DECISION MAKING UNDER UNCERTAINTY IN TOOL CONDITION MONITORING

K. Kelly, G. E. O’Donnell

School Of Engineering, Trinity College Dublin, Dublin 2, Ireland.

ABSTRACT

Tool Condition Monitoring is a field which sees a significant divergence between the published research and current industrial practice. It has been proposed that the principal reason for this divergence is the inability of laboratory-developed systems to display sufficient robustness in coping with the diverse range of stochastic influences that are found in industrial deployment – ranging from operator influence to variations in the properties of tools and workpieces used. Furthermore, several studies have shown that a majority of tool condition monitoring systems purchased by industrial concerns are switched off. Principal among the reasons given is the number of ‘false alarms’ resulting in costly system downtime.

This work considers the current trends in tool condition monitoring research, in parallel with the requirements of industry, and argues that probabilistic rather than binary decision making systems are required in the next generation of tool condition monitoring systems, if the poor industrial acceptance of the previous generation systems is to be countered. Examples are considered from industry and used to illustrate a suggested approach to both generating and assessing such predictions. These examples explore the range of decision making criteria under which such systems are required to operate – from minimization of downtime to the protection of costly capital equipment or customer resources.

KEYWORDS: Tool Condition Monitoring, Machining, Uncertainty, Probabilistic Prediction

1. INTRODUCTION

Manufacturing operations have evolved from largely labour-intensive endeavours to highly automated processes – a process evolution that has seen the traditional role of the operator evolve to a more supervisory capacity. Furthermore, modern machine tools are often encapsulated for environmental and operator safety reasons, thereby, physically removing the operator from the process.

The benefits to this increased automation can be summarised as:

(i) Improvement of part quality
(ii) Reduction of scrap
(iii) Reduction of labour cost, i.e. the number of operators required for the plant is typically fewer as a single operator can supervise more than one automatic machine tool
(iv) Reduction in machining time, e.g. multiple axes operating simultaneously on multiple workpieces and the use of high speed cutting
(v) Reduction in non-productive time, e.g. more complex tooling can be used, more rapid tool change can be effective
(vi) Reduction of operator error due to fatigue etc
Increased predictability with regard to process capability and output, with knock-on benefits for process planning and manufacturing systems strategies, e.g. JIT.

However, most of these productivity increases are limited by the requirement that such automated operations have a high level of reliability. Several strategies are available to manufacturers to reduce the possibility of such failures. The two main options, apart from returning to manual observation of each operation (not feasible in most modern systems), are to run the machine tool below its maximum capabilities, or to introduce some form of feedback (representing the quality of the process to the system). A more specific case of this monitoring of the process is where attention is focussed on the state of the cutting tool. These are termed 'Tool Condition Monitoring' (TCM) systems.

Broadly speaking, there are two possible methods of determining tool condition. In the direct method, tool condition is measured directly in-situ. Measuring devices based on inductance, capacitance, vision, radiation, or pneumatics can be used to measure the level of wear [1]. While direct methods tend to be accurate, they are more complex and often not applicable to real machining environments. They are also usually only implementable between operations meaning that breakage or excessive wear can only be detected post process [1]. Indirect methods involve the measurement of some phenomenon related to tool wear or breakage. Commonly, the measurement of cutting force, torque, temperature, vibration, spindle motor power, feed motor power, and strain are used to indirectly indicate the level of tool wear. As the measured phenomenon often varies with process conditions it can be difficult to find a good correlation between tool wear and the sensor signal over the full range of operating conditions. Indirect methods do, however, offer the advantages that they: can continuously monitor the process, are less complex than direct methods, and are applicable in industrial environments. For these reasons, the majority of the research carried out in TCM has concentrated on indirect methods.

2. STATE OF THE ART AND CURRENT PRACTICE IN TOOL CONDITION MONITORING

It has been estimated that up to 20% of machine tool down-time in production is due to cutting tool failure [2]. This has provided considerable motivation to researchers and industrial practitioners to develop means of speedily recognising difficulties with the process and/or tool condition. Further motivation is provided by the negative impact that worn tools have in the context of dimensions, finish, and surface integrity.

Monitoring systems can be considered [3] to have a generic structure which can be described as consisting of three, usually distinct, components:
- Sensors
- Signal processing and feature extraction
- Decision making and control

A wide range of sensor types have been investigated in the academic research community [1, 4, 5, 6, 7, 8, 9], including torque, audible noise, acoustic emission, eddy currents and magnetic impedance. Numerous ‘classical’ (e.g. using frequency analysis or statistical parameter extraction) signal processing strategies have been applied in the research [1, 2, 3, 6, 10, 11] as well as more ‘modern’ methods such as adaptive filtering, wavelet analysis and linear discriminant analysis [2, 3, 6, 12, 13].

