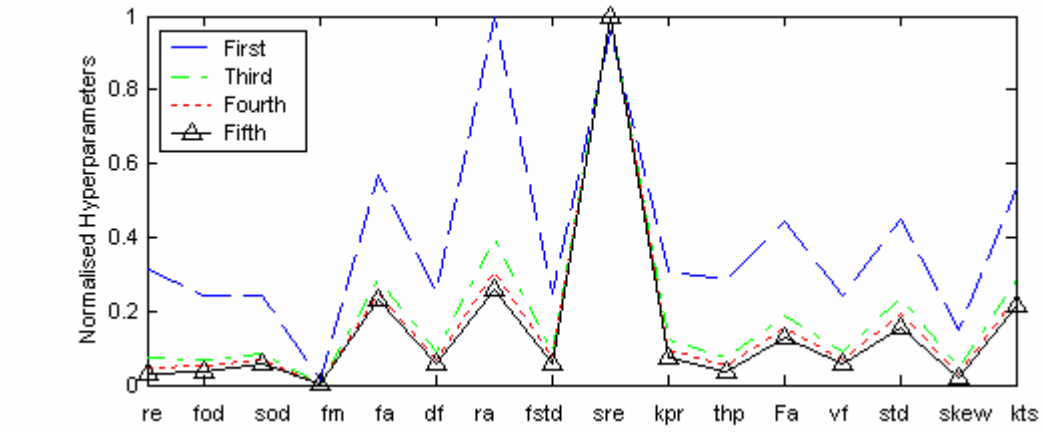


Figure D15. Illustration of the Feature Selection Processes of Test_b4
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

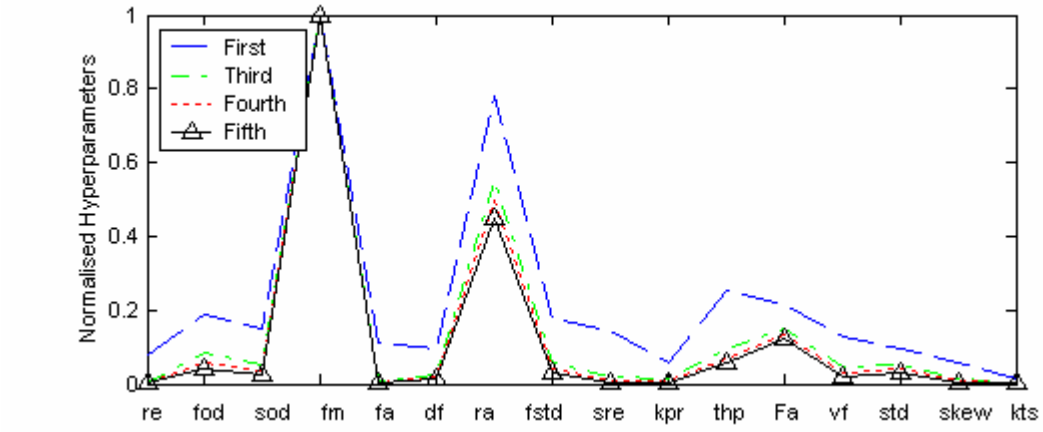


Figure D16. Illustration of the Feature Selection Processes of Test_b5
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

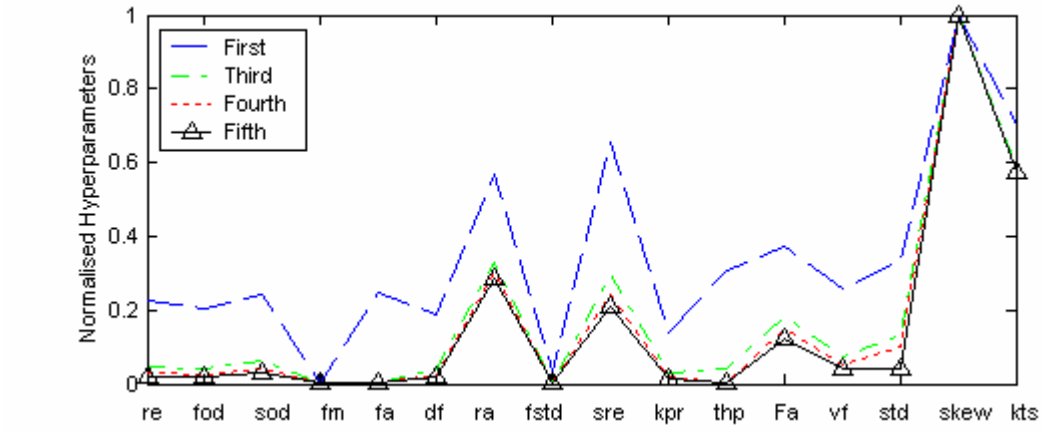


Figure D17. Illustration of the Feature Selection Processes of Test_b6
Cutting conditions: spindle speed = 1200rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: A30N.

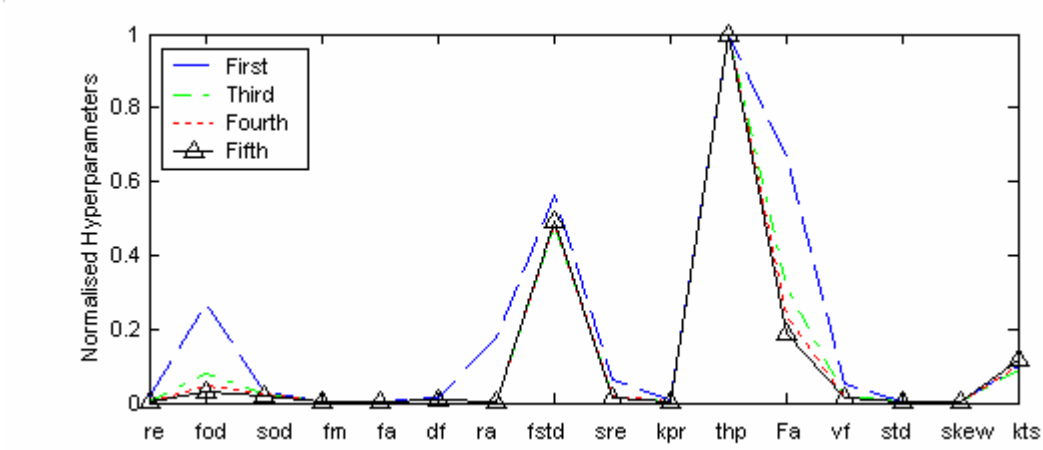


Figure D18. Illustration of the Feature Selection Processes of Test_b7
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

