

Correlation between Machining Monitoring Signals, Cutting Tool Wear and Surface Integrity on High Strength Titanium Alloys



A Thesis submitted to the University of Sheffield
for the degree of Doctor of Engineering in the Faculty of Engineering

by

Nathan Ray

Department of Mechanical Engineering

University of Sheffield

June 2018

ACKNOWLEDGEMENTS

The last four years would not have been possible without such a positive, supportive, encouraging, and kind supervisory team and research group. More specifically, I would like to thank my academic supervisors, friends, and now colleagues, Dr. Elizabeth Cross and Prof. Keith Worden for their consistent help, advice, and insight, which have been key in enabling this work. From in-depth discussions to mid-afternoon coffee breaks, I honestly couldn't have asked for more from a supervisory team. I would also like to thank the wider Dynamics Research Group for providing such an effective research environment, for the numerous insightful conversations had, and for broadening my knowledge of other research areas. Tim Rogers also deserves thanks for his help and advice with the GP code used in this thesis.

I never intended to follow the path of postgraduate study. I had always been set on an industrial career following my undergraduate degree; however, throughout my bachelors there was one key individual who changed my mind and introduced me to the world of research. Professor Brad Wynne is to thank for that initial introduction, after recommending I explore the EngD program and personally introducing me to Professor Sam Turner at the AMRC. It was following discussions with both Brad and Sam that I decided to focus on research, and I am extremely grateful to Lizzy, Keith, and Sam for giving me the chance to take on this project.

Both Germain Forgeoux and Jean-Paul Villain-Chastre from Safran Landing Systems have been instrumental in shaping and providing funding for this work, without which it would certainly not have been possible. I'd like to also thank them both for being so welcoming and accommodating within their premises in Gloucester.

This work was co-funded through the EPSRC Industrial Doctorate Centre (IDC) in Machining Science (EP/I01800X/1) and by Safran-Landing-Systems. I would like to specifically thank both Andrea Haworth and Clare Clarke from the IDC for their continued support behind the scenes for this work, which is vital to the success of all EngD program projects.

Lastly, but by no means least, I cannot express my thanks enough to my family and friends. You have provided continuous support and have all contributed to maintaining my happiness and well-being during the last four years. Particular thanks goes to Jess for her continued positivity and belief in me, and for always making time for me when I need it.

ABSTRACT

It is widely accepted that tool wear has a direct impact on a machining process, playing a key part in surface integrity, part quality, and therefore, process efficiency. By establishing the state of a tool during a machining process, it should be possible to estimate both the surface properties and the optimal process parameters, while allowing intelligent predictions about the future state of the process to be made; thus ultimately reducing unexpected component damage. This thesis intends to address the problem of tool wear prediction during machining where wear rates vary between components; for instance, due to the relatively large size of the component forging and, therefore, inherent material variations when compared to existing research. In this case, the industrial partner, Safran Landing Systems, is interested in the ability to predict tool wear during the finish milling of large, curved, titanium components, despite differing material properties and, therefore, tool wear rates.

This thesis is split into four key parts, the first of which describes in detail the formulation and implementation of an experimental procedure, intended to provide a working set of industrially representative monitoring data that can be used throughout the remainder of the thesis. This part includes development of a relevant machining strategy, material specimen extraction, sensor selection and placement, and 3D tool geometry measurement, all of which have been completed at industrial partners' facilities. It finishes with a preliminary investigation into the data collected during the machining process from the tools, material specimens, and sensors placed in close proximity to the cutting zone.

The second, third, and fourth parts follow logically from one-another, beginning with a state classification problem, and ending with a full dynamic model prediction of wear during the machining of large landing gear components; this method, however,

is applicable to many other machining scenarios using the new technique applied in this thesis. The state classification chapter is a necessary first step in developing a predictive model, as it aims to prove the data is indeed separable based upon the generating wear state. Once confirmed, given the sequential nature of tool wear, the order of observations can be included in the modelling, in an attempt to improve classification accuracy. This forms the basis of the state-tracking chapter, and leads naturally into the full dynamic model prediction in the final part. This is a promising result for the machining community, as process monitoring often relies on operator expertise to detect wear rate fluctuations and, in turn, results in over-conservative tool usage limits, adding time and expense to many complex machining processes. Wear prediction also presents the opportunity to predict part quality through pre-existing relationships between the acquired signals and material surface finish - correlations which are explored and presented as part of this thesis.

The solution to predicting a varying wear rate within a harsh machining environment introduced in this thesis is based around the application of a Gaussian process (GP) NARX (Nonlinear Auto-Regressive with eXogenous inputs) model borrowed from the machine learning prediction and, more recently, structural health monitoring (SHM) communities. The GP-NARX approach is found to be well suited to the application of wear prediction during machining, and forms a promising contribution to the development of autonomous manufacturing processes.

Contents

1	Introduction	1
1.1	Condition monitoring	1
1.2	Industrial context	2
1.2.1	Safran Landing Systems	2
1.2.2	Titanium-5Al-5Mo-5V-3Cr	4
1.2.3	Ball nose finish milling	5
1.3	Motivation	7
1.4	The engineering doctorate	9
1.4.1	Aims and objectives	9
1.5	Brief outline of thesis	10
2	Literature review	11
2.1	Industrial monitoring applications	11
2.2	Sensor selection	14
2.3	Industrial monitoring equipment	20
2.4	Signal processing	22
2.5	Classification techniques	23
2.6	Sequential modelling	25
2.7	Predictive modelling	25
2.8	Overview of literature	26
3	Experimental methods	29
3.1	Machining trial concept	29
3.2	Specimen extraction	32
3.3	Machining process parameters	34
3.4	Tool path analysis and programming	35
3.4.1	Trial element One - tool wear	35
3.4.2	Trial element Two - data collection	38
3.5	Sensor selection and data acquisition	40

3.6	Machine specifics and dynamical testing	43
3.7	3D tool geometry measurement	46
3.8	Overview of data collected	48
4	Preliminary data analysis	49
4.1	Acoustic emission data preparation	49
4.2	Time domain feature extraction	52
4.3	Material specimen features	56
4.4	Tool scan features	58
4.5	Correlations between material, tool, and AE data	67
4.6	Other acoustic emission-related features	70
4.7	Principal component analysis	71
4.8	Overview of preliminary analysis	73
5	Tool state classification	75
5.1	Classification methods	75
5.2	Support vector machines	79
5.2.1	SVM overview	79
5.2.2	SVM mathematical formulation	80
5.3	Training and testing sets	83
5.4	Results	83
5.5	Overview of tool state classification	92
6	Tool state tracking	93
6.1	Markov models	93
6.2	Hidden Markov models	97
6.3	Training and state sequence prediction	100
6.4	Results	101
6.5	Overview of tool state tracking	107
7	Tool state prediction	109
7.1	NARX models	109
7.2	The Gaussian process	111
7.3	GP-NARX	116
7.4	Results	117
7.5	Overview of tool state prediction	125
8	Conclusions and future work	127

8.1	Experimental data collection	127
8.2	Wear modelling and prediction	129
8.3	Industrial implementation	131
8.4	Limitations	132
8.5	Future work	133
Bibliography		135
A Coupon technical drawing		147
B Kistler 8152B1 datasheet		149
C Kistler 9255B datasheet		155
D PCB 355B02 datasheet		159
E Tool inspection fixture		161

INTRODUCTION

This chapter begins by introducing condition monitoring and its application within machining. This is then followed by presenting the specific industrial context of tool condition monitoring used throughout this thesis, the motivation behind this research, and finally a summary of the layout of the remaining chapters.

1.1 Condition monitoring

Condition monitoring is the process of observing the changes in parameters relating to the state of a system or process, often in order to identify or quantify a deviation from normal operating conditions. Such strategies are often implemented as an alternative to corrective and preventative maintenance, in an aim to reduce cost and give better process understanding, ultimately leading to predictive maintenance strategies. A common application of condition monitoring is to monitor the status of a piece of machinery through parameters such as vibration, temperature or power usage to name a few [1, 2]. Monitoring of electrical motors is one of the most common applications of condition monitoring [1], along with other items of rotating machinery, for example wind turbines [3], which use carefully placed sensors to measure parameters of interest.

Condition monitoring is especially useful in situations where access to machinery may be restricted or impossible when in operation, as it provides sensor data which can be accessed at a safe distance or from a remote location. A prime example of

such a scenario is the monitoring of a cutting process during machining of metallic components, where an operator may find it beneficial to observe a number of phenomena, measured within close proximity to the cutting zone, in order to ensure safe working levels and avoid potential damage [4]. One of the most critical factors in machining is tool wear [5], a subject that has been covered extensively in previous literature; however, a solution for accurate and reliable diagnosis across a range of machining processes and conditions is still not available. This reality will be discussed further in the following chapters, yet it provides an adequate starting point to discuss the need for further research effort in the area of tool condition monitoring.

1.2 Industrial context

1.2.1 Safran Landing Systems

Safran Landing Systems (SLS) is a subsidiary of Safran S.A., a French-born company specialising in aerospace design, manufacture, maintenance and defence. While Safran has a wide range of interests in the aerospace sector, SLS has a primary focus on aircraft landing gear. More specifically, SLS manage the entire life-cycle of a landing gear from design and development to manufacture, installation, and after-care.

As suppliers to a number of major aircraft manufacturers such as Airbus and Boeing, SLS have a large focus on research and development, ensuring that their products and processes are constantly improving and developing to meet the demand of increased efficiency both in service and during production. As such, they have been realising a steady transition towards the use of titanium components in preference to steels over the past fifty years, much in keeping with the rest of the aerospace sector. Figure 1.1 shows how the percentage by weight of titanium has been steadily increasing for common Airbus and Boeing aircraft over the last seventy years.

New civil aircraft platforms such as the Boeing 787 Dreamliner and Airbus A350 XWB, bring new manufacturing challenges, such as the ability to machine titanium alloys at rates that meet the increasing demands of reduced cost targets, increased volume and increased process capability. For the advanced manufacturing sector,

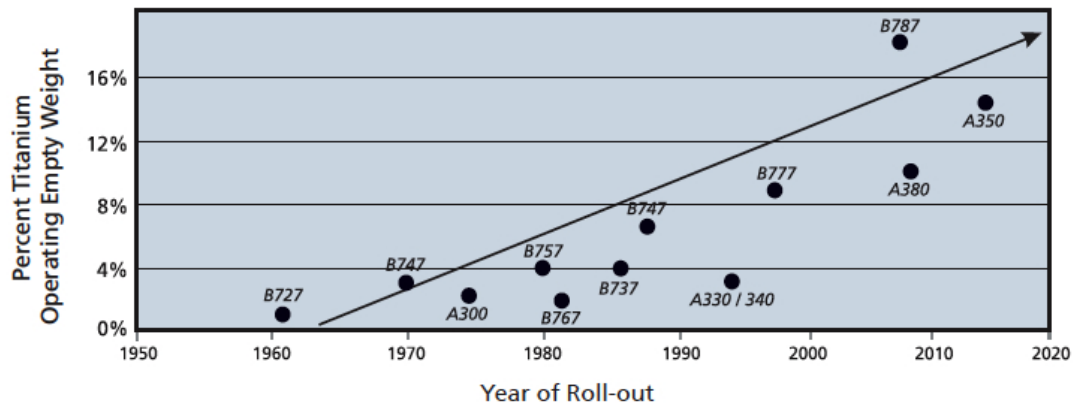


Figure 1.1: Plot to show increasing use of titanium in common aircraft over seventy years [6].

this global drive towards increased productivity equates not only to higher surface speeds and machining rates, but the need for better process understanding and the ability to avoid unnecessary machine downtime. Safran Landing Systems develop, produce, and maintain both the nose and main landing gear for the 787 (Figure 1.2), with approximately 50-60% of the total cost of production of these components resulting from the machining stage [7]. It is therefore evident, that any time saving during machining will have a significant impact on the overall part cost to manufacture.

One product in particular, the 787 truck beam, is machined from a solid forging of Ti-5Al-5Mo-5V-3Cr, which brings with it new challenges due to the lower machinability of this material when compared to the high-strength steels used previously [9]. Such factors include the low modulus of elasticity, hardening characteristics, low thermal conductivity, and chemical reactivity at elevated temperatures [9]. Another example of one of the challenges faced when machining this alloy comes from the forging process. The forgings used for the 787 truck beams are some of the largest titanium forgings in the world, and as a result, can have a slight variation in material properties from end-to-end and around the circumference. This presents an issue when machining, as calculating tool life and cutting parameters is directly dependant upon the material properties. Varying properties result in conservative cutting conditions and an element of the unknown when predicting tool life, as the harder the material, the faster the tools wear [9].

For Safran Landing Systems, removing this uncertainty and maximising tool use is of the utmost importance as it brings with it direct cost-saving benefits and



Figure 1.2: Image of a Boeing 787 landing gear [8].

improved process efficiency. Utilising tools to their full potential without operator intervention makes for a more automated, streamlined process, which is the end goal for any production environment. By making use of in-process monitoring and observed relationships between process variables, SLS aim to be able to perform online prediction of a tool's state while gaining an understanding of the material properties of each individual truck beam that leaves the production line - something which has not yet been successfully implemented in the industry.

1.2.2 Titanium-5Al-5Mo-5V-3Cr

The ongoing trend towards increased titanium usage in the aerospace sector is down to a number of key factors. Despite considering the costly and time-consuming manufacturing processes required for titanium components, titanium does have the advantage that it is currently the fourth most abundant metallic element in the Earth's crust [10]. Not only is titanium less dense than the steels it is replacing, it also has a higher ultimate tensile strength (UTS) than most commonly-used steels. Table 1.1 shows a comparison between common steel and titanium alloys [11].

It is both its beneficial strength-to-weight ratio and its fatigue resistance that are really driving the push behind titanium components, particularly in landing gear applications. For the majority of flight times, be it short or long haul, the landing

Material	Yield strength (MPa)	Ultimate tensile strength (MPa)	Density (kg/m ³)
Titanium 5Al-5Mo-5V-3Cr	1310	1364	4500
Titanium 10V-2Fe-3Al	1135	1240	4816
Titanium 6Al-4V	895	975	4430
Steel AISI 4130	951	1110	7850
Steel ASTM A514	690	760	7800

Table 1.1: Comparison between mechanical properties of common steel and titanium alloys [11].

gear is considered dead weight, serving no purpose other than a brief operation at takeoff and landing. Weight saving is therefore highly significant, as it has a considerable impact on overall efficiency. The landing gear, however, is also a safety-critical component, supporting the entire airframe when grounded and experiencing large impacts upon landing, and as such, fatigue performance is the dominant factor in design.

More specifically, the 5Al-5Mo-5V-3Cr (Ti-5553) alloy is that chosen by SLS for a number of structural landing gear components. Ti-5553 was developed jointly between VSMPO and Boeing in the mid 1990s, and was intended as a higher strength, lower cost alternative to the existing 10V-2Fe-3Al alloy [12]. It has the structural benefits of being less prone to segregation, having better hardenability in thick sections (lowering machining costs), being air cooled rather than water cooled (less distortion) during forming and retains an approximately 15% higher strength [12]. On the other hand, when considering machining, Ti-5553 does have the major downfall that current cutters experience roughly 10-20% less tool life than when used to machine the more common 6Al-4V alloy [12], leading to considerable effort being put towards maximising tool use and hence the requirement of this project.

1.2.3 Ball nose finish milling

Within the machining stage of the aforementioned truck beam, there are both roughing and finishing stages, each containing seven sub-categories of operations. The seven categories of interest are defined as (which are also highlighted in Figure 1.4):

- Surfaces to centre / quads,
- Steady bands,

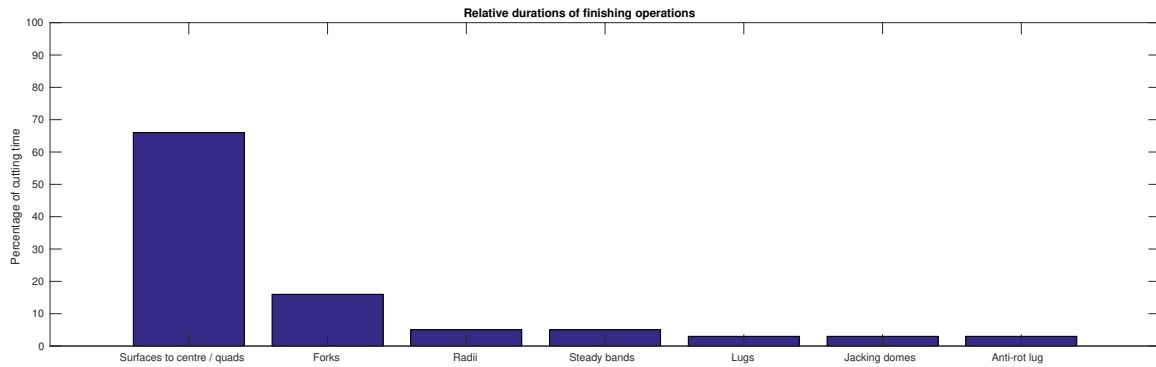


Figure 1.3: Plot showing relative cutting durations for finishing operations.

- Radii,
- Forks,
- Jacking domes,
- Lugs,
- Anti-rot lug.

Between these machining zones, 51.6% of the initial forging weight is removed, made up of 33% from roughing and 18.6% finishing, or similarly 63.8% and 36.2% respectively of the total material removed. While it appears at first that there is much more to be gained from streamlining the roughing stage, in reality, the machine time taken for roughing and finishing is very similar due to the larger depth of cut in roughing compared to finishing. In practice, roughing takes approximately 57% of the cutting time as opposed to 43% taken by finishing. Again, one would easily make the conclusion here that there are greater savings to be made by focussing research effort on roughing, however, this is still not the case. In roughing, each process uses a different tool, resulting in a large number of individual operations to optimise. On the other hand, finishing uses only two different tools and uses these at a ratio of 95% to 5% - 20mm ball nose and radius end mills respectively. It is therefore evident that the finishing operation using the same tool for 95% of its duration, provides the greatest return in terms of potential savings per operation optimisation. Figure 1.3 shows the relative cutting times per finishing operation, with the surface to centre/quads procedure being of the most interest. The areas of the truck beam machined during this operation are shown in Figure 1.4.



Figure 1.4: Image showing, in dark blue, the areas of a 787 truck beam machined during the surfaces operation. Other operations of interest are also indicated.



Figure 1.5: Image of 20mm ball nose tool with cutting regions indicated.

More specifically, this operation uses a 20mm solid carbide ball nose end mill with six cutting flutes, an example of which can be seen in Figure 1.5. Specific details of this tool will be discussed in Chapter 3; however, it is worth noting here that the useful cutting radius of the tool is limited to that region where all six flutes are used, excluding the tool tip where the flutes converge (shown in green and red respectively).

1.3 Motivation

Over the past decade or so, Acoustic Emission (AE) has received, and continues to receive, a significant level of interest in the field of condition monitoring and process

control. In machining, AE is caused by phenomena such as sliding friction at a flank-workpiece interface and the breaking of chips [13]; both of which are directly related to a cutting process and corresponding tool condition. AE therefore, has been used in a considerable amount of previous work for the monitoring of tool condition [14–19], surface roughness [20,21], and event detection [13], to name but a few. AE-based monitoring is also available from a number of commercial monitoring system suppliers [22]; however, it is often used alongside a range of different sensors in a data fusion approach [13], the benefits of which include increased reliability, more robust decision making, and increased noise rejection. Such commercial systems are rarely successfully implemented, however, as they tend to target a wide range of processes with simple data processing methods and often fall short of the requirements set by industry; they also follow a diagnostic approach rather than prognosis of events before they happen [22]. This thesis aims to provide a solution to future tool wear prediction, which has not before been possible with existing tool measurement and predictive modelling techniques.

In finish milling, the focus is largely on the surface of the part. Required roughness values are regularly given in product specifications, and maintenance of the surface finish within certain limits is considered crucial. As tool condition plays a significant role in surface generation, tools are therefore, often replaced conservatively and rarely experience breakage. It is consequently desirable to determine surface properties and tool state in-process, made possible using measured AE and the correlation between tool condition and workpiece surface integrity. By building relationships between the tool geometry, AE, and the workpiece surface, it is believed to be possible to begin to predict the future state of the process and make intelligent decisions about tool replacement, for instance. Recent developments in 3D scanning of tools, allow a wide range of data to be collected relating to the tool's condition, providing a firm foundation for exploring correlations with gathered AE signals and in-process prediction of wear rates.

Historically, tool wear has been measured by examining the wear scar on a given tool through the use of either a microscope or stylus-based instrument [5, 23, 24]. In the case of flank wear (the most commonly-used wear type for measurement), this is generally adequate, however, these techniques have the distinct disadvantage that they only provide two-dimensional data. Considering that flank wear is only one of a number of different wear types [5] (which will be explored further in the following chapter), these methods are limited in the insight they can provide when wear affects

further regions of the tool. State of the art three-dimensional tool measuring has enabled this thesis to gather a greater understanding of cutting edge wear than has previously been possible, allowing quantification of wear features and averaging of a large number of scans along the length of a single cutting edge.

1.4 The engineering doctorate

The engineering doctorate (EngD) is a four year long programme that aims to provide PhD-level research projects to postgraduate engineers, with the added requirement that projects are based around a real business case as identified by an industrial partner. Each EngD project consists of a first year containing taught modules and background research, followed by three years that follow a similar structure to a more typical PhD. In the case of the Industrial Doctorate Centre (IDC) in machining science, the research engineer spends approximately 25% of their time at university, with the remaining 75% spent working directly with the projects sponsoring company and the Advanced Manufacturing Research Centre (AMRC) with Boeing.

This research project has been co-funded by the EPSRC (EP/101800X/1) and Safran Landing Systems - the project's industrial sponsor. The first year of this specific project has included three mini-projects focussing on exploring previous literature, design and implementation of an industrially relevant machining trial, and preliminary analysis of acquired data, which have then formed the basis for the remainder of the work presented in this thesis.

Given the applied and industrial nature of this project, the main aims and objectives are:

1.4.1 Aims and objectives

- To produce direct production benefits including cost savings while maintaining part quality.
- To gain an understanding of the correlations between monitoring signals, cutting tool state, and material surface condition,

- To provide a method of on-line prediction of tool state.
- To ensure that through intelligent process monitoring, cutting tools are fully utilised whilst maintaining surface integrity.

1.5 Brief outline of thesis

- Chapter 2 reviews a range of previous works conducted in the condition monitoring, machining, and machine learning communities. The relevance to this thesis is explored and the suitability for application to tool wear monitoring is discussed.
- Chapter 3 presents an experimental trial design and implementation, used to provide an appropriate dataset for use throughout the remainder of the thesis.
- Chapter 4 details the preliminary processing and analysis of the data collected in Chapter 3. Useful features are identified and extracted from recorded AE data, tool scans, and material analysis.
- Chapter 5 explores a static classification problem, separating observed data points into a number of discrete clusters through the use of a Support Vector Machine (SVM).
- Chapter 6 builds upon the findings of Chapter 5 to take advantage of the sequential nature of tool wear. It explores the use of Markov models to probabilistically predict wear states.
- Chapter 7 introduces the Gaussian process (GP) Nonlinear Auto-Regressive with eXogenous inputs (NARX) model and applies it to the experimental dataset in order to predict wear level with natural confidence bounds.
- Chapter 8 concludes the thesis. Industrial implementation is discussed, along with future work.

LITERATURE REVIEW

Following the industrial motivation and context behind this work in Chapter 1, it is paramount to explore and review previous and current work being conducted in the field, to ensure a need for this research, while providing an insight into potential methods of solving the wear prediction problem. This chapter aims to provide such a review of existing work, focussing primarily on monitoring applications within and outside of machining, sensor selection, signal processing techniques, and predictive modelling methods from both the machine learning and structural health monitoring (SHM) communities. It also aims to present a comparison between different works, highlighting some key differences and contradictions in conclusions drawn in them, as is often the case in tool wear prediction research [13].

A convenient starting point here is to highlight a number of existing reviews of current monitoring technologies within the context of machining. Three of the most appropriate are those in [13], [25], and [26]; each maintaining a slightly different focus, yet all sharing many of the same ideas in relation to sensor selection and data processing methodologies.

2.1 Industrial monitoring applications

In the context of machining, monitoring schemes have been primarily implemented to infer, through some observed parameter, the state of wear of a cutting tool, as found in references [15, 19, 22, 24, 27–31] to name a few. As an extension of this, monitoring

schemes tend to aim to detect collisions and tool breakage also, alongside surface roughness [20, 32], given the critical nature of these features to final part quality. More often than not, due to the harsh environmental conditions machining processes occur under, these features are inferred through indirect means, as opposed to being measured directly, a detail that will be explored further in the following section.

It is worth drawing attention at this stage, to a number of different wear types experienced by cutting tools during machining. In general, most previous work in the area of tool wear monitoring focuses on flank wear [5], due to it being the most common and preferred wear type (given its ease of measurement and ability to regrind tools). There are, however, a range of wear types, which if not considered, may be cause for monitoring strategies to be suited to only a handful of processes, or worse, fail entirely. A typical example of this would be a blunt, used tool, chipping, to form a sharper cutting edge. While the tool has not become less worn, only measuring the edge radius could cause confusion in predictions. Similarly, built-up edge is common during machining of titanium, which also appears to have the effect of reducing the flank wear level through adding material to the tool. Figure 2.1 shows a selection of eleven different wear types, the most interesting in terms of this thesis being flank wear, crater wear, built-up edge, and gross fracture. A full explanation of each wear type is given in [5], although it is worth highlighting that the most common tool life prediction model, Taylor's tool life equation [33], predicts only flank wear.

It is readily agreed in the machine monitoring community that there is still no clear and recognised methodology to designing a successful monitoring system that is universally suitable to a range of processes. There is a general methodology of monitoring system development presented in [13], yet this provides very little assistance given its inherent vagueness. At this stage, therefore, such development requires investigation on a per-process basis; a costly and time-consuming operation. An example of this reality can be seen by comparing the work of Haber [34] and Lan [17]. Haber portrays the view that AE sensors attached to soft chuck jaws are more sensitive to wear dependent AE than those connected to the spindle assembly, however, this is in direct contrast to that which is reported by Lan, despite both focusing on flank wear. This appears to be a common finding between individuals, and is easily explainable by the vast differences in experimental processes and test rigs, with sensor locations and test materials playing a key role in data quality. Continuing the AE example above, the recorded values are heavily dependent upon

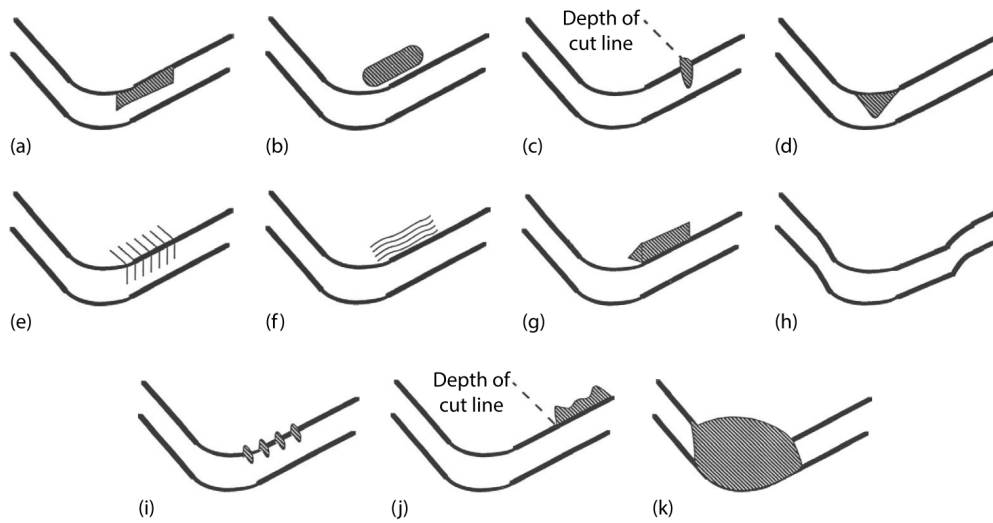


Figure 2.1: Types of wear on cutting tools: (a) flank wear; (b) crater wear; (c) notch wear; (d) nose radius wear; (e) comb (thermal) cracks; (f) parallel (mechanical) cracks; (g) built-up edge; (h) gross plastic deformation; (i) edge chipping or frittering; (j) chip hammering; (k) gross fracture [5]

the transmission path between cutting zone and sensor, with material, machine-workpiece interface, and spindle assembly providing a few key differences between experimental works of the past.

While there are many features of a particular experimental set-up that can be unique, probably the most important to draw attention to, is the distinction between observing a continuous [24] or an intermittent [28] process, in the case of machining, turning or milling respectively. The majority of previous literature focusses solely on one or the other, and despite their similarities in terms of metal cutting, they are fundamentally different; the key distinction being that an intermittent process creates a segmented signal as chips are formed, due to the tool's contact with the workpiece being interrupted. A continuous process, however, maintains constant tool-workpiece contact during a cutting cycle and, therefore, produces a continuous output signal. While this may sound like a simpler process to observe, it is not without its challenges, primarily due to the fact that chip formation and tool pickup can be inconsistent throughout the process. The main consequence for monitoring an intermittent process is that further signal processing is needed to extract the useful in-cut information, and to discard that from the period where the cutter is not in contact with the workpiece [28].

Certain applications lend themselves to successful in-process monitoring system im-

plementation, such as those commonly found in the automotive manufacturing sector. The biggest contrast between such automotive manufacturing processes and those found in machining, is in the relative simplicity and repeatability of the operations involved. Take, for example, two common procedures performed during manufacture; spot welding and hole drilling. Spot welding is a process where two metallic surfaces are joined at a point, by the heat obtained from resistance to a brief electrical current. When compared to a lengthy finish milling process, each individual weld or hole takes a mere fraction of the time to complete. In both cases, detection of wear is much less important than detecting breakage or irregularities between specimens, allowing implementation of a simple, comparative monitoring system [22]. The comparative approach simply compares data or features obtained during a reference operation to that obtained during each subsequent repetition. This method, while perfectly adequate for such simple tasks, has a number of severe limitations. Primarily, as the length of operation grows, so does the volume of data needed for comparison. When considering processes completed over a number of days, this method becomes impractical, not to mention that it relies on the process remaining consistent between datasets used for comparison. This type of system relies on pre-set limits of allowable diversity between measurements, which is fundamentally flawed when considering a process that naturally develops over time, often at differing rates. For the example of tool wear monitoring, this flaw is easily highlighted when considering a process in which wear rate can vary. If a tool wears more quickly than during the training set, an observation at time t may be well above the acceptable deviation from the control set at the same time, despite still being within the tool's overall useful life. This presents a trade off between the accuracy of such systems, and the number of potential false alarms. Jemielniak [22] covers in more detail the commonly applied monitoring strategies in industry, including the use of fixed limits and process signature comparison. The key point to note here is that all of these strategies only provide an indication that the process has changed in some way. If the change is due to tool or component damage, in many applications, this is simply too late to be notified and a predictive technique is required.

2.2 Sensor selection

At this point, it is necessary to draw a distinction between direct and indirect sensing methods. Direct methods can be considered as those which measure some variable

directly, whereas indirect methods are those which infer the state of certain process variables by monitoring a secondary phenomenon which is affected by the variable of interest. As an example, using an optical sensor to measure surface roughness of a machined surface would be considered a direct method, while the use of a dynamometer to measure tool condition through cutting forces is a common application of an indirect method. Direct methods often have the benefit of providing more reliable data over a range of conditions, when compared to indirect methods; however, they are often more expensive and complex in their design. The major drawback of most direct methods is their susceptibility to damage and interference due to the harsh environment in which nearly all manufacturing processes occur, including extremes of temperature, humidity, and high-pressure coolant delivery to the cutting zone. More specifically for this thesis, SLS use coolant delivery pressures of upwards of 80 bar, which not only pose a risk to direct methods from coolant ingress, but also restrict any view of the cutting zone that may be possible during dry cutting. For this reason, the majority of previous work has taken advantage of the accessibility of indirect sensing methods to monitor interesting aspects during the machining process. It is worth noting here that the combination of both methods provides possibly the most complete solution, using indirect methods during machining and direct methods during idle time between operations to reinforce the indirect data obtained. Tools in the tool carousel, for instance, can have their wear state directly assessed to help confirm the understanding that may have been inferred previously, minimising any error between processes. This method provides decision robustness, but if used independently, provides no information of the tool state during machining [25].

Based upon the choice of direct or indirect sensing methods, there is a wide range of different sensing technologies available, depending upon each specific application. There is a general consensus among review authors: Abellan-Nebot [13], Byrne et al. [25], and Teti et al. [26], that cutting force is the most characteristic variable for providing information about the cutting action, chip formation and therefore, wear, and each make the point that the dynamometer often gives the most accurate representation of this. It is unsurprising that the cutting force is considered highly relevant when predicting tool wear and surface finish, as the mechanism results in blunting of the cutting edges as more material is cut, increasing the friction between cutting surfaces. The measured forces, therefore, increase throughout this stage of wear when compared to a fresh tool earlier in the process. Similarly, tool breakage results in a characteristic signature in the cutting force, in the form of a temporary

increase followed by a dramatic drop [35]. The dynamometer has the added benefit of providing data relating to both material surface roughness and vibration generated during the machining process, given that the measured vibrations are often of a higher frequency range than that of the cutting force [13]. Dynamometers are not, however, without their downsides. The main issues following dynamometer implementation in a production environment are their relatively high cost (approximately 40 times that of a typical AE sensor), limited frequency response, and probably most importantly, their intrusive nature into the process [4, 36, 37]. Figure 2.2 highlights the application of a typical dynamometer beneath the fixture used throughout the experimental work in this thesis. As the component and fixture size start to increase to similar scales as found within SLS, it becomes impractical to mount the assembly to such a device, without making the dynamometer so large that it lacks sensitivity to the effects of tool wear on the process.

To overcome this limiting fact, work has been conducted successfully into integrating force measurement into the spindle assembly of a machine tool or between the tool and its holder [38]. This is significantly less intrusive, has a much better frequency response than conventional dynamometer plates, and presents the opportunity to take measurements from a much wider range of operations. The issue of high cost still remains, however, and the fact that the monitoring equipment is now an integral part of the tool-spindle assembly can cause further issues relating to tool changing systems and coolant delivery.

The dynamometer is only one of a number of sensor types nonetheless, some of which prove more advantageous in certain circumstances. Other commonly-used sensors in the context of machining are: power feedback, AE, and accelerometers; depending upon what phenomenon is intended to be observed and quantified. All of these sensors fall into the category of indirect sensing, and an overview of their frequency of use in the previous research literature can be seen in Figure 2.3.

In applications where chatter detection and reduction is critical, accelerometers and microphone-based systems are most common, measuring vibration levels in the audible frequency range generally released when a cutting process becomes unstable. This methodology has been proven successful extensively in both turning [40, 41] and milling [28], where high material removal rates can cause a process to be close to instability under normal conditions. Considerable work has been conducted previously into machining process dynamics and chatter reduction [42]; however, in the case of this thesis, the modest depth of cut ensures that chatter is highly unlikely,

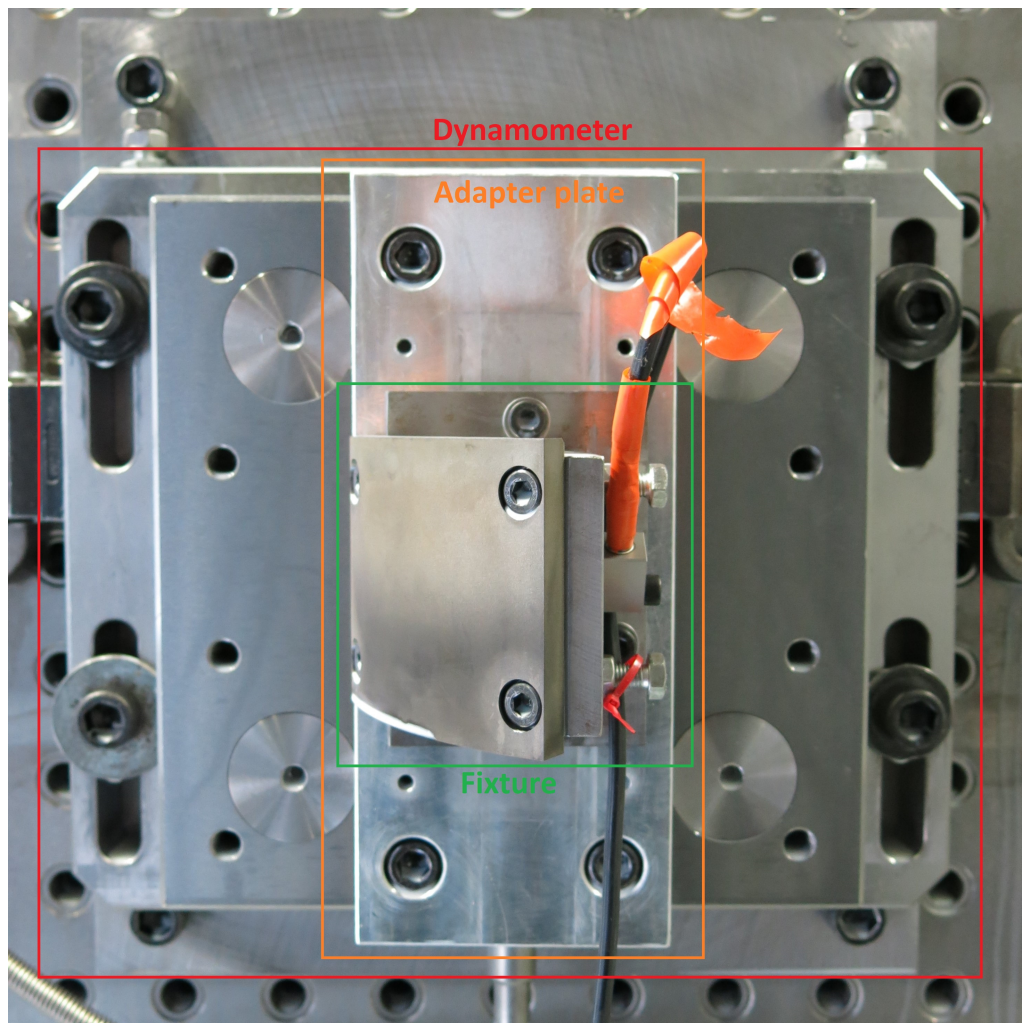


Figure 2.2: Example of dynamometer use, indicating fixture and adapter plate.

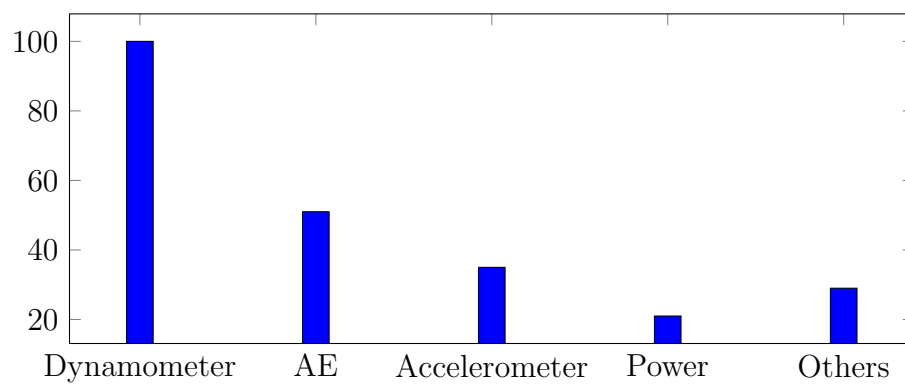


Figure 2.3: Plot to show relative use of sensors in monitoring systems adapted from [13,39].

with the focus being very much aimed towards wear prognosis. While accelerometers have been used previously for wear detection [43], primarily for their relative simplicity and low cost, it is believed that AE sensors provide a more promising option [43], given their capacity for considerably higher frequency response, which often spans the range of frequencies released during chip formation, and is less sensitive than accelerometers to noise generated during machining [31]. In machining, accelerometers are more commonly used for prediction of surface roughness, since surface roughness is the superposition of the feed per tooth mark and the displacements of the cutting tool due to vibrations [13]. Accelerometer use has been more successful in turning when compared to milling, as vibrations in milling are less correlated to surface roughness due to the intermittent nature of the process and runout effects [44]. Vibration measurements, however, can be used for tool wear diagnosis based on the fact that during flank wear, the contact area between tool and workpiece increases. Initially, the increased area leads to increasing frictional damping, and, therefore, reduced vibrations [13]; yet past a certain point, the larger cutting force becomes more dominant and results in stronger excitation. Correlations between vibration-related signal features and tool wear are discussed in more detail in [45], despite the fact that AE is believed to be more accurate and reliable [46].

AE is simply the waves of vibrational energy released from a material when it undergoes stress, commonly as a result of deformation in the shear zone, deformation and sliding friction at the chip-tool interface, sliding friction at the flank-workpiece interface, and the breaking of chips [13]. Due to the fundamental relationship between these mechanisms and the tool state, AE receives a considerable amount of interest from a number of research communities. In addition to the direct applicability to the cutting process, AE frequencies are usually in the range of 10kHz to 10MHz, which is generally above the range of normal machine vibrations and, therefore, reduces interference when compared to accelerometer data [13]. Such high-frequency data, however, does require very high data capture rates in order to accurately represent those signals observed and, as such, can potentially pose an issue when considering real-time processing. While the data capture is computationally expensive and the volume of data is vast, careful feature selection can ensure redundant data is discarded and AE-based monitoring strategies are feasible [25]. An important point of consideration when assessing the feasibility of using AE sensors is the proximity of the sensor to the source. In order to achieve a maximal signal to noise ratio, the sensor must be located along a suitable AE transmission path; however, the exact path is often unknown, which can result in attenuated signals and poor quality

data; distance also has a large impact on the quality of the measured signal, and so it follows that keeping the sensor close to the point of generation is beneficial [26].

In order for an AE sensor to detect signals with minimum attenuation, it must be coupled to the surface on which it is mounted. A crucial aspect of AE is their ability to travel through a variety of coupling materials, and as such, sensors can be located and coupled to rotating parts using a highly transmissible fluid [16]. The ability of AE to travel through different materials presents the opportunity to couple a sensor to the spindle assembly, theoretically detecting AE from the cutting zone directly through the cutting tool, at a fixed distance; although this is out of the scope of this work, it is certainly an area of interest for future industrial implementation and will be discussed further in Chapter 8. Previous works have also explored using a coolant stream as an AE transmission path [47]; however, following process-specific testing as prior work to this thesis, this appears better suited to tool breakage rather than wear detection, due to the effect multiple high-pressure coolant streams have on the ability to detect already weak signals. Jemielniak makes an important point that one must be cautious when setting up AE sensors, as energy spikes from a number of sources can cause overloading of the in-line amplifier, and therefore, poor signal quality [48]. Care must be taken to set the gain of the buffer amplifier as low as possible, while still allowing detection of the necessary signals [48].

Power monitoring systems are another widely-used technology in machine monitoring due to their low cost and ease of retrofitting to existing machinery. The basic premise of such a system is that by putting these sensors on spindle and/or axis drives of a CNC machine, it is possible to infer the cutting forces throughout the process as power draw is proportional to the torque of the drive; a property directly related to the cutting forces in a given direction. In applications where spindle power utilisation is a proportionately high percentage, this is much more feasible than a scenario where usage is low, given that the sensing bandwidth is hugely limited by the inertia of the spindle acting as a low-pass filter [25]. Spindle inertia severely limits the full potential of this technology, especially when taking shallow cuts [36]. When power monitoring is attached to the axis drives, this is less of an issue, however, machine ball screws now present a source of interference as lubrication and wear levels have a direct impact on the torque profile exerted by the drive [13]. Overcoming this interference requires modelling the characteristics of the drive assemblies and presents a new challenge in itself.