In the area of decision making and control a large body of work exists on the use of fuzzy logic and neural networks [2, 13, 14, 15, 16, 17].
However, when one considers the industrial deployment of such systems it is striking that very little of the academic research has migrated successfully. A range of companies have been selling tool condition monitoring systems since the early 1990’s. During this time there has been progress on the integration with the machine controller and the factory networks [5, 11] but in terms of the three ‘components’ listed above, the majority of current systems are markedly similar to those available 15 years ago. [1, 2, 3, 18].

The range of sensors employed in commercial monitoring systems has been summarised by various researchers [5, 18].

<table>
<thead>
<tr>
<th>Physical quantity (sensor type)</th>
<th>Suppliers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Manuf. 1</td>
</tr>
<tr>
<td>Force related quantities</td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td></td>
</tr>
<tr>
<td>Torque</td>
<td></td>
</tr>
<tr>
<td>Strain</td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td></td>
</tr>
<tr>
<td>1-3 axis force sensor</td>
<td></td>
</tr>
<tr>
<td>Measuring plate</td>
<td></td>
</tr>
<tr>
<td>AE</td>
<td></td>
</tr>
<tr>
<td>Acoustic emission</td>
<td></td>
</tr>
<tr>
<td>AE fluid sensor</td>
<td></td>
</tr>
<tr>
<td>Rotating AE sensor</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
</tr>
<tr>
<td>Vibration/ultra sound</td>
<td></td>
</tr>
<tr>
<td>Camera</td>
<td></td>
</tr>
<tr>
<td>Laser</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Sensors used in Commercial Monitoring Systems, after [18]

Of these sensor types the use of spindle power sensors is most prevalent, probably due to the relatively low cost, ease of retro-fitting/integrating and the relatively intuitive relationship between the signal obtained and the machining process.

Monitoring strategies available from commercial TCM system manufacturers can be grouped into six approximate categories [1, 5, 18], the first five being used for breakage/collision/missing tool detection and the sixth for wear detection.

(i) static limits - where an ‘alarm’ is generated when the signal exceeds (or drops below) a given amplitude
(ii) time defined limits – as above, but where different limits are used at different stages of the process
(iii) part-signature – where a ‘teach-in’ operation is used to generate a reference signal. This reference signal is often then scaled to give and upper and lower ‘envelope’ on the process signal
(iv) dynamic limits - where the rate of change of the sensor signal is used in conjunction with one of the above strategies
(v) pattern recognition – typically where a simple series of rules are used to identify patterns in the sensor signal (e.g. a large spike followed a big drop in level) characteristic of a breakage or other catastrophic event for that process
(vi) wear – where typically an extracted parameter (e.g. area under the spindle motor power curve) are monitored over time – an exceeding of the pre-set limit indicating that the tool is worn and should be changed

It might naively be imagined that the slow migration of methods, strategies and technologies from the tool condition monitoring research to industry indicates a cautious attitude
on behalf of both the end-users and the monitoring system manufacturers. However, many researchers have found that the satisfaction rating of end-users with commercial monitoring systems is low [1, 2, 3, 5, 19, 20]. Indeed Byrne et al. [1] report that a large percentage of monitoring systems are switched off, while Ketteler [19] reports satisfaction ratings of less than 30% for machine tool manufacturers and 40% for end-users with currently available monitoring systems. The primary reasons for dissatisfaction were the occurrence of false alarms, and the occurrence of missed alarms where the monitoring system would fail to detect a tool breakage. It was also noted that the operation of the systems was deemed to be too complex, strategies were deemed to be too simple, and sensor performance was not adequate. Some of these points echoed the review carried out by Byrne et al [1] which would indicate that there has not been much progress in the advancement of TCM and process monitoring as a technology. Other researchers [3, 5, 15, 16] have speculated on the reasons for the poor performance and concluded that the academic research has traditionally been conducted in a ‘clean’ environment where the range of stochastic influences (e.g. operator error, preceding operations, workpiece material fluctuations, tool supplier changes etc) is greatly lessened. The strategies developed are often not robust enough (either in their fundamental design, or in the way in which they are actually deployed at the end-user site) to cope with these influences. In particular, some extensive testing at end-user facilities demonstrated that a large percentage of false alarms in that particular operation could have been obviated with better set-up, training and use of the monitoring system [21].