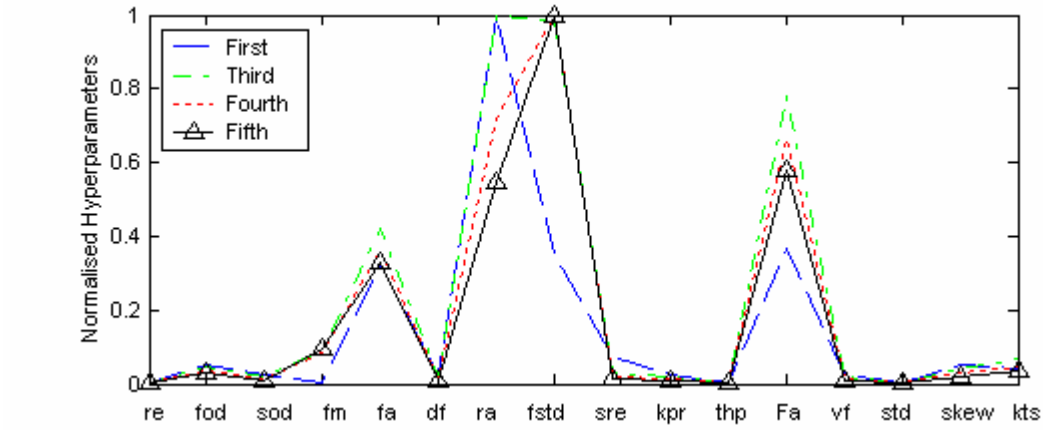
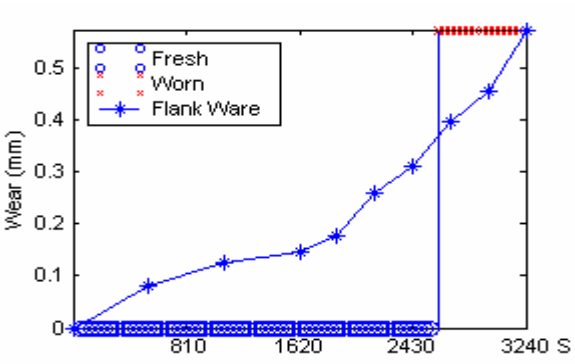
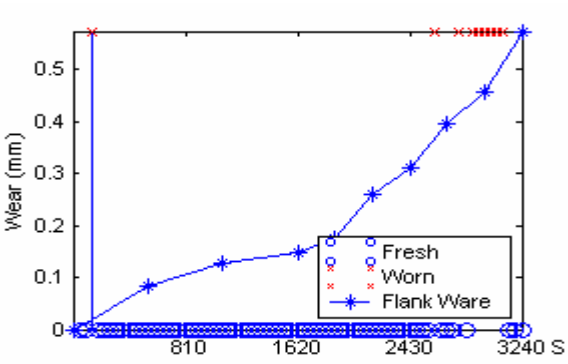


Figure D19. Illustration of the Feature Selection Processes of Test_b8
Cutting conditions: spindle speed = 800rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

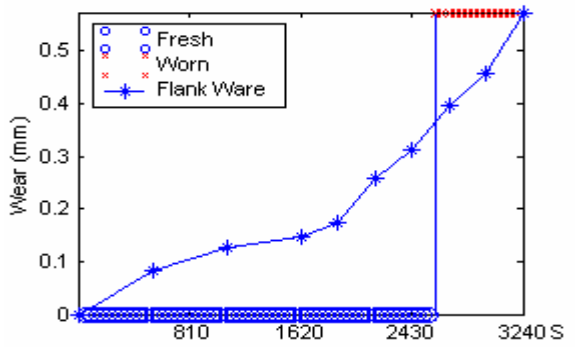
Appendix E
Tool Wear Recognition Results



(a)

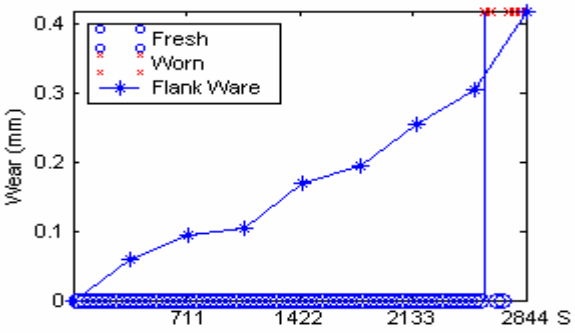


(b)

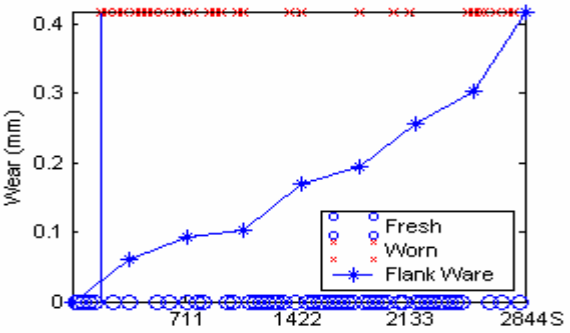


(c)

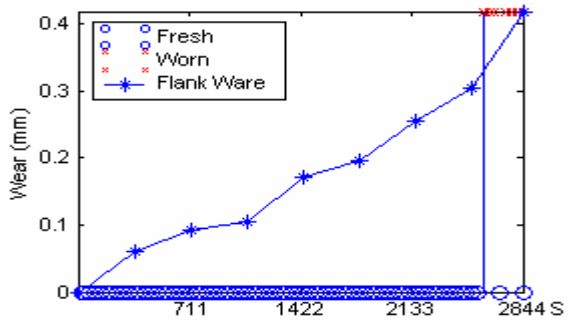
Figure E1. TWR Results of T2
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.