These technologies are the most commonly found in the previous literature, yet

there are a number of other options not discussed yet in this work, due to their unsuitability to the specific process of interest here. Attention should be drawn to [13] and [25], where a number of these sensor types are touched upon in slightly more detail, including ultrasonics, optical, and temperature sensors to name a few.

2.3 Industrial monitoring equipment

While laboratory-based experiments aim to reproduce the conditions found during production as closely as possible, the industrial environment is still often very different to the small-scale test environments found in research. As a result, specific hardware has been constructed for industrial implementation, intended to provide a more robust and reliable solution to that used in development work. Such systems are sold commercially as complete process monitoring solutions by companies such as: Artis Marposs, Brankamp, Montronix, and Nordmann [22]; consisting of both sensors and data acquisition/analysis packages in most cases. SLS currently use the Artis system in particular; however, all of the aforementioned systems aim to alert the operator of any deviation in the process from some predefined set-point or pre-observed example process. Probably the most simple example of a process deviation would be breakage of a tool, resulting in a drastic change in observed sensor signals. In recent years, these systems have become increasingly more integrated with existing numerical controls present on many automated machining centres, providing the ability to share two-way data from internal sources, and to pause machining in an alarm situation. This data sharing allows validation of additional sensor outputs, while providing a low-cost method of monitoring parameters not before available.

Taking the Artis system as a typical example, two main strategies for condition detection are implemented. The first and most straightforward approach to detecting process deviations is to simply compare observed signals to those recorded from an ideal representation as a reference. This method relies primarily on the repeatability of a certain operation, having minimal variation between realisations considered to be within the normal operating limits. The Artis system implements this methodology by comparing both a signal's amplitude and a definite integral to determine the current state, using user-defined upper and lower percentage limits to indicate abnormal conditions and raise an alarm. For short, repeatable, operations under constant conditions - for example, hole drilling and drill breakage - this system is

entirely applicable; however, as a process's duration increases, so does the volume of data which must be stored and loaded for comparison. While this may sound like a trivial problem in many cases, when considering high sample rate or multiple data channels collected over many hours, it becomes ineffective and impractical to implement due to the naive approach to feature extraction. The other significant flaw with using such a method to monitor a milling process is that the reference recording is only valid for a single set of operating conditions; any variation would require the timely procedure of re-recording a reference set, while the predefined static limits can often struggle to distinguish wear from expected process deviations due to material, for example.

To combat some of these issues raised with the classic industrial monitoring scheme, Artis implement their second method, known as their "dxdt" real-time strategy. This approach combines static and dynamic limits again, not only comparing signal amplitude values to pre-set limits, but also the value of signal gradients across a 10ms window. An example of such limits can be seen in Figure 2.4, adapted from the Artis technical literature. Static limits are user-defined to detect tool wear and breakage, whereas dynamic limits are used to detect any sharp, high amplitude changes in signal value. The dynamic limits are adjusted on-the-fly by using the previous 10ms of samples to calculate the limits for the current sample period, having the added benefit of needing no reference and, therefore, being applicable across a variety of machining parameters. For a tool with a gradual wear profile, this method may be incapable of classifying a tool as worn until past a certain threshold, hence is only appropriate for diagnosis as opposed to prediction of remaining useful life. This further relies on expert knowledge of the wear profile in order to set accurate static limits - a time consuming problem SLS have previously encountered.

As the above discussion highlights, there are some distinct limitations of the state-of-the-art industrial monitoring systems; however, it should be noted that in many situations, they are perfectly suitable. To be commercially viable, these systems are designed as a modular one-size-fits-all type of system, where they can be adapted to monitor a variety of different processes and operations, and as such, have a very general capability as opposed to a more targeted and specific goal. The point here is that, depending upon the process of interest and how safety critical detecting tool breakage is, these type of systems may be the most straightforward approach, on the assumption that the limitations are well understood.

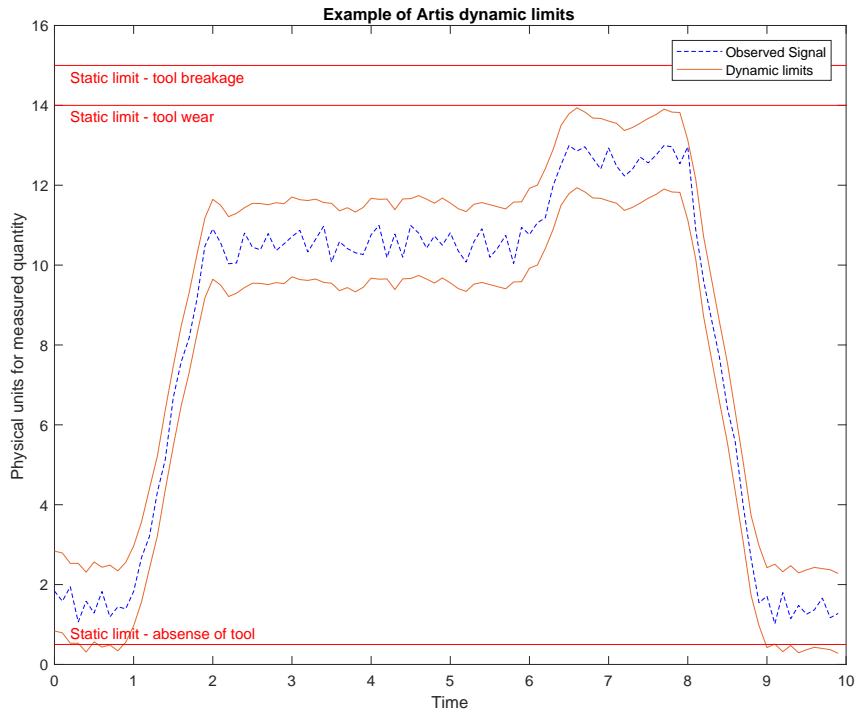


Figure 2.4: Example of Artis dynamical limits method.

2.4 Signal processing

In most cases, simply collecting sensor data is of little use without correlating measurement features to phenomena of interest, as no sensor is available that can measure any type of damage directly. A raw, indirect measurement is unlikely to be practically useful for wear measurement on its own, due to the sheer volume of information that an unprocessed data stream contains [49]. Many processing techniques exist to analyse waveform and multidimensional data, with the aim to extract only that information which is of use in a given situation and to discard the rest. The procedure of extracting this useful information from raw signals is more commonly known as *feature extraction* [2], and often includes an element of dimensional reduction, to ease computational requirements that follow. Before feature extraction, signals from the in-cut portion of an intermittent process need to be identified and isolated, as previously mentioned, although this could arguably be included as an element of the feature extraction process. This process ensures that any established features relate solely to the portion of data collected when in contact with the workpiece, and is well documented in [28]. The choice of useful features, or feature

selection, follows extraction and intends to identify those features which carry the majority of the useful information.

Worden et al. provide an in depth discussion of feature extraction and selection in the time, frequency, and time-frequency domains in a general condition monitoring context [50]; and this is a useful point of reference at this stage. Frequency domain analysis is commonly used in the structural health monitoring (SHM) community where analysis of vibration is predominant in detecting and preventing damage of structures; however, in machining, and more specifically when using AE data, time domain features are commonly used in preference. Preference of time-domain features is essentially due to the fact that in both abrasive and adhesive wear, the amplitude of AE signals is found to be directly influenced by wear mechanism progression [51], a finding that is agreed between previous works over the course of a decade [18]. Once a range of potential features have been extracted, methods exist to automatically identify those which are particularly informative and again, help to reduce redundant information [15].

Common time domain features used in previous monitoring schemes include descriptive statistics such as: mean, peak, standard deviation, root mean square (RMS), skewness and kurtosis, to name only a few [2]. One of the benefits in using such statistics, taking the RMS as a typical example, is the ability to calculate them in hardware circuitry before sampling occurs [52]. This method of dealing with analogue signals directly for filtering and feature extraction saves computational effort, while also minimising data loss, and is therefore practical for a real-time industrial implementation.

2.5 Classification techniques

Following determination and selection of meaningful features from the observed signals, it is possible to begin using these features to make informed inferences about the underlying process. Probably the most obvious starting point is to attempt to establish the state of the underlying, generating process, by assigning the full range of possible observations into distinct classes. Each class can then be given a meaningful label, with new instances being sorted accordingly by some informed mechanism. Machine learning classifiers can be applied to almost any data problem, and example implementations are readily available in almost all communities where

data-based decisions are made. An in-depth review of classification techniques can be found in [53], including comparisons of common methods, and combining classifiers. A key point is raised here that selecting a particular classifier for a given problem is not a case of whether that algorithm is superior or not, yet more a decision based on which method is likely to outperform the others under the given conditions in the problem.

Identifying groupings in the data is necessary before new points can be labelled, a task which can be completed in a number of ways. In general, such methods fall into two categories, either unsupervised clustering or supervised assignment for classification. Clustering techniques are well covered in [54]; however, supervised techniques remove a level of uncertainty when the required information is available. In many cases, principal component analysis (PCA) has become a frequently-used tool, preceding class labelling, to help maximise the distance between possible clusters of points [55]. An example of a typical strategy, making use of the aforementioned tools can be found in [27].

More specifically to machining, classification techniques have been used to identify tool, process, and material state in a number of previous works. In some instances, such classification systems have been paired with existing empirical methods to help improve classification accuracy and minimise incorrect predictions. For instance, Silva [56] makes use of two types of neural network to classify statistical features from AE and dynamometer signals, in an attempt to predict flank wear. Including the Taylor model of tool wear is found to improve predictions in all cases, raising the correlation coefficient between prediction and test sets from 0.691 to 0.872 in the example given, indicating that there is still room for improvement, however. Another example can be found in [57], where the author compares three machine learning classifiers (support vector machines, multilayer perceptron neural network, and radial-basis function neural network) in their ability to classify data features acquired from multiple sensors during end milling, into one of five classes, each representing an individual wear state. Cho concludes that the support vector machine (SVM) consistently outperforms the other neural network-based algorithms tested, although an ensemble approach can improve accuracy further. Although focussing on a different machining process to that of interest in this work, an in-depth review of classifiers and their use in the previous literature relevant to machining is presented in [24].

2.6 Sequential modelling

State tracking with a sequential model is a natural progression from the classification techniques presented previously, taking into account that the current state of a process often provides useful information on the likely next state to be observed. The most common method of dealing with data in this way is to construct a Markov chain of observations, developing into a hidden Markov model (HMM), when these observations are used to correspond to hidden latent variables. Considering data in its sequential form is not a new practice, and a review of general methods can be found in [58].

Hidden Markov models have previously been commonplace in speech recognition applications [59], and in more recent years, have found their way into a number of different research areas including tool wear detection and diagnosis [2]. HMMs have been proven a versatile tool for tool condition monitoring, given the wide range of applied processes and sensor information that they have been used to model in previous works. Examples include accelerometer-based [60] and acoustic-based [61] monitoring of end milling operations, and both dynamometer-based [62] and strain-based [63] monitoring of wear in drilling. In all of the aforementioned examples, the model formulation follows the same procedure given in [61], albeit with varying model topologies. Boutros [61] presents a 95% successful classification rate when using an HMM to predict wear state; however, this is limited given that flank wear is the only wear type used, and the inclusion of only three states (sharp, worn, and broken respectively) does not give any indication of the level of wear, and therefore, rate of progression. With the ultimate aim to predict remaining useful life, this method requires extension to include a greater number of wear levels. This is similar to the work by Miao [64], in which HMMs are used to distinguish between normal and failure tool conditions, neglecting to explore wear development, and any observations leading to failure.

2.7 Predictive modelling

When compared to diagnostics, the literature surrounding prognosis of tool condition is significantly more limited [2]; this is likely due to the fact that predicting remaining useful life is no trivial task when subject to changing environmental and

operational conditions, requiring either adequate mechanistic knowledge of the process, or significant computational power and data resource. A number of previous works in prognostics can be found in [2], however, recent advances in removing environmental effects [49] and non-linear dynamical system identification [65] present a new opportunity for prediction of remaining useful life with natural confidence intervals.

More specifically, prognostics of remaining useful tool life within machining is still a relatively unexplored application of predictive modelling. A typical example of a mechanical system is provided by Yan et al. [66], where a logistic regression model is developed to calculate the probability of failure of an automated door at any given point, and an ARMA model implemented to trend the condition variables for failure prediction. The results, however, are not well quantified and are susceptible to error given small variations in operating conditions. A common problem when searching for literature regarding tool life prognostics is highlighted by Zhou et al. [67], where the title of the paper indicates "...prognosis of tool wear...", however, at no point is a future prediction of wear state considered. Exploring the previous literature indicates an issue with terminology, where the machining community appear to class wear prognosis as the prediction of wear state given some observation, rather than the prediction of a future state given past observations. Taking into account the lack of application and success of remaining useful life models to the tool wear problem, it is probably unsurprising that industry in general currently either adopts a run-to-failure, or, as-frequent-as-possible maintenance policy [2]. Engel et al. [68] provide a discussion on a number of issues involved with predicting remaining useful life, primarily the management of uncertainty, and conclude by explaining that the success of a prognostic method heavily relies on feature selection and process-specifics. For tool wear, the variation between cutting processes and conditions is likely the reason for the lack of progress in this area.

2.8 Overview of literature

This chapter has presented and reviewed a selection of previous works thought to be relevant to the application explored in this thesis. While it is found that some of the key areas have been investigated in great detail before, it is evident that there is still significant room for improvement in the area of predictive tool condition monitoring,

given new and emerging methods from more recent years. This chapter forms the basis for those which follow, with individual methods being examined in more detail at the appropriate stages.

EXPERIMENTAL METHODS

After assessing previous work in the field and taking into account the particular challenges faced with the task outlined in Chapter 1, it is apparent that example data needs to be collected in order to determine trends developing throughout the machining process, and moreover, useful features within the gathered data which give an insight into the state of the process at any given time. Considering the absence of suitable existing data, this chapter therefore focusses on detailing a machining trial design and implementation, intended to provide the required dataset for in-depth analysis in the following chapters.

3.1 Machining trial concept

Obtaining a clear understanding of the time-dependant behaviour of cutting tools for a particular machining process is vitally important for improving part quality, production efficiency, and keeping costs to a minimum. Completing a tool-life trial provides clear information relating to the condition of tools at various time intervals, and provides data which can help to increase process efficiency by ensuring tools are used to their maximum potential. Once tool usage limits are established, the tool's cutting edge features can be measured along with material surface properties at various wear conditions, giving a clear indication of how tool wear affects the surface, and therefore part quality. In addition to this, it is then possible to monitor various process parameters and conditions during the cutting operation which can

be used to infer tool condition and surface quality during manufacturing.

The common experimental approach taken in tool wear trials, is to define process parameters, begin machining at these parameters, and pause at predefined intervals to take measurements of the tool wear [67]. This is prevalent across the majority of previous works in literature, primarily because this method ensures that the duration of machining is limited to that of the entire life of a single tool, keeping time and costs to a minimum. The main problem with this approach is that, especially during milling, the tool is removed from its holder during the wear measurements. The result is that the tool is unlikely to be in an identical position to that from which it was removed from the holder, altering the cutting action through change in run-out and tool length. Any such change in cutting action has the potential to introduce error to the test, and would not be found in a typical industrial scenario. The second issue with this method is that, once a wear measurement is taken, the tool is then used for another extended duration, eliminating the opportunity for future measurements of the tool at any past state.

To overcome these problems, the experimental machining trial used to explore the desired correlations in this thesis was split into two parts. Firstly, a selection of tools were used to machine a cylindrical billet of Titanium-5553 for predetermined time intervals, consequently generating a selection of tools in different wear states. These tools could then be measured using an Alicona 3D optical microscope system to classify the various levels of wear and explore how the wear characteristics develop with time, although, a rotating cylindrical billet makes sensor cable routing difficult. In order to avoid cabling issues, once the state of the tools was confirmed, the second half of the experimental trial was conducted - using these tools to cut a smaller, static material specimen, while collecting monitoring signals from the process. This two-stage methodology provides information on the tool state, how wear develops with time, the effect tool state has on the signals generated during cutting, and the effect tool state has on the material surface. The main benefit of this approach being that, providing the second machining phase is of a short duration, the tools used remain in the same state to which they were first measured. This enables future measurements to be conducted, and the possibility to collect further data should it be needed. It should, however, be noted that the cost of both time and tooling is significantly greater taking this approach rather than the common one.

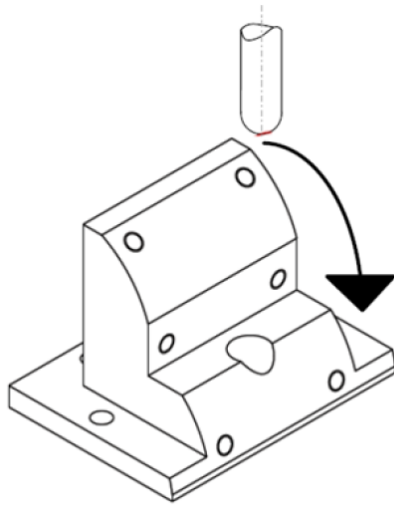
Table 3.1 shows the tool numbers and the duration of machining experienced by each tool in the first stage of experimentation, prior to 3D geometry measurement.

Tool Number	Tool Duration (hours)
1	0.00
2	0.50
3	1.00
4	1.50
5	1.75
6	2.00
7	2.25
8	2.50
9	3.00
Total	14.50

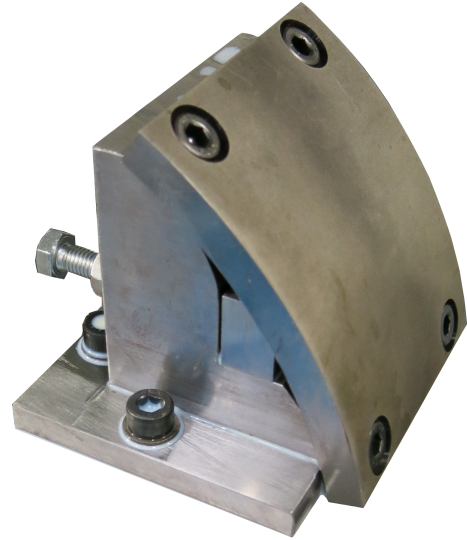
Table 3.1: Table to show test tools and durations

These figures were chosen based on current usage data from the industrial process, and extended beyond the current limits to provide an overview of tool wear above and beyond that experienced during normal usage. The choice of nine tools was based around providing enough data points that a wear curve could be estimated, while keeping costs to a minimum by not using an excessive number of tools and limiting machine time to two full days (14.5 hours). Moreover, the current industrial process in SLS is based on operator experience and uses tools for approximately two hours. The machining in this thesis was, therefore, extended 50% further (to three hours), to ensure that these usage guidelines were reasonable, and to explore pushing wear past the limits that are commonly experienced during machining of production parts, meanwhile, collecting monitoring data under these conditions. Finally, the distribution of tool usage durations over this three-hour period is more dense around the two-hour tool, as previous operator experience would tell us that this is the period towards the end of tool life, where accelerated wear occurs.

The second pseudo five-axis process machined a selection of test specimens taken from a Ti-5553 landing gear component, using the previously worn tools and collecting data from a number of different sources during the operation. An AE sensor was attached to the back of the workpiece fixture and the workpiece assembly was placed on a dynamometer, allowing both AE and force data to be collected simultaneously (a more detailed description and image can be found in Section 3.5). It should be noted again, however, that in an industrial environment, a dynamometer is not appropriate due to its intrusive nature and size limitations. Here, the dynamometer has been used to corroborate AE sensor data to provide confirmation that AE features are sensitive to phenomena observed previously in the literature,



(a) CAD model of the workpiece fixture with cutting direction indicated.



(b) Fixture and workpiece for second element of experimental trial.

Figure 3.1: Test specimen workpiece and fixture for trial element two.

through dynamometer data. Figure 3.1 shows the fixture and workpiece which have been designed and manufactured specifically for this process, with the cutting direction indicated. The cutting operation results in gathered data and material surfaces relating to each of the nine tool conditions, allowing further analysis to explore correlations between the tools, material, and monitoring data.

3.2 Specimen extraction

For any experimental work to be industrially relevant, the tools, material, and tool paths must match those found in real manufacturing processes as closely as possible. This is a trivial task when considering tool selection; however, obtaining exemplar material specimens that have undergone the same pre-processing stages as a finish-ready part is more challenging. For a part to be ready for finish machining, it must have first undergone four fundamental procedures. Beginning with a bespoke forging, a part is then heat-treated, rough-machined, and semi-finish machined before preparing for the finishing operation. From an industrial point of view, each processing step adds value to the part in question and so processing a unique sample of material is a costly option, even more so when considering the time taken fitting such a specimen into a running production line and the required forging die. For the



Figure 3.2: CAD model of a Boeing 787 truck beam - the end-cut shown circled in red.

same reason, it is also not feasible to obtain a production part, as the value is too great from a business perspective. The solution to obtaining representative material samples is to use discarded waste material from a critical stage of the production process to form test pieces, resulting in material samples that have undergone the exact same processing route as a legitimate flight-worthy part.

To aid in the accurate positioning and clamping of a truck beam during its various processing stages, each forging contains a keyway (see Figure 3.3b) which is aligned with various chuck and fixturing solutions to provide a mechanical linkage and ensure correct orientation during manufacture is maintained. Post-manufacture, this keyway is redundant, and is therefore removed in the form of an ‘end-cut’ in an effort to maintain maximal weight savings (Figures 3.2 and 3.3a). This end-cut provides an ideal source of material for test specimens to be extracted from and allows experimental work to be designed around the available usable material; three individual test specimens can be removed from each end-cut as indicated in Figure 3.3c, to make optimal workpieces for the second-element of the trial process. Their curved profile is beneficial, as it allows tool paths to be designed that simulate those found on the production line, ensuring the following results are industrially relevant.

In order to remove the complex geometries of the specimens from the equally complex end-cuts, a wire electrical discharge machine (EDM) must be used to extract samples at a 135mm radius. The wire EDM has the benefit of easily removing material samples with very little waste, although it can be a time-consuming process. Similar machines are commonly used in research where material samples are required; however, it is worth noting that the electrical discharge process generates a re-cast layer on the surface of the sample, which must be removed by a pre-cut, to ensure uniform material properties from the onset of the monitored machining. A technical drawing of the coupons produced can be found in Appendix A.

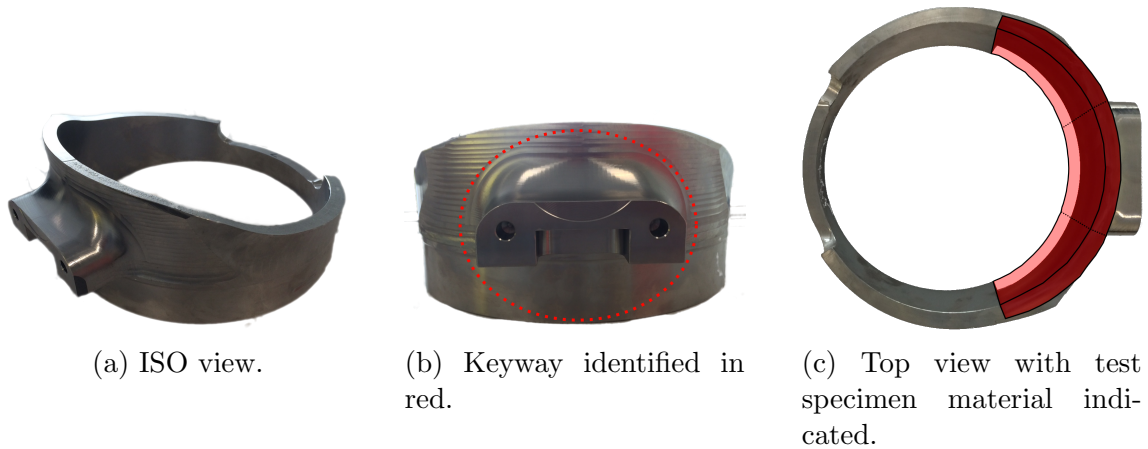


Figure 3.3: Truck beam end cut as removed from 787 truck beam, used for trial element two.

3.3 Machining process parameters

The test parameters used in this work are taken from the industrial finishing process within SLS, and are intended to be representative of the operation finishing used in manufacture. Using these parameters allows any gathered data and developed models to be directly applied and validated in a typical industrial environment. The parameters used are an axial depth of cut (a_p) of 0.3mm, radial depth of cut (a_e) of 0.7mm, spindle speed (N) of 5082rpm, and feed rate (V_c) of 4.268m/min. The experimental tool paths are also designed to be representative of the industrial process.

The tools used here are solid carbide, 20mm diameter 6-flute Kennametal ball-nose cutters which have their lead angle (α) varied between -70 and -50 degrees, resulting in an effective cutting diameter of between 15.32 and 6.84 millimetres. The changing lead angle ensures that the entire useful cutting radius of the tool is used, maximising tool life; however, this has an impact on the gathered signals and must be accounted for when analysing the data samples. The useful cutting radius can be seen in Figure 3.4.

Throughout all cutting processes, Hocut 795B coolant must be used for both cooling and lubrication, as dry machining of titanium can result in very high temperatures at the cutting zone, and a potential fire risk. This was applied through both flood and through-tool supply at 100 bar and 70 l/min.

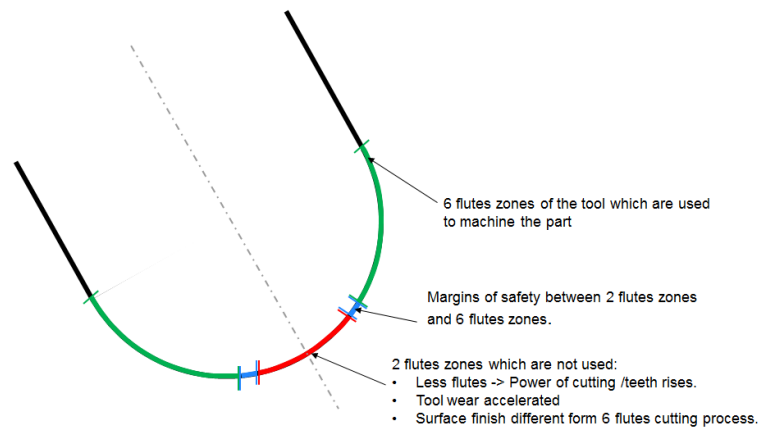


Figure 3.4: Useful cutting radius of Kennametal ball-nose cutters.

3.4 Tool path analysis and programming

As previously mentioned, the tool paths were designed to accurately represent those found in manufacture, resulting in the full cutting radius of the tools being utilised and maximising tool life. Both trial elements require dedicated tool paths based on those found in industry and so are developed accordingly.

3.4.1 Trial element One - tool wear

To achieve an even wear profile, the tool's lead angle is constantly changed while machining rings around a circular billet. Figure 3.5 shows this path for the primary stage of machining, intended to provide example tool wear states. Engage and retract motions are needed to ensure both a smooth beginning to each cut and that the tool is clear of the workpiece before any rapid moves are performed and potential damage caused. These are indicated in red, while the general steady-state tool path is shown in green. The tool interpolation axis is indicated in blue which correspond to varying the tool's lead angle between those values given in Section 3.3.

This tool path is adapted from the steady band machining operation as found in Safran and mentioned in Chapter 1.2.3. More specifically, this process finish machines cylindrical bands around a truck beam section to aid in fixturing during further operations. Figure 3.6 shows a CAD model with the tool path visible. By comparing Figures 3.5 and 3.6, it can be seen that the experimental tool path follows a very similar course to that experienced by a tool in industry, albeit for a smaller

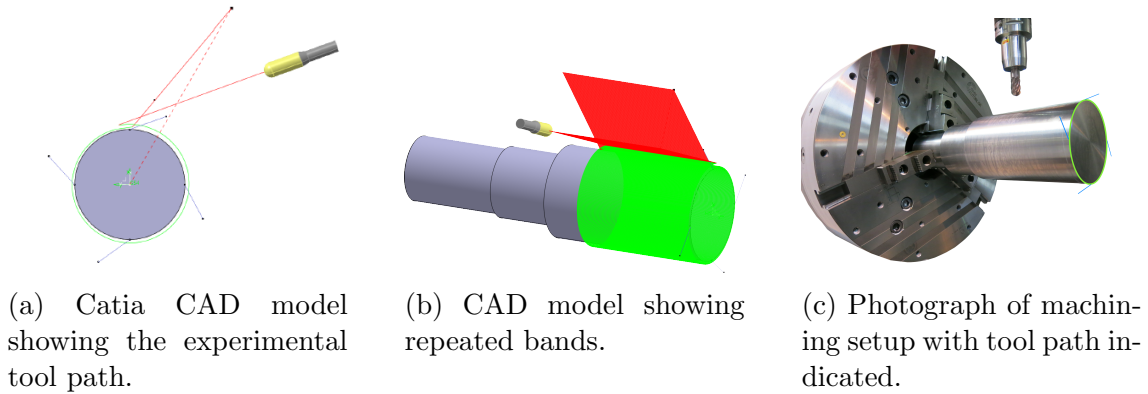


Figure 3.5: Images showing the process tool path for trial element one in green, and the tool interpolation axis in blue. The engage and retract motions are shown in red.

diameter, and in this case is considered a representative example of the industrial machining process.

The cylindrical billet of Ti-5553 used to generate the nine different levels of wear has a length of 270.2mm, allowing a total of 386 complete bands to be machined along its length when stepping over by 0.7mm, an example of which can be seen in Figure 3.5b. It becomes apparent when considering the 150mm starting diameter that a number of passes over the billet's length are necessary to complete the total of 14.5 hours of machining time required from Table 3.1. For example, beginning at 150mm, a single ring equates to a distance cut of

$$\begin{aligned}
 D &= \pi(d_s - 2a_p) \\
 &= \pi(0.15 - 0.0006) \\
 &= 0.4694m
 \end{aligned}
 \tag{3.1}$$

where D is the distance travelled and d_s is the initial diameter. At a surface speed (V_c) of 4.268m/min, this cut would last for,

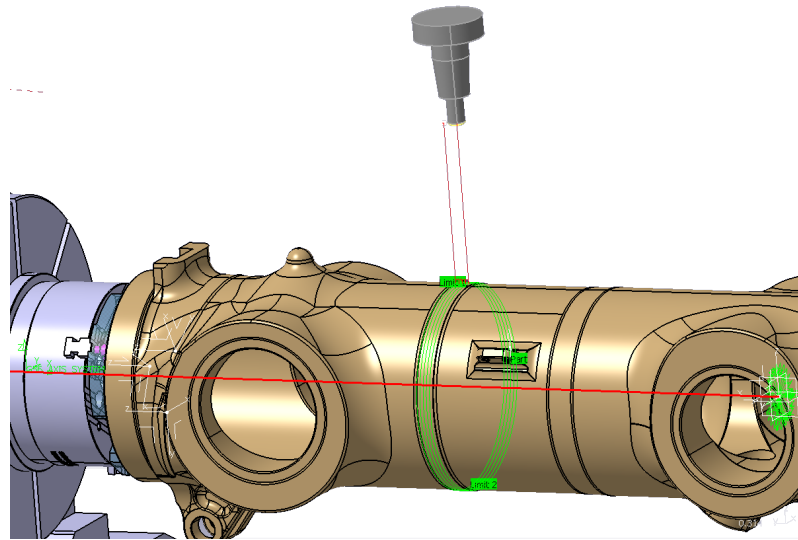


Figure 3.6: CAD model showing the industrial steady band machining tool path.

$$\begin{aligned}
 t &= \frac{D}{V_c} \\
 &= \frac{0.4694}{4.268} \\
 &= 0.1100mins = 6.6seconds
 \end{aligned}
 \tag{3.2}$$

For a required cutting time of 30 minutes, a total of 272 rings are needed, using the majority of the 386 available per pass. Once 386 rings have been cut, it is required that another step in a_p be performed and the process continue, taking into account the now smaller diameter. To aid in these calculations, a MATLAB script is used to provide both the number of axial and radial steps required for a given cutting duration, while also allowing for new starting positions along the workpiece due to cumulative cutting durations. These values can later be used when programming the Catia V5 Computer-Aided Manufacture (CAM) process, with the CAM software also calculating the expected cutting duration, providing an effective verification of the computed values used.

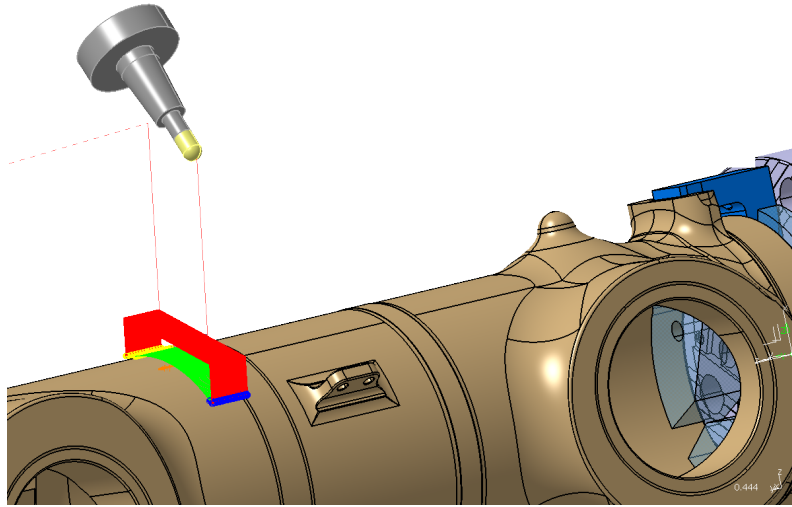


Figure 3.7: CAD model showing the industrial surface to center / quads machining tool path.

3.4.2 Trial element Two - data collection

The tool path used for machining the test specimens is pseudo five-axis, in that it simulates a five-axis operation, but limits the working area and orientation of the specimen to allow three-axis operation. The tool path itself strongly resembles that of Trial element One (using four axes); however, the machined surface is not rotated as the workpiece is now a sectional specimen, as opposed to a cylindrical billet. This method has the added benefit of allowing sensor wires to be routed near the sample without becoming twisted or tangled as the operation progresses.

As explained in Chapter 1, the surface to centre/quads operation is the most time-consuming process, and therefore has the potential for the greatest benefit from any performance improvements. The tool path used for this trial element consequently focuses on that used for the surface to centre/quads operation, ensuring that any subsequent results have the possibility of maximising their efficiency gains throughout the entire manufacturing cycle. Figure 3.7 shows a CAD model of this process, with tools paths indicated in the same manner as before; in this figure, it is clear to see the engage and retract motions shown in yellow and blue respectively. It can also be seen that this process follows a very similar path to the machining of the steady band to the right; however, it is limited to a smaller arc on the truck beam's circumference.

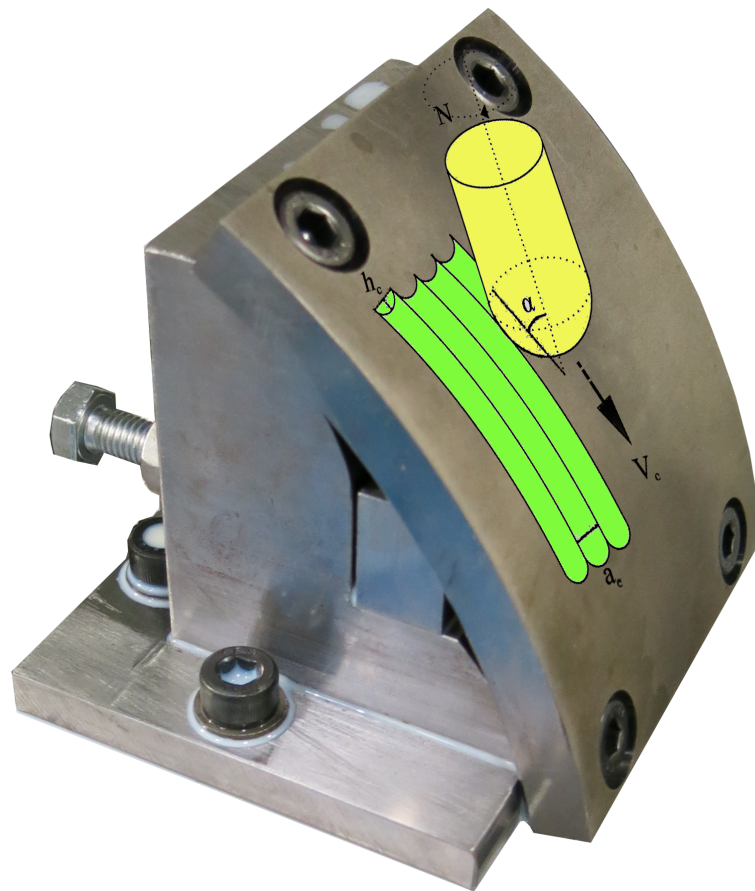


Figure 3.8: Workpiece and fixture with experimental tool path indicated.

Figure 3.8 shows the same fixture as in Figure 3.1b, with the experimental tool path indicated. Again, comparison between Figures 3.7 and 3.8 shows that the experimental tool path follows the same procedure as the industrial process after a 90-degree rotation. The process parameters from Chapter 3.3 are also indicated, with the addition of the cusp height (h_c) between passes. Given that the coupon is 90mm wide (as shown in Appendix A) and allowing for 10mm clearance at each edge, the working area is 70mm wide. With a radial depth of cut of 0.7mm, it is possible to perform a total of 100 passes per coupon.

As mentioned previously, the coupons are extracted using wire EDM and therefore must have their recast layer removed from the working area before being used for any testing. For this reason, two CAM processes are used for removing the recast layer and performing the test cuts respectively, using a single tool for preparation and the nine worn tools for the test cuts. Removing the recast layer ensures that the material surface of each specimen is consistent and that they are geometrically homogeneous when the test cuts are performed, regardless of fixturing accuracy, thus providing comparable data between cuts and tools.

3.5 Sensor selection and data acquisition

As touched upon in Chapter 1.3, sensor selection for in-process monitoring is a topic that has been widely covered in a vast amount of previous work [14–19], with the benefits and drawbacks of both individual and fusion approaches also well documented. Again, it is readily agreed that there is still no widely adopted methodology to designing a successful monitoring system that is applicable to every potential machining application, given the extensive scope of experimental methods used in previous work.

Chapter 2.2 shows that the literature indicates the most commonly-used sensor is a dynamometer, used to indirectly measure cutting forces, yet this is not always the most practically appropriate solution. When specifying a sensing system, it is critical to consider the environment that such a system is operating in alongside the placement location of each sensor. An industrial production environment is not always the ideal situation for sensor placement, primarily due to the potential of water ingress into supposedly sealed connections when subject to high-pressure coolant application. Additionally, industrial machinery tends to contain a number

of noise sources which can interfere with the ability to extract desired information from acquired data, such as spindle motors, servo drives, and swarf removal devices to name but a few. For this reason, and to maximise the signal to noise ratio, it is necessary to place sensors as close to the cutting action as possible with a clear transmission path from source to sensor.

In this work, AE is explored as the monitored variable, with small piezoelectric-ceramic sensors used to continuously ‘listen for’ the bursts of AE energy released as waves of vibrational energy from the material when it undergoes stress. This stress can be due to deformation in the shear zone, deformation and sliding friction at the chip-tool interface, sliding friction at the flank-workpiece interface, and the breaking of chips [13]; all of which provide a key insight into the behaviour of the underlying process and its state at any given time. More specifically, a Kistler 8152B1 Piezotron sensor was used, as this is intended for use within machining structures and is of a rugged design, rated at IP67 resistance to dust and water ingress. The frequency response of an ideal 8152B1 sensor is shown in Figure 3.9, with the full datasheet available in Appendix B. During all experimental machining trials, both dynamometer and acceleration data were collected alongside the principal AE data for completeness. These data were collected using a Kistler 9255B large plate dynamometer and PCB 355B02 single-axis accelerometer respectively, details of which can be found in Appendices C and D. The dynamometer served a single purpose - to validate and corroborate the AE sensor data, and to confirm that the data collected was meaningful and not simply noise. A brief study was conducted as part of the first trial to determine the former. The sensors were located as shown in Figure 3.10; the key point to note being the proximity of the AE sensor to the workpiece surface. A number of different locations could have been chosen; however, this layout provides adequate safety from the cutting process while maintaining a continuous transmission path for signals to reach the sensors. Further development work relating to sensor placement and spindle attachment can be found in Chapter 8.5. As briefly discussed in Chapter 2.2, use of a Coolant-based AE sensor was explored, however, appears better suited to tool breakage rather than wear detection, due to the effect multiple high-pressure coolant streams have on the ability to detect already weak signals.

All sensor data were collected using a National Instruments (NI) CompactDAQ cDAQ-9174 chassis complete with NI 9201, NI 9223, and NI 9234 modules. More specifically, a low sample-rate analog input module (10KHz), a high sample-rate

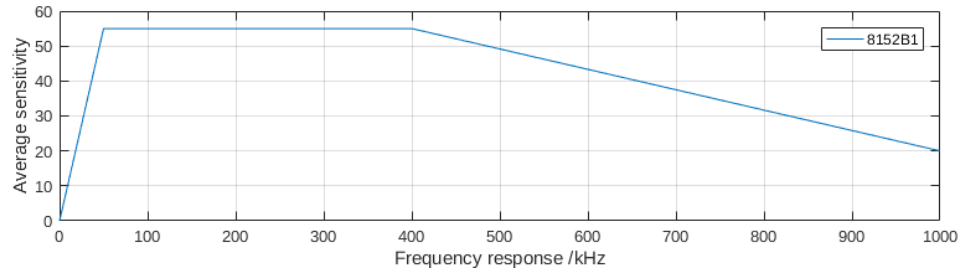


Figure 3.9: Ideal frequency response of a typical Kistler 8152B1 sensor [69]

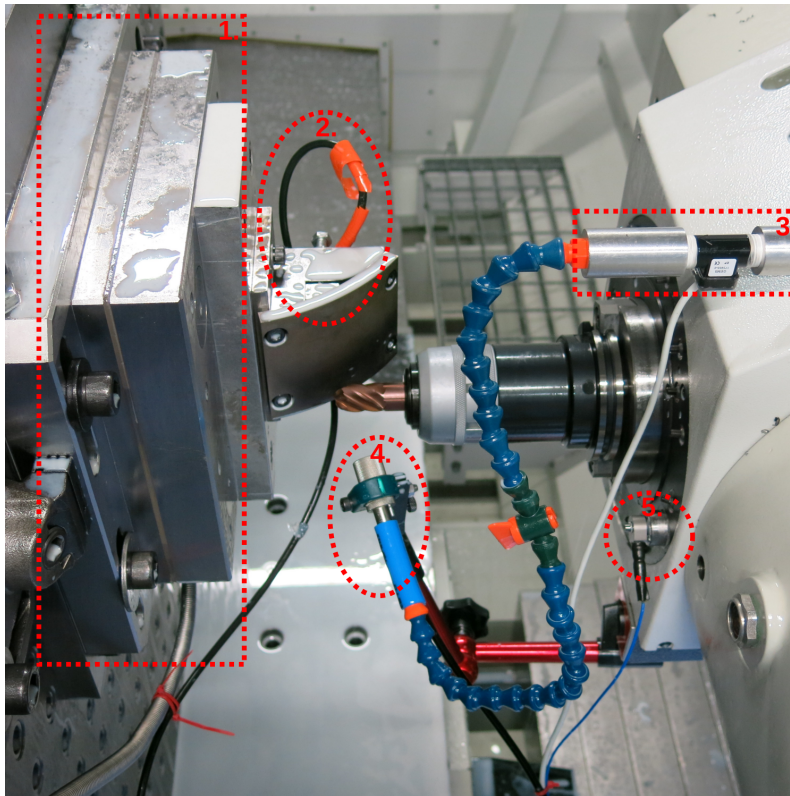


Figure 3.10: Photograph of experimental setup with sensors included. 1. Dynamometer, 2. AE sensor, 3. Coolant flow rate sensor, 4. Coolant-based AE sensor, 5. Single-axis accelerometer.