3. CURRENT TRENDS IN MANUFACTURING

The field of manufacturing technology continues to develop apace, as the drive to bring down costs, lead-times and to increase efficiencies continues. This development is encapsulated in the goal of ‘High Performance Cutting’ [22]. A non-exhaustive list of current trends and some of their implications for tool condition monitoring is given below:

- Higher speed machining – more stress on cutting tools, reaction time must be quicker, downtime is more expensive, machines are typically more expensive and therefore require better protection
- Advances in cutting tool technology (enabling higher speeds, facilitating use of ‘difficult to machine’ materials, e.g. Inconel, and dry machining) – tools are more expensive and require better protection, damage to workpieces (e.g. aerospace components) may incur huge costs
- Eco-friendly machining (e.g. dry machining, MQL) – the absence/reduction of cooling leads to temperature increases, increased adhesion between tool and workpiece, chip clogging – all of which can influence process measurands (e.g. cutting force, acoustic emission) in unpredictable ways.
- Minimisation of non-productive time – use of combination tooling (more complex process signals), reduction in off-line monitoring (greater dependence upon TCM systems), automation (partial or total absence of operator means that ‘false alarm’ stoppages must be avoided

The implications of these developments for TCM are many and varied, but principal among them is the need for more robust systems which are capable of dealing with a greater (than heretofore) range of operating conditions, and often in the partial or total absence of an operator. It is imperative in such a scenario that production managers can have confidence in TCM systems, as a failure (which is the status quo as noted above!) to have such confidence acts as a bottleneck in the race to increasing productivity and flexibility in the manufacturing environment.
4. **ASSESSMENT OF MONITORING SYSTEMS – A NEW APPROACH**

An analogous area to tool breakage detection is the science of rare event forecasting in meteorology – e.g. prediction of floods, hurricanes, earthquakes etc. Stanski et al [23] give a comprehensive review of many methods of forecast verification. Two categories of verification are of particular interest here:

- Dichotomous forecasting (e.g. “yes, the tool is broken”, or “no, the tool is not broken”)
- Probabilistic forecasting (e.g. there is a 60% chance that the tool is broken, and a 40% chance that it is not)

It should be immediately obvious that it is possible to move from probabilistic to dichotomous forecasting (but not the reverse!) by using a threshold. As noted above, the majority of assessment of monitoring system performance as recorded in the literature corresponds to the case of dichotomous forecasting.

### 4.1 Dichotomous Forecast/Prediction Assessment

As reported in Stanski [23], there are a large variety of categorical statistics which can be used in the assessment of such forecasts. A sample of such methods is given in table 2 below:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Calculation</th>
<th>Details</th>
<th>Range</th>
<th>Perfect Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>[ \frac{\text{hits} + \text{correct negatives}}{\text{total}} ]</td>
<td>What fraction of the forecasts was correct?</td>
<td>0 to 1</td>
<td>1</td>
</tr>
<tr>
<td>Bias</td>
<td>[ \frac{\text{hits} + \text{false alarms}}{\text{hits} + \text{misses}} ]</td>
<td>How did the forecast frequency of ‘yes’ events compare to the observed frequency of ‘yes’ events?</td>
<td>0 to ∞</td>
<td>1</td>
</tr>
<tr>
<td>Probability of detection</td>
<td>[ \frac{\text{hits}}{\text{hits} + \text{misses}} ]</td>
<td>What fraction of the observed ‘yes’ events were correctly forecast?</td>
<td>0 to 1</td>
<td>1</td>
</tr>
<tr>
<td>False Alarm Ratio</td>
<td>[ \frac{\text{false alarms}}{\text{hits} + \text{false alarms}} ]</td>
<td>What fraction of the predicted ‘yes’ events were actually false alarms?</td>
<td>0 to 1</td>
<td>0</td>
</tr>
<tr>
<td>Hanssen and Kuipers’ discriminant (True Skill Score)</td>
<td>[ \frac{\text{hits} - \text{false alarms}}{\text{false alarms} + \text{correct negatives}} ]</td>
<td>How well did the forecast separate the “yes” events from the “no” events? A score of zero indicates ‘no skill’</td>
<td>-1 to 1</td>
<td>1</td>
</tr>
<tr>
<td>Heidke Skill Score</td>
<td>[ \frac{[\frac{1}{2} (\text{Hits} \times \text{Correct Negatives}) + (\text{Misses} \times \text{False Alarms})]}{[\text{Hits} + \text{Misses} \times (\text{Misses} + \text{Correct Negatives}) + \text{Hits} \times \text{False Alarms} + \text{Correct Negatives}]} ]</td>
<td>What was the accuracy of the forecast relative to that of random chance? A score of zero indicates ‘no skill’. More useful than the ‘true skill score’ when dealing with rare events.</td>
<td>-∞ to 1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 Categorical Statistics for Forecast Verification, after [23]
4.2 Probabilistic Forecast/Prediction Assessment