(a)

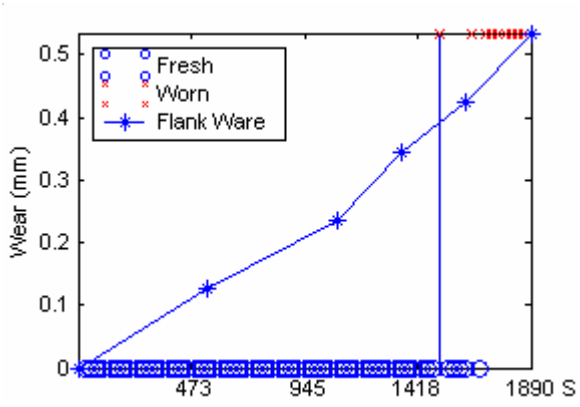


(b)

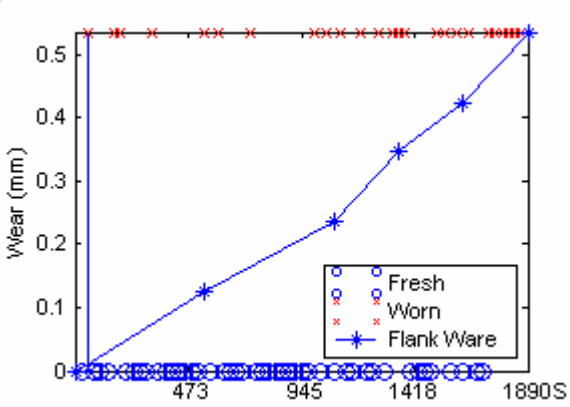


(c)

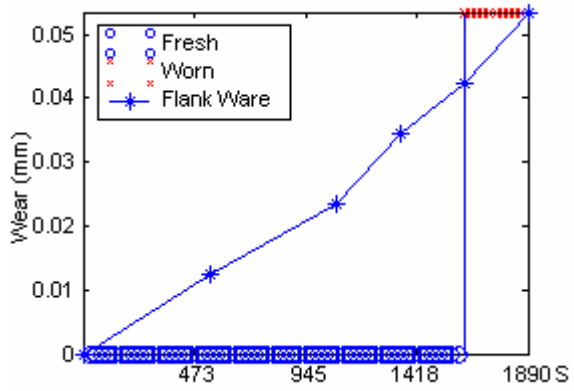
Figure E2. TWR Results of T3
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.



(a)

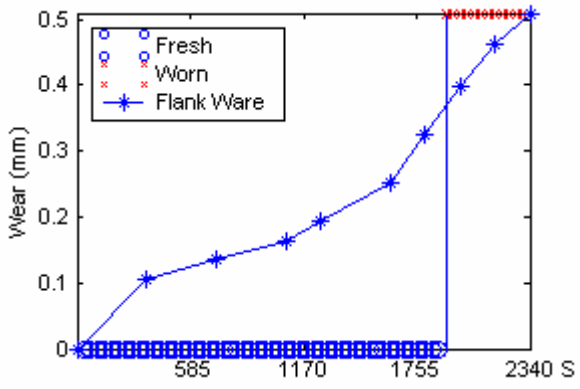


(b)

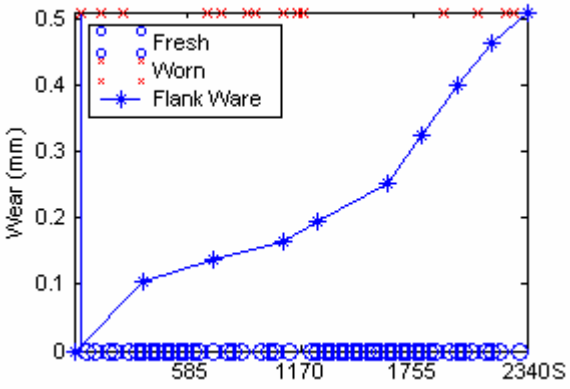


(c)

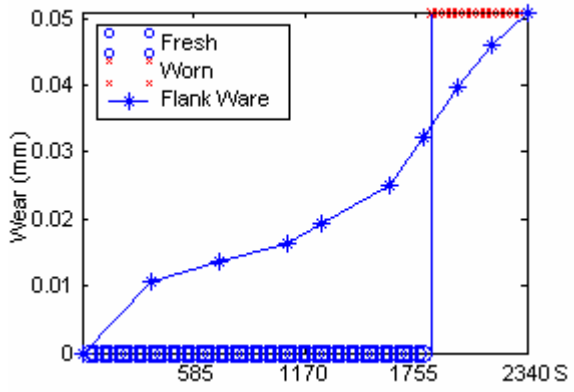
Figure E3. TWR Results of T4
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: AC325.