Figure 3.11: Image showing National Instruments cDAQ-9174 with 9201, 9223, and 9234 modules.

analog input module (1MHz), and an IEPE capable accelerometer module (51.2KHz) respectively. The main advantage to using such a system is the ability to synchronise data streams between modules, regardless of individual sample rates (each 9174 chassis contains 3 on-board timers). Not only is the NI hardware extremely flexible with regards to the variety of different data types and sensors it can support, but it is also possible to program within a number of different environments. Primarily LabView is used for this task, but it is also possible to control such systems from within MATLAB directly. Figure 3.11 shows an example cDAQ chassis complete with data acquisition modules.

3.6 Machine specifics and dynamical testing

For both previously-described trial elements, a specific machine tool was selected from those available at the Advanced Manufacturing Research Centre (AMRC) with Boeing in Rotherham, South Yorkshire. The crucial capabilities required for both trial parts are high pressure (>80bar) coolant delivery through the tool holder, and a stable region of operation around the defined process parameters. As both elements use different tool paths and strategies, primarily a mill-turn followed by a milling operation, two distinct machining centres were chosen.

For the initial tool wear trial, the DMG Mori NT 5400 was selected as the ideal candidate; it is a five-axis horizontal mill-turn machine, consisting of twin lathe chucks and a single milling turret providing the multi-axis positioning. The NT 5400 provides a maximum spindle speed of 2,400 rpm per lathe spindle and 6,000 rpm at the milling head, making it more than capable of meeting the requested

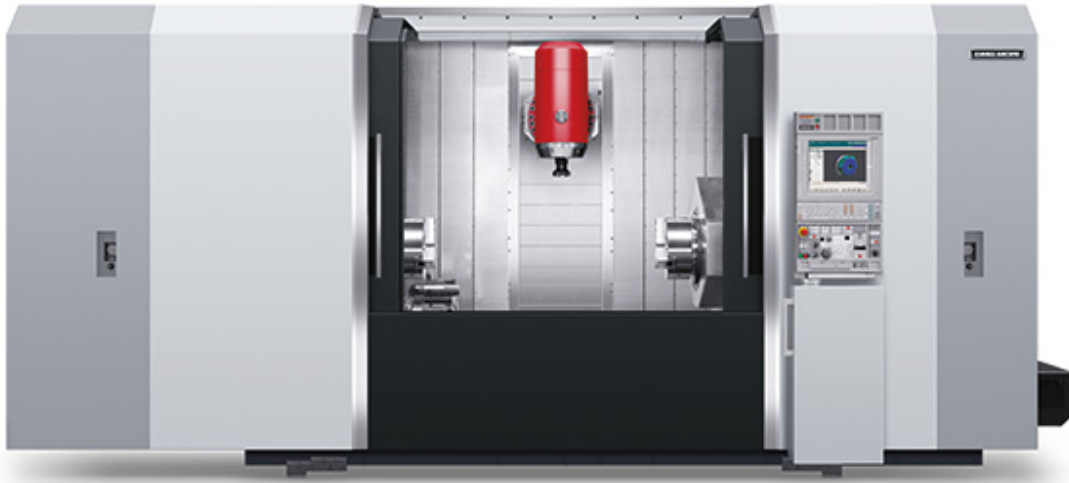


Figure 3.12: Image showing DMG Mori NT 5400 mill-turn machining centre.

process parameters, while also providing high pressure through tool coolant delivery. Figure 3.12 shows an image of an NT5400, supplied by DMG Mori.

Prior to cutting any material, the machine was fully set up and the tool tip subject to experimental modal analysis following a tap test. This is simply to ensure the stability of the cutting process at the chosen parameters, and to guarantee analogous cutting conditions between machines during research and manufacture. For more information regarding cutting tool dynamics and stability prediction, see the work by Altintas [42, 70] on the subject. The data from the experimental modal testing is read and stability calculations performed within METAL-MAX TXF software [71], resulting in a stability lobe plot, as can be seen in Figure 3.13 for the NT 5400.

Similarly, for the data collection portion of the experimental work, the Starrag STC 1250 was selected. The STC 1250 is the AMRC's dedicated titanium milling centre due to its high structural rigidity, and is also capable of providing high-pressure through-tool coolant. An added benefit of the STC 1250 is the additional functionality fitted within the control system, allowing for communication with internal monitoring software and sensors. The same process of modal testing was performed for the experimental setup in the Starrag, and the resulting stability lobe plot can be seen in Figure 3.14.

It is clear from comparing both of these plots, that due to the shallow depth of cut (DOC) used in this work, the process will remain well within the bounds of stability

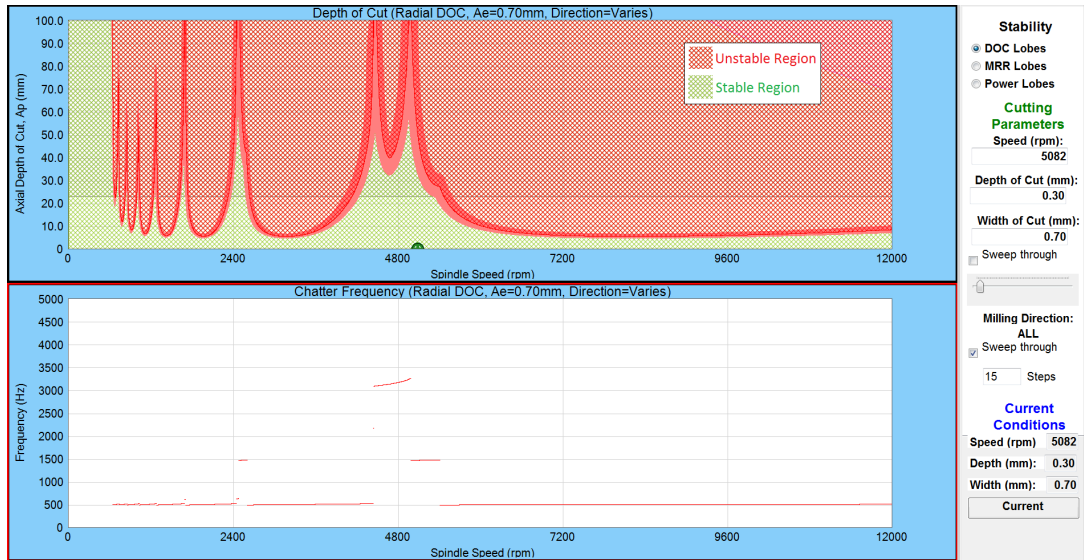


Figure 3.13: TXF results showing stability lobes for the experimental process on the NT 5400

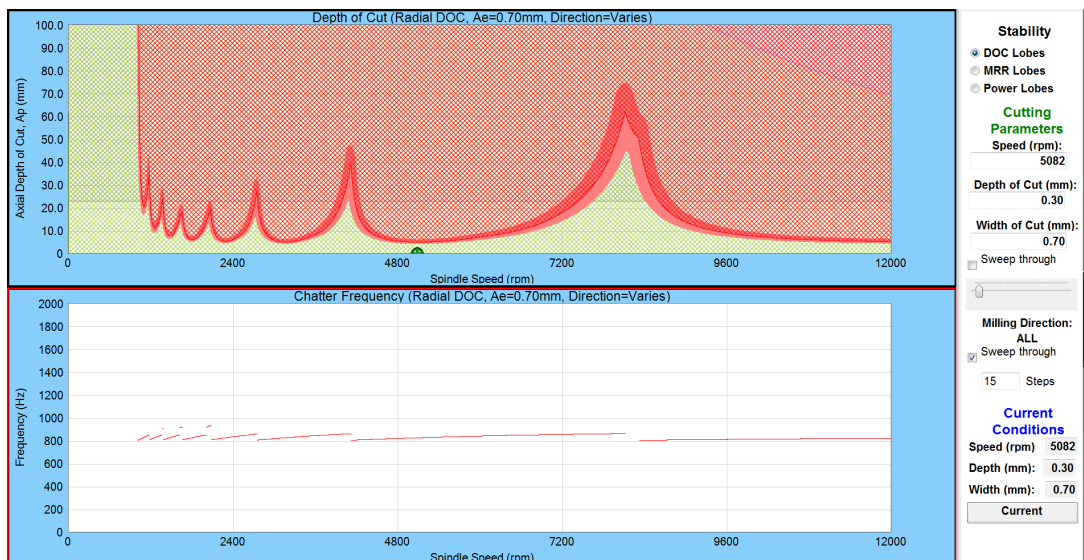


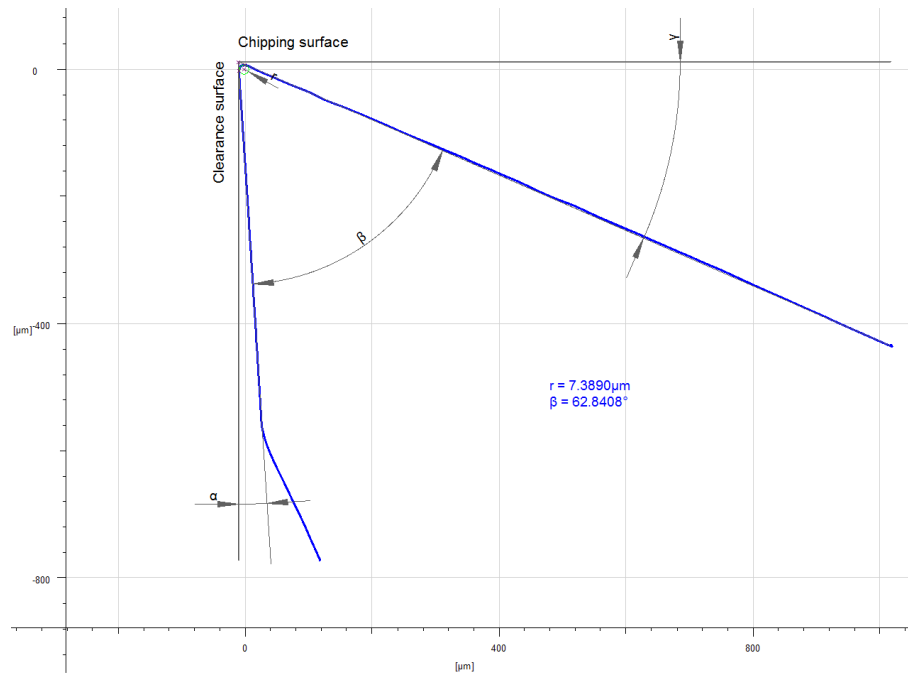
Figure 3.14: TXF results showing stability lobes for the experimental process on the STC 1250

at all times. The cutting process should, therefore, be unaffected by the change in machining centre used when process parameters are maintained consistent. It is worth noting, however, that the two scenarios do still have very different dynamic behaviours when depth-of-cut (DOC) is increased, primarily due to the extended tool length required with the NT 5400. Should there be a requirement for deeper cuts, careful consideration of cutting parameters would be needed if maintaining the same physical system.

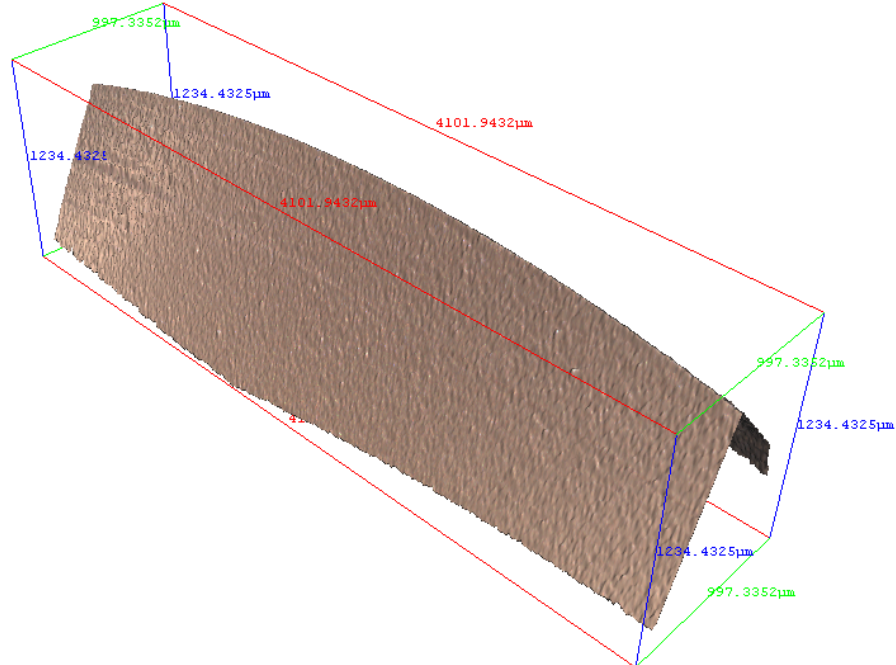
3.7 3D tool geometry measurement

All experimental tools were scanned using a newly developed Alicona InfiniteFocusSL; an optical measurement device which provides 3D analysis of form and finish (including surface roughness and cutting edge features) by varying its focus over the cutting edge scan area. By comparing a scanned 3D tool profile to a previous reference profile, it is possible to gain an insight into how the tool state develops with use when compared to a new tool. To achieve this, ten new tools were measured and the average of these used as a reference which could be studied against each worn tool. Table 3.2 shows a selection of tool features measured by the Alicona during a standard cutting-edge scan operation and difference measurement. Each tool is held in a bespoke fixture during scanning to ensure correct alignment to the same location on each flute of each tool. This fixture was designed and manufactured specifically for this thesis, details of it can be found in Appendix E.

Only a selection of these features are relevant when considering the cutting edge of a ball-nosed milling cutter. The most commonly trusted feature as an indication of tool wear is the cutting-edge radius, which is expected to increase with wear, suggesting a blunting of the cutting edge as it experiences more use [5]. Blunting of the cutting edge is certainly expected during cutting, where flank wear is the dominant mechanism; however, crater wear could disrupt this trend by dramatically changing the form of the cutting surface. The wedge angle, therefore, is also used as another indicator of the state of the cutting edge. Figure 3.15 shows both a 2D and 3D flute profile of a new tool scan.



(a) Example Alicona scan showing 2D flute profile of a new tool.



(b) Example Alicona scan showing 3D flute profile of a new tool.

Figure 3.15: Example Alicona scans.

Name	Units	Description
r	μm	Mean radius of mean edge
α	$^{\circ}$	Clearance angle
β	$^{\circ}$	Wedge angle
γ	$^{\circ}$	Chip angle
Ecq	μm	Form deviation of circle (RMS)
Dpos	μm	Max. deviation above reference surface
Dmean	μm	Mean deviation
Vp	μm^3	Volume of peaks above reference surface
Vv	μm^3	Volume of valleys below reference surface
SIMcd	μm	Greatest depth of defects (ISO 8785)
SIMch	μm	Greatest height of defects (ISO 8785)
SIMt	μm^2	Whole Area of defects (ISO 8785)

Table 3.2: Table to show Alicona measured tool features [72].

3.8 Overview of data collected

This chapter has detailed an experimental procedure consisting of two independent machining trials as required to collect data for further analysis in the coming chapters. Following the testing procedure outlined above, two sample datasets are gathered by repeating the experimental method twice throughout the duration of the project, resulting in both training and testing/validation data under slightly differing conditions. Each dataset contains nine tools of varying wear states, nine machined material samples, and 450 AE recordings (50 per tool). Once compiled, the collected data is ready for pre-processing, to ensure that it is in the correct format, with the most relevant and descriptive features identified for model development.

PRELIMINARY DATA ANALYSIS

Now that the experimental procedure has been defined and accomplished, the data collected throughout the operation can begin to be analysed and prepared ready for characterisation and modelling. The experimental process has been repeated twice and, as seen throughout Chapter 3, the data collected consists of AE data captured during the machining of a number of material samples with a number of unique tools in each case. Each tool and material specimen provides key information relating to the state of the cutting process which can be correlated to that gained from the AE released. This chapter concentrates on the preparation, pre-processing, and selection of indicative features required to correlate the obtained data to one another. All data processing is carried out using the MATLAB software environment.

4.1 Acoustic emission data preparation

Figure 4.1 shows a typical example of the raw AE signals collected during a 190-second experimental machining program. There are a few key attributes to draw attention to at this stage. Firstly, the signal is periodic and cuts can be seen evenly spaced from one another (each peak represents a single curved tool path), and secondly, the first cut has a large amplitude relative to the remaining cuts. In a real-world situation, this would not always be the case, given that there is always the potential for operator intervention and fluctuations in machine speed; however, by extracting only the in-cut information, this potential lack of periodicity is no

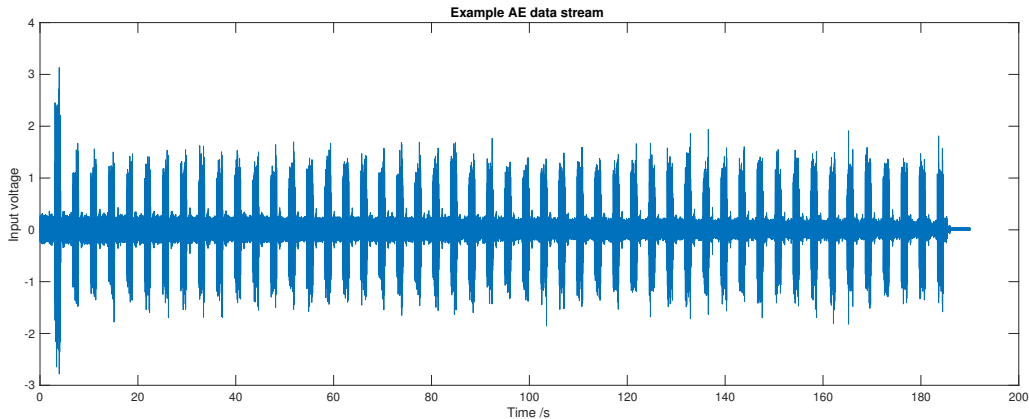


Figure 4.1: Plot showing example AE dataset.

longer a concern. The distinct first cut is simply due to the fact that the initial cut in a fresh piece of material uses a greater area of the tool's surface when compared to the 0.7mm step experienced by the cuts that follow. In reality, any surface undergoing a finishing operation has already had numerous roughing procedures performed first, allowing the primary finishing cut to start with minimal workpiece engagement, so as to minimise the risk of damage. In this experimental setup, the first cut information is merely discarded as irrelevant and unrepresentative of the common scenario experienced by such a tool.

The acquisition rate for all AE data presented here is 1 MHz or one million samples per second; this is adequate for representation of signals up to the top of the consistent portion of the sensor's frequency response at 400 kHz. While the high sample rate is beneficial in allowing the full band of stable sensitivity to be used, it also results in a huge amount of data capture. Again, this provides a further basis for extracting only those in-cut data from the full continuous dataset. The most straightforward method of extracting individual cuts is through detecting a rising edge in the signal that exceeds a preset threshold, keeping only the data from this point until a similar falling edge is detected. Setting a detection limit of 0.5V results in clear identification of the in-cut region, as shown in the 15 second example in Figure 4.2. The noise present due to rapid machine moves and high-pressure coolant is also more easily visible in this figure, which can be seen as the non-zero data portion between cuts. An example plot of the in-cut region is shown in Figure 4.3, with a more zoomed-in view of this shown in Figure 4.4.

By zooming further in to the data collected during the cutting operation, it is clear to see the bursts of AE released as each of the six flutes cuts a chip during a single

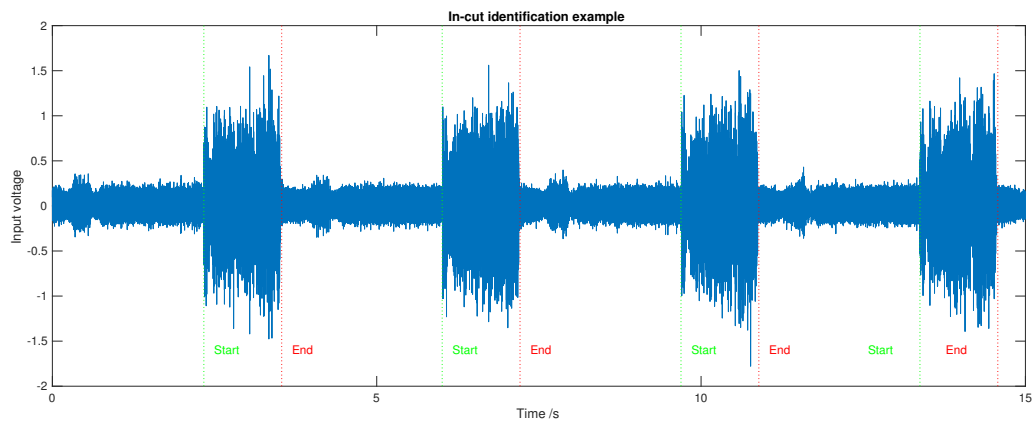


Figure 4.2: Plot showing example AE dataset with in-cut regions identified.

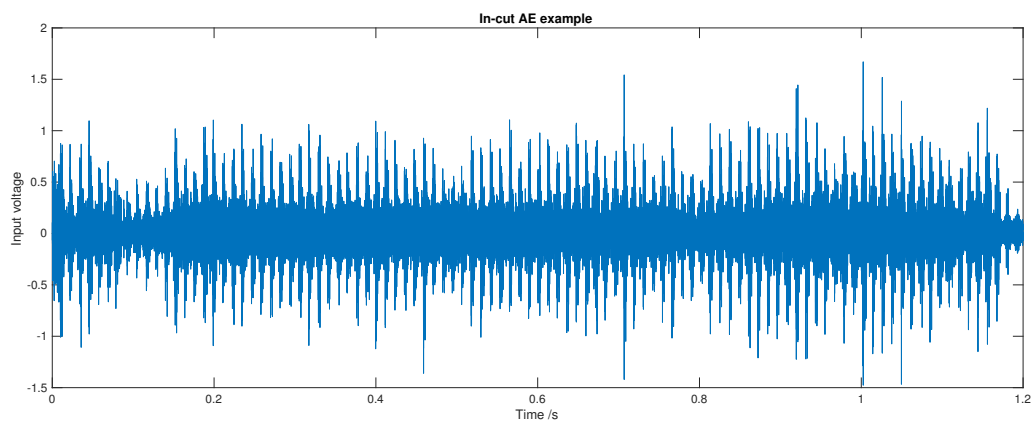


Figure 4.3: Plot showing example in-cut AE data during a single cut.

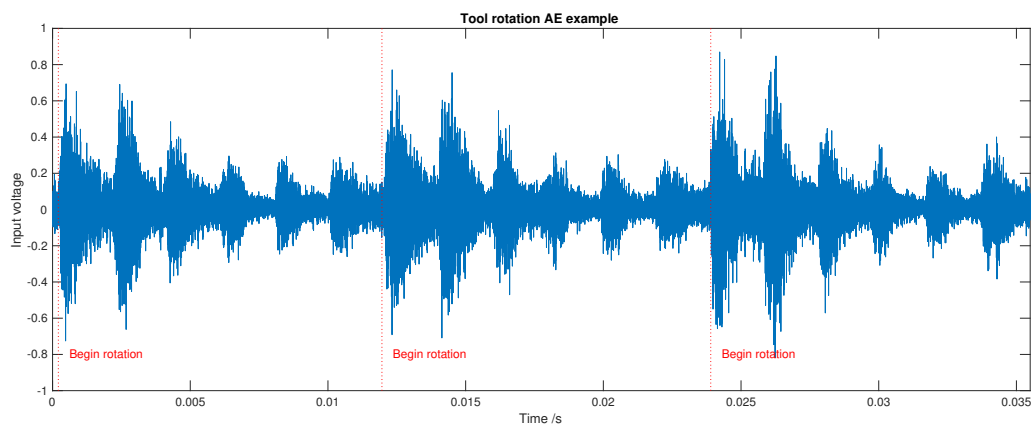


Figure 4.4: Plot showing zoomed in-cut data with tool rotations indicated.

rotation of the tool, with tool runout also indicated by the change in amplitude between flutes of a new tool. Identifying individual flute passes can be confirmed by measuring the time between each AE burst recorded and comparing this to the theoretical tooth-passing frequency (tpf). Measuring the average time between six AE bursts results in a period of 0.0118 seconds, giving a rotational frequency of 85.1 Hz and tpi of 14.18 Hz. The theoretical tpi is given by,

$$\begin{aligned}
 tpf &= \frac{N}{60 * T_n} \\
 &= \frac{5082}{60 * 6} \\
 &= 14.12Hz
 \end{aligned}
 \tag{4.1}$$

where T_n is the number of flutes or teeth present on a given tool, and N is the spindle speed in revolutions per minute. Comparing both the measured and theoretical tpf indicates that the bursts visible in Figure 4.4 are indeed due to the individual flute's cutting action as the tool rotates, given the minimal 0.4% error. Each experimental procedure dataset is split into separate files containing only in-cut data, as this retains all relevant information while discarding that which is extraneous. While performing analysis on a flute-by-flute basis is an interesting avenue for exploration, it is not a focus here due to the likely difference in runout between each tool, the curved coupon nature, and inability to determine which flute is first to engage the workpiece all having further effects on the gathered data which would require modelling.

The Kistler 8152B1 Piezotron sensor features both internal 50 kHz high-pass and 1MHz low-pass second-order Butterworth filters, ensuring only information from within the sensor's optimal frequency response region is collected. The result of this internal filtering is that minimal signal processing is required to remove noise prior to feature extraction.

4.2 Time domain feature extraction

Once the captured AE data has been split into its 49 individual cuts per tool, potentially meaningful features can be explored to investigate trends between the nine

tools used for each trial. As discussed in Chapter 2, it is common practice when exploring signals, to focus on features in the time, frequency, and time-frequency domains depending upon specific applications and the information required. In this case, time domain features are considered the most logical starting point considering that they are independent of the frequency components caused due to process parameters that may change. It is also expected that as a tool experiences wear, the changes in chip formation will affect the energy released in the form of acoustic emissions, thus having a direct impact on the amplitude of the signals obtained.

It has previously been shown that time-domain features (such as a signal's root-mean-square (RMS)), increase in proportion to the amount of flank wear present [18], a phenomenon which is to be expected, as the RMS is a measure of power within a signal, and a worn tool is likely to generate increased friction in the cutting zone. It is also worth noting here, that due to crater wear posing a significant change to the effective rake angle of a tool, any cratering can mask the effect progressive flank wear has on the AE signal's energy content [18]. 3D tool scans indicate that flank wear, however, is the dominant wear mechanism in this process, ensuring that tools can be readily reground and reused rather than discarded after each cycle. That being said, it is entirely possible to observe and characterise a change in wear mechanism if required, through careful observation of tool wear features.

As AE power content is proportional to the strain rate of the cutting operation [18], any classification limits set based on one set of parameters, will need updating as changes to these are made. Fortunately, there is ongoing work on strain-rate calculation based on machining parameters [73] and software packages such as the CutPro machining simulation application, which present a potential opportunity for any proportional changes to be made as process parameters are altered, although this is yet to be conducted to the authors knowledge.

The time domain features of interest are peak AE value (AE_{pk}), mean AE value (AE_{avg}) or DC offset, AE RMS (AE_{rms}), skewness of the AE signal (AE_{skew}), and the kurtosis of the AE signal (AE_{kurt}). Each feature is calculated over a single cut period, as shown in Figure 4.2, to ensure effects due to runout are minimised and the resulting value is a culmination of a number of chips, reducing uncertainty due to chip variations and also computational effort. Taking an amplitude-based envelope of the signal and calculating the aforementioned features of this provides a clear way to visualise how the AE_{pk} and AE_{rms} vary with differing energy levels released during chip formation.

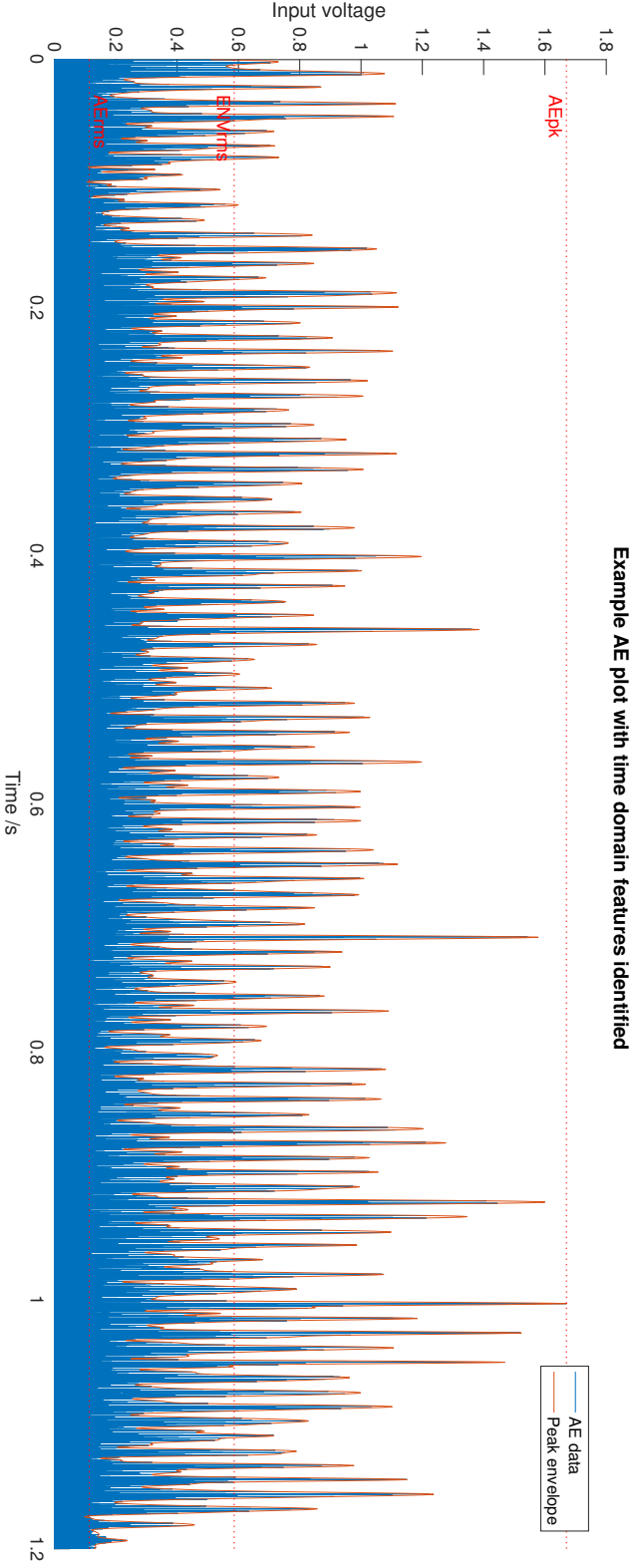


Figure 4.5: Plot showing example in-cut AE with envelope and features indicated.

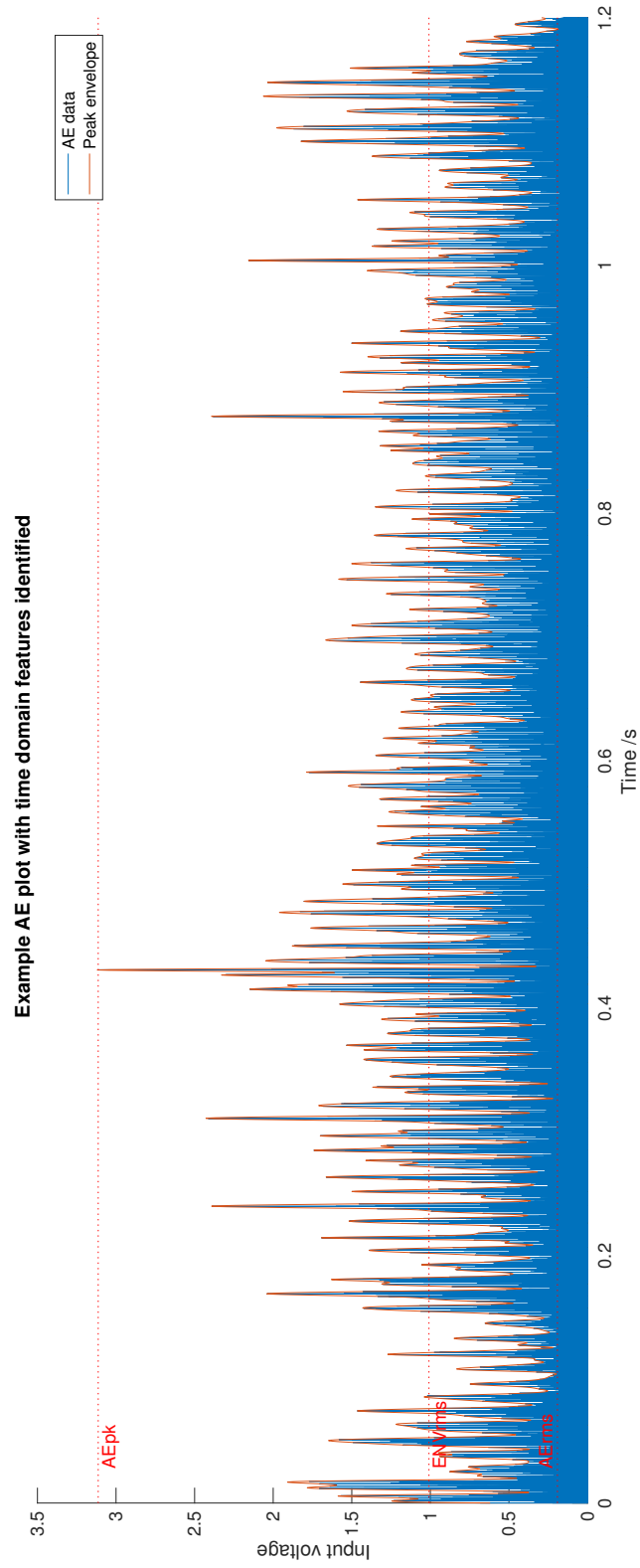


Figure 4.6: Plot showing example worn in-cut AE with envelope and features indicated.

Figure 4.5 shows example in-cut data from a single cut with a relatively unworn tool, while Figure 4.6 shows the same example from a more worn tool. In both plots, the AE_{pk} and AE_{rms} are indicated for both the signal and a peak based envelope, allowing a few key differences to be readily identified. While the envelope does not contain any further information than is present in the raw signal, it serves two purposes. Firstly, it acts as a low-pass filter, aiding in identification of signal peaks by removing the effects of the dense, high-frequency, low-amplitude signal components. A visual example can be seen by comparing both the AE_{rms} and EN_{rms} in Figures 4.5 and 4.6. Secondly, the envelope provides a method of down-sampling the data by discarding high-frequency information, in turn, reducing computational load. Both the peak and RMS values can be seen to increase with a more worn tool specimen, which is clearly visible by considering the RMS lines of the signal envelope varying from approximately 0.6 to 1.0 V, and the peak lines from 1.7 to 3.1 V correspondingly. Correlations between these features, material surface, and tool wear statistics will be explored further in Section 4.5, after specifying features of interest relating to both tool wear and surface integrity. The overall shape of both plots follow a similar form, such as the drop in amplitude at around 0.1 seconds through the cut, which is due to the consistent profile of the surface being machined and the tool diameter changing in the same manner as the process progresses.

4.3 Material specimen features

A requirement of all parts produced by SLS, is that they meet certain criteria detailed in strict specifications to ensure part quality and uniformity. While specific details of these documents are restricted, it is common knowledge that they define a number of parameters relating to the surface finish of the component which is a direct result of the machining process. While some of the parameters are a result of part geometry and tool paths, others are influenced more by the tool state and level of wear such as surface texture requirements. Surface texture is often quantified in terms of deviation from an ideal form with larger values indicating rougher surfaces. Common roughness measures include the arithmetical mean deviation of the roughness profile (Ra), the root mean square of the roughness profile over the evaluation length (Rq), and the maximum height of the roughness profile (Rz) [74]. Such measures are, however, two-dimensional and unfortunately do not fully describe the complex geometries of the cutting effects created with a ball nosed profile milling op-

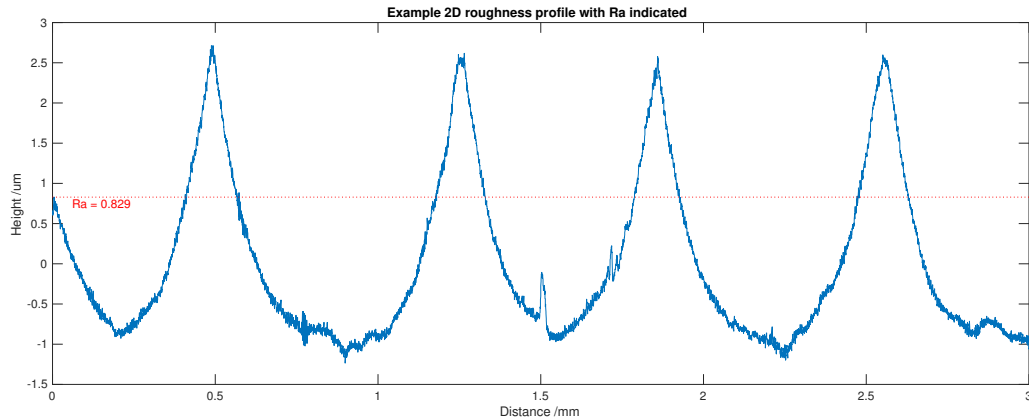


Figure 4.7: Plot showing example 2D roughness profile with Ra indicated.

eration [75], and consequently, three-dimensional roughness statistics (denoted with an S rather than an R) are considered more reliable in this scenario. The drawback of such statistics is that they are not currently governed by an ISO standard in the same way as two-dimensional statistics are, and coupled with the fact that more complex scanning equipment is needed to acquire measurements, they are seldom used in industry.

Both 2D and 3D surface measurements were taken with a stylus-based profiler and optical measurement device respectively, allowing comparison of features from each. Figure 4.7 provides an example profile measured with a 2D profiler across three cusps, the Ra value indicated having been calculated from an average of five measurements. Similarly, Figure 4.8 provides two example images of 3D surface profiles over a single cusp each, where a 2D profile is simply a single slice in y through either of these. Both Figure 4.8(a) and Figure 4.8(b) would provide a similar 2D profile if measured across the centre of the image; however, the shape of the two profiles are largely different over the remaining area, which can only be measured with a 3D scan or careful positioning of the profiling stylus. This issue is overcome in this thesis by taking multiple 2D profiles across a predetermined area and averaging these, ensuring scans are taken over the full radial depth of cut. Relationships between cutting parameters and surface roughness profiles are described in more detail in [21], where it is also highlighted that tool wear has a profound impact on surface roughness through deflections and thermal effects.

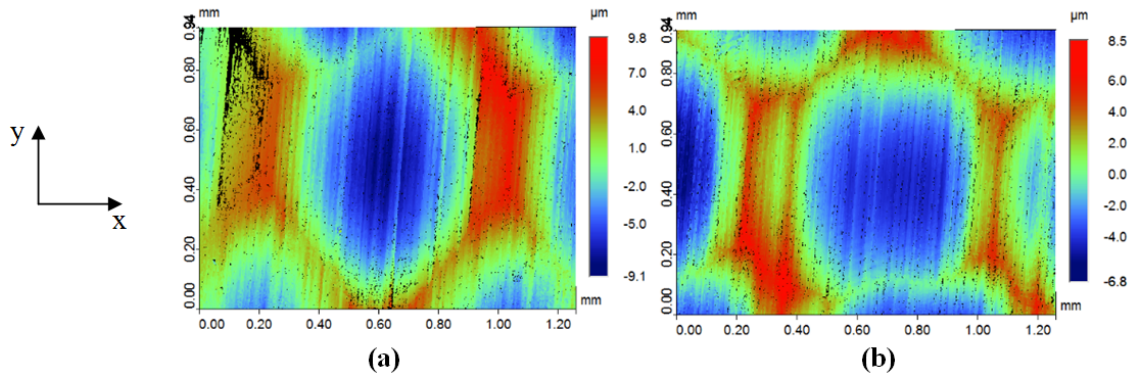


Figure 4.8: Images showing example 3D roughness profiles from a surface machined by a less worn (a) and more worn (b) tool respectively.

4.4 Tool scan features

As discussed in Chapter 3, the Alicona InfiniteFocusSL is used for all tool edge measurement, providing a 3D surface similar to that created from 3D material profiles, where individual wear profiles can be extracted at numerous points along the cutting edge. Each flute was scanned five times, and these scans were averaged, as with the material specimens, to ensure minimum chance of anomalous measurements. Each measurement is also an average of one hundred slices through the cutting edge, in an effort to minimise the effect of possible pick-up on results. Table 3.2 contains a list of the calculated features reported by the Alicona for each tool inspection. These features can then be correlated with both those extracted from the AE signals and from material scans to identify the most closely-related parameters.

Figure 4.9 shows a 3D scan of a flute from a new tool, with the corresponding profile shown in Figure 4.10. The cutting radius r of this flute is equal to 14.8 microns, and the wedge angle β equal to 61 degrees. Looking at the 3D image, a uniform, sharp cutting edge can be seen across the length of the flute, and the clearance and rake faces are smooth. This is to be expected from an unused tool, given that no wear should be present.

Comparing Figures 4.9 and 4.10 with Figures 4.11 and 4.12, the effect of short usage duration on the cutting surfaces can be seen. Flank wear begins, and material is abrasively removed from the flank face of the tool. As the tool is used further, flank wear continues, resulting in the image shown in Figure 4.13. This progression of flank wear indicates that this is the dominant wear type for this process in the early

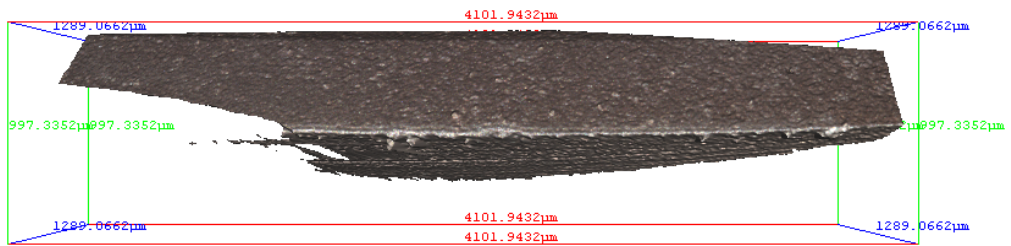


Figure 4.9: 3D scan of a single flute from a new cutting tool.