A probabilistic prediction specifies a value between 0 and 1, representing the likelihood of the event in question taking place. In the context of TCM, this would represent the conditional probability that a tool is broken given the nature of the measured signal(s). Probabilistic predictions, with the limiting exceptions of when 0 or 1 is the output, can never be definitively right or wrong – e.g. a forecast of 50% likelihood of rain tomorrow is neither endorsed nor discredited (solely) by the absence of presence of rain tomorrow! Instead a set of probabilistic forecasts, \( p_i \), is verified using observations that those events either occurred \( (o_i = 1) \) or did not occur \( (o_i = 0) \).

The most basic method of probabilistic forecast assessment is probably a comparison with the ‘climatological frequency’, e.g. if cutting tools break during 1.5% of machining operations, then TCM system should, on average, generate a 1.5% probability of failure. A significant (in the standard statistical sense) deviation from this frequency would indicate an over- or under-prediction bias in the system. However, it is possible to have a system which simply predicts a constant failure rate equivalent to the climatological frequency. Such a system is said to have zero resolution. A so-called reliability diagram may be used to graphically assess the performance of a probabilistic prediction, by plotting the observed frequency against the forecast probability. An example is shown below:

![Reliability Diagram](image)

**Figure 1. Reliability Diagram, after [24]**

The diagram illustrates another key concept in probabilistic prediction assessment – the reliability. A system with perfect reliability would have a slope of 1 (the solid diagonal line in the diagram), indicating a perfect match between the observed frequency of occurrence and the predicted probability. The ‘no-skill’ line indicates an average between the climatological average and the perfect reliability lines – anything between this no-skill line and the perfect reliability line indicates that the forecasting system displays some insight or skill.

A formulaic companion to the reliability diagram is the so-called ‘Brier Score’:
\[ BS = \frac{1}{N} \sum_{i=1}^{N} (p_i - o_i) \] (1)

It should be readily apparent however that both the reliability diagram and the Brier score will be highly sensitive to the climatological frequency – i.e. if the rate of tool breakage is quite low, it is possible to have a ‘respectable’ Brier score, but a poorly performing system. Similarly a fundamentally useful system may, due to sampling issues, generate a ‘poor’ reliability diagram. A derived parameter, the Brier Skill Score, is sometimes used in lieu of the above, to account for climatological frequency.

A useful companion to the reliability diagram is the ‘Relative Operating Characteristic’ which considers the impact of decisions made using various (increasing) probability thresholds. In a TCM context this would equate to taking action (usually halting the process) when a breakage is identified with a probability greater than the threshold value. The probability of detection is then plotted against the false alarm rate, as shown in the diagram below:

![Relative Operating Characteristic](image)

Figure 2. ROC Diagram, after [24]

The ROC curve, in a TCM application, displays the ability of the system to discriminate between breakages and ‘normal’ operations. It also provides a powerful reminder of an underlying difficulty with TCM systems – the price to be paid for an increase in detection rate is an increase in false alarms. An ideal system would have an ROC curve shaped like an inverted ‘L’ climbing with infinite slop to a probability of detection of 1 for a false alarm rate of zero. A ‘no-skill’ system is indicated by the dotted line – where there is a direct proportionality between false alarm rate and the probability of detection. The further away the ROC curve is from this skill line (assuming it lies above it of course!) the more useful the system can be considered to be. A measure of the usefulness of the system therefore is the area under the ROC curve.

Some recent work, e.g. Jewson [25] has looked at alternative measures for probabilistic assessment. In Jewson’s paper he proposes the use of a likelihood measure, where the likelihood has its traditional statistical meaning (i.e. the probability or probability density, of the observations given the model and the parameters of the model). Formally, using the more tractable log-likelihood:
where \( o_i \) is a binary variable referring to the observation (i.e. tool broken or not broken) and \( f_i \) refers to the forecast probability that the tool is broken. As the value in the square bracket will always be less than unity, the log-likelihood will always have a negative value. The better the forecast, the smaller will be the absolute value of this parameter.