(a)

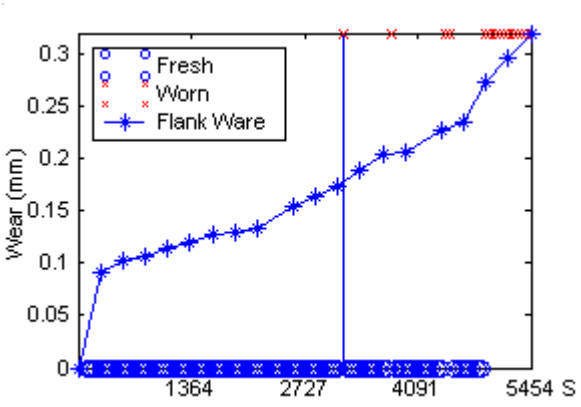


(b)

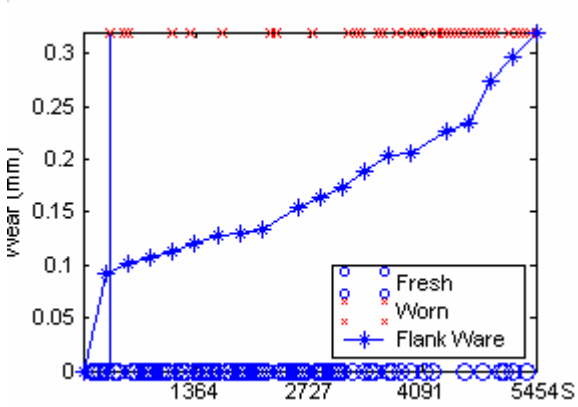


(c)

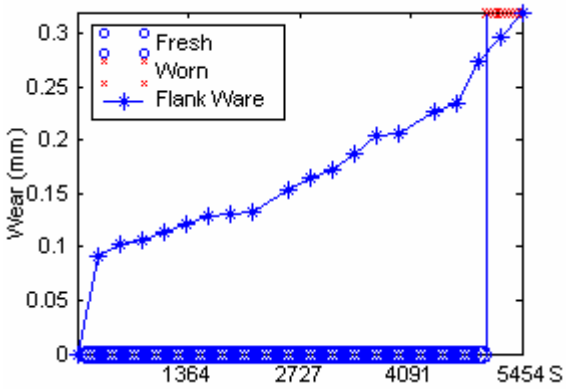
Figure E4. TWR Results of T5
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1200rpm, feed rate = 300mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: AC325.



(a)

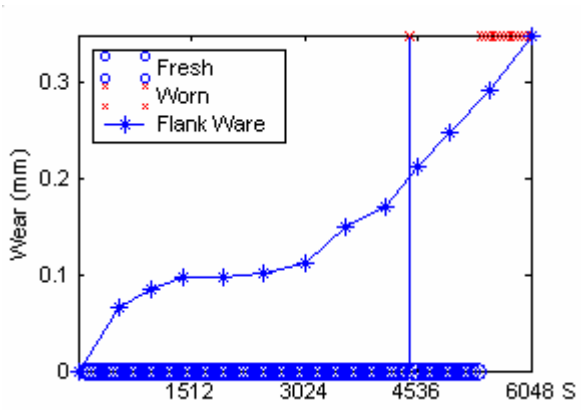


(b)

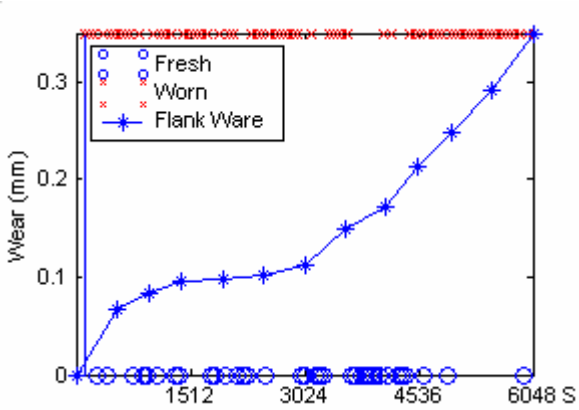


(c)

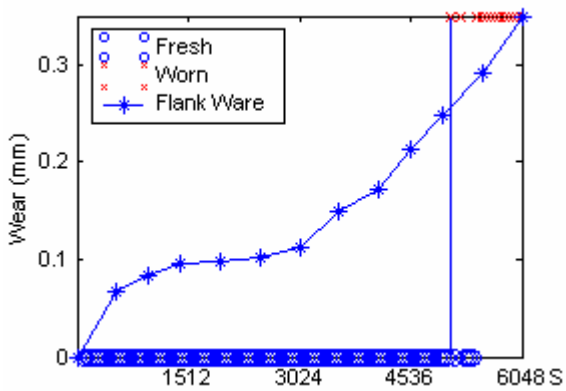
Figure E5. TWR Results of T6
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 600rpm, feed rate = 200mm/min, depth of cut = 2 mm, insert number = 4, immersion rate: FULL, insert type: AC325.



(a)

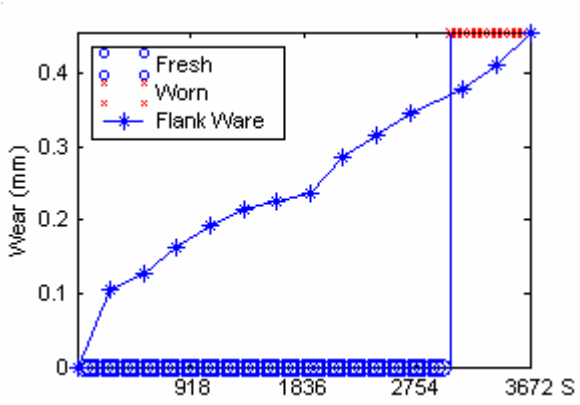


(b)

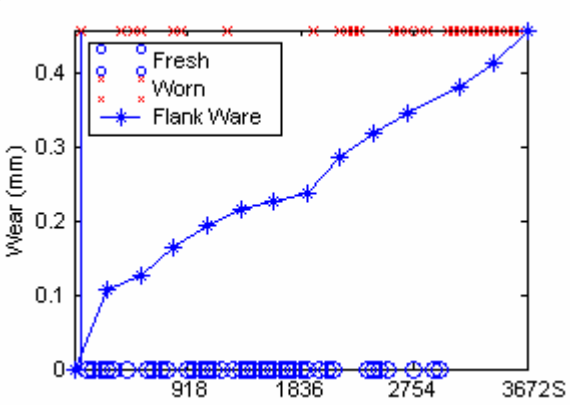


(c)