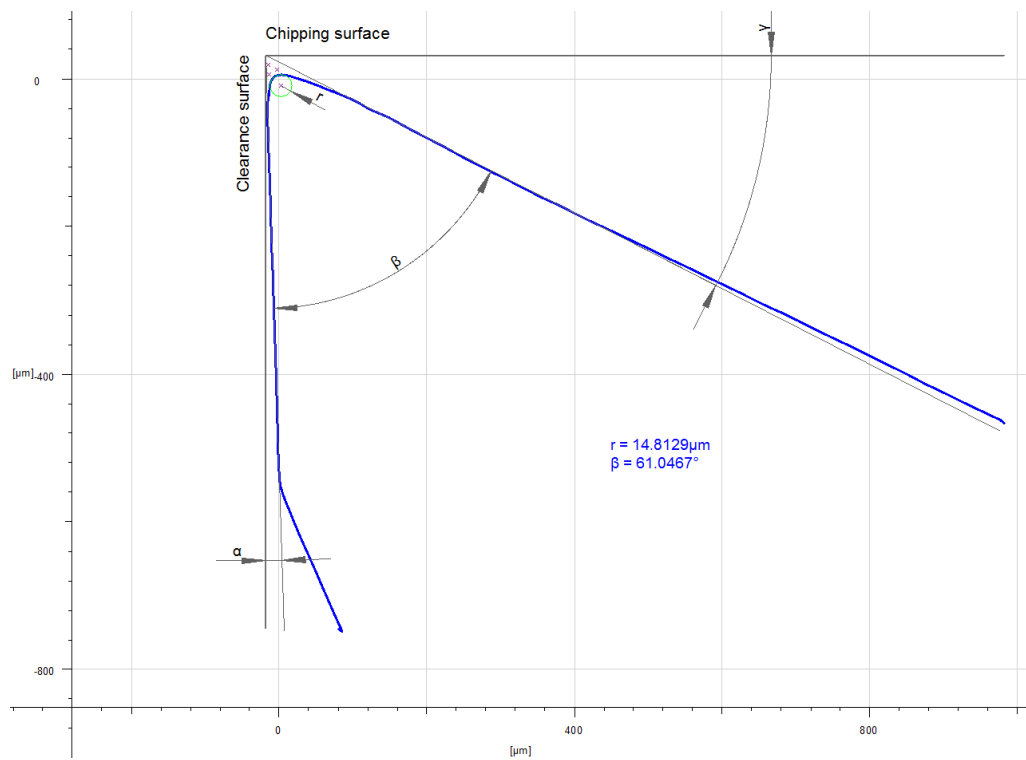


Figure 4.10: Average profile from 100 scans of a single flute from a new cutting tool.

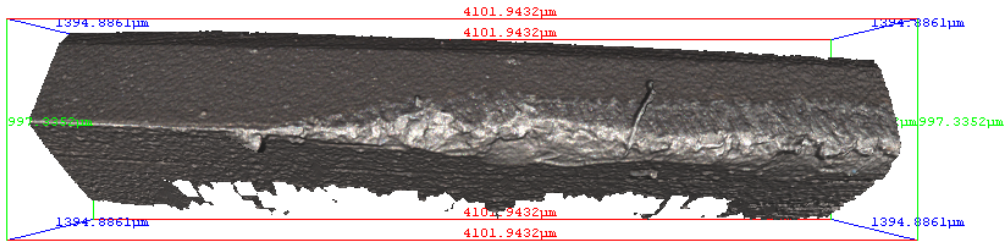


Figure 4.11: 3D scan of a single flute showing minor flank wear.

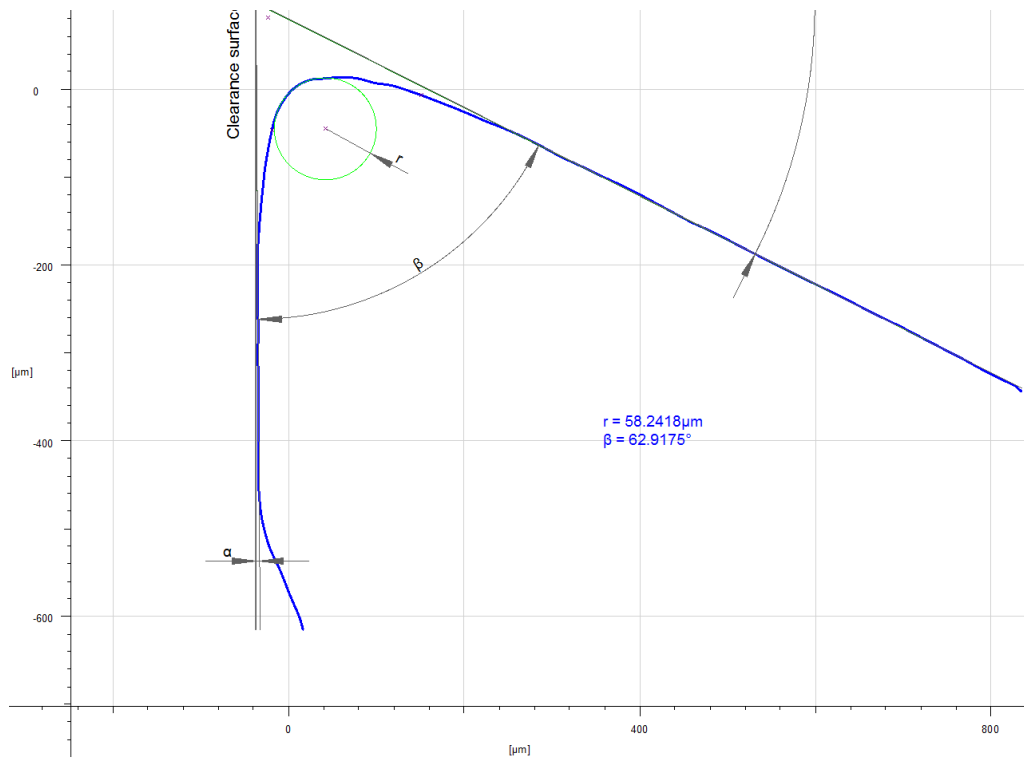


Figure 4.12: Average profile from 100 scans of a single flute with minor flank wear.

stages of use.

Historically, tool wear has been quantified by measuring the distance between the tool's edge and the bottom of the flank worn area [5, 23], however, the method of obtaining 3D scans used in this thesis, provides a level of information that has not been readily available before. Sick [24] states that approximately 69% of previous work uses the width of flank wear area as the primary wear measure. The profiles in Figures 4.10, 4.12, and 4.14 show that this flank wear can also be determined through measurement of the cutting-edge radius. It is to be expected that the cutting edge becomes blunted as wear progresses, and this can be seen through the increase in cutting-edge radius from 15 microns, to 65 microns. The wedge angle,

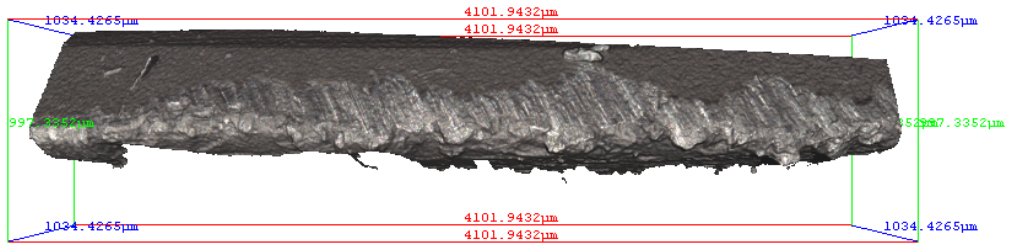


Figure 4.13: 3D scan of a single flute showing further flank wear.

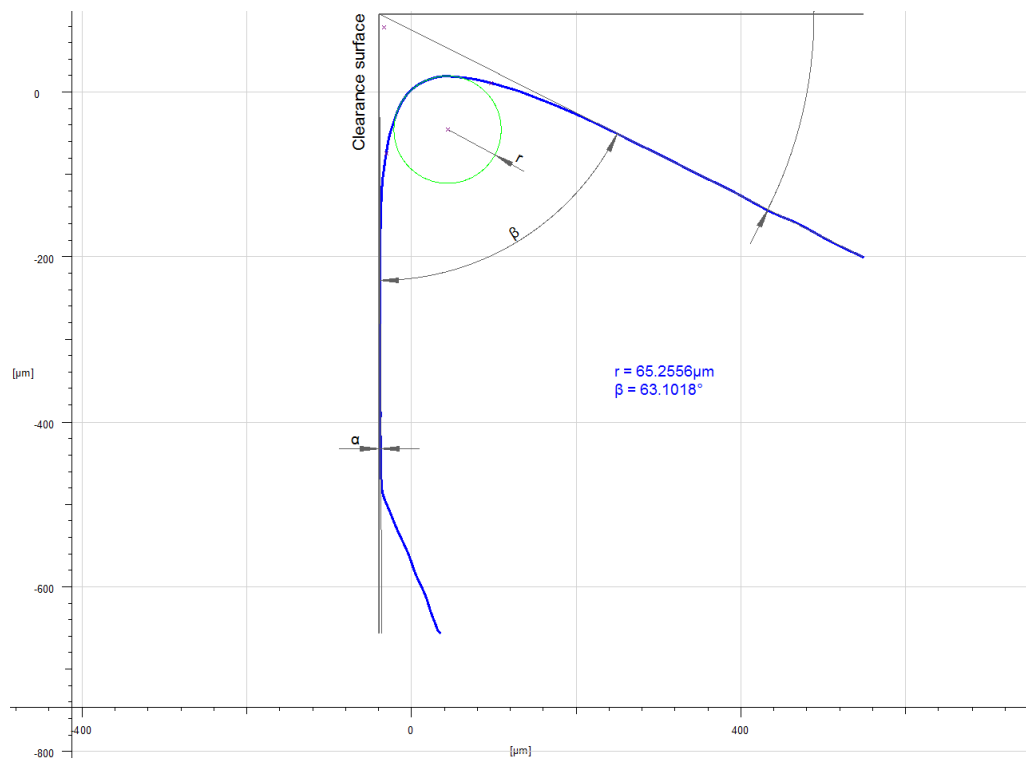


Figure 4.14: Average profile from 100 scans of a single flute with further flank wear.

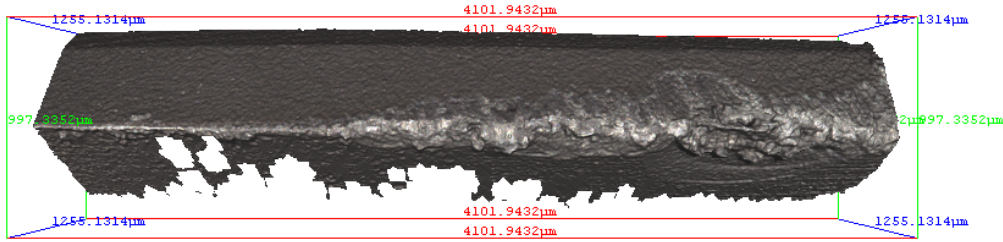


Figure 4.15: 3D scan of a single flute showing the beginning of crater wear.

β , also emerges as a useful measure of tool wear in this thesis, and can be seen to increase with tool wear in a similar fashion to the cutting-edge radius. Considering Figures 4.15 and 4.16, crater wear can be seen to develop as the wear type changes from extended use. This type of wear also has the effect of increasing the measured wedge angle, although, if the cutting-edge radius measurement is taken across the crater, it appears as if the value reduces, indicating a sharpening of the tool. This phenomenon can be seen in the profile in Figure 4.17, which is taken across the crater shown in Figure 4.16. Despite the tool being used for longer than that in Figure 4.14, the cutting-edge radius value is lower. This highlights the importance of careful measurement of each flute, and the necessity to take an average of measurements across the flutes length.

Once crater wear has begun, the tool rapidly deteriorates, with crater wear leading to fracturing of the tool's cutting edge. Figures 4.18 and 4.19 show a tool which has progressed to advanced crater wear; the corresponding profile measurement is shown in Figure 4.20. Again, both the cutting-edge radius and wedge angle have further increased as the tool is used further. Finally, Figure 4.21 shows a tool which has worn past both flank and crater wear, and has now begun to fracture. A tool in this state cannot be used any further, and would be discarded if found during manufacture, due to the likely negative effect it would have on dimensional accuracy of the finished part. This figure shows clearly the difference between flank wear, crater wear, and final fracture from left to right, and on the different faces of the tool on which they occur.

As shown in the preceding example figures, the tools used throughout the experimental work in this thesis are dominated by flank wear for the most part, with crater wear presenting as usage is continued, finally leading to failure. Both the cutting-edge radius and wedge angle are found to describe wear level despite differing dominant wear types, and so throughout the remainder of this thesis, the ex-

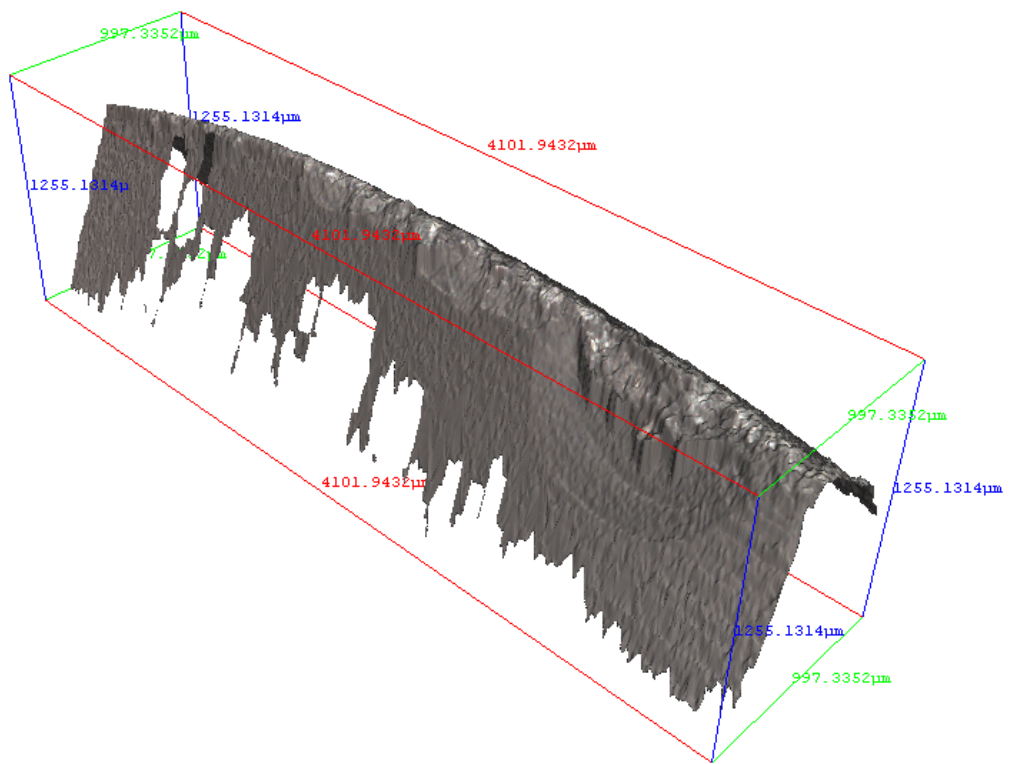


Figure 4.16: 3D scan of a single flute showing the beginning of crater wear, viewed from the side.

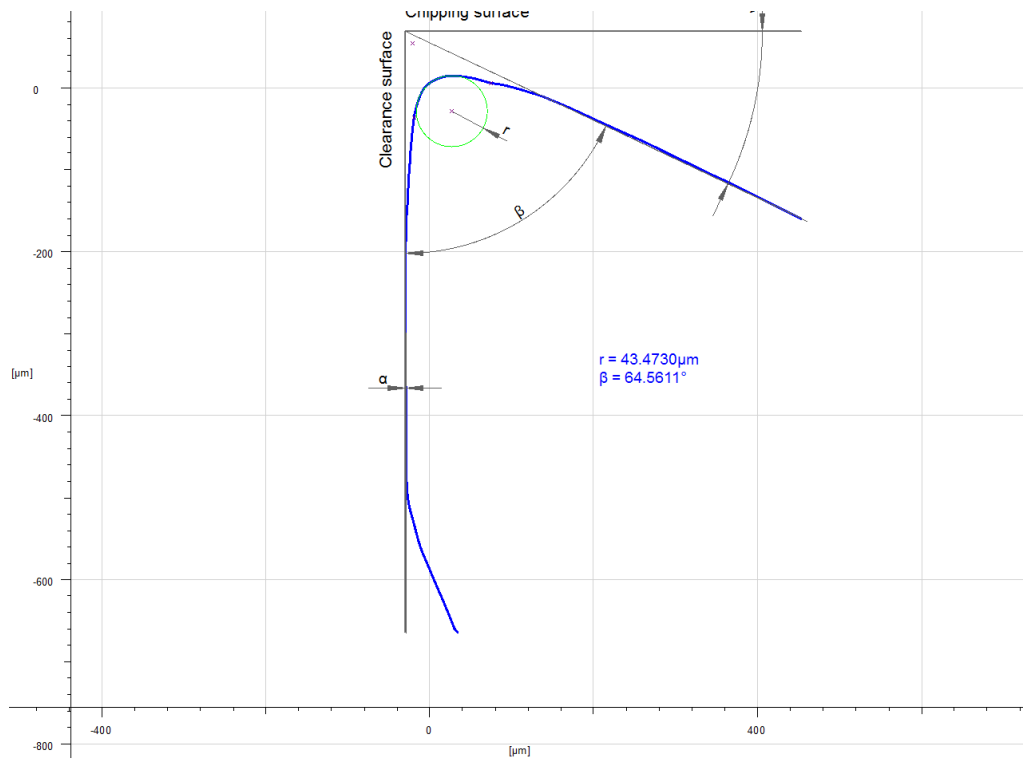


Figure 4.17: Profile from scan of a single flute with crater wear, taken across the crater.

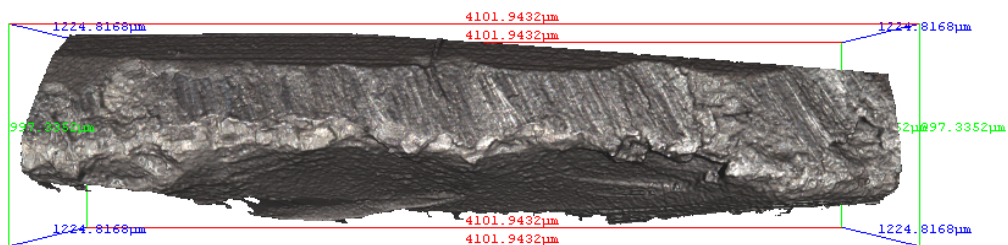


Figure 4.18: 3D scan of a single flute showing advanced crater wear.

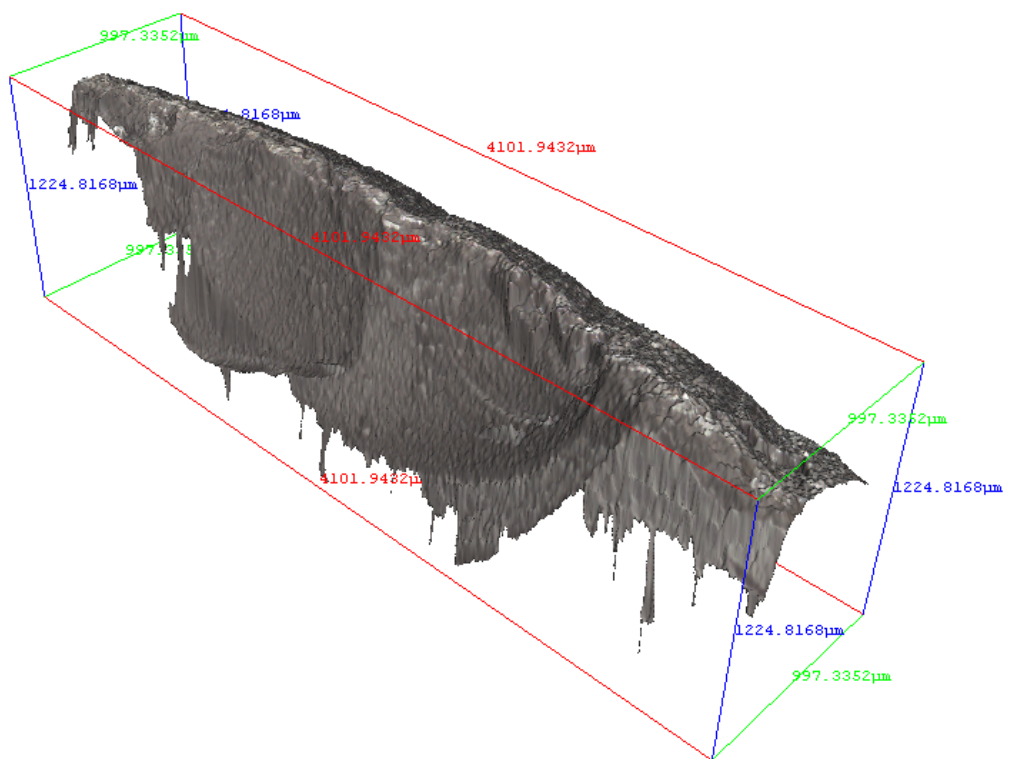


Figure 4.19: 3D scan of a single flute showing advanced crater wear, viewed from the side.

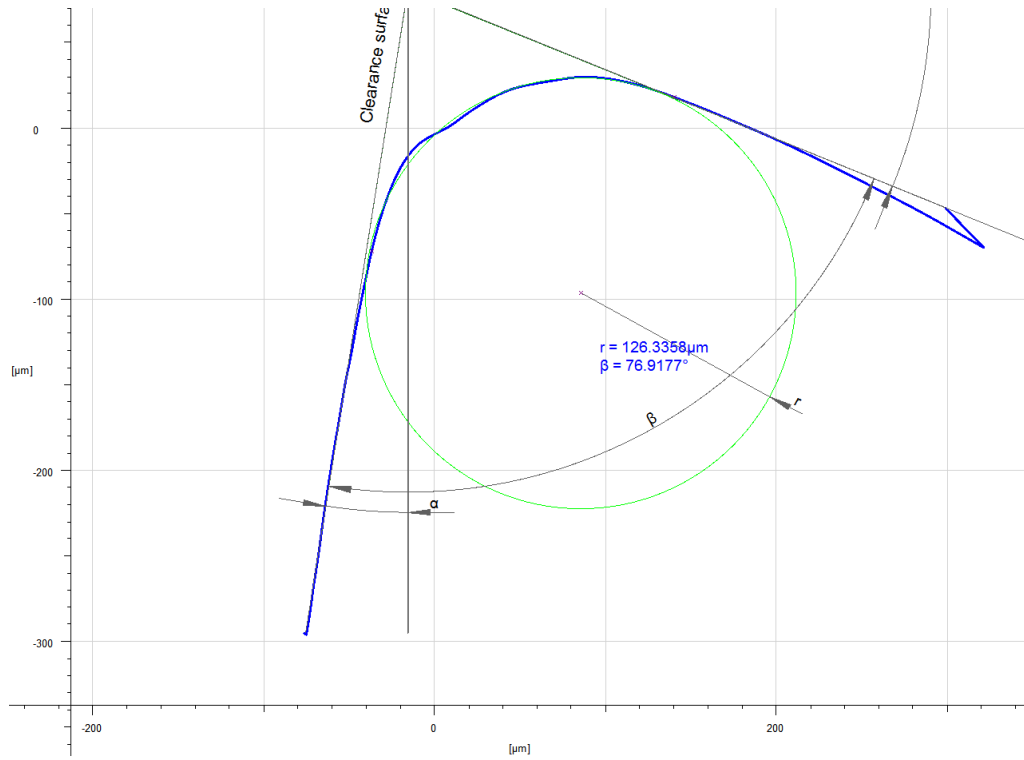


Figure 4.20: Average profile from 100 scans of a single flute with advanced crater wear.

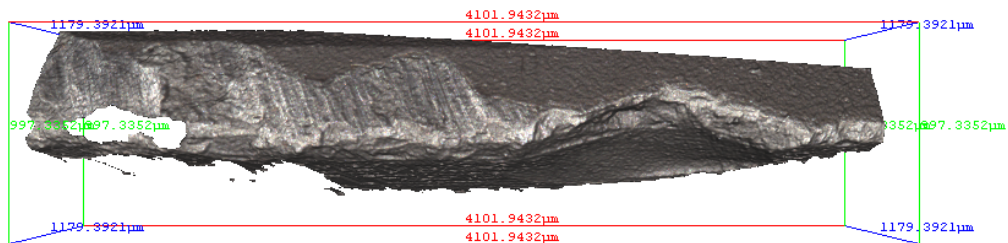


Figure 4.21: 3D scan of a single flute showing fracturing.

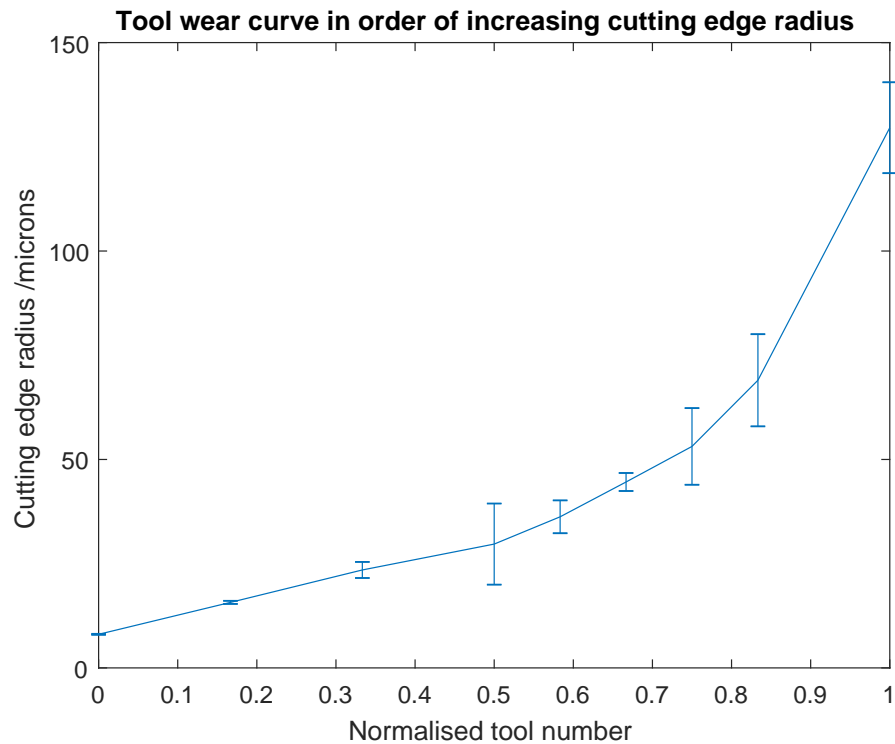


Figure 4.22: Wear curve from first experimental trial. Each point represents an average from 5 individual scans, the error bars represent the standard error between the 5 scans.

perimental tools are treated in order of wear level accordingly, determined through measurement of cutting-edge radius and wedge angle from 3D scans. A wear curve, based upon cutting-edge radius from an average of 5 scans, and ordered by increasing wear level, can be seen in Figure 4.22. Each point represents an average from five individual scans, and the error bars represent the standard error between the five scans.

4.5 Correlations between material, tool, and AE data

Given that a number of potentially interesting time-series features have been identified, it is possible to normalise these and perform an exhaustive search for features with the highest correlation coefficients between data types over the life cycle of a tool. For the purpose of this work, correlations are considered strong if the correla-

Tool	AE	Material
r	AE_{pk}	Ra
α	AE_{rms}	Rq
β	AE_{skew}	Rz
γ	AE_{kurt}	Sa
Ecq		Sq
Dpos		Sz
Dmean		
Vp		
Vv		
SIMcd		
SIMch		
SIMt		

Table 4.1: Table to show features used to explore correlations. Tool measures as defined in Chapter 3.7. Material surface measures as defined in Chapter 4.3

tion coefficient between datasets is greater than 0.67 with a corresponding p-value of less than 0.05 when testing the hypothesis of no correlation - this is in agreement with the critical values of the Pearson product-moment correlation coefficient [76]. The p-value is the probability of finding more extreme results than those observed when the null hypothesis is true; in this case, providing an indication that the observations are truly correlated.

The correlation coefficient is a measure of linear dependence between two random variables x and y , and returns a value between -1 and +1 relating to perfect negative and perfect positive linear correlations respectively. When applied to a sample of n observations, it is defined as:

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right) \quad (4.2)$$

where \bar{x} is the mean and σ_x the standard deviation of x ; similarly, \bar{y} and σ_y are the mean and standard deviation of y .

Performing this calculation between all AE, tool, and material features provides a method of identifying those features which best relate to one another and are, therefore, considered the most beneficial to explore further. Table 4.1 shows the inputs into this comprehensive search.

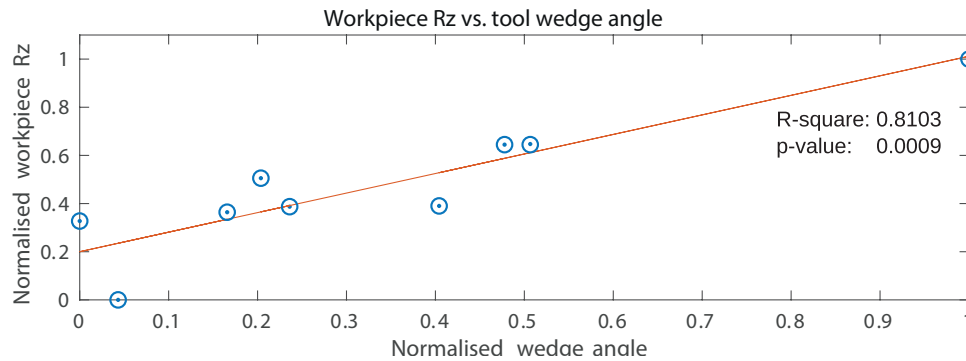


Figure 4.23: Plot of workpiece Rz against tool wedge angle (scan location 1).

By calculating both r and p values, it is a straightforward task to order these in ascending order, indicating the features with greatest correlations between them. Between tool and workpiece features, it is found that the 2D roughness features provide greater correlation with obtained tool scan measurements when compared to their 3D counterparts. This is most likely due to the greater scan area used for 2D measurements, when compared to 3D, and the greater generality this provides. It is also found that the tool's wedge angle bears the greatest relation to these same roughness parameters, an unsurprising result considering the fundamental effect that changing rake and clearance angles have on chip formation and, therefore, surface generation. Figure 4.23 shows this relationship and linear fit with the R^2 and p -value of 0.81 and 0.0009 indicated as an example of this relationship. In addition, similar correlations are observed with both Ra and Rq values as the material roughness measures are closely related to one another for an individual specimen.

In a similar fashion, the features obtained from a dataset of nine tools can be compared to a selection of those taken from in-cut recorded AE taken from machining with each of the aforementioned tools. An example correlation matrix plot can be seen in Figure 4.24, where the peak AE measurements are shown to bear the strongest relationship to the observed wear features - more specifically, the wedge angle and form deviation from a circle. As in Section 4.4, the tool's wedge angle is a useful feature for quantifying tool wear in this thesis, and the peak of the AE signal during cutting is found to correlate well to this. Figure 4.25 indicates this linear relationship with an R^2 value of 0.744 and p -value of 0.0028 shown. Peak values taken from each AE sample providing useful information is again, as expected, given that AE signals are released in bursts during material deformation and will be recorded as spikes in the signals accordingly, having negligible effect on the un-enveloped signal RMS for a relatively long sample length.

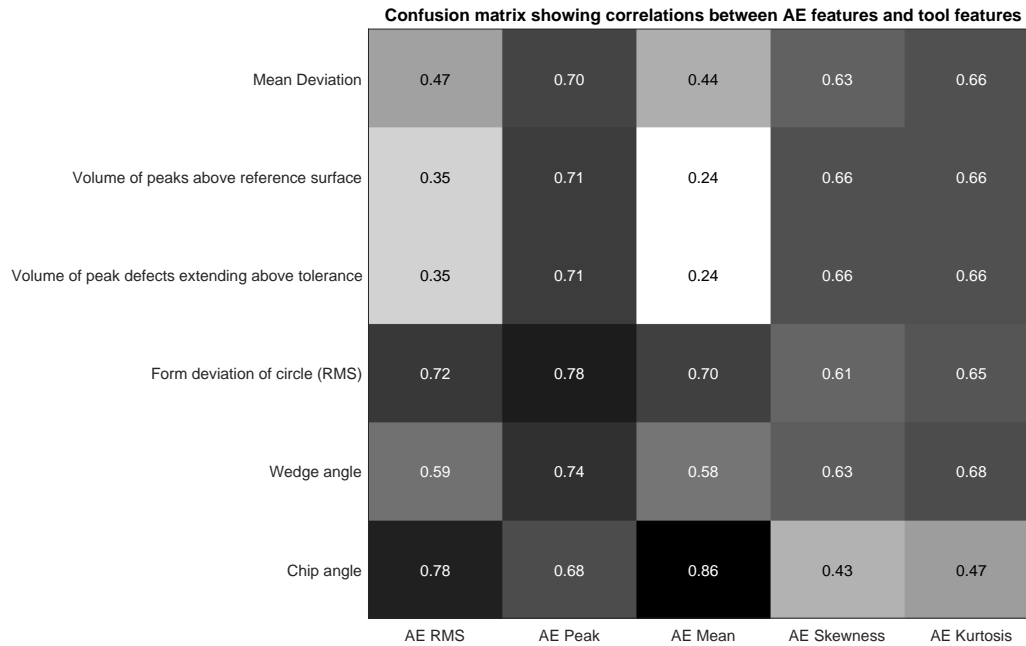


Figure 4.24: Correlation matrix showing relationships between a selection of tool features and AE features.

Both Figures 4.23 and 4.25 indicate linear relationships between a select few features, which is by no means the full extent of the relationships between tool, workpiece, and recorded AE. The presented relationships do, however, provide essential evidence that the belief that these features must be connected has substance, despite the limited sample size. It is also worth noting that each data point presented in these figures is from a unique tool, and subsequently, has inherent variations due to manufacturing tolerances and tool-holding accuracy; a result of the test procedure that would not be present in a production environment. The presence of such correlations provides grounds to continue exploring connections between datasets, aiming for the ability to predict tool state (and, therefore, material surface roughness) given a sample of in-process recorded AE data.

4.6 Other acoustic emission-related features

For completeness, it is necessary to discuss a select few frequency domain features while also exploring signal development during the relatively short machining period experienced by each tool during data collection. Computing the discrete Fourier

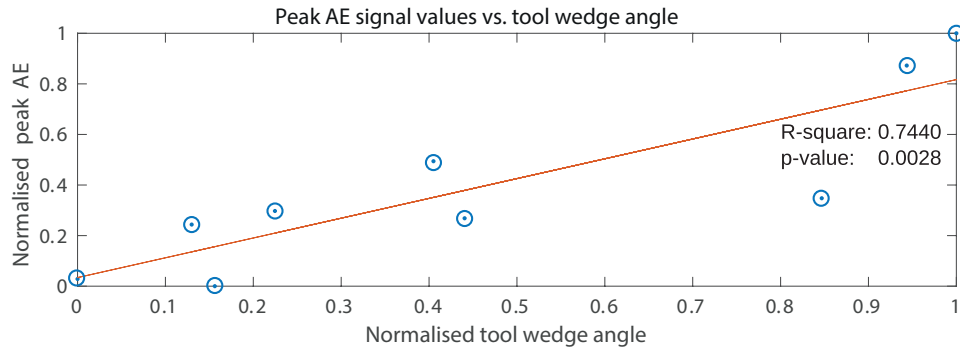


Figure 4.25: Plot of peak of AE signals against tool wedge angle (scan location 2).

transform (DFT) over each cut using the fast Fourier transform (FFT) algorithm provides a clear method of visualising the frequency content of each recording. By plotting these transformed data in order, the plot in Figure 4.26 can be obtained in a similar way to a spectrogram. This figure shows that the magnitude of the frequency content between 70 to 100kHz increases with wear, indicating more energy is released at these frequencies as tools become more worn. While this thesis will continue to focus on time-domain features, understanding how the frequency content of the measured AE varies with wear, provides another avenue for exploration.

4.7 Principal component analysis

Principal component analysis (PCA) is a widely used tool in exploratory data analysis due to its wide range of possible uses. Primarily, PCA is a tool for identifying patterns in data sets of high dimension, where simply displaying the data in a graphical form is not possible. Once any significant patterns have been found in the data, redundant data can be discarded, enabling PCA to be used as a tool for dimensional reduction, data compression, variable selection, and data classification to name a few uses [55]. In this work, where a number of feature variables have been calculated, it is possible to perform a PCA to aid in identification and generation of fewer, more meaningful features, and to then help cluster data into possible tool states. Once potential clusters have been identified, a statistical learning methodology such as a support vector machine (SVM) can be implemented to classify incoming data. This work is well suited to such a supervised learning model given that the in-cut recorded AE data correspond to a single tool from a set of nine, providing an intuitive method of labelling nine categories for data separation.

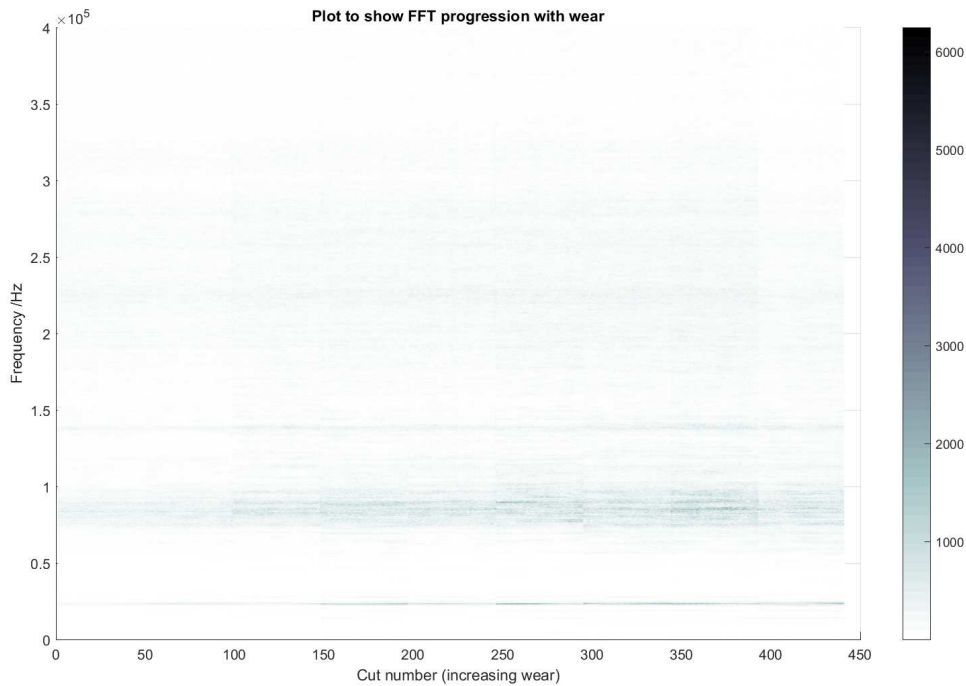


Figure 4.26: Plot to show frequency content of AE signals as wear progresses.

Figure 4.27 shows the first and second principal component scores, calculated from the time-domain AE features previously described. Each of the 9 clusters corresponds to AE data points generated using a single, individual tool, and can, therefore, be treated as being a grouping of points generated from a similar wear level. Naturally, nine unique tools result in nine unique clusters, and given that a number of the clusters can be clearly identified, with a reasonable degree of separation, this problem becomes well suited to a classification problem and is explored further in the following chapter. It should be noted, however, that if the colours were removed from the plot, separating of the data in this space becomes a more difficult task. This is also to be expected in a production environment, where tool wear is a progressive process, and clusters are, therefore, likely to be very close and potentially hard to create decision boundaries from. Nonetheless, treating the acquired data as a classification problem provides confirmation that repeatable relationships exist between observed AE data and measured tool wear values.

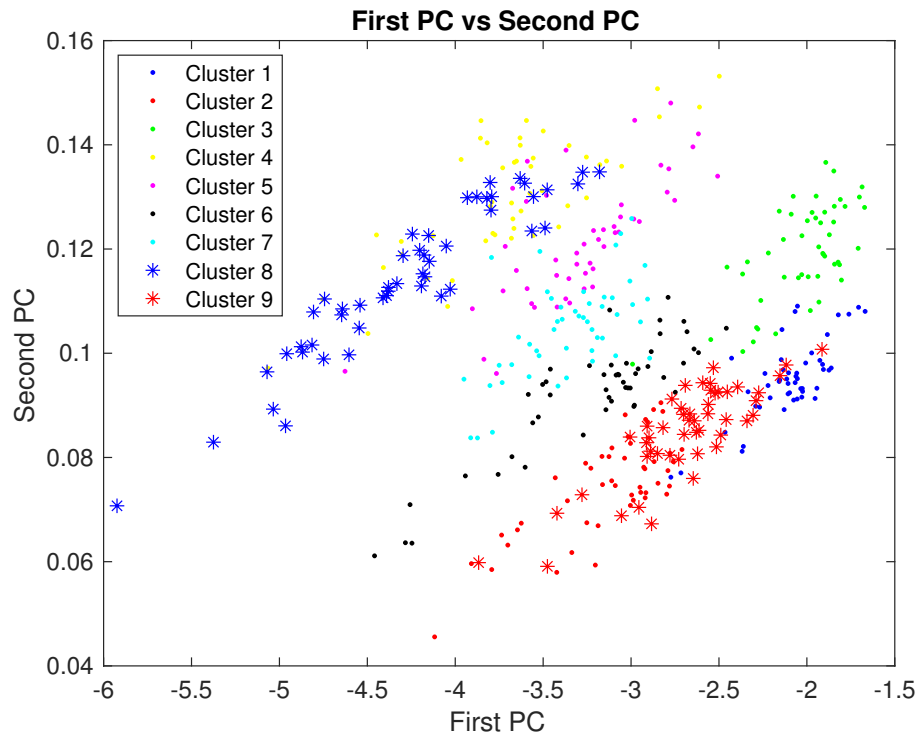


Figure 4.27: Plot showing first vs. second principal component of AE features.

4.8 Overview of preliminary analysis

This chapter has given an overview of the data collected during machining, detailing that obtained from tool scans, material topography, and AE recordings. It is shown that there is a large amount of information relating to the cutting process within the in-cut AE data recordings, and it is entirely possible to correlate a number of features based around this with both tool and material properties. Although this chapter only touches on a small selection of potential linear relationships between features, it provides grounds for further exploration into more complex interactions between features, both linear and non-linear in nature, while suggesting a method of wear diagnosis based on tool state clustering.

TOOL STATE CLASSIFICATION

Given that a selection of relationships have been established between various key process-generated properties, the first logical step in building a probabilistic model is to begin with exploring whether a learning algorithm is capable of classifying the incoming acoustic emission features into a discrete number of clusters based upon the tool generating these data. If this state diagnosis is possible, it provides firm evidence that the signals develop in a unique and distinct manner as the process continues, providing grounds to explore predictive or prognostic modelling further. This chapter, therefore, explores a static classification problem and examines whether it is possible to accurately separate a number of AE recordings based upon the associated wear levels.

5.1 Classification methods

State classification is not a ‘one-size-fits-all’ approach, rather it is possible with a number of different methods and techniques, each varying in suitability for a given problem. Primarily, the classification problem can be split into two subcategories; those requiring supervised and unsupervised learning approaches respectively, the latter also being known as clustering or cluster analysis in some communities.

The supervised approach relies on training a classifier based upon a training set containing a range of example data akin to that which the model is expected to work upon when implemented. In this case, each vector of input features is paired

with a correctly identified cluster to which it belongs, often identified manually through pre-existing knowledge of the underlying data-generating process. Once the classifier is trained and validated, it can be used to identify which pre-existing group a new observation belongs to, assuming it is similar to that which the model has experienced during training and is from a similar generating process.

Clustering, on the other hand, is an unsupervised approach that aims to identify groupings of points in a dataset based on their similarity to their neighbours and dissimilarity to more distant points, without any previous knowledge of potential categories. Such algorithms tend to explore spacings between data points using some chosen distance measure, sorting new members based upon the minimum distance between them and other, existing members. Once a number of clusters are identified, the problem becomes very similar to the classification problem described previously; however, clustering tends to be an iterative process whereby cluster boundaries are updated as new data are experienced.