5. STRATEGIC DIRECTIONS FOR FUTURE DEVELOPMENTS

Some of the deficiencies of currently available industrial tool condition monitoring systems have been highlighted in section 2 above, as has the gap between the industrial practice and the academic research. Reliability and robustness are key requirements for the further transition of the fundamental research into commercial application. In order to effect these improvements, operators will need to have increased confidence in the decision making capabilities of TCM systems. Demonstrably the ‘black box’ approach to tool breakage recognition (and associated reaction by the machine controller) has not worked. It has been widely recognised that effective human judgment relies on both objective and subjective judgement. The mushrooming field of fuzzy logic and its applications to decision making, since Zadeh’s seminal paper [26], is testament to the human requirement to incorporate degrees of uncertainty into their decision making processes. In the light of this, it is somewhat puzzling that dichotomous forecasting dominates the TCM system market. Intuitively, notwithstanding the possibly increased complexity of designing such a system, a system which gives probabilistic predictions of tool breakage would be of more value to (and therefore less likely to be misused by) machine operatives.

There is however, a more fundamental reason why TCM systems will need to consider using a probabilistic output. A perfect TCM system would recognise all breakages and produce no false alarms. Alas, such a system exists only as an idealisation and real systems will have to compromise (as per the ROC curve above) between false alarms and missed breakages. However the location of that ‘sweet spot’ will vary depending on the application. Consider by way of example the following two processes:

a) Mass production of (relatively) low value item such as a cast iron housing for a fuel injection pump in the automotive sector.

b) Maintenance of an Inconel™ component for an aircraft engine

The requirements of a TCM system for each will be radically different. In (a) the tools are likely to be of greater value than the workpieces, so the main priorities of the system will be to protect the expensive capital equipment, minimize downtime, maximise the tool life, and prevent damage to the workpiece. Provided that damage to the machines can be avoided, an amount of missed breakages will usually be tolerable, provided that the total downtime due to missed breakages and false alarms is minimized. However, in (b), the cost of the workpiece is typically several orders of magnitude bigger than that of the tool, so the priority is protection of the workpiece, protection of capital equipment, protection of the cutting tool and reduction of downtime in that order. Here, an amount of false alarms will be tolerated, provided that no missed breakages (with attendant workpiece damage) occur.

The above scenario provides an illustration of the costs of errors and the value of information in the decision making process. One can easily conceptually extend the scope of the decision making process to incorporate system or enterprise wide data, where the decisions to be made by the monitoring system must be considered in the context of wider costs and context.
The costs and benefits of a decision making system (in this case the TCM system) can be evaluated using the cost-loss value method, as outlined by Richardson [27] and Wilks [28]. In this method, the value of a forecasting system can be defined as follows:

\[
Value = \begin{cases} 
\frac{C}{L} \left( \frac{\text{hits} + \text{false alarms} - 1}{P} - 1 \right) + \text{misses} & \text{if } \frac{C}{L} < P_c \\
\frac{C}{L} \left( \frac{\text{hits} + \text{false alarms}}{1 - P} + \text{misses} \right) - P_c & \text{if } \frac{C}{L} \geq P_c
\end{cases}
\]

where \( C \) is the cost associated with taking an action (e.g. downtime due to stopping the process), \( L \) is the loss associated with not taking an action (e.g. damage to the workpiece and/or machine due to a missed breakage) and \( P_c \) is the climatological frequency (e.g. the average probability of breakage per operation).

The value of a particular monitoring strategy may therefore be assessed in terms of the actual costs and losses relevant to any end-user, but plotting a graph of cost-loss ratio (as the abscissa) versus Value. The same methodology may be adapted for probabilistic systems, except that in this case there will be family of curves (each curve representing a probability threshold as per ROC analysis), with the optimal value obtained from the overall envelope of the individual curves as shown in figure 3. Mylne [29] provides some examples of the practical application of such techniques in weather forecasting, including a direct comparison between the benefits of a dichotomous system and a probabilistic one.

It is suggested by the authors that future TCM systems will need to incorporate such decision making processes directly and/or provide a facility for bi-directional communication within factory/enterprise wide decision-making systems.

Figure 3. Cost-Loss Ratio Envelope, after [24]
6. CONCLUSIONS

An overview of the state of the art in TCM research and industrial practice has been given. In particular the gap between the two has been explored and reasons for this gap put forward. Attention has been drawn to the hitherto untapped (for TCM) research in the area of climatological forecasting and its applicability to TCM systems in the industrial environment. The power and flexibility of such systems will be required if the increasingly sophisticated measurement and analysis techniques are to migrate from the laboratory to the factory floor. In parallel the increased pressure on production systems to operate at higher speeds, with greater precision, greater agility and increasingly with less direct human intervention, present urgency to this required migration.

7. ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of all their colleagues in Trinity College and to recognise the input of many other collaborators and colleagues over the years, including, but not limited to: - University College Dublin, Robert Bosch, Kistler Instruments, Scania, Huller-Hille & Artis.

8. REFERENCES