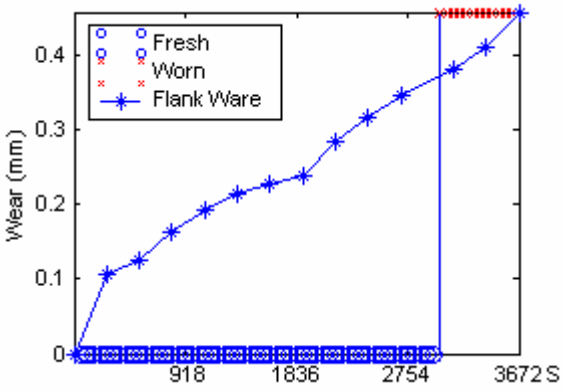
Figure E6. TWR Results of T7
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 800rpm, feed rate = 100mm/min, depth of cut = 2 mm, insert number = 2, immersion rate: FULL, insert type: AC325.



(a)

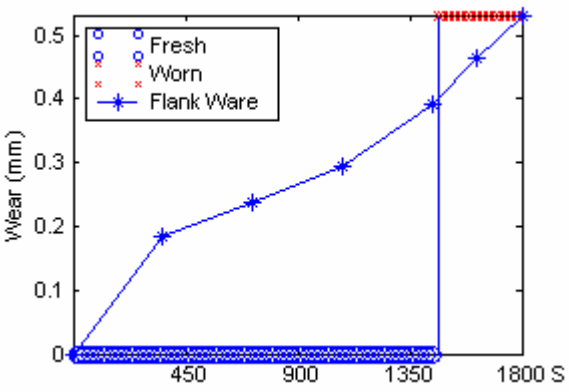


(b)

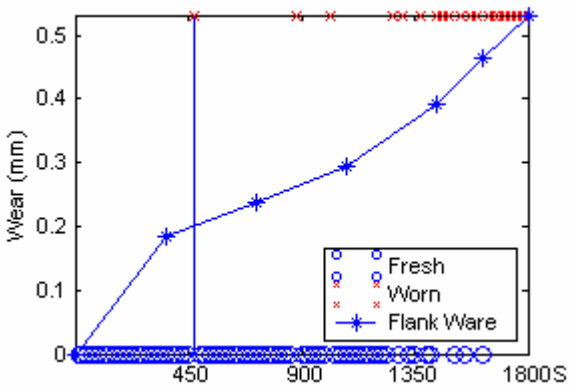


(c)

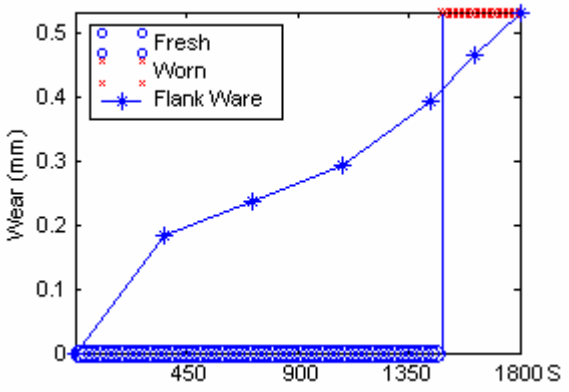
Figure E7. TWR Results of T8
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 200mm/min, depth of cut = 1mm, insert number = 2, immersion rate: FULL, insert type: A30N.



(a)

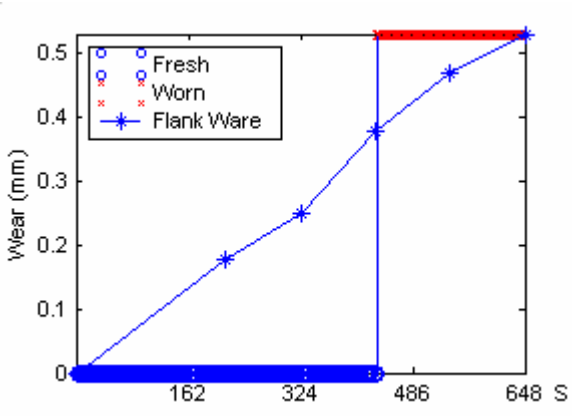


(b)

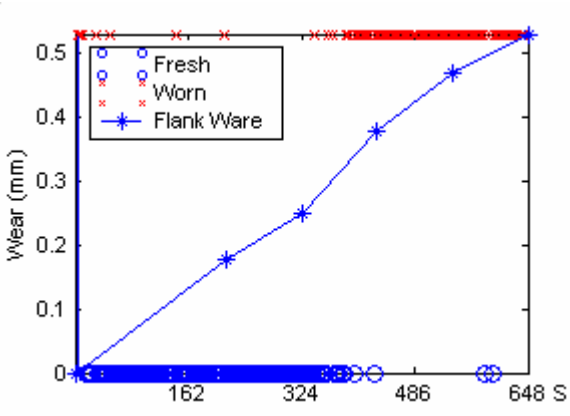


(c)

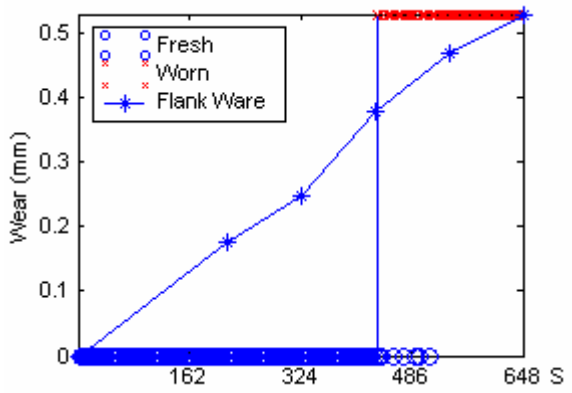
Figure E8. TWR Results of T9
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1000rpm, feed rate = 300mm/min, depth of cut = 2mm, insert number = 4, immersion rate: FULL, insert type: A30N.