Based upon the data described in Chapter 4, this work is well suited to a classification-based approach to state diagnosis, given that each AE data point obtained can be grouped with others generated using that same tool and, therefore, level of wear. It follows that, in this case, incoming data vectors can be assigned to one of nine possible states when considering two independent trial repeats, or one of eighteen states when merging both datasets obtained for this thesis. Hence, clustering will not be touched on further here; the focus will be on classification algorithms, given their relevance to machining as discussed in Chapter 2.5.

A visual example of data groupings can be obtained by performing a principal component analysis (PCA) on the selected AE features given in Chapter 4, reducing the dimensionality to two dimensions (in this case) that contain over 95% of the variance present (Figure 5.1), and can be easily represented graphically such as previously in Figure 4.27. This figure clearly shows how data generated with uniquely worn tools forms a number of groups which can be readily identified by the reader. While there is some overlap present in this particular 2D example, it should be noted that minimal overlap between clusters makes for a simpler solution and, therefore, input features should be chosen that maximise the distance between groups.

A range of algorithms are available for performing such a classification given an example training set, each with their own benefits and drawbacks. A few examples of such algorithms are: decision trees, maximum margin classifiers, and nearest

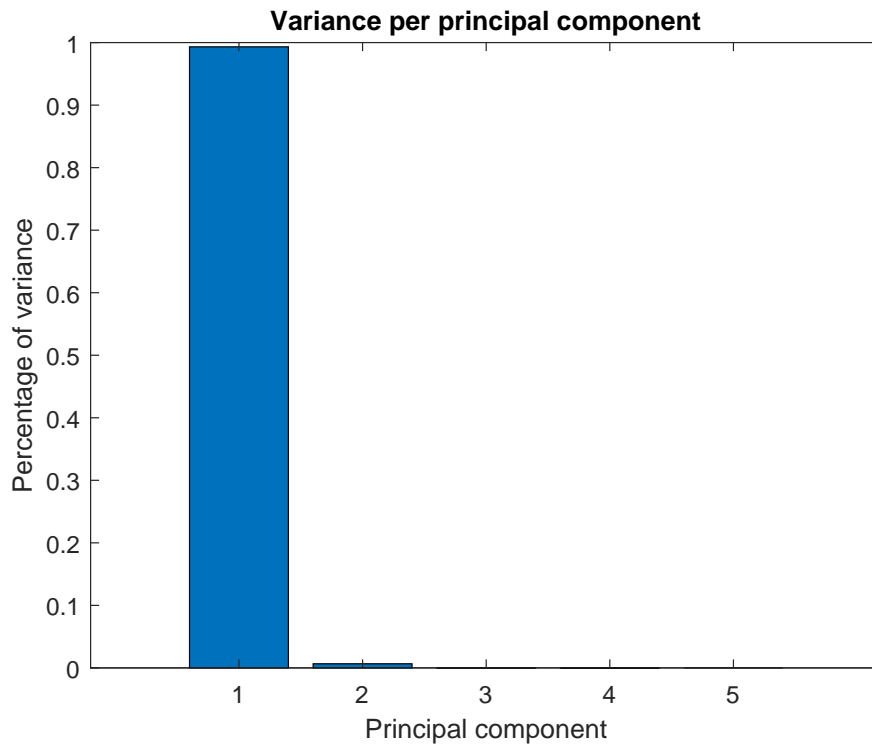


Figure 5.1: Plot showing percentage of variance per principal component.

neighbour classifiers, although this list is by no means exhaustive [54].

Decision trees are a relatively straightforward concept, beginning at the tree root with a binary decision regarding the data one wishes to classify. This decision then determines the branch to follow to the next node/decision and so on until the resulting leaf node or category is reached. While this is a very logical and clear-to-follow process, it does have its pitfalls. The benefits of such a method are the speed of both training and prediction, while consuming a small amount of memory to complete the operation in its most basic form. Decision trees are also trivial to interpret on the most part, with each decision and branch easy to follow visually by an observing party. The downside, however, is that they can suffer from low predictive accuracy [77]. The more leaf nodes (categories) present, the greater the risk of over-fitting to training data, resulting in simpler, more robust trees being chosen on the most part with reduced accuracy.

Nearest neighbour (NN or k NN) algorithms are often considered the benchmark for machine learning classifiers and are well defined in the literature such as [78]. They are non-parametric learning algorithms, indicating that no assumptions are made about the underlying data distribution while the model itself does not perform

any generalisation using the training set. The brief learning phase is, therefore, very fast, as computation is performed at classification. The downside of this is that the classification speed is comparatively slow, and memory usage high due to the fact that the entirety of the training set must be maintained for locating the nearest neighbours to a new member. In classification, the k NN algorithm returns a predicted class membership based upon a majority vote of the classes of its k nearest neighbours. One can select a value of k based upon the specific problem, along with specifying a distance measure and weighting function to suit.

Maximum-margin classifiers are a group of classifiers whereby a margin is defined as the perpendicular distance between a decision boundary and the closest of the data points present in that group. Maximising this margin between groups provides an optimal choice of hyperplane, and is determined by a small subset of the data points known as support vectors [54]. This leads directly to the concept of the support vector machine (SVM) which will be explored in detail thanks to its potential classification accuracy and ease of training with limited training sets.

The SVM has been used previously in machining, for the classification of wear into one of a number of possible states. Sun *et al* [79] use an SVM to classify data from a cutting operation into one of three states, corresponding to a sharp, usable, or worn tool respectively. This is relatively limited in its application, as the work only explores flank wear, and provides no information on wear level beyond the most basic states mentioned previously. Cho *et al* [57] use a similar method, however, extend the scope to include further classes for low, medium, and severe wear, as well as identifying chipping and breakage. This method is a much more industrially relevant method, and results in [57] indicate around a 90% accuracy using data from force, vibration, AE, and spindle power sensors combined. While this thesis intends to use features from single sensor data, the results from [57] provide evidence that the SVM is an appropriate tool for tool wear classification.

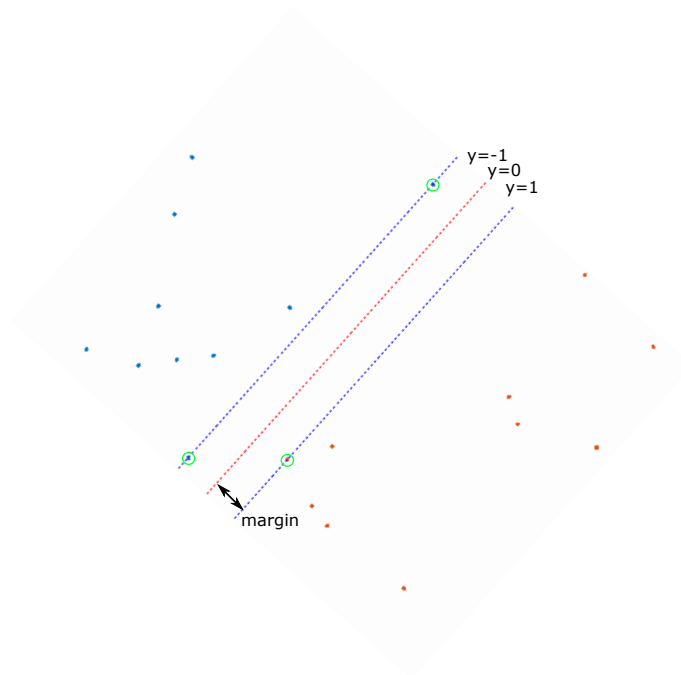


Figure 5.2: Illustration to show the margin, hyperplane, and support vectors. Hyperplane shown in red, support vectors as circled points.

5.2 Support vector machines

5.2.1 SVM overview

As briefly mentioned previously, an SVM makes use of a distance measure and support vectors to identify an optimal hyperplane between data clusters. In its simplest case, the SVM uses linear boundaries which can be illustrated as shown in Figure 5.2.

While this is the simplest form of a SVM, in reality, classification problems often have more than two classes and are inseparable by a simple straight line in the input space. In many cases, class data in the input space can also overlap between classes causing poor generalisation if training is implemented to achieve exact separation of the provided data.

When dealing with datasets of more than two categories, the training process simply becomes a combination of single, binary classifiers, isolating specific parts of the training set to establish the first decision boundary and regrouping to find the rest. The two most common methodologies of implementation are the one-against-all and

one-against-one (or pairwise) approaches. The former breaks the problem down by isolating a single category and finding the line that best separates this from the rest, while the latter uses pairs of categories to establish a number of unique decision boundaries. The one-against-all approach typically selects a class for a new member based on a single class accepting it while all others reject it, whereas the one-against-one method compares classifications of each pair and selects the class with the maximum pairwise votes. It can be expected, therefore, that a pairwise system is likely to require greater processing power and hence time to classify a new member, although this does generally result in improved classification rates.

In the case where the data are not linearly separable in the input space, the data can be mapped to a higher-dimensional space where linear separation is possible, allowing the classifier to perform the same action as before. Consider the fixed feature-space transformation:

$$x \rightarrow \phi(x), \quad \text{where } \phi : R^d \rightarrow R^D \quad (5.1)$$

where the data is now linearly separable in the mapped feature space.

5.2.2 SVM mathematical formulation

This discussion closely follows that by Bishop [54]. Starting with the two-class classification problem, using linear models of the form,

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (5.2)$$

where $\phi(\mathbf{x})$ denotes the transform described in (5.1) and the bias parameter, b , is made explicit. Note that this formulation can eventually progress to a dual representation expressed in terms of kernel functions, avoiding the need to work explicitly in the feature space. The training data set comprises N input vectors $\mathbf{x}_1, \dots, \mathbf{x}_N$, with corresponding target values t_1, \dots, t_N where $t_n \in \{-1, 1\}$, and new data points \mathbf{x} are classified according to the sign of $y(\mathbf{x})$.

It can be assumed at this stage that the members of the training set are linearly separable in feature space, so that by definition, there exists at least a single choice

of parameters \mathbf{w} and b such that a function of the form (5.2) satisfies $y(\mathbf{x}_n) > 0$ for points where $t_n = +1$ and $y(\mathbf{x}_n) < 0$ for points where $t_n = -1$, so that $t_n y(\mathbf{x}_n) > 0$ for all training data points.

It is entirely possible that there are a number of solutions that accurately separate the classes exactly; however, it is beneficial to select the solution with the smallest generalisation error. This is briefly touched on in Section 5.1 where, in the case of an SVM, the concept of maximising the margin is used to identify the optimal decision boundary. For an insight into the origins of maximum margin theory, the reader's attention is drawn to [80].

Figure 5.2 shows that the perpendicular distance of a point \mathbf{x} from a hyperplane defined by $y(\mathbf{x}) = 0$, where $y(\mathbf{x})$ takes the form (5.2), is given by $|y(\mathbf{x})|/\|\mathbf{w}\|$ (where $|y(\mathbf{x})|$ is the absolute value of $y(\mathbf{x})$, and $\|\mathbf{w}\|$ is the norm of \mathbf{w}). In addition, only those solutions for which all data points are correctly classified so that $t_n y(\mathbf{x}_n) > 0$ for all n are of interest. The resulting distance of a point \mathbf{x}_n to the decision surface is given by

$$\frac{t_n y(\mathbf{x}_n)}{\|\mathbf{w}\|} = \frac{t_n (\mathbf{w}^T \phi(\mathbf{x}_n) + b)}{\|\mathbf{w}\|}. \quad (5.3)$$

The margin is given by the perpendicular distance to the closest data point \mathbf{x}_n from the set, and it is necessary to optimise the parameters \mathbf{w} and b in order to maximise this distance. It follows, therefore, that the maximum margin solution is found by solving

$$\arg \max_{\mathbf{w}, b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_n [t_n (\mathbf{w}^T \phi(\mathbf{x}_n) + b)] \right\} \quad (5.4)$$

where the factor $1/\|\mathbf{w}\|$ is taken outside the optimisation over n because \mathbf{w} does not depend on n . Solving this optimisation problem directly would be very complex [54], and so it is required that it is converted into a simpler-to-solve equivalent problem. To do this relies on the fact that performing the rescaling $\mathbf{w} \rightarrow k\mathbf{w}$ and $b \rightarrow kb$ leaves the distance from any point \mathbf{x}_n to the decision surface unchanged. That is, $t_n y(\mathbf{x}_n)/\|\mathbf{w}\|$ remains constant. This freedom can be used to set

$$t_n (\mathbf{w}^T \phi(\mathbf{x}_n) + b) = 1 \quad (5.5)$$

for the point that is closest to the surface. In this case, all data points will satisfy the constraints

$$t_n(\mathbf{w}^T \phi(\mathbf{x}_n) + b) \geq 1, \quad n = 1, \dots, N. \quad (5.6)$$

This is known as the canonical representation of the decision hyperplane. In the case where the equality holds true, the constraints are said to be active; it follows that for other points, they are said to be inactive. By definition, there will always be a closest point resulting in a single active constraint and at least two active constraints once the margin has been maximised. The problem then simplifies to maximising $\|\mathbf{w}\|^{-1}$ or similarly, minimising $\|\mathbf{w}\|^2$, leaving the optimisation problem as

$$\underset{\mathbf{w}, b}{\operatorname{arg\,min}} \|\mathbf{w}\|^2 \quad (5.7)$$

subject to the constraints given by (5.6). As explained by Bishop [54], this is an example of a quadratic programming problem where it is required to minimise a quadratic function subject to a set of linear inequality constraints. While it may appear initially that the bias parameter b has been removed from the optimisation, it is in fact determined implicitly via the imposed constraints as changes to $\|\mathbf{w}\|$ must be compensated by changes to b . The classification of new data points once the model is trained is done by evaluating the sign of $y(\mathbf{x})$ defined in (5.2).

This provides an overview of the SVM, whereby the training data points are linearly separable in the feature space $\phi(\mathbf{x})$; the resulting SVM providing an exact separation of the training data in the input space \mathbf{x} regardless of the decision boundary not necessarily being linear. In reality, this is not always the case, and there often exists an overlap between class distributions such that exact separation of the training data can lead to poor generalisation. In order to compensate for this and allow a selection of training points to be misclassified, the SVM is modified to include a slack variable for each data point which acts as a measure of distance away from the boundary. The problem then becomes a minimisation problem including minimising the sum of slack variables as a trade-off against the margin. The readers attention is drawn to Chapter 7 of [54] if they wish to follow the inclusion of slack variables, as this is an established technique.

5.3 Training and testing sets

When training a support vector machine, specifically with a relatively small dataset, k -fold cross validation provides a method of evaluating model accuracy while ensuring all data points are used for both training and validation.

To begin with, the original data set is split into k equal size sub-samples, in this case containing equal distributions of class labels. Of these k sub-sets, a single one is retained and used as validation data against a model trained on the other $k - 1$ samples. Repeating this process k times, using each sub-set for validation only once, provides k accuracy evaluations which can be averaged to give a single measure. The major benefit of this method is that it provides a robust evaluation given a relatively modest data set, ensuring that all data points are used for both training and validation; each being used a single time for the latter.

In this work, a 5-fold cross validation has been used as a trade off between evaluation accuracy and speed, while ensuring each sub-sample contains enough individual points from each class to train a SVM which can generalise well with larger datasets.

5.4 Results

For its relative simplicity and fast prediction speed, a simple linear SVM is the first candidate for training and validation of each trials dataset, with the aim of classifying and separating each individual tool's data from that generated by the other eight. The input vectors represent the first three principal components of the time-domain features generated from in-cut AE data (representing 95% of the variance in the dataset), each containing 441 points generated from 9 tools.

Considering the first trial dataset independently, the trained model returns an accuracy of 99.1% when using a one-against-one multiclass method and five-fold cross validation. To achieve this, the training time was 1.78 seconds in MATLAB with the classifier being capable of assessing approximately 4400 observations per second on the chosen hardware; considering a new observation occurs every 1.1 seconds, this is considered more than sufficient in terms of prediction speed. In comparison, using a one-against-all method on the same hardware, results in an accuracy of 91.8% taking a mere 1.24 seconds to train. This method is also capable of predicting at

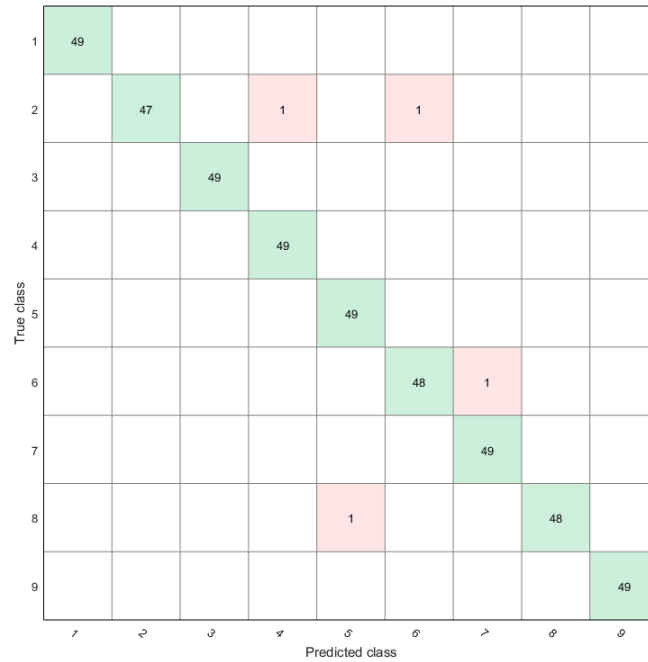


Figure 5.3: Confusion matrix of linear SVM classifications for trial one.

twice the rate of the one-against-one method; however, in this case the accuracy is more important than the prediction speed due to the relatively modest observation rate. For this reason, in all further models the one-against-one method will be used exclusively.

The confusion matrix for this SVM can be seen in Figure 5.3 and provides a clear indication of classification errors in much more detail than the prediction accuracy alone. As expected, correct classifications are shown in green along the diagonal of the matrix with misclassifications shown in red. As the model prediction accuracy is so high, there are only four incorrect points, which are most likely caused by built-up material on the tool's cutting edge. Built-up material has been observed in a select few 3D tool scans (see Figure 5.10), however, can't be measured for each data point, given that it is a temporary phenomenon during machining in this process. Incorrectly classified points beside the diagonal such as (7,6) are the most commonly expected misclassification type, indicating that there is likely to be some overlap between states either side of one another. This type of error can be expected when the generating process for two classes is very similar and therefore difficult to distinguish when compared to other classes.

In a similar manner, considering the second trial dataset independently, results in

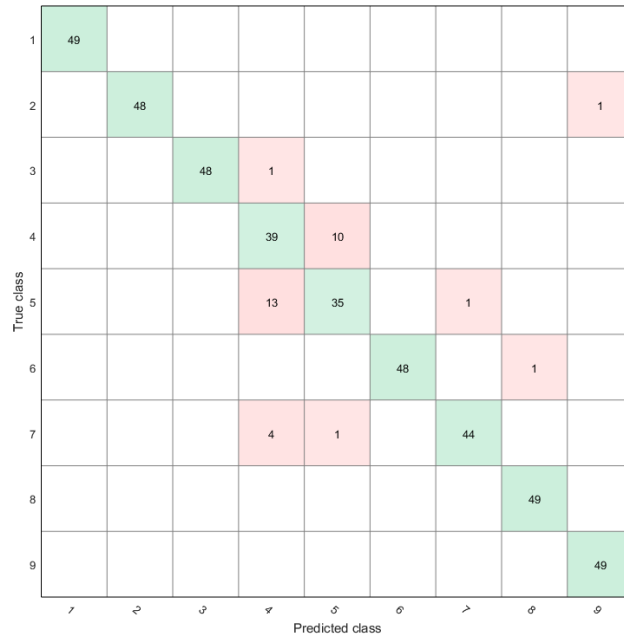


Figure 5.4: Confusion matrix of linear SVM classifications for trial two.

a prediction accuracy of 92.7% through training and validation taking 3.26 seconds on the same hardware. This is significantly lower than the 99.1% obtained using the trial one data and, therefore, requires further attention. The confusion matrix for this model can be seen in Figure 5.4 and follows the same specification as before. Interestingly, the majority of the errors associated with this model arises from the distinction between classes four and five and is shown by the misclassified observations in (4,5) and (5,4). As mentioned previously, this would indicate that classes four and five in this case are generated from a similar process state and therefore have resultant data that is difficult to distinguish from one another. As the classes in these models represent unique tools, the possibility arises of using the tool wear levels as class labels allowing observation of those classes which are expected to be in close proximity.

Using the cutting-edge radius (in microns), correct to the nearest whole number as class labels, has very little effect on the model results from the first trial, with accuracy remaining at 99.1% and the confusion matrix as shown in Figure 5.5. Conversely, the results from the trial two dataset tell a different story. Comparing the new class labels reveals that both classes four and five, while generated with unique tools, are a result of the same cutting edge radius, and as discussed in Chapter 4, the same level of wear. The similar cutting edge radii would explain the

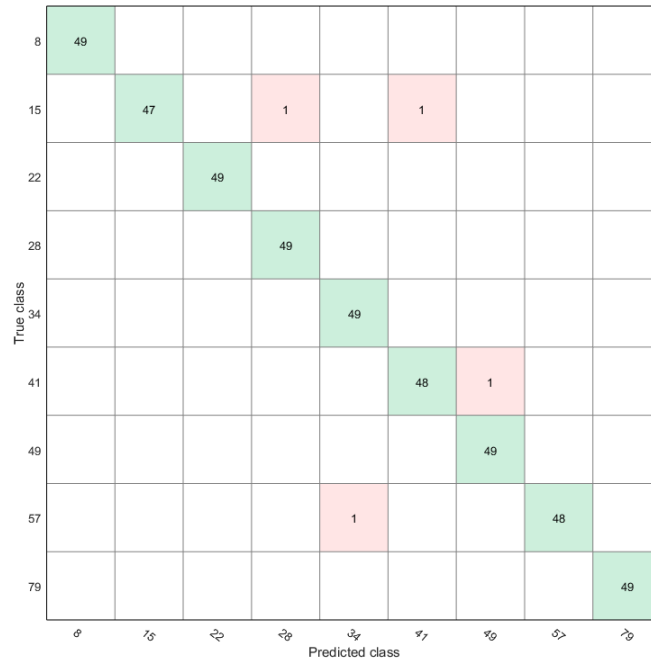


Figure 5.5: Confusion matrix of linear SVM classifications for trial one using wear measurements as class labels.

difficulty the linear SVM found in separating these classes previously, as in reality, they form a single class of twice as many observations as the others. Retraining and validating the SVM using the now eight classes, results in a much improved accuracy of 98.2% and the confusion matrix shown in Figure 5.6. As this figure presents, the class containing observations created with a 77 micron cutting edge radius contains 98 unique observations and improves accuracy by 5.5%. The concept of using wear measurements as class labels also begins to demonstrate how, despite having a single limited measurement per 49 observations, the problem can begin to be formulated into a regression analysis.

Figures 5.7a and 5.7b show, for instance, the wear curves of the two independent trials by focussing on the chosen wear measurement from Chapter 4 of cutting-edge radius. Comparing these figures to the sequential classification of data points using the previous models, shows a clear connection between the two. One point to note here is that the curves are dissimilar in shape despite the same method being used to obtain them, a clear illustration of the necessity of this project over merely following typical cubic wear curves. The fundamental reasoning behind the variation in wear rate observed here is the diversity in material properties through the cylindrical billet used to experimentally wear the tools. In industry, the same experience is present

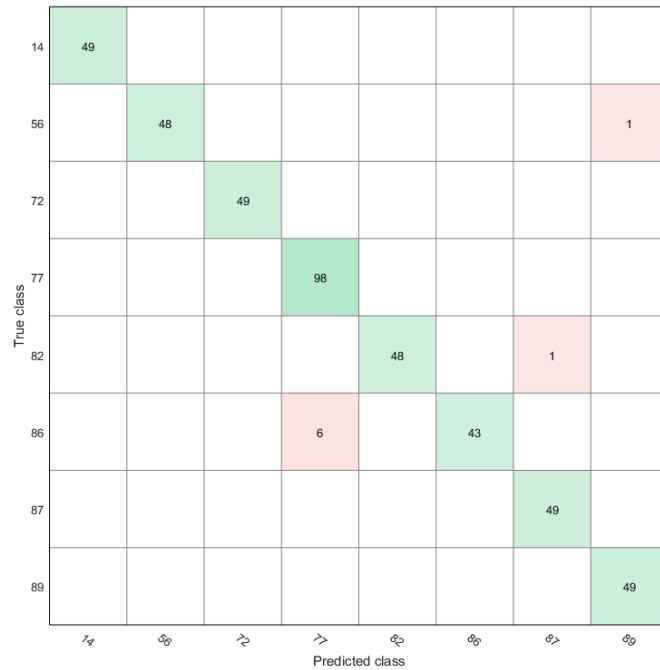
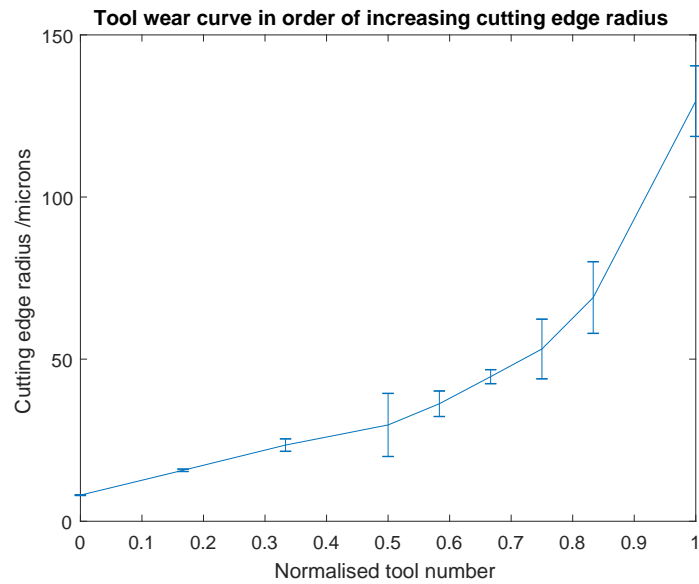


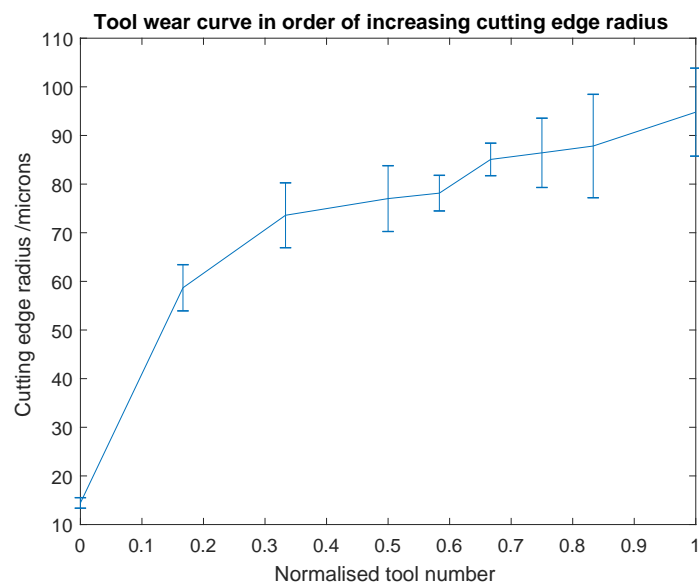
Figure 5.6: Confusion matrix of linear SVM classifications for trial two using wear measurements as class labels.

due to the large forgings and relative inaccuracy of the forging and heat treating processes. The effects of this material variation to in-service performance are often insignificant, providing the bulk of the material meets specification following shot-peening - a cold working process used to strengthen and relieve tensile stress in components, modifying mechanical properties through plastic deformation at the material's surface. Compressive stresses are beneficial in the surface of machined parts, as cracks are held closed and struggle to propagate, increasing fatigue life dramatically. Shot peening, however, occurs after finish machining, and so the lack of uniform properties is still present to affect tool wear rates.

Figures 5.8a and 5.8b highlight this relationship by presenting the sequentially classified data points in blue alongside the mean wear measurement for each tool as an orange line, misclassifications are circled in red. As expected, by using discrete classes to track wear progression, the classifier acts as a sample-and-hold filter when compared to the underlying nine measurements, switching class and remaining there until data from the next class is experienced. Provided class transitions follow a common sequence, this reality indicates that the problem could be well suited to a sequential model form, which will be explored in Chapter 6. As can be seen in Figure 5.8, the misclassified points are very few and should it be necessary, could be

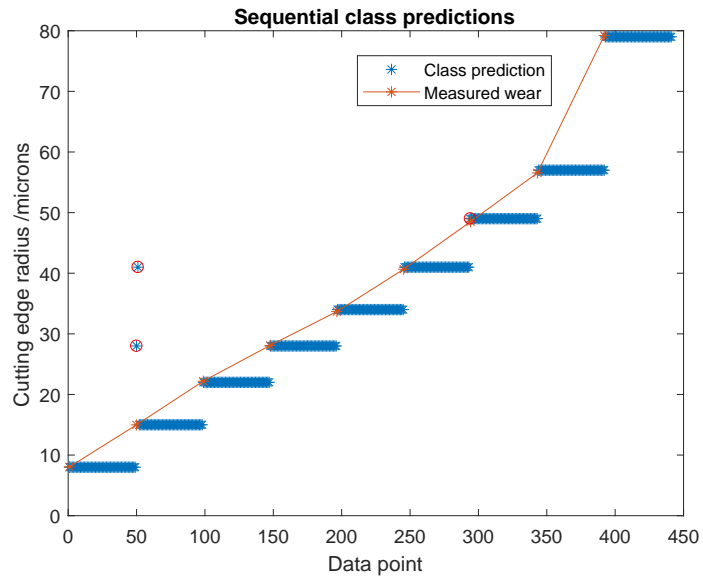


(a) Wear curve from trial one.

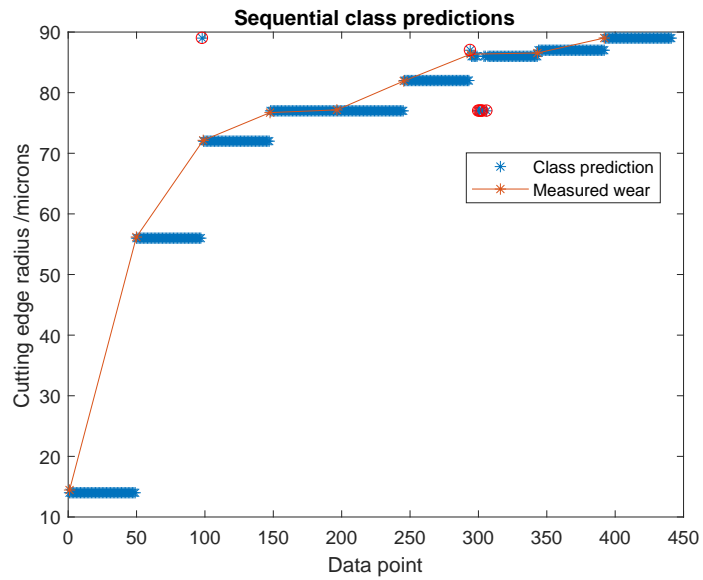


(b) Wear curve from trial two.

Figure 5.7: Cutting edge radius wear curves from each trial. Error bars indicate standard error across 5 tool scans.



(a) Sequentially classified states from trial one.



(b) Sequentially classified states from trial two.

Figure 5.8: Sequentially classified states plotted in order. Misclassifications are circled in red.

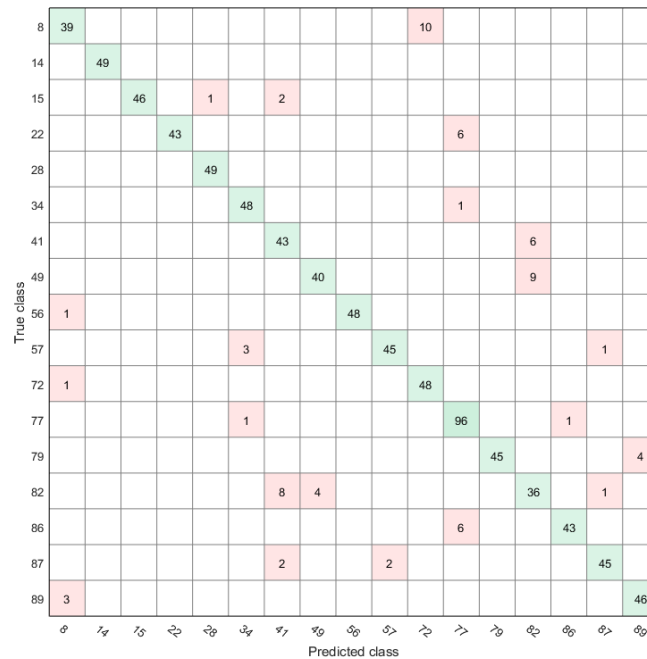


Figure 5.9: Confusion matrix of linear SVM classifications for all trial data combined.

removed by treating the problem as sequential in nature, and removing predictions that do not move to the next class in increasing order. This is feasible in an industrial context provided the process is continually monitored, ensuring that each tool progresses through every wear state in order, albeit at potentially different rates.

As a more rigorous test of the SVM method, a single model can be trained and validated with all available data from both trials together as opposed to each individually. Given the greater number of potential classes, it is expected that prediction accuracy will drop due to more potential overlap between them, however, this should manifest as misclassifications along the diagonal of the confusion matrix for acceptable errors between neighbouring states. Including data from two different tests is also likely to have a negative effect on the prediction accuracy again, possibly due to class overlap, but also different environmental conditions on each test day. This is a typical problem experienced in industry and, therefore, is worth consideration at this stage.

As can be seen in Figure 5.9, the prediction accuracy of this model drops down to 91.7% as expected; however, there is no confusion between adjacent classes as anticipated. This indicates possible confusion between datasets causing the increase in error, and is most likely due to built-up edge forming occasionally on the tool's

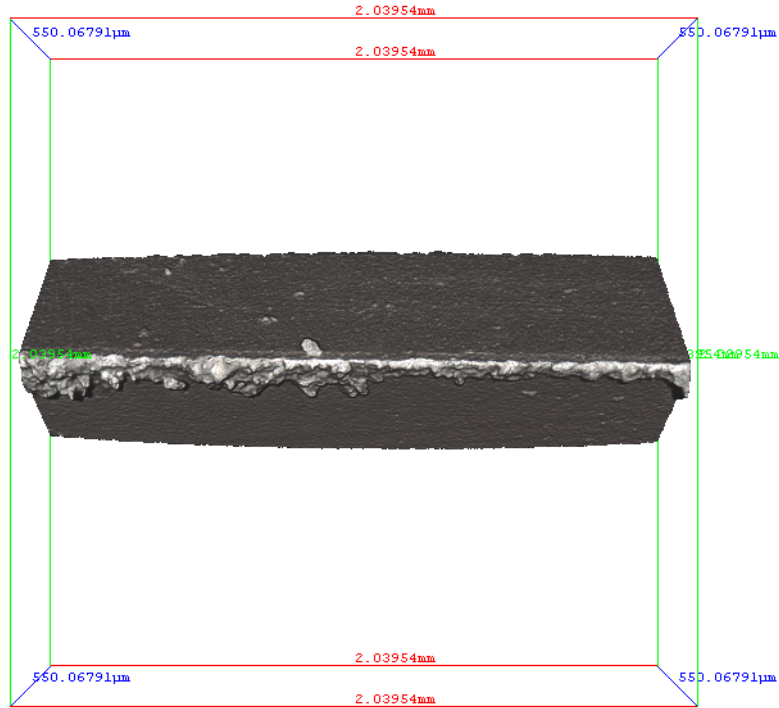


Figure 5.10: 3D scan of a tool showing built-up edge.

flute(s), before detaching again. Figure 5.10 shows an example of a tool displaying material built-up on the cutting edge. Considering a new tool, this addition of material can result in an apparent blunting of the cutting edge, while a worn tool may have material filling craters that have formed. It follows that, while briefly attached, such material could alter AE recordings and cause occasional erroneous classifications. This phenomenon is likely exaggerated by the increased number of classes spanning the same range, as the margin between classes across the tool life is decreased, and classes become closer. Environmental effects could also account for some of the overlap between datasets, and one possible improvement to this would be to take a reference observation each time a trial is started and normalise across datasets. The key then is the progression of wear from a known reference, such as which states a tool moves through and in what manner. Considering the order of this data proves critical to understanding wear progression, it follows that treating the data as a sequential problem would be beneficial. Given that a prior observation provides greater insight into the likelihood of the current points state, development of a probabilistic model which selects a class based on the previous point appears a logical step in minimising outliers and constraining classifications to a limited number of possibilities at each step.

5.5 Overview of tool state classification

This chapter has discussed a selection of classification algorithms and presented the benefits and drawbacks of a number of these before proceeding to explore the support vector machine in more detail. After presenting the mathematical formulation of the SVM, a linear example is trained for each dataset with prediction accuracies of 99.1% and 98.2% for trials one and two respectively. This result is advantageous, as it indicates that, given in-process recorded AE data, tool state can be accurately diagnosed and assigned to a specific class. In an industrial context, this represents an effective method of indicating tool wear level to a machine operator, without the need for interrupting the process. Currently, to ensure part safety, regular tool observations are taken during machining, and can result in an 8% increase in machine time if the process is paused for even 5 minutes per hour. Implementing an SVM as discussed in this chapter, removes the need for regular stoppages for tool wear observations as data is acquired during machining, and therefore offers direct time and cost savings to the production process. The limiting factor of this method is that it does not provide the operator with any insight into the wear rate of the current tool, as it simply indicates a group in which the tool is believed to belong. This method, however, is still a benefit when compared to the methods currently applied in industry.

It is found that if one plots the state classifications in sequential order of observation and uses the class labels as wear measurements, the classified data points indicate position on the wear curve. Following the progression between classes can provide information relating to wear rate, indicating that a sequential model may be an appropriate next-step. This is in agreement with the results from a single model trained on all available data which shows a decrease in accuracy and greater confusion between non-adjointing classes, highlighting the need to consider a given points' existing class when predicting its next most likely position.

TOOL STATE TRACKING

As discussed at the end of Chapter 5, wear classification has been proved possible on a per-trial basis, using AE observations and a support vector machine to predict a tool's cutting-edge radius class. A problem arises when combining data from multiple trials, as environmental factors and material properties result in varying wear curves. One solution to this problem is to treat the order in which a sequence of observations arrive, as a fundamental insight into the most likely next observation, rather than each point being independent. This chapter, therefore, focuses on sequential data in the form of Markov models, following a probabilistic approach to predicting the most likely next state of a tool given its current level of wear.

6.1 Markov models

The most straightforward way to treat data follows that in Chapter 5, where the sequential aspects of the data are ignored and observations are treated as independent and identically distributed (i.i.d.). Figure 6.1 presents a graphical model of independent observations $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \dots\}$, shown as independent due to the lack of connecting lines between them. The issue with such an approach is that it fails to take advantage of the sequential nature and patterns of the generating process. Take the tool wear problem for instance, where the aim is to accurately predict the level of wear present on a tool's cutting edge; the assumption can be made that a tool must become more worn during use as negative wear is irrational, and also that, in

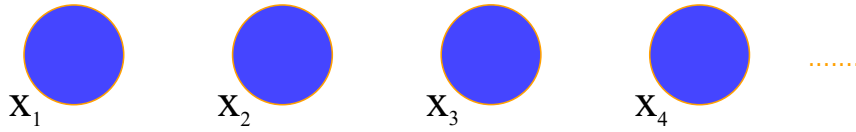


Figure 6.1: Observations as independent, corresponding to a graph with no links.

general, wear is a steady and progressive process rather than a set of rapid, instantaneous changes. It is worth noting that in a damage scenario, rapid wear progression is entirely possible; however, this situation will be touched on more later. As discussed previously in Chapter 5, built-up edge can cause an indication that a tool is wearing negatively, however, its effects can be ignored by selection of an appropriate model topology, given that the usage duration and underlying wear is still known to increase with machine time. If the data is treated as i.i.d., then the relative frequency of classifications into each state is all that is gained. Taking the preceding assumptions into account though, when based on a tool's current classification, wear is only likely to remain in its current class or move to a higher one, as moving backwards through classes is not possible without resharpening a tool. Knowledge of a tool's current class is therefore a significant piece of information when predicting which class the next observation is likely to belong to. This method has been successfully implemented previously to track the progress of tool wear in drilling [62]; however, this work only uses three states to indicate sharp, workable, and dull tools respectively, and is based around measured cutting-force data. The main issue with the work in [62], is that wear level per tool is not quantified, and is instead, treated as a direct function of usage duration and inferred from the measured signals. Similarly, previous work has been conducted with a focus on sequential tool wear during milling, yet the success of this work is severely limited due to the sparse training data and resultant binary classifier [81]. Probably the most promising work in this area can be found in [61], where a 95% successful classification rate is obtained in predicting sharp, worn, and broken tools during milling.

Following the procedure in [54], the i.i.d. assumption must be eased in order to express this concept in a probabilistic model. One of the simplest ways to do this is to consider the Markov model, in which the probability of being in a state at any time, relies only on the state at the previous time interval. For machining, this translates as the probability of the current wear state, relies only on that of the wear state previously obtained - a logical assumption given that wear is cumulative in nature. To begin with, the product rule is used to express the joint distribution of a sequence of observations as,

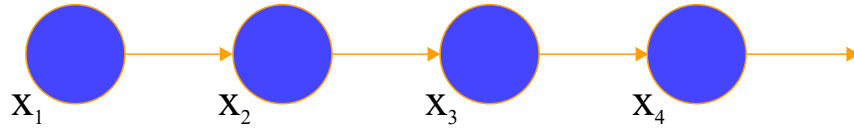


Figure 6.2: First order Markov chain of observations $\{\mathbf{x}_n\}$ where the distribution $p(\mathbf{x}_n|\mathbf{x}_{n-1})$ of a single observation \mathbf{x}_n is conditioned only on the previous observation \mathbf{x}_{n-1} .

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N) = \prod_{n=1}^N p(\mathbf{x}_n|\mathbf{x}_1, \dots, \mathbf{x}_{n-1}). \quad (6.1)$$

Assuming then that each of the conditional distributions on the right-hand side is only reliant upon the most recent previous observation and independent of all others, the first-order Markov chain is obtained as shown in Figure 6.2. This then simplifies Equation 6.1 to,

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N) = p(\mathbf{x}_1) \prod_{n=2}^N p(\mathbf{x}_n|\mathbf{x}_{n-1}). \quad (6.2)$$

Through evaluation of (6.2) and application of the product rule, it follows that the conditional distribution for a single observation \mathbf{x}_n , given all previous observations up to time n , is given by

$$p(\mathbf{x}_n|\mathbf{x}_1, \dots, \mathbf{x}_{n-1}) = p(\mathbf{x}_n|\mathbf{x}_{n-1}) \quad (6.3)$$

and therefore, the distribution of predictions made about the next observation will be solely dependent on the immediately preceding observation and independent of all other earlier observations. This leads directly to the application to tool wear where, due to its cumulative nature, it is anticipated that each measurement will encode the previous information before it and, consequently, the conditional distribution for each new observation only relies on that before it. Generally, the conditional distributions $p(\mathbf{x}_n|\mathbf{x}_{n-1})$ are always equal and bear no dependence on n , following the assumption of a stationary time series. In other words, at each new state prediction the conditional probabilities are the same as for previous predictions and any adjustable parameters are often fixed through some learning process. The Markov chain can be extended to include the effects of more than one previous observa-

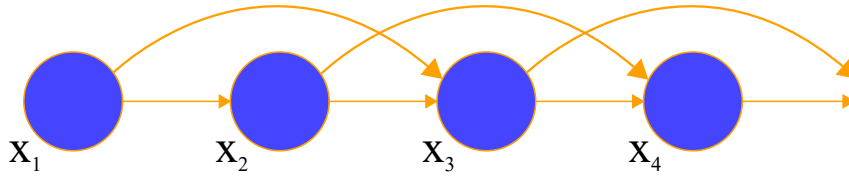


Figure 6.3: Second order Markov chain of observations where the distribution of a single observation \mathbf{x}_n is conditioned on the previous two observations \mathbf{x}_{n-1} and \mathbf{x}_{n-2} .