(a)

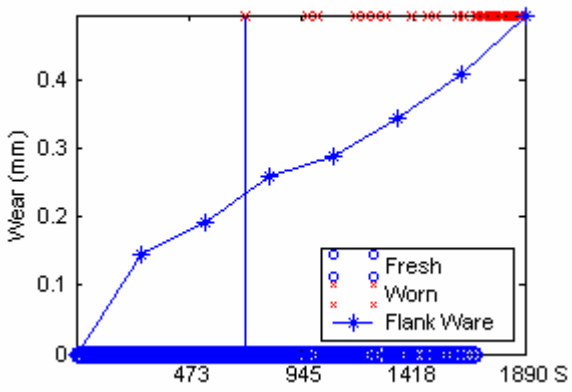


(b)

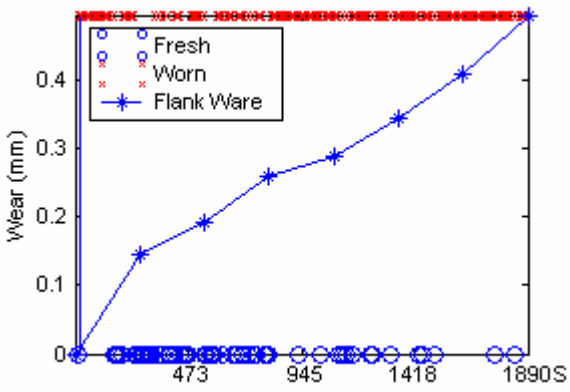


(c)

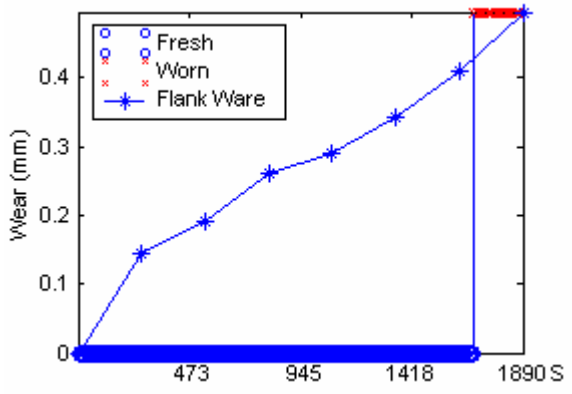
Figure E9. TWR Results of T10
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1200rpm, feed rate = 100mm/min, depth of cut = 1 mm, insert number = 2, immersion rate: FULL, insert type: A30N.



(a)



(b)



(c)

Figure E10. TWR Results of T11
(a) Entire, (b) Rejected, (c) Selected Set
Cutting conditions: spindle speed = 1200rpm, feed rate = 200mm/min, depth of cut = 1 mm, insert number = 4, immersion rate: FULL, insert type: A30N.

Appendix F Miscellaneous

Figure F1. Force Measurement System

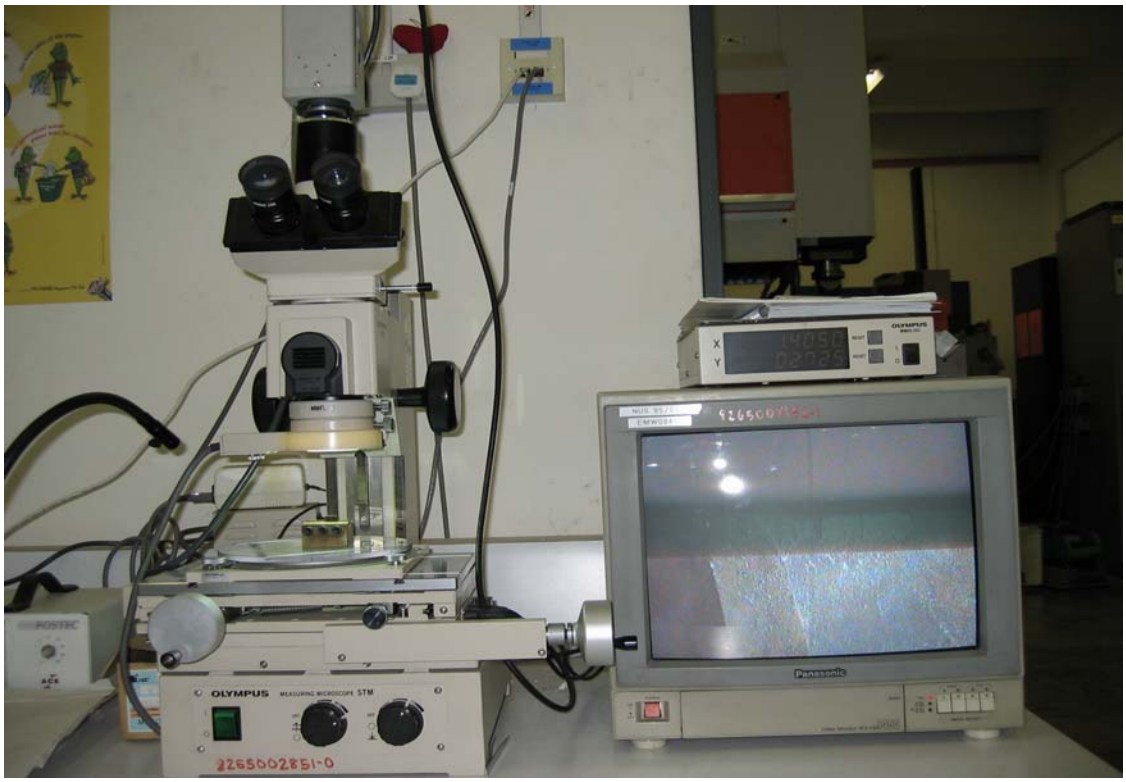
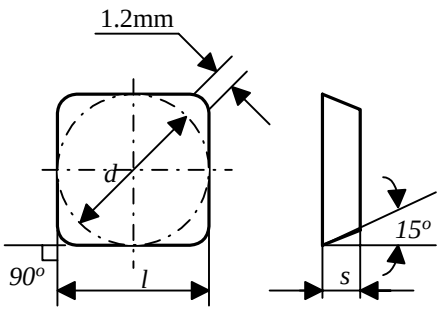
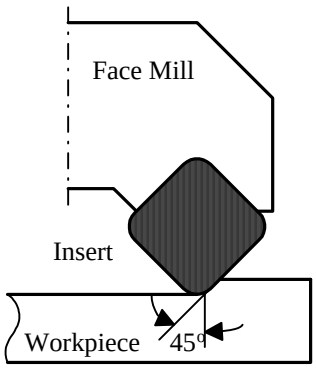


Figure F2. Tool Wear Measurement System



Type: Sumitomo SDKN42MT	
Relief Angle = 15°	Cutting Edge Length *l* = *d* = 12.7mm
Thickness *s* = 3.18mm	Nose Width = 1.2mm

Figure F3. Insert Geometry



Type: TUNGALOY EGD4450R	
Rake Angle: A.R. = +15°, R.R. = -3°	Cutting Diameter = 50mm
Number of Inserts = 4	Stock: Right Hand

Figure F4. Face Mill Geometry

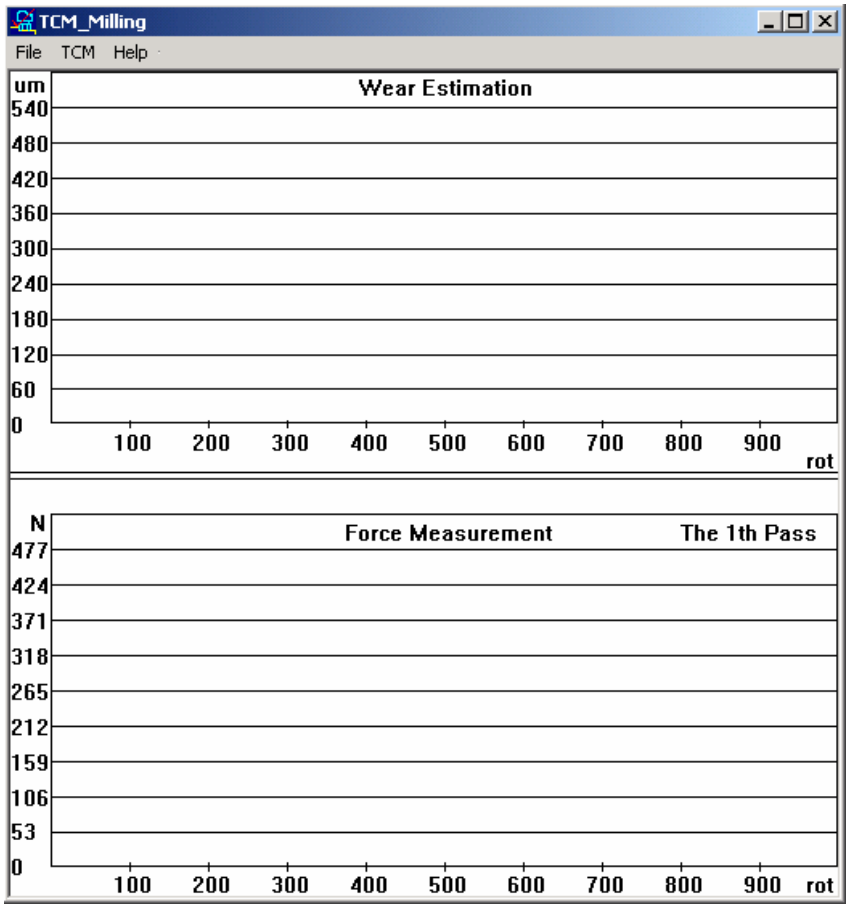


Figure F5. View Window of the Online TCM Software

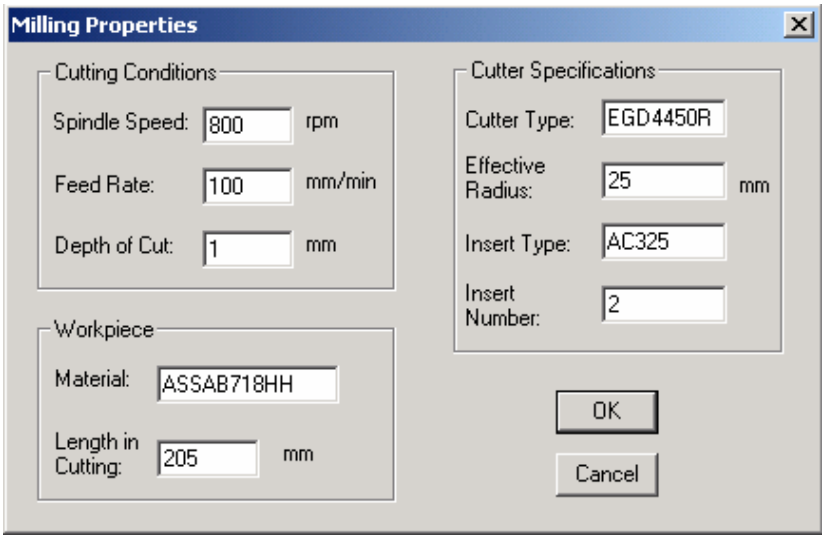


Figure F6. Milling Properties Dialog