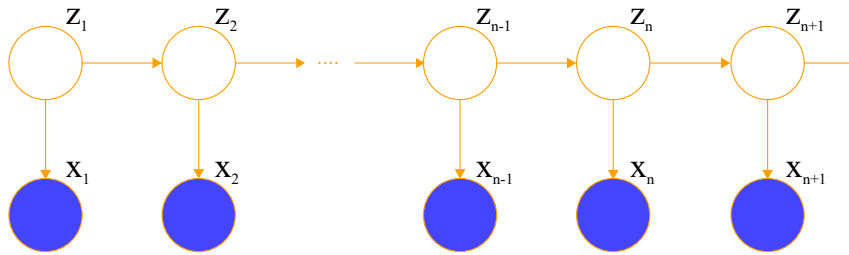


Figure 6.4: Markov chain of latent variables known as a state space model.

tion, creating an M^{th} order chain in which the conditional distribution for a chosen observation relies on the previous M observations. An example of this would be the second -order Markov chain in Figure 6.3, where the predictions rely on both the previous and previous-but-one variables. The downside of this approach is that the number of parameters grows exponentially with the model order M and so becomes impractical very quickly. Assuming that the observations fall into K discrete classes as previously, then each conditional distribution relies upon $K^{M-1}(K - 1)$ parameters.

To build a model for sequential data that can be specified using a limited number of parameters, while not being restricted to the Markov assumption of any order, latent variables are introduced as the simple building blocks of more complex models. Rather than using each observation to form a chain as in Figures 6.2 and 6.3, each observation \mathbf{x}_n is assigned an underlying latent variable \mathbf{z}_n which is used to form the Markov chain. These latent variables can, if necessary, be of different type or dimensionality than the observed variable, yet the observations are conditioned on the state of the corresponding latent variable. In a graphical form, this results in the model shown in Figure 6.4, which satisfies the key conditional independence property that \mathbf{z}_{n-1} and \mathbf{z}_{n+1} are independent given \mathbf{z}_n . Consider the case of an indirect monitoring strategy implemented on a machine tool; the underlying process forms the latent wear variables \mathbf{z}_n whereas the features calculated based on sensor outputs correspond the the observations \mathbf{x}_n . The observed features are a direct result of the underlying process, however, the actual wear state remains unseen.

The joint distribution for this model is an extension of (6.2) and is given by

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{z}_1, \dots, \mathbf{z}_N) = p(\mathbf{z}_1) \left[\prod_{n=2}^N p(\mathbf{z}_n | \mathbf{z}_{n-1}) \right] \prod_{n=1}^N p(\mathbf{x}_n | \mathbf{z}_n). \quad (6.4)$$

This model describes the hidden Markov model (HMM) provided that the latent variables are discrete, as is the case with single point wear measurements.

6.2 Hidden Markov models

As has been done previously, the model shown in Figure 6.4 is taken and the probability distribution of \mathbf{z}_n allowed to rely only on the state of the previous latent variable \mathbf{z}_{n-1} through a conditional distribution $p(\mathbf{z}_n | \mathbf{z}_{n-1})$. To form an HMM, the latent variables must be discrete and it follows, therefore, that the conditional distribution can be defined by a table of numbers known as the transition matrix \mathbf{A} , with each element defining a single transition probability. They are given by $A_{jk} \equiv p(z_{nk} = 1 | z_{n-1,j} = 1)$, and are probabilities and thus satisfy $0 \leq A_{jk} \leq 1$ where $\sum_k A_{jk} = 1$ so that the matrix \mathbf{A} has $K(K - 1)$ independent parameters. This allows the conditional distribution to be rewritten in the form,

$$p(\mathbf{z}_n | \mathbf{z}_{n-1}, \mathbf{A}) = \prod_{k=1}^K \prod_{j=1}^K A_{jk}^{z_{n-1,j} z_{nk}}. \quad (6.5)$$

It is worth noting here that the first latent node \mathbf{z}_1 is unique, as it has no preceding node, and so has a marginal distribution of $p(\mathbf{z}_1)$ given similarly by a vector of probabilities π with elements $\pi_k \equiv p(z_{1k} = 1)$ so that

$$p(\mathbf{z}_1 | \pi) = \prod_{k=1}^K \pi_k^{z_{1k}} \quad (6.6)$$

where $\sum_k \pi_k = 1$.

The transition matrix for a single latent variable is shown graphically in Figure 6.5 where $K = 3$. Squares are used to distinguish between variables (like in Figure 6.3), and the states of a single variable shown here. The arrows in this figure represent

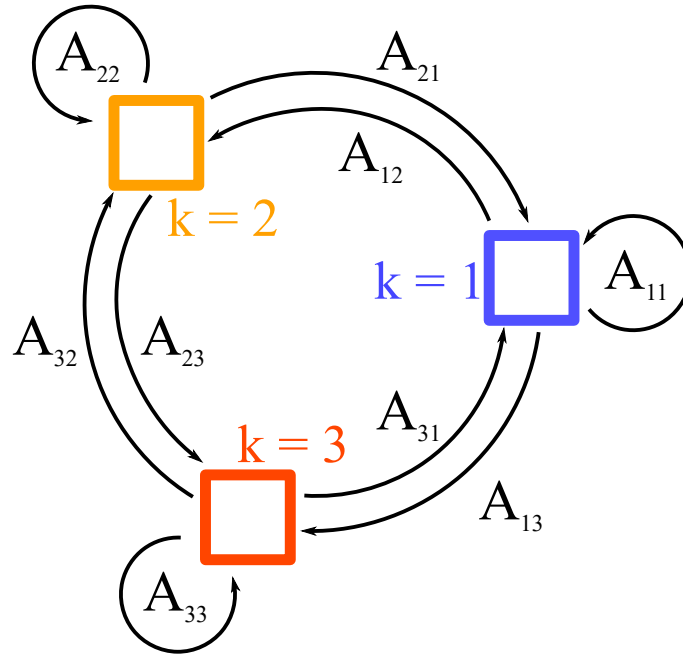


Figure 6.5: Transition diagram showing three possible states of a single variable.

the elements of the transition matrix A_{jk} . Similarly, if the latent states were considered in order, the representation in Figure 6.5 can be extended to form the lattice representation of latent states shown in Figure 6.6, where each column represents one of the latent variables \mathbf{z}_n .

The probabilistic model formulation is completed by defining the conditional distributions of the observed variables $p(\mathbf{x}_n|\mathbf{z}_n, \phi)$, where ϕ is a set of parameters governing the distribution. These parameters are defined as the *emission probabilities* and for discrete values of \mathbf{x} are specified in conditional probability tables much like A . For observations \mathbf{x} , of a continuous nature, these emission probabilities can be represented by a Gaussian distribution for example. As \mathbf{x}_n is observed, the distribution $p(\mathbf{x}_n|\mathbf{z}_n, \phi)$ consists of a vector of K numbers corresponding to the K possible states of the binary vector \mathbf{z}_n for a given ϕ . A simple example of Gaussian emission probabilities is that of an observed binary signal, where the signal observed contains Gaussian noise on the underlying steady signal. Given the underlying signal, the probability distribution of observed points would also be Gaussian.

These emission probabilities can be expressed as

$$p(\mathbf{x}_n|\mathbf{z}_n, \phi) = \prod_{k=1}^K p(\mathbf{x}_n|\phi_k)^{z_{nk}}. \quad (6.7)$$

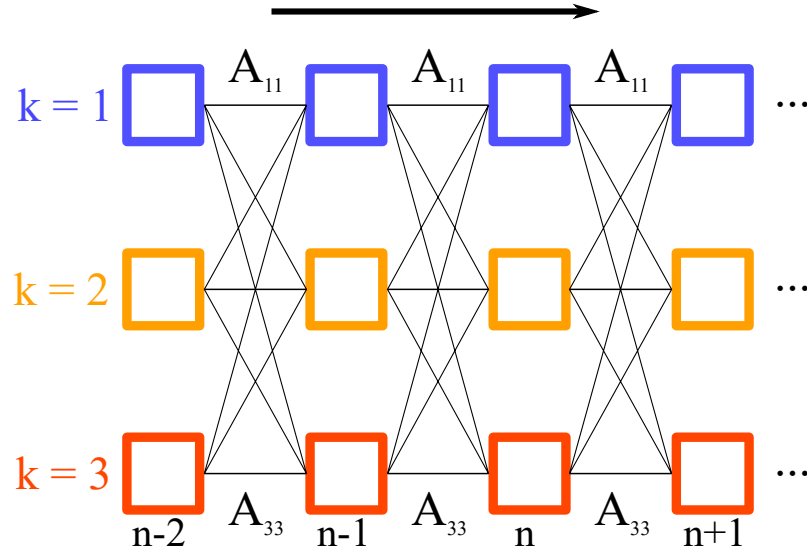


Figure 6.6: Lattice representation of the latent states including possible states and transitions.

This allows the joint probability distribution over both latent and observed variables to be written as

$$p(\mathbf{X}, \mathbf{Z}|\theta) = p(\mathbf{z}_1|\pi) \left[\prod_{n=2}^N p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) \right] \prod_{m=1}^N p(\mathbf{x}_m|\mathbf{z}_m, \phi) \quad (6.8)$$

where $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$, and $\theta = \{\pi, \mathbf{A}, \phi\}$ denotes the set of parameters governing the model.

Hidden Markov models can form a number of different structures by carefully selecting the transition matrix values A_{jk} . A common example of this, relevant to progressive tool wear, is the left-to-right HMM. This model constrains the movement between states to only the forward direction, moving to states of a higher value and restricting movement backwards. Many processes follow this assumption, such as the case of tool wear mentioned earlier where tools can always be assumed to increase in wear as they are used; the idea of a used tool becoming less worn being nonsensical. This structure is easily obtained by setting the elements A_{jk} of \mathbf{A} to zero if $k < j$ and results in the diagram shown in Figure 6.7. The HMM is also capable of a degree of time-independence where stretching or compressing of the time axis naturally results in a change in the number of transitions to the same state before jumping to the next. This allows model predictions of wear to be largely independent of the rate of wear, a result which is advantageous in the case where

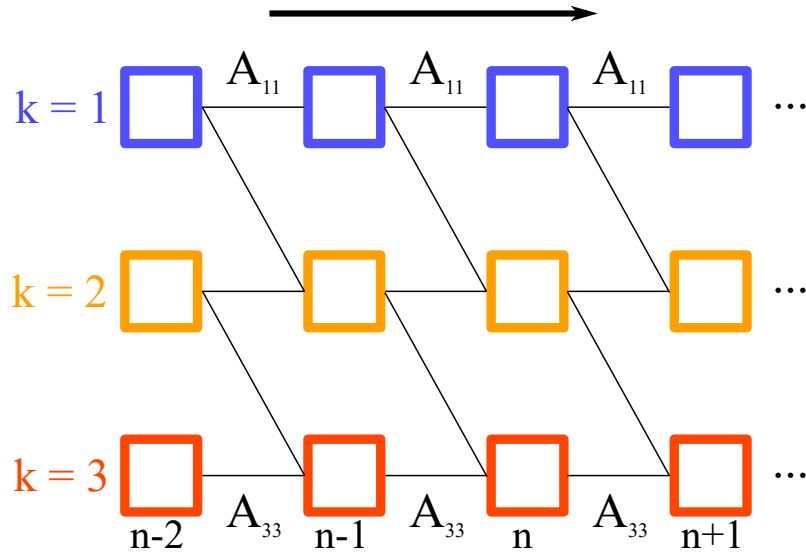


Figure 6.7: Lattice representation of the three-state left-to-right HMM in which the state index k is restricted to an increase of 1 at each transition.

material properties and conditions can effect wear rate across a single part.

6.3 Training and state sequence prediction

Given a previously observed dataset $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, the parameters of the HMM can be obtained using maximum likelihood. The likelihood function is obtained by marginalising over the latent variables from the joint distribution

$$p(\mathbf{X}|\theta) = \sum_{\mathbf{z}} p(\mathbf{X}, \mathbf{Z}|\theta). \quad (6.9)$$

As the joint distribution $p(\mathbf{X}, \mathbf{Z}|\theta)$ doesn't factorise over n , the summations over \mathbf{z}_N cannot be treated independently and with N variables to be summed over, each with K states, the total number of terms is K^N . This is impractical as the number of terms grow exponentially with chain length and so the expectation maximisation (EM) algorithm is used as an efficient solution. For a full explanation and derivation of this method, please continue to follow the procedure in Chapter 13 of [54]. As a summary, however, the EM algorithm is split into its component parts and calculations are performed alternating between the E and M steps until some convergence criterion is met, such as a minimal change in the likelihood function. In this case,

the algorithm runs until the log likelihood that the input sequence is generated by the currently estimated transition and emission matrices, the change in norm of the transition matrix (normalised by the size of the matrix), and the change in norm of the emission matrix (normalised by the size of the matrix) are all below 10^{-6} .

In the majority of cases, the latent variables have some meaningful interpretation and so it is often beneficial to calculate the most probable sequence of states given a sequence of observations. It is worth noting here that finding the most probable sequence of states is not the same problem as finding the most probable set of states individually and, in general, such a set of states will not correlate well to the most probable sequence of states. The solution to the most probable sequence of states is calculated efficiently using the Viterbi algorithm [54] which searches the space of paths through a lattice structure to find the most probable path; its computational cost growing linearly with chain length despite the number of possible paths growing exponentially. Again, this follows the well documented procedure in Chapter 13 of [54] for those wishing to explore in more detail.

6.4 Results

Following a similar methodology to Chapter 5, a unique HMM has first been trained and evaluated for each trial dataset, followed by a single model trained and evaluated on data from a combination of both. Again, as applied in Section 5.3, a five-fold cross validation is used given the modest dataset size, and ensures that the full dataset can be used for both testing and training, while ensuring each data point is never used for both training and testing a single model. In both cases, the model was restricted to a left-to-right architecture as in Figure 6.7, although in this case there are nine possible states corresponding to each unique tool measurement. As the states of the training data are known, the parameter estimation becomes a simple task of calculating the maximum likelihood estimate of the transition and emission probabilities. Once these parameters have been defined, the Viterbi algorithm is used to calculate the most probable path through the model given a set of testing data.

Considering the first trial dataset independently, the trained model returns an overall accuracy of 85.7% when comparing the resultant most probable sequence of states to that actually measured. The training process in this instance takes 0.0005 seconds

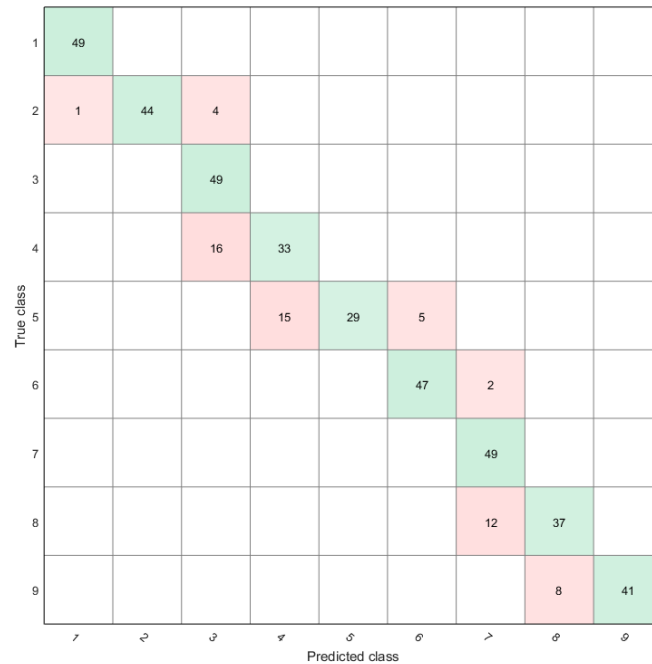


Figure 6.8: Confusion matrix of HMM sequence predictions for trial one.

on the same hardware as used in Chapter 5, making it a significantly faster algorithm than that used for SVM training. The estimated transition matrix is, as expected, a diagonal matrix containing probabilities greater than zero in (k, k) and $(k, k + 1)$, and less than one for all but the last state whose probability of remaining in the same state is unity.

The confusion matrix for this HMM can be seen in Figure 6.8 which provides a straightforward comparison with the models trained in Chapter 5 and shown in Figures 5.3 and 5.4. While it is clear to see that the resulting prediction accuracy has dropped between model types (from 99.1% to 85.7%), it is worth noting that not only does the HMM have the added benefit of being a probabilistic model, but any misclassifications are between neighbouring states as opposed to being spread through the state space. In general, the confusion matrix shows a greater number of misclassifications below the ideal diagonal, followed by lower correct predictions in the following state, indicating that the model is, in the case of states three, four, seven, and eight, remaining in the current state beyond the instance at which a transition should occur.

Now, considering the second trial dataset independently, a corresponding prediction accuracy of 87.5% is obtained using a sequence prediction obtained in 0.0007 seconds.

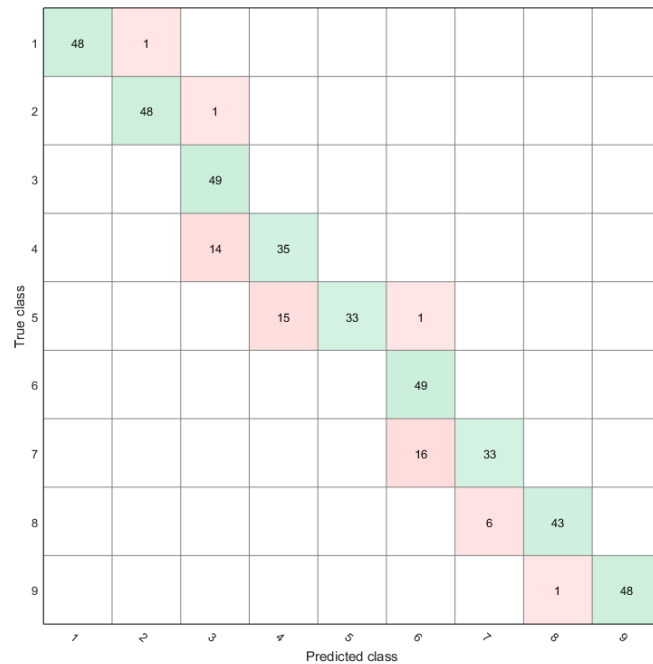


Figure 6.9: Confusion matrix of HMM sequence predictions for trial two.

Again, when comparing this to the 92.7% given from a linear SVM it appears at first as if this method is lacking in benefit; however, assessing the resultant confusion matrix provides more of an insight into the advantage of this method. Comparing that in Figure 6.9 with that from the SVM in Figure 5.4, the main difference is the tightness of predictions to the diagonal in the HMM generated predictions. While the overall error may be worse, in reality, a greater number of predictions to either side of the correct state is preferential over misclassifications in distant states.

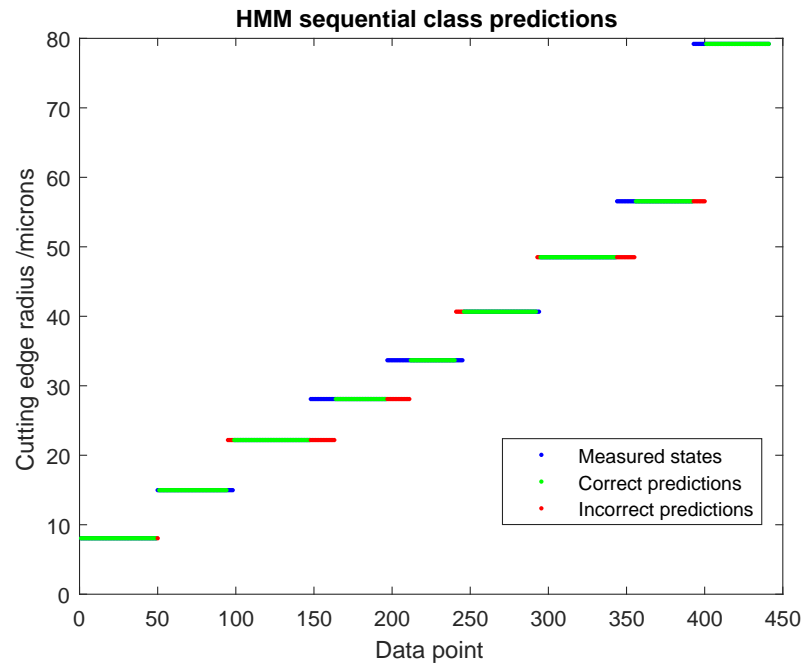
Another way to visualise how the predictions fit with progressive tool wear is to plot the predictions alongside the measured state, using wear values as the state labels in much the same way as in Chapter 5. As this method relies on the sequential nature of the data, the correct predictions are shown in green, whereas incorrect predictions are shown in red with their corresponding correct value in blue to enable comparison with the measured sequence. These plots for both trials are shown in Figures 6.10a and 6.10b for trials one and two respectively. The first thing to note is that the general shape of the curves follow closely those in Figure 5.7; secondly the lack of scattered points away from the central trend show a reduced number of largely misclassified points.

In both cases, as indicated by the confusion matrices, the models tend to continue

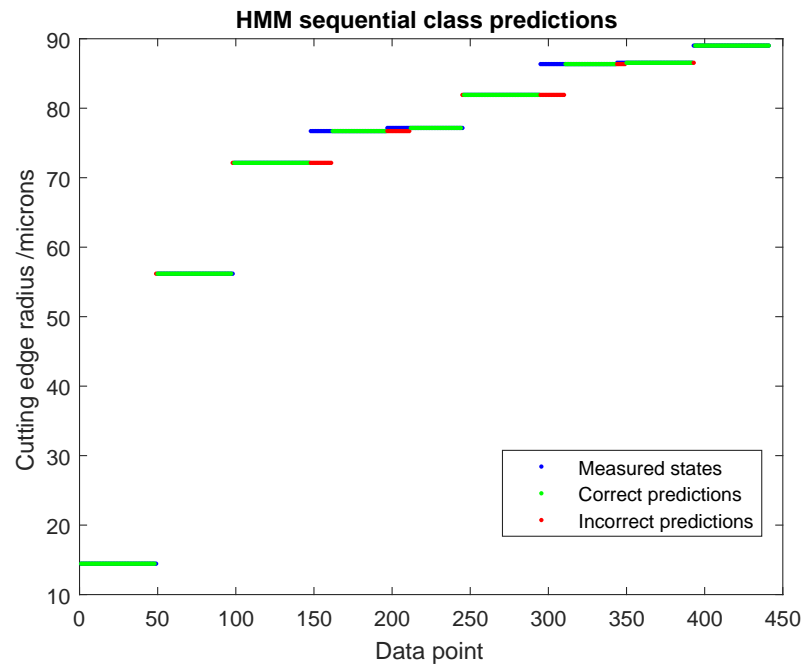
to return current state predictions for longer than necessary before the transition to the next state. This could be corrected by a simple adjustment of the probabilities in the transition matrices; however, it poses little issue, given that the general trends remain the same and the more worn states are reached correctly. This is likely due to the fact that distinguishing excessively worn tools in the input space is easier than those in the middle of their life.

Due to the sequential nature of the input data and HMM algorithm, it doesn't follow to consider the data from both trials together, as in Chapter 5, considering this will fundamentally alter the sequence of observations and cause the resulting model to be inappropriate to any new dataset. An interesting test of this type of model is to train the HMM on the data from one trial and test on the other. While it is expected that the overall prediction accuracy will be poor due to the two different wear curves of these datasets, it is interesting to see the resulting predicted state sequence and how this compares to that which the model is trained with.

Figure 6.11 shows a plot similar to those in Figure 6.10, based on an HMM trained and tested on independent datasets, in this case plotted as linear state labels rather than wear measurements. As anticipated, due to the different wear curves, the resulting model accuracy is only 23.6% - a poor result in any sense. The figure, however, shows a curious phenomenon in that the predictions appear to simply be a lagged realisation of the expected curve. Considering the inherent behaviour of the HMM, this would suggest that progression through states in trial one is slower than that in trial two due to the distance between state predictions in Figure 6.11 increasing. Towards the end of life, this gap begins to close again implying that the rate of progression of trial one has increased or similarly, trial two has decreased with respect to one another. Comparing the wear curves in Figure 6.12 with Figure 6.11 confirms this suspicion. Figure 6.12 shows firstly that the curve for trial one begins in a lower wear state, reaching the same state as trial two at $t + 1$. Paying attention to the curve gradients, it also shows that the wear rate in trial 2 is much higher in the initial stages than trial one, reaching a roughly similar rate during the mid-life, with the wear in the latter accelerating at the end of life. These characteristics correspond directly to Figure 6.11, where the initial state prediction of trial one unsurprisingly remains in the initial state for twice as long when compared to the second dataset used for training. As the wear rates and, therefore levels differ, the gap between predictions grows, shrinking again as the wear levels begin to coincide in the final state. This provides evidence that given a baseline dataset, the HMM can not only



(a) Sequentially classified states from trial one.



(b) Sequentially classified states from trial two.

Figure 6.10: Sequential predictions from HMM.

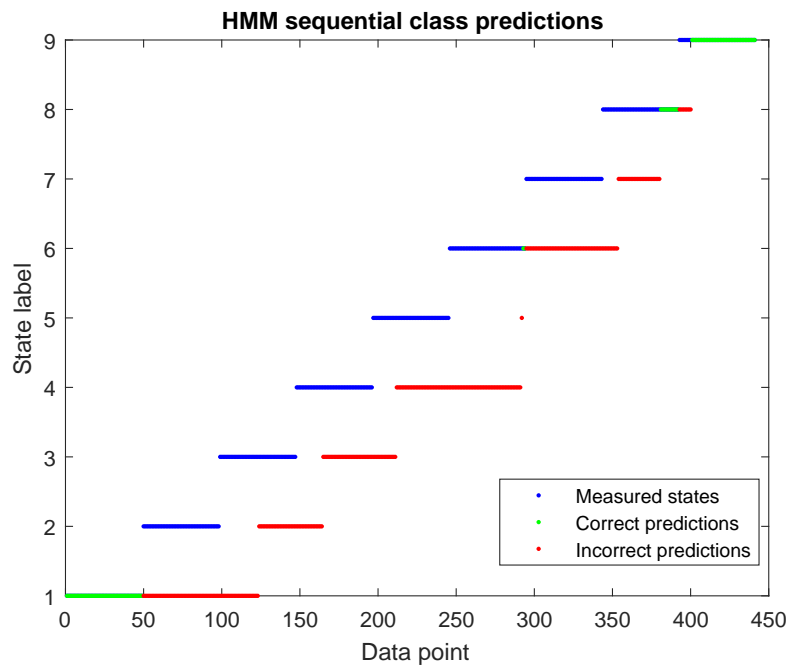


Figure 6.11: Sequential predictions from HMM trained on dataset two and tested on dataset one.

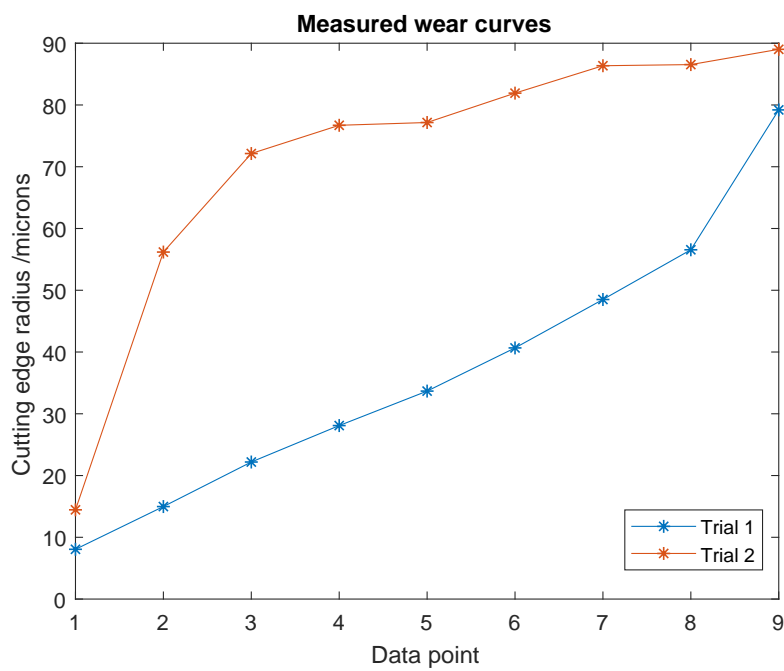


Figure 6.12: Wear curves of both trial datasets.

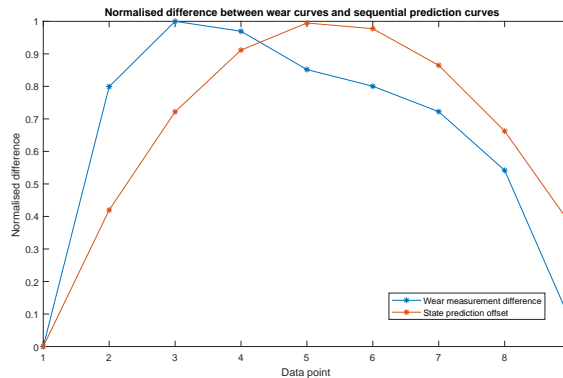


Figure 6.13: Normalised difference between wear curves and sequential prediction curves.

be used to probabilistically classify tool wear, but to also give an indication of wear rate in relation to the training set, should it vary.

A further way of visualising this concept is shown in Figure 6.13, where the difference between wear curves in Figure 6.12 and predicted state curves in Figure 6.11 are normalised and plotted beside each other. A reasonable correlation can be observed between these plots, and further work into measuring correlations would aim to confirm the idea that the time between HMM state predictions and a predefined baseline can be used to estimate wear rate alongside wear level.

6.5 Overview of tool state tracking

This chapter has explored the use of the hidden Markov model as a tool to enable tracking of tool wear state, exploiting the sequential, progressive nature of wear and the HMM. After presenting the mathematical formulation of the HMM and its graphical form, two independent models are trained and validated corresponding to each separate trial. These models result in prediction accuracies of 85.7% and 87.5% respectively, which while indicating poorer performance than the previously discussed SVM, have the benefit that errors are in neighbouring states and are, therefore, favourable in comparison to scattered predictions.

The model trained on a single dataset is then tested with the data collected from an independent trial, as would be the case in industry, and rather unsurprisingly does not give the same results as testing with data taken from the same set. It does,

however, provide predictions of wear that account for the varying wear rate between trials, allowing both wear prediction and rate estimation from an observed input sequence by calculating the lag in state progression from a base set.

The HMM, therefore, proves to be a useful tool in wear tracking given its ability to withstand stretching or compressing of the time axis corresponding to decreasing or increasing wear rates respectively. This resilience does, however, rely on the observed data remaining in the same input space between tools. More specifically, this method relies on machine process parameters remaining constant. A change in parameters can, depending on scale, result in observed signals varying out of the range experienced during training and would require retraining on this new parameter set accordingly.

In order to further validate these findings and implement such a model industrially within SLS, the requirement is mostly one of more data. By monitoring the industrial process and taking regular tool wear measurements, similarly to the experimental method in this thesis, it is expected that the tools will wear at differing rates and hence provide comparable data to that within this chapter, further confirming the previous findings. In its current form, the aforementioned model can be implemented into production, given a training set is first provided under fixed process parameters and conditions. As an extension, running complementary trials in which process parameters are modified incrementally would increase the size of the input space, and further exploratory work in this area would give a greater understanding of parameter correlations with variations in observed features.

TOOL STATE PREDICTION

Chapter 6 has shown that it is possible, based on a previous observation, to predict a tool state corresponding to the following AE observation in a sequence. This provides the machine operator with the ability to track the current tool's wear level during machining. Given the ability of the HMM to track wear state and, therefore, infer wear rate, the final task is to develop a method of predicting future wear state, given a sequence of past observations and a model trained on a wide range of data.

This chapter focusses on the implementation of Gaussian process (GP) based models, more specifically a Nonlinear Auto-Regressive with eXogenous inputs (NARX) approach based on Gaussian processes to predict wear level with natural confidence bounds for predictions.

7.1 NARX models

The problem of predicting future wear state based on process feedback has been relatively unexplored in previous, machining-focussed work. Generally, detecting a current state is adequate, given a repeatable process and operator expertise. Artificial neural networks are the most common methodology found in the previous literature [24], and an in-depth review of previous implementations can be found in [24], although these methods have still not been widely adopted, due to poor performance and model over-fitting on limited datasets. The concept of future prediction of tool wear, therefore, is considered a new area of exploration with severely

limited previous work to build upon.

In addition to the neural networks used before, a number of regression tools exist for estimating future values of time series data, the most common being simple polynomial regression or calculation of the moving average. Polynomial regression aims to model a relationship between input and output variable as a polynomial of N degrees in the input, while the moving average method uses an average of previous outputs to calculate a potential next value of the series. In both cases alone, the tools would use a fraction of the data available in this work, considering it is believed that past inputs (of multiple dimensions) and outputs provide insight into the future state of the process. A model which can incorporate various inputs, outputs, and lagged instances of both, therefore, is desirable.

Following its introduction in 1985 [82,83], the NARMAX (Nonlinear Auto-Regressive Moving Average with eXogenous inputs) model has been one of the most versatile and widely used time series models for nonlinear system identification [84,85]. The NARMAX model form encompasses both nonlinear discrete-time process and noise models, however, by assuming the noise process to be white Gaussian the simpler NARX model can be used.

Previously, the NARMAX model has been used in areas such as modelling of gas turbines [86], flexible robots [87], electron flux evolution [88], and financial trends [85] to name a few. The common detail here is the ability to measure both inputs and outputs, and makes the NARMAX model suitable for applications such as industrial processes, control systems, economic data (and financial systems), biology, medicine, and social systems [85]; however, this list is by no means exhaustive. Specifically to the work in this thesis, the NARX model is useful as it enables a model built around the full set of observations and features available, including past wear states, AE features, and noise. The inclusion of past outputs reduces the number of past inputs which need to be included in the model, in turn decreasing model complexity. The importance of this fact is explained and presented in [85].

The NARX model, therefore, takes a form whereby the current system output value is predicted based on a nonlinear function F of previous inputs (x_n) and outputs (y_n)

$$y_i = F(y_{i-1}, \dots, y_{i-n_y}; x_i, \dots, x_{i-n_x+1}) + \epsilon_i \quad (7.1)$$

where the residual sequence ϵ_i is white Gaussian. The number of output lags is denoted n_y , similarly, the number of input lags is denoted n_x . This formulation of a NARX model differs a little from the original [82, 83], in that it allows the use of the present input x_i [84].

It is usual to find NARX models based around a multivariate polynomial expansion basis for the function F where the expansion coefficients are learned using linear least-squares methods. In addition, nonparametric NARX model forms are possible based around a number of machine learning techniques such as neural networks for example [89, 90].

Considering the multivariate polynomial case, the problem of model training can be split into two stages. First, establishing the structure of the problem and selecting which multinomial terms to include in the model, and secondly, establishing the expansion parameters for those terms. Nonpolynomial variants of the NARX model, however, often disregard the primary step and simply include all expansion terms consistent with certain hyperparameters of the model form. In this case, the only consideration is the number of terms included in the model, with too many leading to overfitting through increased model complexity, and therefore, the model learning the noise included with the signal of interest [84].

As previously mentioned, another recently established form of the NARX model is that based upon Gaussian processes [91]. The GP-NARX model has the distinct advantage in that it consists of a Bayesian framework which produces natural confidence intervals for the resulting predictions. The GP will be discussed further in the following section; however [92], provides an overview of the literature surrounding the GP-NARX form.

7.2 The Gaussian process

A Gaussian process is simply a generalisation of the Gaussian probability distribution [93]. A probability distribution is a mathematical function that describes the probability of outcomes of a given event, and can take the form of either a discrete or continuous distribution given the format of the possible outcomes. A stochastic process, on the other hand, governs the properties of functions, with a Gaussian process consisting of elements of $F(x)$ at x which form a Gaussian distribution. The

basic premise here being to perform inference over functions directly, as opposed to inference over parameters of functions [84].

For simplicity, the work here will assume a single output variable for any system of interest [84]. As in the preceding chapters, and for the case of GPs in general, the matrix of multivariate training inputs is denoted as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_N]^T$, while the corresponding vector of training outputs is given as \mathbf{y} . Similarly, the input vector for a testing point is denoted by the column vector \mathbf{x}^* with the unknown resultant output denoted by y^* .

The key regression relationship behind the Gaussian process is,

$$\mathbf{y} = f(\mathbf{x}) + \epsilon \quad (7.2)$$

where ϵ provides a noise term which is assumed to be a zero-mean random variable, given as

$$\epsilon \sim \mathcal{N}(0, \sigma_n^2). \quad (7.3)$$

The noise variance σ_n^2 is classed as a hyperparameter of the model which requires estimation. It therefore follows that

$$\mathbf{y} \sim \mathcal{N}(\mathbf{f}, \sigma_n^2) \quad (7.4)$$

where f is a resulting unobserved or latent variable.

A Gaussian process prior is formed by assuming a Gaussian distribution over functions for the latent f ,

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x})) \quad (7.5)$$

where $m(\mathbf{x})$ is the mean function and $k(\mathbf{x}, \mathbf{x}')$ is a positive-definite covariance function. As a visual example, drawing samples from a Gaussian process prior distribution with a linear covariance function and zero mean, results in the plot shown in Figure 7.1. A second, less trivial example can be seen in Figure 7.2 which similarly

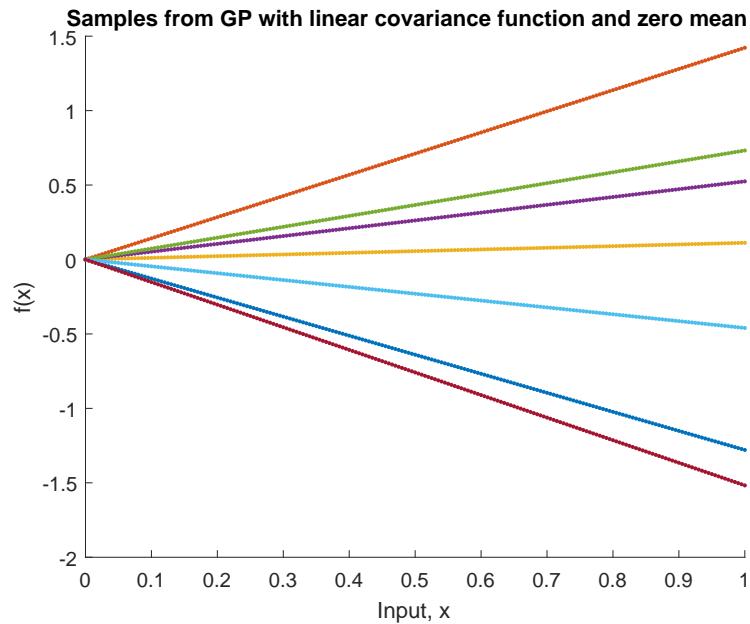


Figure 7.1: Samples from a Gaussian process with linear covariance function and zero mean.

uses a zero mean prior, however, the covariance function in this case is a squared exponential (SE). The SE example contains a characteristic length scale which highlights the importance of the specification of the prior, as it fixes the properties of the functions considered for inference. A number of covariance functions are available depending upon the particular application and prior beliefs, and manipulation of parameters such as the length scale allow, for example, changing the rate of variation of the functions to suit the specific situation. The problem of learning in GPs is simply that of finding suitable parameters for the covariance function, providing a model of the data which one can interpret [93].

Given a dataset of training points, one can consider only those functions which pass through these points exactly, reducing uncertainty close to the observations. The combination of the prior and training data results in the posterior distribution over functions. The addition of more training data points would see the mean function adjust itself to pass through these new points also, and the posterior uncertainty would continue to reduce close to the additional points.

This reduction is made possible given the defining property of the GP, that the density of a finite number of outputs from the process is multivariate normal. Using this property alongside the known marginalisation properties of the Gaussian density

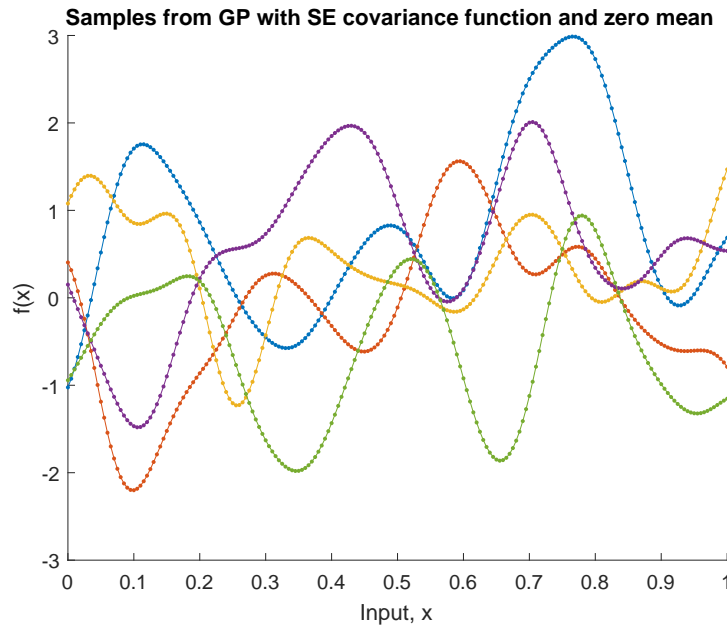


Figure 7.2: Samples from a Gaussian process with squared exponential covariance function and zero mean.

allows one to simply consider just the values of the function in Equation 7.5 at the points of interest, consisting of the training points and predictions. Allowing \mathbf{f} to denote the function values at the training points \mathbf{X} , and f^* to denote the predicted function value at a new point \mathbf{x}^* , one obtains,

$$\begin{pmatrix} \mathbf{f} \\ f^* \end{pmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) & K(\mathbf{X}, \mathbf{x}^*) \\ K(\mathbf{x}^*, \mathbf{X}) & K(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix} \right) \quad (7.6)$$

where a zero-mean prior is chosen for simplicity, and $K(\mathbf{X}, \mathbf{X})$ is a matrix whose $(i, j)^{th}$ element is equal to $k(\mathbf{x}_i, \mathbf{x}_j)$. It follows that $K(\mathbf{X}, \mathbf{x}^*)$ is a column vector whose i^{th} element is equal to $k(\mathbf{x}_i, \mathbf{x}^*)$, and $K(\mathbf{x}^*, \mathbf{X})$ is the transpose of the same [84].

Considering that the unobserved variable \mathbf{f} is of little interest, it can be integrated out from Equation 7.4 [93], as the relevant integral,

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f})d\mathbf{f} \quad (7.7)$$

is over a multivariate Gaussian and has a closed-form solution. The result of this is

the joint distribution for the training and testing target values for the observed y

$$\begin{pmatrix} \mathbf{y} \\ y^* \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I & K(\mathbf{X}, \mathbf{x}^*) \\ K(\mathbf{x}^*, \mathbf{X}) & K(\mathbf{x}^*, \mathbf{x}^*) + \sigma_n^2 \end{bmatrix}\right). \quad (7.8)$$

Once the joint distribution $p(\mathbf{y}, y^*)$ is converted into a conditional distribution $p(y^*|\mathbf{y})$, using standard results for the conditional properties of a Gaussian, the final expression is obtained [84, 93]

$$y^* \sim \mathcal{N}(m^*(\mathbf{x}^*), k^*(\mathbf{x}^*, \mathbf{x}^*)) \quad (7.9)$$

where,

$$m^*(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1} \mathbf{y} \quad (7.10)$$

is the posterior predictive mean, and

$$k^*(\mathbf{x}^*, \mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - K(\mathbf{x}^*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1} K(\mathbf{X}, \mathbf{x}^*) + \sigma_n^2 \quad (7.11)$$

is the posterior predictive variance, again expressed for the observed y .

The result of this definition is that the GP model provides a posterior distribution for the unknown y^* . The mean from Equation 7.9 can be taken as the best estimate of the value of y^* , while the variance can be used to define confidence intervals.

As previously mentioned, when using a covariance function such as the SE function of the form

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2l^2} \|\mathbf{x} - \mathbf{x}'\|^2\right), \quad (7.12)$$

it is necessary to select the optimal hyperparameters for a given problem. Such parameters l and σ_f^2 , together with the noise variance σ_n^2 , form the set θ which can be optimised through a number of methods. In this case, maximising the function

$$f(\theta) = -\frac{1}{2}\mathbf{y}^T [K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I] \mathbf{y} - \frac{1}{2} \log |K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I| \quad (7.13)$$

provides the hyperparameters and can be conducted using a simple gradient descent, however, more powerful algorithms are available such as that in [94].

7.3 GP-NARX

The GP models discussed in Section 7.2, are essentially static maps having learnt the relationship between point inputs and outputs through a training stage. Continuing to follow the procedure in [84], it is relatively straightforward to adapt the model to learn dynamical system behaviour by applying the NARX framework. The NARX form in Equation 7.1 is used with the function F replaced with a GP.

In the previous chapters, model assessment has simply consisted of comparing model predicted state classifications with a predefined validation set and determining the number of correct classifications. As this work has moved towards a regression problem, the model targets are developed from the discrete wear measurements to vectors, and are formed using cubic-spline interpolation on class labels used in Chapters 5 and 6. There are a number of options to determine the validity of the dynamic model, the most basic method of validation is to compute one step ahead (OSA) predictions, where the training data is used to predict outputs for a given time using observed inputs and outputs prior to that instance. Once predicted outputs are obtained, it is possible to compare the predicted and observed outputs at each point in much the same way as in the previous chapters. This can be defined as

$$y_i^* = F(y_{i-1}, \dots, y_{i-n_y}; x_i, \dots, x_{i-n_x+1}). \quad (7.14)$$

As a value for comparison, the Normalised Mean-Square Error (NMSE) is used and given by

$$\text{NMSE}(\hat{y}) = \frac{100}{N\sigma_y^2} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (7.15)$$

As mentioned in [84], previous experience has shown that a NMSE of less than 5.0 indicates good agreement, while below 1.0 indicates an excellent fit.

A more stringent test of the model is to compute the Model Predicted Output (MPO) which is given by,

$$y_i^* = F(y_{i-1}^*, \dots, y_{i-n_y}^*; x_i, \dots, x_{i-n_x+1}) \quad (7.16)$$

where rather than previously-observed outputs being used, the previously-predicted outputs are used for the following predictions. It is worth noting here that this test can be conducted on both testing and training data.

The GP algorithm does have a couple of drawbacks which are worth bringing to attention here. The first being that the algorithm relies on the inversion of the covariance matrix, which is inherently computationally expensive, costing $O(N^3)$ multiplications, where N is the number of training points. The prediction of a new output requires $O(N)$ multiplications for the predictive mean and $O(N^2)$ for the corresponding variance. While this is not an issue considering the relatively modest size of the training set used in this work, it should be noted that this fact can be prohibitive when considering larger datasets. The second obstacle is due to noise on the training set when making multi-step ahead predictions, as the predictive outputs are fed back as inputs and carry their predictive uncertainty with them. This results in any noise on the inputs being amplified by future predictions. For further detail of potential solutions to these drawbacks, the reader is directed towards [84].

7.4 Results

Since tool wear itself is an intrinsically continuous quantity (despite wear measurements forming discrete classes), the implementation of the GP-NARX methodology here aims to solve a regression problem as opposed to the classification approach of previous chapters. A requirement of the regression problem is paired inputs and outputs which, given the limited number of wear measurements, presents a challenge in obtaining sufficient data. One solution to this problem is to use ordinal regression [95], another is to simply repeat the experimental work, taking tool scans at a higher frequency. This concept, however, owing to the time taken to obtain such

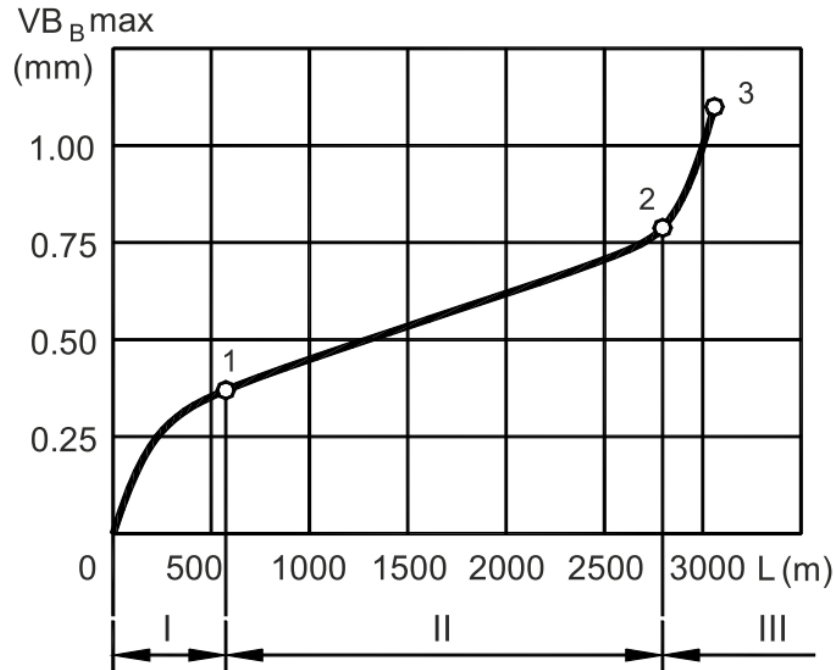
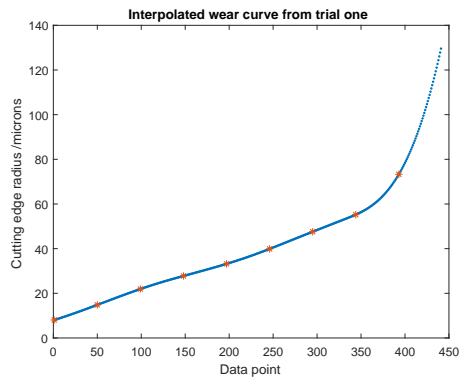


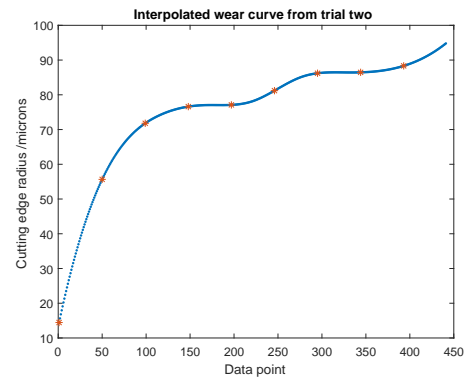
Figure 7.3: Cubic wear curve of flank wear vs. length of the cutting path. The break-in period (I), steady-state wear region (II), and the failure region (III) are indicated.

measurements, becomes impractical in reality. A more feasible approach is to use previous knowledge of tool wear curves to interpolate between the currently measured points to increase the size of the dataset to match that of the AE features. A classically accepted tool wear curve consists of three distinct areas, together forming a cubic curve [23]; the break-in period (I), steady-state wear region (II), and the failure region (III) as shown in Figure 7.3. Given that wear is generally a progressive mechanism, it follows that a spline interpolation between points is adequate for this purpose and provides a smooth curve in keeping with previous works. The resulting curves can be seen in Figure 7.4 and will be used as the output datasets from this point. An example of one of the input feature vectors for each trial can be seen in Figure 7.5, where the moving average is also plotted alongside.

The second challenge faced with such a regression problem, is the splitting of data into both training and testing sets. Selection of a suitable training set is of fundamental importance to the success of such a model, as the Gaussian process is incapable of extrapolation with any degree of certainty and, therefore, must experience training data of the full range to be expected as an input during prediction. For this reason, both trial datasets are split in two by taking alternate input fea-

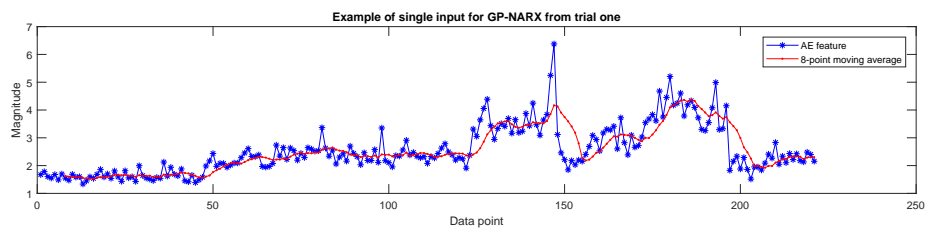


(a) Interpolated wear curve from trial one.

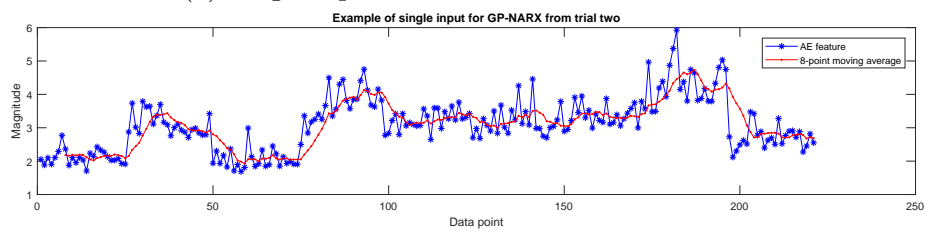


(b) Interpolated wear curve from trial two.

Figure 7.4: Interpolated wear curves.



(a) Single input feature vector from trial 1.



(b) Single input feature vector from trial 2.

Figure 7.5: Examples of single input feature vector for GP-NARX from each trial.

ture vectors for each training set and assigning the missed vectors to testing sets. The common method of splitting trial datasets in half, for example, is inappropriate considering the nature of the curves. The selection of a training set in this case is mainly limited by the fact that, due to experimental budget and timescale, only two example datasets are present, each with profoundly different wear curves. Merging these two datasets disturbs the time-dependent nature of the signals and, therefore, each dataset will initially be considered independently. Another option is to concatenate datasets to simulate a tool change (with a sharp change in wear state) which will be explored subsequently.

To begin with, the individual trial datasets are split into two subsets for training and testing respectively. Considering the universal nature and limited number of parameters to optimise, the squared-exponential covariance function is chosen, however, in contrast to that in Equation 7.12, automatic relevance determination (ARD) is used to learn an individual length-scale for each dimension of data depending upon its relevancy to the regression [96]. The SE ARD kernel function is defined as

$$K(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp \left[-\frac{1}{2} \sum_{d=1}^D \left(\frac{x_d - x'_d}{l_d} \right)^2 \right] \quad (7.17)$$

where l_d is an individual length-scale hyperparameter for each input dimension x_d .

As previously discussed, the hyperparameters l , σ_f^2 , and σ_n^2 are to be optimised to suit this particular problem and in this case, a population based optimiser using a quantum-behaved particle swarm [97] is used to maximise the log marginal evidence in Equation 7.13. This method is based upon that discussed, and code used, in [94]. The last requirement before training and validation can commence, is to select the number of input and output lags required for the model, which are also hyperparameters of the GP-NARX model. A simple search based around minimising the predictive error of the model resulted in input and output lags of $n_x = 8$ and $n_y = 3$ respectively.

Using $n_x = 8$ and $n_y = 3$ along with those optimal GP hyperparameters obtained, the GP-NARX model can be trained and validated on separate and unique sets taken from the total dataset of trial one. Focussing only on the MPO test data predictions, due to their more stringent nature, the NMSE calculated is 1.21; a satisfying result when also considering the confidence intervals (± 3 standard de-

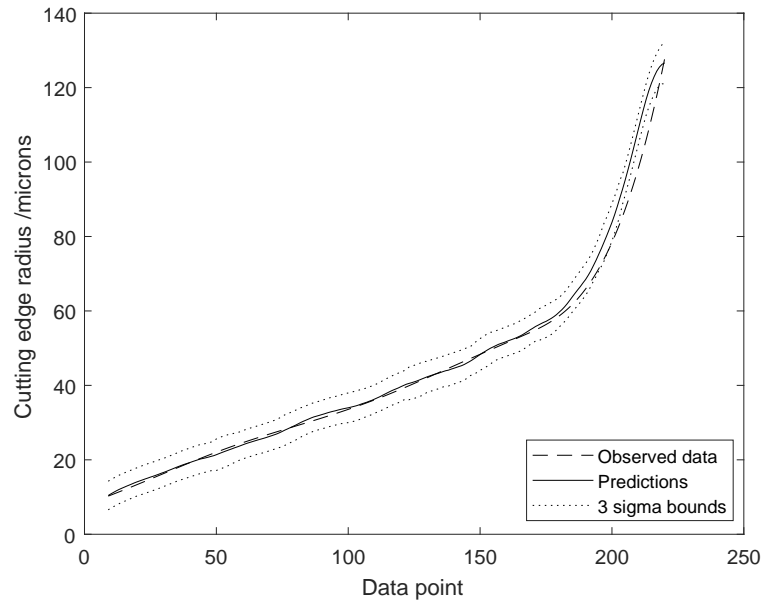


Figure 7.6: GP-NARX model predicted output based on training and validation sets taken from trial one data.

viations) due to parameter uncertainty are very close to the data. The resulting prediction comparison can be seen in Figure 7.6. Following the same procedure for the second trial, the NMSE is 0.187 which results in the plot shown in Figure 7.7.

Both of these plots show that the model predicted outputs are close to the observed data, featuring tight confidence intervals around predictions. In either case, the target curves are relatively slow moving and are not difficult for such a dynamical model to grasp, especially when using lagged outputs. Figure 7.6 clearly shows the effect of feeding predictions back into the model when errors are present as at approximately point 175 the predictions begin to slowly drift away from the observed data, increasing in error as more predictions are made. The risk of using automatic relevance determination in the NARX case (especially where training and testing data are similar, and previous outputs are used in prediction) is that the model can effectively focus entirely on the previous output points rather than input data, learning only the curve shape and appearing valid under testing. In this case, considering the previously proven relationships between input vectors and output classes, it is assumed that this is not the case; however, the best way to prove validity is through increased testing set size; something which is out of scope of this work due to resource availability and budget.

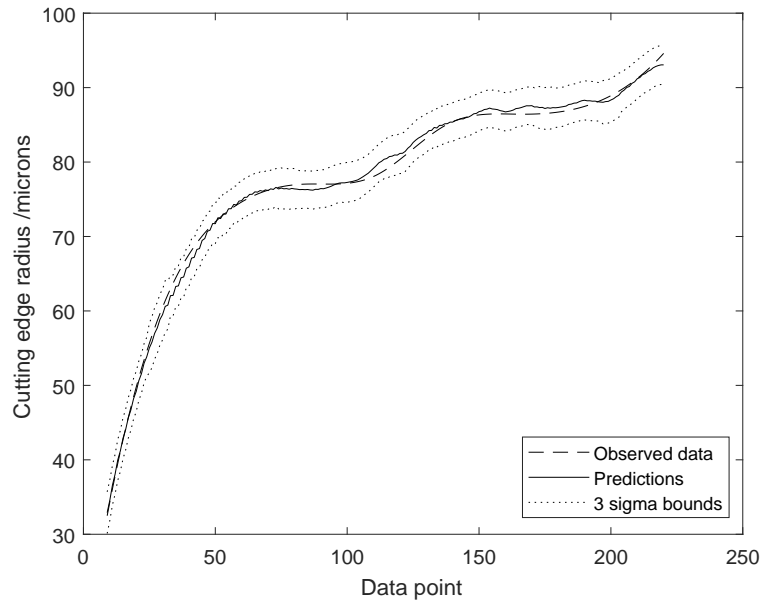


Figure 7.7: GP-NARX model predicted output based on training and validation sets taken from trial two data.

The smooth nature of the wear curve is no difficult task for the GP-NARX model; however, combining wear curves to simulate a tool change poses another challenge for such a model. An example of such a scenario can be seen in Figure 7.8 where the observed output curves are concatenated. The sharp drop at point 220 breaks the smooth, slow moving function and requires modification of the length scale hyperparameter to accommodate the sudden change in direction.

Training the GP-NARX using the previously adopted procedure, and testing on the other half of the combined points results in the plot shown in Figure 7.9 with associated MPO NMSE of 2.79. It is clear to see that the predicted function is more of a coarse function, lacking the smoothness of Figures 7.6 and 7.7 due to the decreased length scale required to capture the abrupt discontinuity. This figure also shows that the resulting model manages to capture the transition between tools well, albeit overshooting slightly. It is worth noting that while the NMSE is greater and, therefore, indicates a poorer model predicted output, the benefit is that the trained model is more general, having experienced a wider range of training inputs.

Another benefit of including multiple trial data in this way is that it enables further validation of the model, and confirmation of previous assumptions, by switching the order of concatenation between training and testing sets. This ensures that the

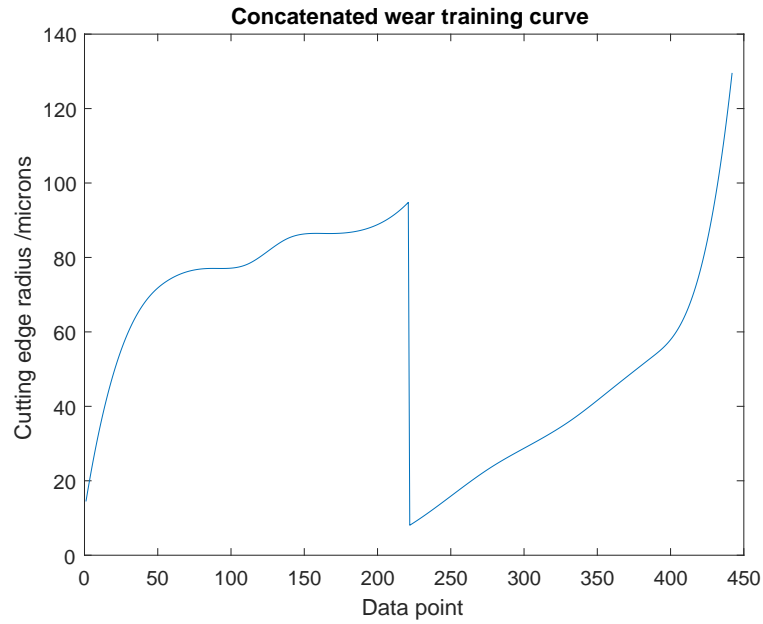


Figure 7.8: Observed output curves concatenated into single training dataset.

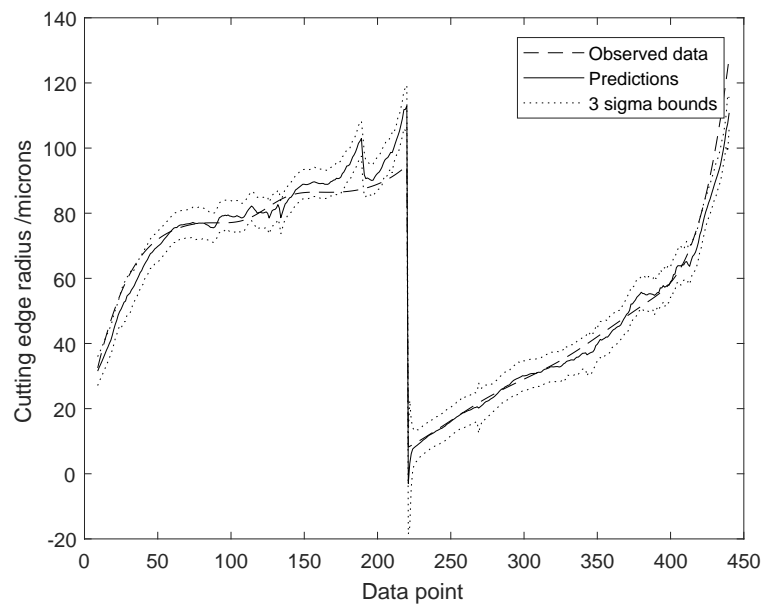


Figure 7.9: GP-NARX model predicted output based on training and validation sets taken from concatenated trial data.

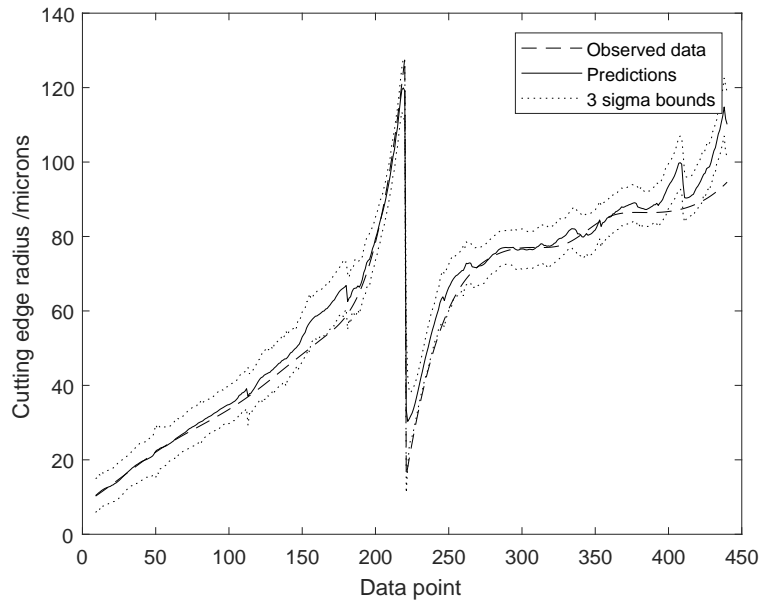


Figure 7.10: GP-NARX model predicted output based on training and validation sets taken from concatenated trial data.

training process isn't simply learning the output curve shape, but rather learning the desired relationships between model inputs and output in a way that it is flexible between tools. Again, training using that data shown in Figure 7.8, however, this time testing using a reversed order, results in the plot shown in Figure 7.10.

This is an encouraging result, as the MPO NMSE here is 2.81, representing a good fit to the observed data despite the validation data being of a different shape to that the model was trained upon. Not only does this provide further evidence to support that the models trained on each trial dataset independently are valid, but also implies that given enough training data, such a model would be robust to tool changes and varying wear rates.

While this is a positive result, there is still certainly room for improvement. Taking both Figures 7.9 and 7.10 into consideration, there are areas of both where the predictions begin to deviate from the observed data, albeit being corrected at a later point. This implies prediction error fed back as input noise, and indicates a need for an improved training set, alongside an exploration into kernel selection. For the purpose of this work, all trained models have used the previously described SE ARD kernel function, given the relatively slow moving nature of the wear output and tunable length-scale hyperparameter. There may, however, be a more suitable

kernel function or combination of functions depending upon the specific process and the expected behaviour of the output. This concept is touched upon in [98], given that the issue of optimal kernel selection in general is beyond the scope of this thesis.

7.5 Overview of tool state prediction

This chapter has detailed the Gaussian process and its combination with the NARX model to form a useful tool for wear prediction with associated confidence bounds; it details selection of training and testing data, based upon interpolated wear curves for both independent trials, followed by a combination of both.

The models trained on half of the data from trial one and trial two return NMSEs of 1.21 and 0.187 respectively, indicating a healthy agreement between model predicted outputs and observed testing data. It is noted, however, that due to the similar nature of the testing and training data in both of these cases, the results could potentially be misleading if the model is heavily biased to learning the output curve over inputs. Ideally, to confirm this isn't the case, a greater volume of data is required, albeit this is not practical at this stage of the work. As a substitute, the training curves are concatenated and a model is trained on a combination of data from both trials. This method has the benefit of a larger training set and, therefore, more general model, resulting in a NMSE of 2.79. While this suggests a lower quality fit than treating each trial independently, it does provide the opportunity to reorder the wear curves of the testing set to ensure that the model isn't simply duplicating the output seen during training. By doing this, a NMSE of 2.81 is obtained, confirming that the model is valid, and capable of performing wear prediction throughout tool changes and under differing wear rates.

It is concluded that this method is a capable technique for industrial tool wear prediction, where a greater quantity of data would be beneficial to allow further testing and validation. As mentioned in the chapter, Gaussian processes are very good at interpolation but are incapable of extrapolation in this application and as such, training data is needed which contains a full range of likely signals throughout whichever process is being observed.

In an industrial context, in-process data is readily available, and given access to this, such a model could be trained on each specific process within a manufacturing

plant such as SLS. Following the initial training phase, continuous observation of a machining process can occur and the GP-NARX model can provide the operator with tool life predictions for each individual part, and the corresponding confidence intervals of each prediction. The ultimate goal of such a method, is to remove the need for constant human observation of the machining process - a goal which the work presented here takes a step towards achieving. Implemented across a number of machining centres, a single operator could be responsible for a greater number of machining instances, with each machine informing the operator of when the next tool change is likely to be required. Pairing this method with a scheduling system naturally leads to implementing scheduled tool maintenance, in an optimised manner, based on the real measured conditions within the factory.

CONCLUSIONS AND FUTURE WORK

Now that the industrial motivation, experimental setup, data collection, and wear modelling have been detailed, it follows to formally conclude this work, including explaining any limitations and future work that could be undertaken, along with how industrial implementation of the predictive models developed could be achieved. The work here has attempted to present the steps in developing a predictive tool wear model in a logical manner, beginning with justification of the particular process chosen and describing the methodology of acquiring representative data, following with applying three increasingly-capable modelling techniques, and ending with testing a working predictive model on the data obtained. This concluding chapter will act as an overview of the conclusions drawn at the end of each preceding chapter, with additional focus on industrial implementation and future work.

8.1 Experimental data collection

The experimental procedure used, has consisted of two independent machining trials, intended to provide unique datasets for comparison and to provide adequate data for modelling. The trials each consisted of machining a number of test coupons while collecting AE data primarily, and resulted in nine tools at different wear states from each trial. These tools were then measured and correlated with the AE features generated by cutting with each tool.

This experimental methodology has the main benefit that for each trial completed,

it provides a full selection of related and measurable components. Using individual tools and material specimens ensures that once the developed machining process is complete and accompanying AE data collected, the used tools and cut material is available indefinitely for further analysis and repeated measurements if required. Splitting the experimental trials into two stages also has the added benefit of reducing cost and machining time, as machining a cylindrical billet is an industrial representative process which provides a continuous supply of material for extended cutting durations.

The main downside of this method is the associated cost of using multiple tools, each being a costly purchase when considering their solid carbide construction. Not only that, but the experimental procedure is incapable of providing a continuous dataset from new to failed tool, as the tools are removed for measurement. Continuous measurement of a single tool would also result in such a large volume of data that it may be infeasible to store (due to the high data sample rate needed for AE recording), and would be impossible to take wear measurements in-process due to the use of high pressure coolant obscuring any view of the cutting edge, and the rotational speed not providing adequate scan time. Any chosen methodology, therefore, requires an interrupted process and the associated limitations this brings. That being said, when monitoring a continuous industrial process, the data volume limitation is lifted, as samples can be taken for feature calculations and then discarded once a wear prediction is made.

In an ideal world, if budget and timescale were no issue, the trial procedure would have been carried out more than twice, with each instance split into more than nine tools. This would improve the limited number of points present on each wear curve and reduce the interpolation needed in Chapter 7; however, nine tools is considered throughout this work to provide an adequate number of class labels and, coupled with previous knowledge of common wear curves, a reasonable insight into the shape of the wear curve for each trial.

Looking at the wear curves gained from tool measurements following each trial, reaffirms the motivation behind this work entirely. It is clear to see that despite following the same testing procedure with identical tools, the material variations result in vastly different wear curves in each instance. This variation is beneficial in the sense that it clearly demonstrates the need for an intelligent monitoring solution, while providing enough data to explore and validate correlations between AE features and wear. It does, however, pose the questions of how much variation is

likely to be experienced during industrial manufacture of these components, whether due to material variations, environmental conditions, process parameters, or built-up edge interference. For this reason, one of the main areas of future work is to repeat the experimental procedure a number of times, to ensure that enough data is gathered for training a predictive model before industrial implementation. The more different wear curves and greater the range of input features experienced during training, the more accurate a trained model is likely to be. It is worth noting that implementation of any such model becomes a case of obtaining adequate training data over the full range for that particular process. Should one wish to implement the models described in this work for their own industrial application, the experimental procedure (or similar) would need to be carried out to provide training data applicable to that specific situation.

8.2 Wear modelling and prediction

Chapters 5, 6, and 7 follow a natural progression from wear classification to prediction, moving from a time-independent diagnosis to a time-critical prognosis. It is established that a linear SVM is a capable model for classifying tool wear state given a selection of time-domain AE features as inputs. The classification method is well suited to this problem due to the discrete wear measurements taken, each of which corresponded to a number of AE data points; it also further confirms the existence of solid relationships between AE features and wear. A novel method of obtaining tool wear measurements is used, based upon 3D cutting-edge scans, which provides a greater understanding of wear level and progression when compared to existing wear measurement techniques [5, 23, 24]. The key result of Chapter 5 is that diagnosing a tool's state can be improved by including the order that observations are experienced in. A confusion matrix is presented containing a combination of data from both trials, and it is clear to see that a number of misclassifications occur scattered between classes; an undesirable situation due to the possibility of classifying points far away from their intended class. In industry, this could have the consequence of component damage if it results in over-use of a tool, or increased cost, if a tool is removed before its useful life is reached. It is therefore suggested that by taking into account the existing class label when suggesting the most likely class for a newly-observed instance, the accuracy of the model can be improved and errors minimised to within a class either side of the actual. The hidden Markov

model is the ideal candidate for such a situation, providing a probabilistic prediction of the most likely state of an observation based upon the currently occupied class.

The HMM is successful in providing accurate sequential state classifications, with any error remaining alongside the diagonal of the confusion matrix. This is beneficial over that given by the SVM as, while the misclassification rate may be higher, the error of each misclassification is less and provides a less misleading diagnosis. It is also established within Chapter 6, that wear rate can be inferred from HMM class predictions by comparing the class curve to that of a baseline dataset, another novel finding when compared to previous work in the research area. This finding is a positive result as it indicates that although the wear curves (and, therefore, rates) of the two independent trials differ, the relationships between wear and AE features remain present, exclusive of wear rate. For industrial application, this is also a productive result as it provides evidence that a single wear trial could be conducted to gather model training data, with subsequent cutting operations simply being compared to that control set. The downside of this method is that any change in process parameters would require retraining of the model to obtain a new transition and emission matrix, specific to those values. This is obviously a time consuming and costly process that is less than ideal in a changing production environment. The SVM suffers from a similar issue, since data unseen during training and distant from existing points cannot be confidently classified, although it may still fit within the class bounds defined in the SVM and result in successful classification. As an extension of this approach, a predictive model is required that can predict future wear states given a current and potentially changing wear rate; the GP-NARX model proving an ideal candidate for this prediction problem that has not yet been investigated before this work.

The GP-NARX model predicted output is shown to accurately predict continuous wear with natural confidence intervals for each independent trial. The risk of using ARD to weight input features is highlighted when considering the similarity between training and testing sets, given that weighting the previous outputs highly could cause the model to be biased towards output only, providing misleading results. As a more stringent test, where possible without altering the dynamics of the process, the data is concatenated for training, and reordered for testing. This presents the possibility of switching the trial curves order between the training and testing sets to prove independence of wear order and output curve shape. The model is successful in predicting this new shape of output curve and confirms that the GP-NARX model

is an ideal tool for predicting tool wear during machining. It does, however, require training data of the full range likely to be experienced during operation, to maintain prediction confidence. Given the success presented in Chapter 7 and the absence of data-based future tool wear prediction in previous work, this forms a promising contribution to the development of autonomous manufacturing processes.

8.3 Industrial implementation

While the industrial nature of this work has remained an important concern throughout, it is worth drawing attention to the further requirements needed for a full industrial implementation of such a predictive model. The concern here relies on a real-time, on-line system implementation rather than the off-line development work focused on in this work to date. Considering the progressive nature of wear and the relatively long usage duration, this should be a somewhat straightforward task, as the time to process a sample and make a prediction is small in comparison to process length. For instance, say a tool remains useful for approximately two hours; taking a ten-second data sample and allowing five seconds for processing and prediction results in 480 predictions throughout its life; an adequate number given the smooth wear curves expected. Again, data storage can be kept to a minimum by discarding data samples once useful features are taken from them and a prediction made.

The use of an embedded hardware solution is an attractive option given their inherent robustness and speed of bespoke task execution. The compactRIO from National Instruments, for example, contains a field-programmable gate array (FPGA) which can be configured to perform an application specific hardware task at great speed when compared to the everyday computer system. Such a system is also intended to be fully compatible with the compactDAQ hardware used throughout this work, sharing data acquisition modules and therefore providing a logical choice of hardware. Such a device (when programmed to suit the task at hand) has the ability to acquire data, perform feature extraction and wear prediction, and also to export these predicted values in such a way that they can be interfaced with a computer numerical control (CNC) system.

The biggest concern is that touched upon in Section 8.4, regarding sensor location. To implement this strategy in a production environment would require fitment of an appropriate AE sensor to the spindle assembly of whichever machining centre is

being used. In some cases this may be easier than others depending upon workpiece clearance and the ability to run wiring through slip-rings; however, it is unlikely to be an impossible task, with some minor hardware modification. In performing such modifications, the operator should be made aware that proximity of a sensor to the cutting action is highly important for good quality data acquisition.

8.4 Limitations

When transferring such a methodology into an industrial setting, it is important to understand where potential limitations of the research may need overcoming to enable full functionality. In this case, there are two key limitations that should be highlighted, despite their relative ease to resolve.

The main hurdle in implementing the described monitoring strategy industrially is the selection of a suitable sensor location. In this work, the AE sensor has remained attached to the workpiece fixture purely due to ease of affixing, proximity to the cutting zone, and low cost of the solution as it requires no modification of shared machinery. In reality, when machining large components, this would be impractical due to the potential large distance between cutting action and sensor, and also the fact that the distance is constantly varying and affecting signal attenuation. The most elegant solution here is to simply locate the sensor within the spindle assembly, ensuring a fixed and close proximity to the cutting zone. In some instances, this may require a small modification to the spindle head, the running of wires, and potential use of couplant depending on signal strength and sensor gain - a relatively trivial task given the potential benefits. An example of such a sensor can be found in [99].

The other point of consideration focusses on the acquisition of training data for a particular application. For any supervised learning algorithm, a change in process not experienced before by the model will inevitably require some level of training. For the general production environment where a process rarely changes, this is a minor issue as the model can be retrained according to process parameter modifications, as required, by conducting a tool life trial in a similar manner to Chapter 3 of this thesis. More specifically, conducting the common machining process and recording AE data throughout its duration, while also pausing the process at regular intervals to take tool wear measurements. For a process where parameters are regularly changed, this becomes infeasible as the training stage can require a

relatively large time period in comparison to its usage duration. As previously suggested, the simplest solution to this limitation is to train the model on a wide range of data to begin with, including examples of all of the likely conditions to be experienced during operation. This provides a flexible model, robust to parameter changes; however, this is obviously a time-consuming option and likely parameters may not always be known *a priori*. This option also lends itself to providing machine parameters to the predictive model, allowing a clear indication of any change and likely reducing predictive error, although investigation into process parameter effects on signals forms part of the potential future work.

8.5 Future work

Throughout this chapter, future work has been touched upon in a number of places, falling into three distinct areas: industrial implementation, further model training, and exploring the effects of process parameters and material on acquired signals.

The acquisition of further training and validation data is always beneficial for a supervised learning method as it presents further opportunities for model testing and ensuring of robustness. As with many machine learning algorithms, the training set is crucial to the performance of the GP-NARX model and the more complete, varied, and representative the better. While the data used in this work is adequate for a proof of concept, given budget and timescale, further data would enable further confirmation that the points made here are solid. The most straightforward approach to increasing the size of the training set is to simply repeat the experimental procedure detailed in this thesis a number of times. Each realisation would provide a slightly different wear curve, and a culmination of these provides a more general overview of the behaviour that is likely to be observed during machining. Budget and time aside, it would be beneficial to increase the resolution of the tool wear curves used as targets for the trained model. This will inherently require more frequent process interruptions in order to obtain tool scans; however, it would minimise the interpolation needed to form an estimated curve, resulting in a more accurate understanding of wear behaviour. The experimental process could further be improved by utilising a through-spindle type AE sensor, which would remove the requirement for a two-stage experimental process, and remove the need to separate the tool and holder between scans. Both of these facts would result in reduced error

in measurements, and a more industrially representative procedure provided that such an AE sensor is sensitive enough, and remains resistant to noise sources in the vicinity.

As briefly mentioned in Chapter 4, a further extension of the work presented here would be to investigate a wider range of potential features that could be used as model inputs. Both the frequency domain and wavelet domain have been used in previous work relating to tool wear, indicating that there is still valuable information to be obtained more clearly following such transforms. While it has not been necessary to explore any further at this stage, decomposing the input from an AE sensor into various components would have the benefit of reducing computational burden should a descriptive measure be found. Such methods are common in the SHM community and should be directly applicable within machining. Previous examples within machining can be found in [13, 25, 26].

Continuing with the GP-NARX model, the most beneficial next step would be to investigate the effect of kernel selection on the predictive accuracy of the model in the context of this specific tool wear problem. The kernel function defines the behaviour of the underlying process, and as such, should be matched to the mechanism which the model is aiming to predict. As discussed in Chapter 7, the SE covariance function is used in this work, with the length scale l is a hyperparameter that is tuned for optimal results. Considering the concatenated datasets within this chapter, there are both gradual, slow moving wear curve elements, and abrupt, rapid changes between tools. In this case, it may be beneficial to explore a combination of two different kernel functions, paired and weighted in an attempt to better capture such targets. Selection of such functions is discussed further in [98]. Research also continues in a number of communities, focussed around the development of new predictive models, and often made possible given the current advances in computational processing power. There are likely a number of methods in development which could be applicable to the tool wear problem, and future work should include evaluation of these as they become available.

Another area of future work is to separate the recorded AE data further, focussing on the data recorded per tool revolution or flute pass, rather than per tool pass on a given workpiece. It is shown in Chapter 4 that zooming further into the time-series data provides a clear indication of each chip as it forms, and can even be used to detect tool runout. Correlating each flute with specific chip formations is difficult, due to the fact that removing the tool from its holder alters the runout

present, and the first flute to engage with the material during cutting is random. Provided that the tool remains in the spindle, however, further work in this area should provide another level of insight into the process that has yet to be explored, ultimately increasing process understanding and model accuracy. Such a method should also be able to indicate damage to a single flute and during milling of complex, non-periodic profiles, although validation proves difficult without being able to also measure wear directly during machining.

Finally, given adequate machine time, it would be useful to map process parameters to AE features in a way that enables a single predictive model to function despite a change in parameters. At this stage, each dramatic change in parameters requires a sample of training data, however, such a map would enable a single training set to be used and simply manipulated to suit any parameter change. This is obviously a costly endeavour but the benefits it brings are considered crucial for the progression of research in this area. A further extension would be to explore both different materials and different processes, and assess the suitability of the chosen model formulation in this work to other machining processes. Both grinding and turning also suffer from tool wear, along with milling of parts of varying profiles, so further work in this area would also help to create a more general system that is more widely applicable.

BIBLIOGRAPHY

- [1] S. Nandi, H. A. Toliyat, X. Li, *Condition monitoring and fault diagnosis of electrical motors - A review*, IEEE Transactions on Energy Conversion, vol. 20, no. 4, (2005), pp. 719–729.
- [2] A. K. S. Jardine, D. Lin, D. Banjevic, *A review on machinery diagnostics and prognostics implementing condition-based maintenance*, Mechanical Systems and Signal Processing, vol. 20, no. 7, (2006), pp. 1483–1510.
- [3] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, C. K. Song, *Condition monitoring and fault detection of wind turbines and related algorithms: A review*, Renewable and Sustainable Energy Reviews, vol. 13, no. 1, (2009), pp. 1–39.
- [4] K. Jemielniak, J. Kosmol, *Tool and process monitoring - state of art and future prospects*, Scientific papers of the institute of mechanical engineering and automation of the Technical University of Wroclaw, vol. 61, (1995), pp. 90–112.
- [5] D. A. Stephenson, J. S. Agapiou, *Metal cutting theory and practice*, CRC Press, third edit edn. (2016).
- [6] Boeing, *Internatinal Titanium Association report from Boeing Commercial Aircraft Group* (2010).
- [7] P. Crawforth, B. Wynne, S. Turner, M. Jackson, *Subsurface deformation during precision turning of a near-alpha titanium alloy*, Scripta Materialia, vol. 67, no. 10, (2012), pp. 842–845.
- [8] M. V. Copeland, *Fortune.com* (2008).

-
- [9] J. F. Kahles, M. Field, D. Eylon, F. H. Froes, *Machining of Titanium Alloys*, Materials Forming, Machining and Tribology, vol. 7, (2014), pp. 31–56.
- [10] T. Vulcan, *Abundant Titanium's Import To Aerospace Well Known, But Other Industries Could Bring New Demand*, ETF.
- [11] Anonymous, *The engineering toolbox* (2017).
- [12] D. Musi, *An Update on Ti-5Al-5Mo-3Cr*, in *Titanium 2005*, Scottsdale, Arizona USA.
- [13] J. V. Abellan-Nebot, F. Romero Subirón, *A review of machining monitoring systems based on artificial intelligence process models*, The International Journal of Advanced Manufacturing Technology, vol. 47, no. 1-4, (2009), pp. 237–257.
- [14] K. Jemielniak, L. Kwiatkowski, P. L. Wrzosek, *Diagnosis of tool wear based on cutting forces and acoustic emission measures as inputs to a neural network*, Journal of Intelligent Manufacturing, (1998), pp. 447–455.
- [15] J. Sun, G. S. Hong, M. Rahman, Y. S. Wong, *Identification of feature set for effective tool condition monitoring by acoustic emission sensing*, International Journal of Production Research, vol. 42, no. 5, (2004), pp. 901–918.
- [16] D. V. Hutton, F. Hu, *Acoustic emission monitoring of tool wear in end-milling using time-domain averaging*, Journal of Manufacturing Science and Engineering, vol. 121, no. 1, (1999), pp. 8–12.
- [17] M. Lan, D. Dornfield, *Experimental studies of tool wear via acoustic emission analysis*, Proceedings of the 10th NAMRC, SME, (1982), pp. 305–311.
- [18] S. Y. Liang, D. A. Dornfeld, *Tool wear detection using time series analysis of acoustic emission*, Journal of Engineering for Industry, vol. 111, no. 3, (1989), p. 199.
- [19] D. A. Dornfeld, M. F. DeVries, *Neural network sensor fusion for tool condition monitoring*, CIRP Annals - Manufacturing Technology, vol. 39, no. 1, (1990), pp. 101–105.
- [20] P. G. Benardos, G. C. Vosniakos, *Prediction of surface roughness in CNC face milling using neural networks and Taguchi's design of experiments*, Robotics and Computer-Integrated Manufacturing, vol. 18, no. 5-6, (2002), pp. 343–354.