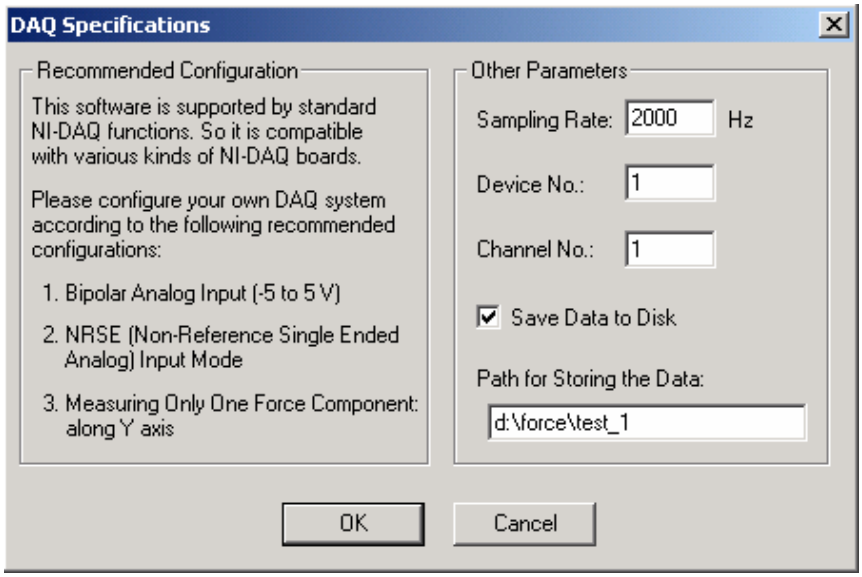


Figure F7. DAQ Specifications Dialog

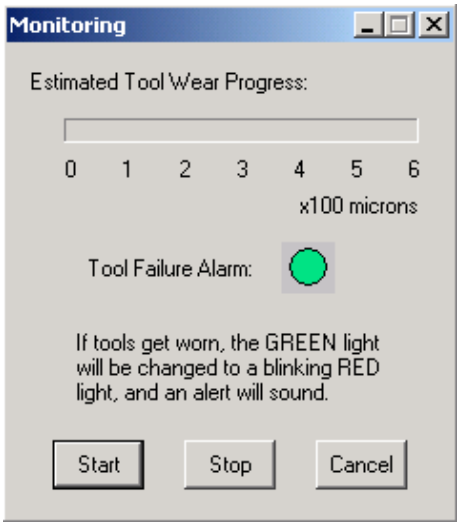


Figure F8. Monitoring Dialog

Tool Condition Monitoring Report

Cutting Conditions:

Cutter Type: EGD4450R Insert Type: AC325 Insert Number: 2

Depth of Cut: 1.0 Feed Rate: 200.0 Spindle Speed: 1000.0

Workpiece Material: ASSAB718HH

Tool failure has been detected at the 48230th rotation.

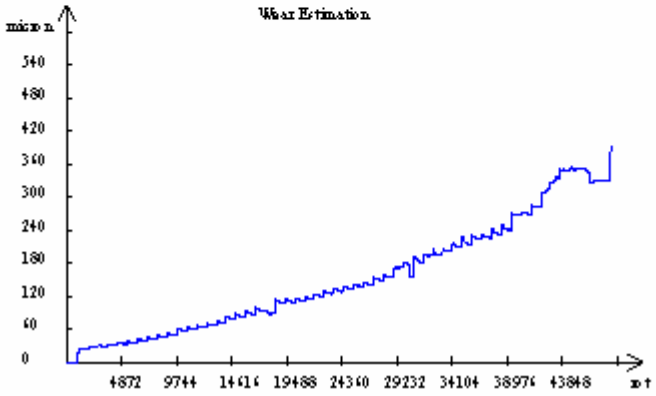


Figure F9. Print TCM Report

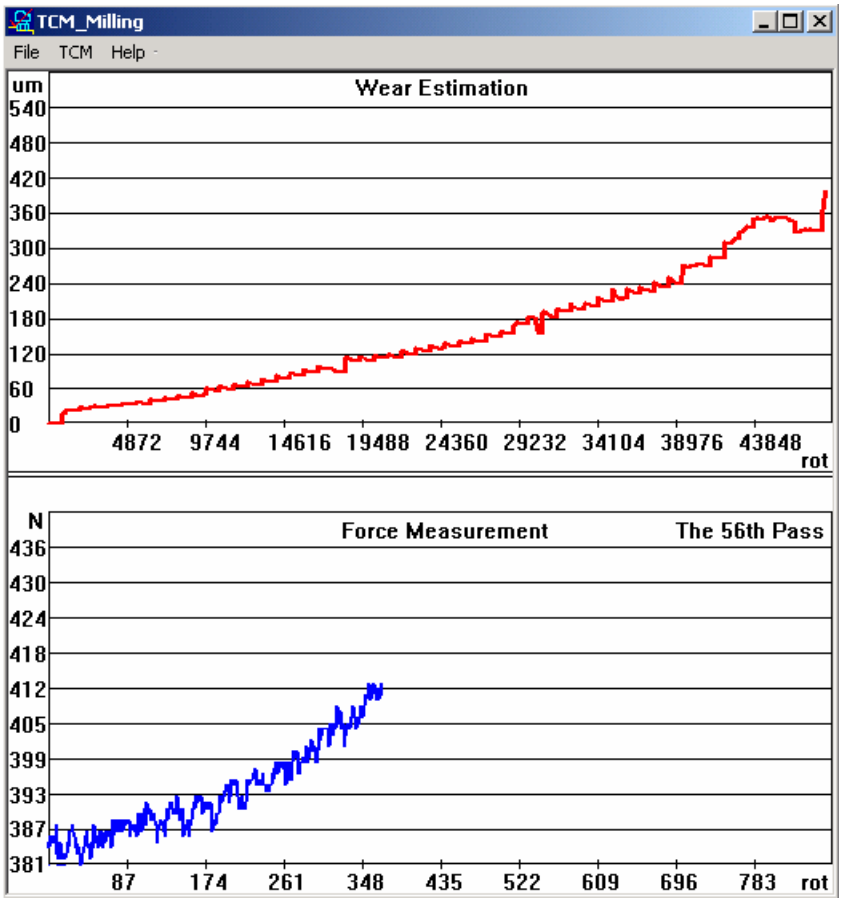


Figure F10. View Window under Working