- [21] P. Benardos, G.-C. Vosniakos, *Predicting surface roughness in machining: a review*, International Journal of Machine Tools and Manufacture, vol. 43, no. 8, (2003), pp. 833–844.
- [22] K. Jemielniak, *Commercial tool condition monitoring systems*, The International Journal of Advanced Manufacturing Technology, vol. 15, no. 10, (1999), pp. 711–721.
- [23] J. P. Davim, *Machining: fundamentals and recent advances*, Springer (2008).
- [24] B. Sick, *On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research*, Mechanical Systems and Signal Processing, vol. 16, no. 4, (2002), pp. 487–546.
- [25] G. Byrne, D. Dornfeld, I. Inasaki, G. Ketteler, W. König, R. Teti, *Tool condition monitoring (TCM) The status of research and industrial application*, CIRP Annals - Manufacturing Technology, vol. 44, no. 2, (1995), pp. 541–567.
- [26] R. Teti, K. Jemielniak, G. O'Donnell, D. Dornfeld, *Advanced monitoring of machining operations*, CIRP Annals - Manufacturing Technology, vol. 59, no. 2, (2010), pp. 717–739.
- [27] I. Deiab, K. Assaleh, F. Hammad, *On modeling of tool wear using sensor fusion and polynomial classifiers*, Mechanical Systems and Signal Processing, vol. 23, no. 5, (2009), pp. 1719–1729.
- [28] N. Ghosh, Y. B. Ravi, a. Patra, S. Mukhopadhyay, S. Paul, a.R. Mohanty, a.B. Chattopadhyay, *Estimation of tool wear during CNC milling using neural network-based sensor fusion*, Mechanical Systems and Signal Processing, vol. 21, no. 1, (2007), pp. 466–479.
- [29] M. Elbestawi, T. Papazafiriou, R. X. Du, *In-process monitoring of tool wear in milling using cutting force signature*, International Journal of Machine Tools and Manufacture, vol. 31, no. 1, (1991), pp. 55–73.
- [30] S. Kurada, C. Bradley, *A review of machine vision sensors for tool condition monitoring*, Computers in Industry, vol. 34, no. 1, (1997), pp. 55–72.
- [31] A. Rivero, L. N. López de Lacalle, M. L. Penalva, *Tool wear detection in dry high-speed milling based upon the analysis of machine internal signals*, Mechatronics, vol. 18, no. 10, (2008), pp. 627–633.

-
- [32] W.-H. Ho, J.-T. Tsai, B.-T. Lin, J.-H. Chou, *Adaptive network-based fuzzy inference system for prediction of surface roughness in end milling process using hybrid Taguchi-genetic learning algorithm*, Expert Systems with Applications, vol. 36, no. 2, (2009), pp. 3216–3222.
- [33] F. W. Taylor, *The art of cutting metals*, Scientific American, vol. 63, no. 1618supp, (1907), pp. 25929–25931.
- [34] R. E. Haber, J. E. Jiménez, C. Peres, J. R. Alique, *An investigation of tool-wear monitoring in a high-speed machining process* (2004).
- [35] S. Y. Liang, R. L. Hecker, R. G. Landers, *Machining process monitoring and control: the state-of-the-art*, Journal of Manufacturing Science and Engineering, vol. 126, (2004), p. 297.
- [36] S. Dey, J. A. Stori, *A Bayesian network approach to root cause diagnosis of process variations*, International Journal of Machine Tools and Manufacture, vol. 45, (2005), pp. 75–91.
- [37] J. C. Chen, W.-L. Chen, *A tool breakage detection system using an accelerometer sensor*, Journal of Intelligent Manufacturing, vol. 10, (1999), pp. 187–197.
- [38] D. A. Smith, S. Smith, J. Tlusty, *High performance milling torque sensor*, Journal of manufacturing science and engineering, vol. 120, no. 3, (1998), pp. 504–514.
- [39] D.-W. Cho, S. J. Lee, C. N. Chu, *The state of machining process monitoring research in Korea*, International Journal of Machine Tools and Manufacture, vol. 39, (1999), pp. 1697–1715.
- [40] O. B. Abouelatta, J. Madl, *Surface roughness prediction based on cutting parameters and tool vibrations in turning operations*, Journal of materials processing technology, vol. 118, no. 1, (2001), pp. 269–277.
- [41] S. S. Lee, J. C. Chen, *On-line surface roughness recognition system using artificial neural networks system in turning operations*, The International Journal of Advanced Manufacturing Technology, vol. 22, no. 7-8, (2003), pp. 498–509.
- [42] Y. Altintas, A. Ber, *Manufacturing automation: metal cutting mechanics, machine tool vibrations, and CNC design*, vol. 54 (2001).

- [43] Y. M. Ertekin, Y. Kwon, T.-L. B. Tseng, *Identification of common sensory features for the control of CNC milling operations under varying cutting conditions*, International Journal of Machine Tools and Manufacture, vol. 43, no. 9, (2003), pp. 897–904.
- [44] Y.-H. Tsai, J. C. Chen, S.-J. Lou, *An in-process surface recognition system based on neural networks in end milling cutting operations*, International Journal of Machine Tools and Manufacture, vol. 39, no. 4, (1999), pp. 583–605.
- [45] D. E. Dimla Snr., *The correlation of vibration signal features to cutting tool wear in a metal turning operation*, International Journal of Advanced Manufacturing Technology, vol. 19, no. 10, (2002), pp. 705–713.
- [46] B. Bahr, S. Motavalli, T. Arfi, *Sensor fusion for monitoring machine tool conditions*, International Journal of Computer Integrated Manufacturing, vol. 10, no. 5, (1997), pp. 314–323.
- [47] I. Inasaki, *Application of acoustic emission sensor for monitoring machining processes*, Ultrasonics, vol. 36, (1998), pp. 273–281.
- [48] K. Jemielniak, *Some aspects of AE application in tool condition monitoring*, Ultrasonics, vol. 38, no. 1, (2000), pp. 604–608.
- [49] E. Cross, *On structural health monitoring in changing environmental and operational conditions*, Ph.D. thesis, University of Sheffield (2012).
- [50] K. Worden, W. J. Staszewski, J. J. Hensman, *Natural computing for mechanical systems research: A tutorial overview*, Mechanical Systems and Signal Processing, vol. 25, no. 1, (2011), pp. 4–111.
- [51] A. Hase, H. Mishina, M. Wada, *Correlation between features of acoustic emission signals and mechanical wear mechanisms*, Wear, vol. 292-293, (2012), pp. 144–150.
- [52] Kistler, *Acoustic Emission - Piezotron Coupler* (2006).
- [53] S. B. Kotsiantis, *Supervised machine learning: a review of classification techniques*, in *Supervised Machine Learning: A Review of classification Techniques* (2007), pp. 249–268.
- [54] C. M. Bishop, *Pattern Recognition and Machine Learning*, no. 4 in Information science and statistics, Springer (2006).

- [55] S. Wold, K. Esbensen, P. Geladi, *Principal component analysis*, Chemometrics and Intelligent Laboratory Systems, vol. 2, no. 1-3, (1987), pp. 37–52.
- [56] R. G. Silva, R. L. Reuben, K. J. Baker, S. J. Wilcox, *Tool wear monitoring of turning operations by neural network and expert system classification of a feature set generated from multiple sensors*, Mechanical Systems and Signal Processing, vol. 12, no. 2, (1998), pp. 319–332.
- [57] S. Cho, S. Binsaeid, S. Asfour, *Design of multisensor fusion-based tool condition monitoring system in end milling*, The International Journal of Advanced Manufacturing Technology, vol. 46, no. 5-8, (2009), pp. 681–694.
- [58] T. Dietterich, *Machine learning for sequential data: A review*, Structural, syntactic, and statistical pattern recognition, (2002), pp. 1–15.
- [59] L. R. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition* (1989).
- [60] L. Atlas, M. Ostendorf, G. D. Bernard, *Hidden Markov models for monitoring machining tool-wear*, 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100), vol. 6, (2000), pp. 3887–3890.
- [61] T. Boutros, M. Liang, *Detection and diagnosis of bearing and cutting tool faults using hidden Markov models*, Mechanical Systems and Signal Processing, vol. 25, no. 6, (2011), pp. 2102–2124.
- [62] H. M. Ertunc, K. A. Loparo, H. Ocaik, *Tool wear condition monitoring in drilling operations using hidden Markov models (HMMs)*, International Journal of Machine Tools and Manufacture, vol. 41, no. 9, (2001), pp. 1363–1384.
- [63] C. Scheffer, H. Engelbrecht, P. S. Heyns, *A comparative evaluation of neural networks and hidden Markov models for monitoring turning tool wear*, Neural Computing and Applications, vol. 14, no. 4, (2005), pp. 325–336.
- [64] Q. Miao, V. Makis, *Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models*, Mechanical Systems and Signal Processing, vol. 21, no. 2, (2007), pp. 840–855.
- [65] K. Worden, G. Manson, E. J. Cross, *On Gaussian process NARX models and their higher-order frequency response functions*, in *Solving Computationally*

- Expensive Engineering Problems: Methods and Applications*, S. Koziel, L. Leifsson, X.-S. Yang, editors, Springer International Publishing, Cham (2014), pp. 315–335.
- [66] J. Yan, M. Koç, J. Lee, *A prognostic algorithm for machine performance assessment and its application*, Production Planning & Control, vol. 15, no. 8, (2004), pp. 796–801.
- [67] J.-H. Zhou, C. K. Pang, F. Lewis, Z.-W. Zhong, *Intelligent diagnosis and prognosis of tool wear using dominant feature identification*, IEEE Transactions on Industrial Informatics, vol. 5, no. 4, (2009), pp. 454–464.
- [68] S. J. Engel, B. J. Gilmartin, K. Bongort, A. Hess, *Prognostics, the real issues involved with predicting life remaining*, IEEE Aerospace Conference Proceedings, vol. 6, (2000), pp. 457–470.
- [69] Kistler, *Piezotron Sensor datasheet* (2007).
- [70] Y. Altintas, E. Budak, *Analytical prediction of stability lobes in milling*, CIRP Annals - Manufacturing Technology, vol. 44, no. 1, (1995), pp. 357–362.
- [71] B. LLC, *METAL- MAX datasheet* (2010).
- [72] Alicona, *IF-EdgeMasterModule* (2014).
- [73] S. Jaspers, J. Dautzenberg, *Material behaviour in metal cutting: strains, strain rates and temperatures in chip formation*, Journal of Materials Processing Technology, vol. 121, no. 1, (2002), pp. 123–135.
- [74] E. S. Gadelmawla, M. M. Koura, T. M. a. Maksoud, I. M. Elewa, H. H. Soliman, *Roughness parameters*, Journal of Materials Processing Technology, vol. 123, no. 1, (2002), pp. 133–145.
- [75] Y. Quinsat, S. Lavernhe, C. Lartigue, *Characterization of 3D surface topography in 5-axis milling*, Wear, vol. 271, no. 3-4, (2011), pp. 590–595.
- [76] *Appendix B, Statistical tables*, Camden Old Series, vol. 81, (1862), p. 452.
- [77] M. Bohanec, I. Bratko, *Trading accuracy for simplicity in decision trees*, Machine Learning, vol. 15, no. 3, (1994), pp. 223–250.
- [78] T. Cover, P. Hart, *Nearest neighbor pattern classification*, IEEE Transactions on Information Theory, vol. 13, no. 1, (1967), pp. 21–27.

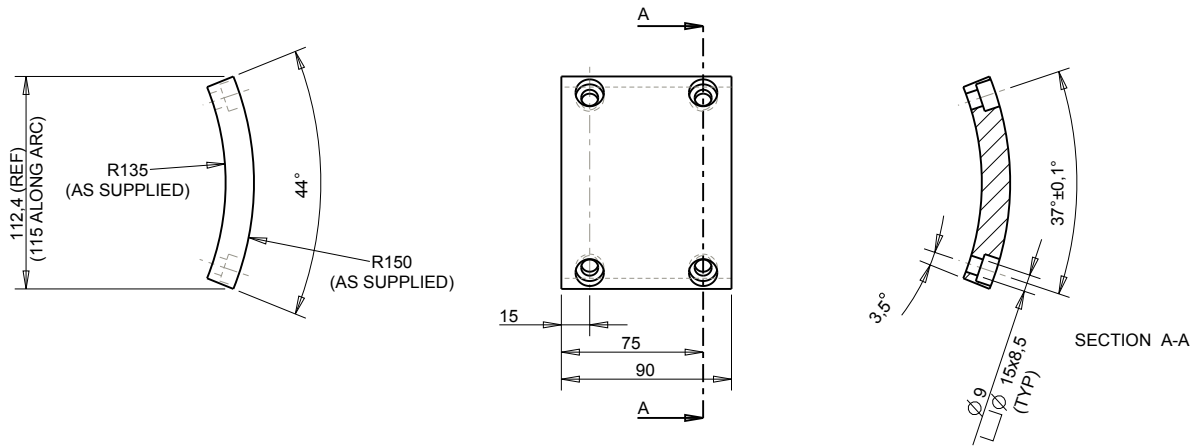
- [79] J. Sun, M. Rahman, Y. S. Wong, G. S. Hong, *Multiclassification of tool wear with support vector machine by manufacturing loss consideration*, International Journal of Machine Tools and Manufacture, vol. 44, no. 11, (2004), pp. 1179–1187.
- [80] S. Tong, D. Koller, *Active learning for parameter estimation in Bayesian networks*, Nips, (2000), pp. 647–653.
- [81] L. Atlas, M. Ostendorf, G. D. Bernard, *Hidden Markov models for monitoring machining tool-wear*, 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100), vol. 6, (2000), pp. 3887–3890.
- [82] I. Leontaritis, S. Billings, *Input-output parametric models for nonlinear systems, part {I}: deterministic nonlinear systems*, International Journal of Control, vol. 41, (1985), pp. 303–328.
- [83] I. Leontaritis, S. Billings, *Input-output parametric models for nonlinear systems, part {II}: stochastic nonlinear systems*, International Journal of Control, vol. 41, (1985), pp. 329–344.
- [84] K. Worden, G. Manson, E. J. Cross, *On Gaussian process NARX models and their higher-order frequency response functions*, Springer Proceedings in Mathematics and Statistics, vol. 97, (2014), pp. 315–335.
- [85] S. Billings, *Nonlinear system identification: NARMAX, methods in the time, frequency, and spatio-temporal domains*, Wiley-Blackwell (2013).
- [86] N. Chiras, C. Evans, D. Rees, *Nonlinear gas turbine modeling using NARMAX structures*, in *IEEE Transactions on Instrumentation and Measurement*, vol. 50, pp. 893–898.
- [87] S. K. Pradhan, B. Subudhi, *NARMAX modeling of a two-link flexible robot*, in *Proceedings - 2011 Annual IEEE India Conference: Engineering Sustainable Solutions, INDICON-2011*.
- [88] M. A. Balikhin, R. J. Boynton, S. N. Walker, J. E. Borovsky, S. A. Billings, H. L. Wei, *Using the NARMAX approach to model the evolution of energetic electrons fluxes at geostationary orbit*, Geophysical Research Letters, vol. 38, no. 18.




- [89] S. Billings, H. Jamaluddin, S. Chen, *Properties of neural networks with applications to modelling non-linear dynamical systems*, International Journal of Control, vol. 55, (1992), pp. 193–1350.
- [90] S. Chen, S. Billings, C. Cowan, P. Grant, *Practical identification of NARMAX models using radial basis functions*, International Journal of Control, vol. 52, (1990), pp. 1327–1350.
- [91] R. Murray-Smith, T. Johansen, R. Shorten, *On transient dynamics, off-equilibrium behaviour and identification in blended multiple model structures*, in *Proceedings of European Control Conference*, pp. BA—14.
- [92] J. Kocijan, *Dynamic GP models: an overview and recent developments*, in *Proceedings of 6th International Conference on Applied Mathematics, Simulation and Modelling*, pp. 38–43.
- [93] C. Rasmussen, C. Williams, *Gaussian processes for machine learning*, The MIT Press (2006).
- [94] T. Rogers, G. Manson, K. Worden, E. J. Cross, *On the choice of optimisation scheme for Gaussian process hyperparameters in SHM problems*, in *11th International Workshop on Structural Health Monitoring*, Stanford University, Palo Alto, CA.
- [95] W. Chu, Z. Ghahramani, *Gaussian processes for ordinal regression*, Journal of Machine Learning Research, vol. 6, no. 2001, (2005), pp. 1019–1041.
- [96] D. J. C. MacKay, *Bayesian nonlinear modeling for the prediction competition*, in *ASHRAE Transactions*, vol. 100, pp. 1053–1062.
- [97] J. Sun, B. Feng, W. Xu, *Particle swarm optimization with particles having quantum behavior*, in *Congress on Evolutionary Computation, 2004. CEC2004*, vol. 1, pp. 325—331 Vol.1.
- [98] C. E. Rasmussen, *Gaussian processes for machine learning*, International journal of neural systems, vol. 14, no. 2, (2006), pp. 69–106.
- [99] Anonymous, *MUSP (pc) presentation*, Tech. rep., Marposs Ltd. (2011).
- [100] Anonymous, *Model 355B02 datasheet*, Tech. rep., PCB Piezotronics Inc.

Appendix A

COUPON TECHNICAL DRAWING

NOTES
 ENSURE 4 HOLES ALIGN CORRECTLY WITH MOUNTING HOLES ON FIXTURE (DRAWING NO: AS644-DR-A-01)
 TO BE CUT FROM MESSIER-BUGATTI-DOWTY LANDING GEAR TRUCK BEAM SECTION (TI-5-5-5-3)



DATE OF ISSUE 15/07/2013	TITLE WORKPIECE	 Advanced Manufacturing Research Centre 		
DOC. TYPE PART	DRAWN BY T WOOD			
<small>UNLESS OTHERWISE SPECIFIED ALL DIMENSIONS ARE IN mm GENERAL TOLERANCES:</small> DIMENSION TOLERANCE X ±0.1 XX ±0.15 XXX ±0.01 XXXX ±0.005 ANGLES ±0.25° SURFACE FINISH µm DEBURR & BREAK ALL SHARP EDGES THIRD ANGLE PROJECTION	CHECKED BY <i>[Signature]</i>	DEPT. DESIGN AND PROTOTYPING CENTRE		
 <small>IF IN DOUBT, ASK. DO NOT SCALE</small>	APPROVED BY <i>[Signature]</i>	PART NO.		
SIZE A3	SHEET 1 / 1	SCALE 1:2	DRAWING NUMBER AS644-DR-A-03	REV. 0
<small>THIS DRAWING IS THE PROPERTY OF THE AMRC AND CANNOT BE REPRODUCED OR COMMUNICATED WITHOUT WRITTEN CONSENT</small>				

KISTLER 8152B1 DATASHEET

[69]

Piezotron® Sensor

Type 8152B...

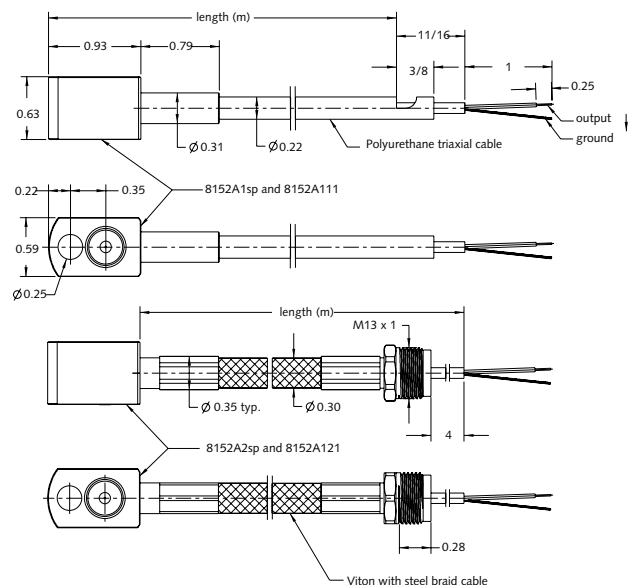
Acoustic Emission Sensor

Piezotron Acoustic Emission Sensor with an integral impedance converter for measuring acoustic emission (AE) above 50 kHz in machine structures. With its small size it mounts easily near the source of emission to optimally capture the signal. The sensor has a very rugged welded housing (degree of protection IP 65 PUR or IP 67 Viton). The small sensor is easily mounted nearby everywhere; an M6 or 1/4-28 bolt is all that is needed.

- High sensitivity and wide frequency range
- Inherent highpass-characteristic
- Insensitive to electric and magnetic noise fields
- Robust, for industrial use (IP65 (PUR), IP67 (Viton))
- Ground isolated: Prevents ground loops
- Conforming to CE

Description

The Piezotron AE Sensor consists of the sensor housing, the piezoelectric sensing element and the built-in impedance converter. The sensing element, made of piezoelectric ceramic, is mounted on a thin steel diaphragm. Its construction determines the sensitivity and frequency response of the sensor. The coupling surface of the diaphragm welded into the housing is slightly protruding to measure the AE signals. Thus a precisely defined coupling force results when mounting. This assures a constant and reproducible coupling for the AE transmission. The sensing element is acoustically isolated from the housing by design and therefore well protected against external noise. The Kistler AE sensors feature a very high sensitivity for surface (Rayleigh) and longitudinal waves over a broad frequency range. Type 8152B1... covers 50 ... 400kHz and Type 8152B2... covers 100 ... 900 kHz. A miniature impedance converter is built into the Piezotron AE Sensor, giving an output low-impedance voltage signal. The AE Piezotron Coupler Type 5125B1, is used to supply power to the sensor and for signal processing. Special highly insulating and low noise connecting cables are not required.



8152B_000-204 a-01.07

Technical Data

Type	Unit	8152B111/121	8152B11/12sp	8152B211/221	8152B21/22sp
Sensitivity	dBref 1V/(m/s)	57	57	48	48
Frequency Range ± 10 dB	kHz	50 ... 400	50 ... 400	100 ... 900	100 ... 900
Ground Isolation	M Ω	>1	>1	>1	>1
Environmental:					
Shock Limit (0.5ms pulse)	gpk	2000	2000	2000	2000
Temperature Range Operating	$^{\circ}$ F	-40 ... 140	-40 ... 140	-40 ... 140	-40 ... 140
Output:					
Bias nom.	VDC	2.2	2.2	2.5	2.5
Impedance	Ω	<10	<10	<10	<10
Voltage full scale	V	± 2	± 2	± 2	± 2
Current	mA	2	2	4	4
Source:					
Voltage (Coupler)	VDC	5 ... 36	5 ... 36	5 ... 36	5 ... 36
Constant Current	mA	3 ... 6	3 ... 6	3 ... 6	3 ... 6
Construction:					
Sensing Element	type	ceramic	ceramic	ceramic	ceramic
Housing/Base	material	stainless steel	stainless steel	stainless steel	stainless steel
Sealing-housing/connector	type	hermetic	hermetic	hermetic	hermetic
Viton Cable Bend Radius, max.	in	0.6	0.6	0.6	0.6
Weight (without cable)	grams	29	29	29	29
Mounting Torque	lbf-in	80 \pm 10	80 \pm 10	80 \pm 10	80 \pm 10

1 g = 9.80665 m/s², 1 inch = 25.4 mm, 1 gram = 0.03527 oz, 1 lbf-in = 0.1129 Nm

Application

The AE Sensor is especially well suited for measuring AE above 50 kHz in the surface of metallic components or structures. Such AE results from plastic deformation of materials, crack formation and growth, fracturing or friction. Application examples are monitoring of processes, tools and machines in metal cutting as well as forming operations. Thanks to its rugged construction and the tightly welded housing this sensor can operate under severe environmental conditions.

Accessories Included

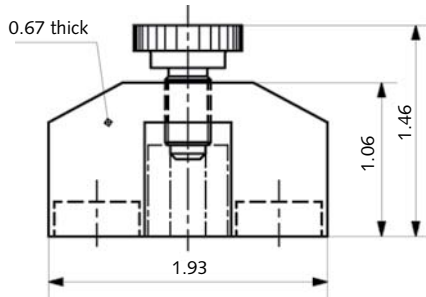
- | | Type |
|--------------------------------|--------------|
| • Mounting screw, 1/4-28 x 1in | 431-0500-001 |
| • Mounting screw, M6 x 25mm | 431-0497-001 |

Optional Accessories

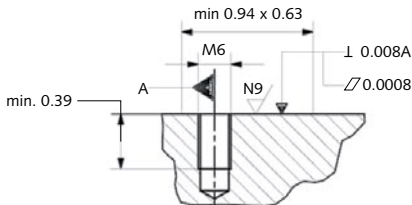
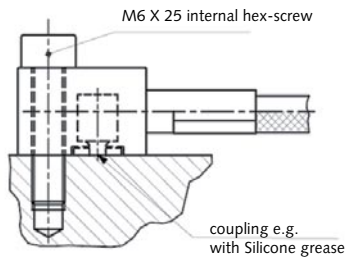
- | | Type |
|------------------------|-------|
| • Magnetic clamp | 8443B |
| • Piezotron AE coupler | 5125B |

Mounting

The AE Sensor is simply mounted with an M6-1/4 screw or a magnetic clamp Type 8443B onto the surface of the structure. A minimum tightening torque is sufficient for a reproducible and constant coupling. The smoother the mounting surface, the better the result. The use of a highly viscous grease (e.g. silicone grease) between the coupling surfaces is recommended.



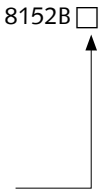
Type 8443B Magnetic clamp



Mounting AE-Sensor

Ordering Key

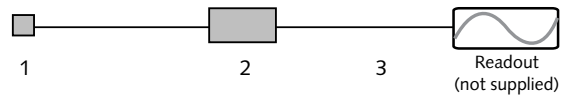
Material	Length	Frequency Range	Ordering Code
PUR	5m	(50 ... 400kHz)	111
PUR	0.3 ... 10m	(50 ... 400kHz)	11sp
PUR	5m	(100 ... 900kHz)	211
PUR	0.3 ... 10m	(100 ... 900kHz)	21sp
Viton	2m	(50 ... 400kHz)	121
Viton	0.3 ... 3m	(50 ... 400kHz)	12sp
Viton	2m	(100 ... 900kHz)	221
Viton	0.3 ... 3m	(100 ... 900kHz)	22sp



Measuring Chain

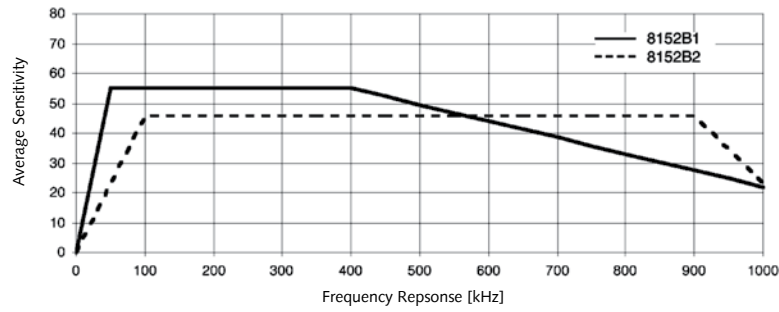
- 1 Acoustic emission sensor
- 2 Piezotron AE Coupler
- 3 Outout cable, made by customer

Type
8152B...
5125B...

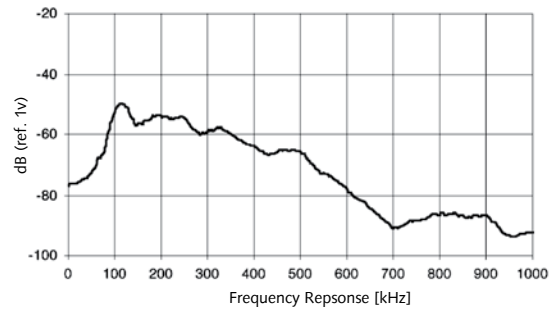
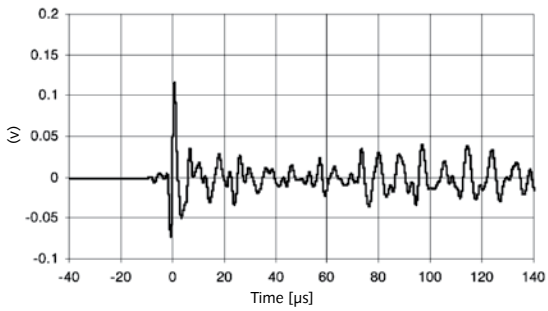


8152B_000-204 a-01.07

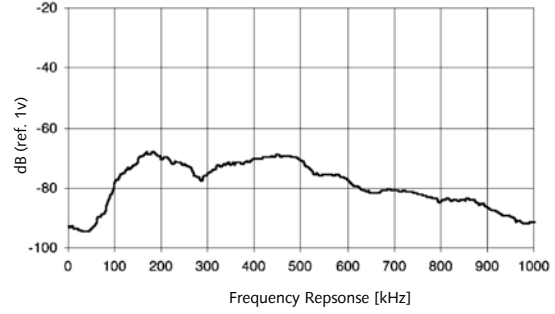
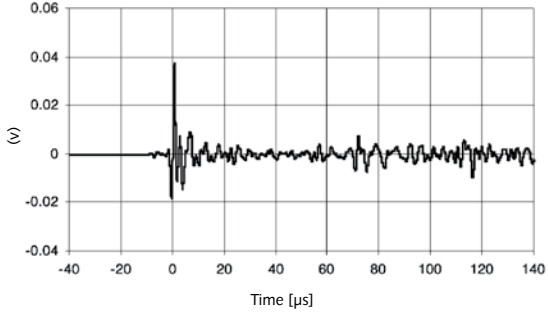
Frequency Response



Type 8152B1



Type 8152B2



8152B_000-204 a-01.07

KISTLER 9255B DATASHEET

[52]

Multicomponent Dynamometer

-20 ... 40 kN, Mounting Plate 260x260 mm

Type 9255B

Quartz 3-component dynamometer for measuring the three orthogonal components of a force. The dynamometer has a great rigidity and consequently a high natural frequency. Its high resolution enables the smallest dynamic changes in large forces to be measured.

- Wide measuring range
- For heavy duty application
- Compact design

Description

The dynamometer consists of four 3-component force sensors fitted under high preload between a baseplate and a top plate. Each sensor contains three pairs of quartz plates, one sensitive to pressure in the z direction and the other two responding to shear in the x and y directions respectively. The force components are measured practically without displacement.

The outputs of the four built-in force sensors are connected inside the dynamometer in a way to allow multicomponent measurements of forces and moments to be performed. The eight output signals are available at the 9-conductor flange socket. The four sensors are mounted ground-insulated. Therefore ground loop problems are largely eliminated.

The dynamometer is rustproof and protected against penetration of splashwater and cooling agents. Together with the connecting cable Type 1687B5/1689B5 and Type 1677A5/1679A5 it corresponds to the protection class IP67.

Application Examples

- Dynamic and quasistatic measurement of the three orthogonal components of a force.
- Cutting force measurements while milling and grinding on larger machines and in machining centers.
- Measurements on stamping machines.
- Measurements on wind tunnel models.
- Measurements of supporting forces at machinery foundations.
- Measurements on rocket propulsion units.



Technical Data

Range	F_x, F_y	kN	-20 ... 20 ¹⁾
	F_z	kN	-10 ... 40 ¹⁾
Calibrated partial range	F_x, F_y	kN	0 ... 2
	F_z	kN	0 ... 4
Overload	F_x, F_y	kN	-24/24
	F_z	kN	-12/48
Threshold		N	<0,01
Sensitivity	F_x, F_y	pC/N	≈-8
	F_z	pC/N	≈-3,7
Linearity, all ranges		%FSO	≤±1
Hysteresis, all ranges		%FSO	≤0,5
Cross talk		%	≤±2
Rigidity	c_x, c_y	kN/μm	>2
	c_z	kN/μm	>3
Natural frequency	$f_n (x, y, z)$	kHz	≈3
Natural frequency (mounted on flanges)	$f_n (x, y)$	kHz	≈1,7
	$f_n (z)$	kHz	≈2
Natural frequency (mounted on flanges and through top plate)	$f_n (x, y)$	kHz	≈2
	$f_n (z)$	kHz	≈3,3
Operating temperature range		°C	0 ... 70
Temperature coefficient of sensitivity		%/°C	-0,02
Capacitance (of channel)		pF	≈500
Insulation resistance (20 °C)		Ω	>10 ¹³
Ground insulation		Ω	>10 ⁸
Protection class EN60529		-	IP67 ²⁾
Weight		kg	52

¹⁾ Application of force inside and max. 100 mm above top plate area.

²⁾ With connecting cable Types 1687B5, 1689B5, 1677A5, 1679A5

1 N (Newton) = 1 kg · m · s⁻² = 0,1019... kp = 0,2248... lbf; 1 inch = 25,4 mm; 1 kg = 2,2046... lb; 1 N·m = 0,73756... lbft

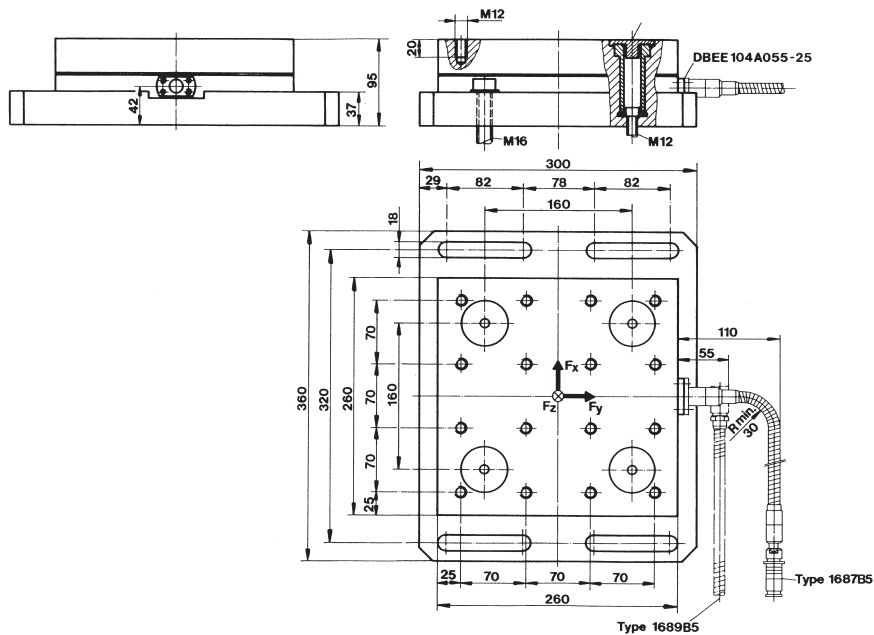


Fig. 1: Dimensions of dynamometer Type 9255B

Mounting

The dynamometer may be mounted with screws or claws on any clean, face-ground supporting surface, such as the table of a machine tool for example.

In order to provide a still better coupling of the measuring instrument with the mounting surface, the dynamometer can, if necessary, additionally be screwed down through the four bores in the top plate. This measure allows to reach a higher resonant frequency of the measuring system. Uneven supporting surface may set up internal stresses, which will impose severe additional loads on the individual measuring elements and may also increase cross talk.

For mounting the force-introducing components, mainly workpieces, sixteen M12 mm blind tap holes in the cover plate are available.

The supporting surfaces for the force-introducing parts must be face-ground to obtain good mechanical coupling to the cover plate.

Signal Conditioning

A multichannel charge amplifier is also needed to build a complete measuring system (i.e. Type 5070A...). The measurement signal is converted into an electrical voltage in the individual channels. The measured value is exactly proportional to the force acting.

Optional Accessories

For 3-Component Force Measurements

F_x, F_y, F_z	Type
• Connecting cable, length $l = 5$ m (3 leads)	1687B5
• Extension cable, length $l = 5$ m (3 leads)	1689B5
	1688B5

For 6-Component Force and Moment Measurements

$F_x, F_y, F_z / M_x, M_y, M_z$	Type
• Connecting cable, length $l = 5$ m (8 leads)	1677A5
• Extension cable, length $l = 5$ m (8 leads)	1679A5
	1678A5

Ordering Code

• Multicomponent Dynamometer	Type 9255B
------------------------------	------------

9255B_000-148e-12.09

PCB 355B02 DATASHEET

[100]

Model Number 355B02	ICP® ACCELEROMETER	Revision: H ECN #: 29040
-------------------------------	---------------------------	-----------------------------

	ENGLISH	SI
Performance		
Sensitivity(± 10 %)	10 mV/g	1.02 mV/(m/s ²)
Measurement Range	± 500 g pk	± 4900 m/s ² pk
Frequency Range(± 5 %)	1 to 10,000 Hz	1 to 10,000 Hz
Frequency Range(± 10 %)	0.6 to 12,000 Hz	0.6 to 12,000 Hz
Frequency Range(± 3 dB)	0.3 to 17,000 Hz	0.3 to 17,000 Hz
Resonant Frequency	≥ 35 kHz	≥ 35 kHz
Broadband Resolution(1 to 10,000 Hz)	0.0005 g rms	0.005 m/s ² rms
Non-Linearity	≤ 1 %	≤ 1 %
Transverse Sensitivity	≤ 5 %	≤ 5 %
Environmental		
Overload Limit	± 5000 g pk	± 49,000 m/s ² pk
Temperature Range	-65 to +250 °F	-54 to +121 °C
Temperature Response	See Graph	See Graph
Base Strain Sensitivity	0.05 g/με	0.49 (m/s ²)/με
Electrical		
Excitation Voltage	22 to 30 VDC	22 to 30 VDC
Constant Current Excitation	2 to 20 mA	2 to 20 mA
Output Impedance	≤ 100 ohm	≤ 100 ohm
Output Bias Voltage	7 to 15 VDC	7 to 15 VDC
Discharge Time Constant	0.5 to 2 sec	0.5 to 2 sec
Settling Time(within 10% of bias)	<10 sec	<10 sec
Spectral Noise(1 Hz)	90 μg/√Hz	880 (μm/s ²)/√Hz
Spectral Noise(10 Hz)	15 μg/√Hz	147 (μm/s ²)/√Hz
Spectral Noise(100 Hz)	5 μg/√Hz	49 (μm/s ²)/√Hz
Spectral Noise(1 kHz)	3 μg/√Hz	29 (μm/s ²)/√Hz
Electrical Isolation(Base)	>10 ⁸ ohm	>10 ⁸ ohm
Physical		
Sensing Element	Ceramic	Ceramic
Sensing Geometry	Shear	Shear
Housing Material	Titanium	Titanium
Sealing	Hermetic	Hermetic
Size (Height x Length x Width)	0.40 in x 0.95 in x 0.63 in	10.2 mm x 24.1 mm x 16.0 mm
Weight	0.35 oz	10 gm
Electrical Connector	10-32 Coaxial Jack	10-32 Coaxial Jack
Electrical Connection Position	Side	Side
Mounting	Through Hole	Through Hole

OPTIONAL VERSIONS
Optional versions have identical specifications and accessories as listed for the standard model except where noted below. More than one option may be used.

A - Adhesive Mount
Supplied Accessory : Model 080A90 Quick Bonding Gel (1)

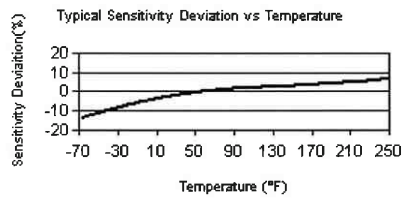
M - Metric Mount
Supplied Accessory : Model M039A22 Allen wrench for use with M3 thread (1) replaces Model 039A22
Supplied Accessory : Model M081B45 off ground screw assembly (1) replaces Model 081B45

W - Water Resistant Cable

Electrical Connector	Sealed Integral Cable	Sealed Integral Cable
Electrical Connection Position	Side	Side

NOTES:
[1] Typical.
[2] Zero-based, least-squares, straight line method.
[3] Transverse sensitivity is typically ≤ 3%.
[4] See PCB Declaration of Conformance PS023 for details.

SUPPLIED ACCESSORIES:
Model 039A22 Allen wrench, 7/64" hex (1)
Model 080A109 Petro Wax (1)
Model 081B45 off ground screw assembly (1)
Model ACS-1 NIST traceable frequency response (10 Hz to upper 5% point). (1)



Entered: BLS	Engineer: JF	Sales: WDC	Approved: BM	Spec Number:
Date: 7/8/08	Date: 7/1/08	Date: 7/1/08	Date: 7/2/08	10534

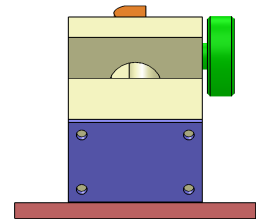
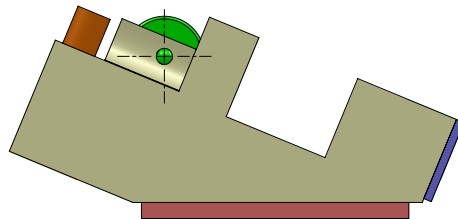
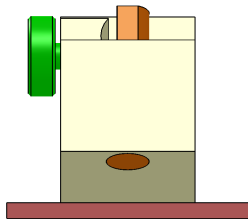
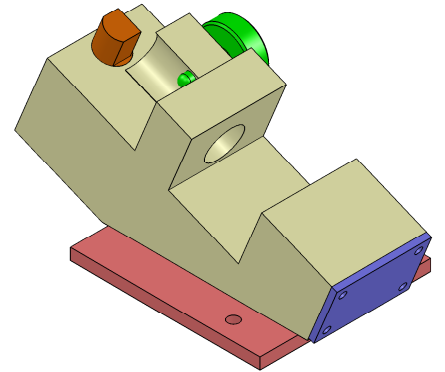
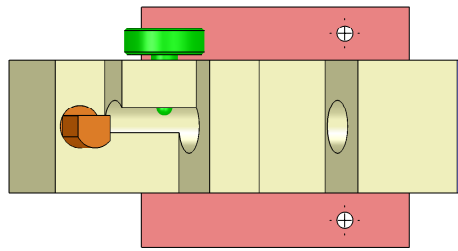
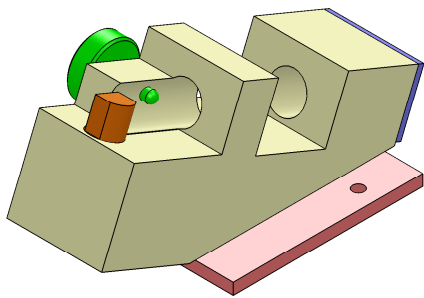
All specifications are at room temperature unless otherwise specified.
In the interest of constant product improvement, we reserve the right to change specifications without notice.
ICP® is a registered trademark of PCB Group, Inc.

PCB PIEZOTRONICS™
VIBRATION DIVISION
3425 Walden Avenue, Depew, NY 14043

Phone: 716-684-0001
Fax: 716-685-3886
E-Mail: vibration@pcb.com

Appendix E

TOOL INSPECTION FIXTURE



UNLESS OTHERWISE SPECIFIED: DIMENSIONS ARE IN MILLIMETERS TOLERANCES UNLESS OTHERWISE SPECIFIED: FRACTIONS DECIMALS ANGLES		FINISH: BLACK	DATE: 20/10/2010	DO NOT SCALE DRAWING	REVISION
DRAWN: SRAJ		SIGNATURE:	DATE:	Raysun Innovative Design Ltd.	
CHECKED: []				TITLE: Cutter inspection Fixture layout	
DATE:				DWG NO.:	
SCALE:				A2	
MATERIAL:				WEIGHT: 2942.983grams	
SCALE: 1:1				SHEET 1 OF 1	