From Data to Diagnosis: Point condition monitoring through machine learning

Richard Parkinson, MEng Cambridge, Balfour Beatty

SUMMARY

'From Data to Diagnosis' presents a framework of processes that lead to the automatic identification of risk level and fault areas for a point machine from raw sensor measurements. Using work on the MJ80 point machine as a case study, each stage of the process is explained and analysed, identifying the benefits, effectiveness and lessons learnt. The framework includes the following processes:

1) *data capture*
2) *feature extraction*
3) *condition assessment*
4) *degradation identification*
5) *diagnosis communication*

In particular, this paper explores how machine learning techniques can be used to achieve results in some of the more complex stages.

1 INTRODUCTION

Points are a critical part of the railway and signalling infrastructure that allow the joining of two tracks. A point machine is an electro-mechanical device that controls the locking and alignment of the points to effect the selection of the desired direction. Point failures have been the cause of the last few major incidents in the UK rail network and on a more regular basis can cause large delays. This makes any improvements to their maintenance regime of high value.

Condition monitoring is the acquisition and analysis of data about a machine's operation to help inform the maintenance strategy. Basic condition monitoring of point machines is already highly prevalent but is still an immature area, with particular opportunities for improvement in the analysis of the data. Standard monitoring techniques like linear trending and thresholding are not always suitable for the highly variant and noisy railway environment, with a low correlation between predicted and actual failures.

2 THE FRAMEWORK CONTEXT

We live in an age where we can collect and store huge amounts of data about anything we want to monitor. A railway infrastructure network could easily have thousands of point machines each moving many times a day, potentially creating an overwhelming amount of data. However, only answers to a few simple questions are important:

- Which machines need attention and when?
- What is the most likely cause of the fault?
- How effective is our maintenance strategy?

In this paper, a new framework for transforming that data into meaningful, high-level answers is explored using the MJ80 as a case study.

The MJ80 is a widely used point machine that is powered by 3 phase electricity. The machine has a two stage locking mechanism for each direction, one for the rails and one in the motor. This makes the machine complex to analyse and therefore an interesting target for condition monitoring.

2.1 The Framework Structure

We propose a new 5 step approach to generating actionable data. The common thread throughout these steps is to refine the data to add value, while often reducing its quantity. These steps provide a structure for the complex task of the conversion process, with each step building upon the last.

Figure 1 shows the framework as applied to the MJ80 demonstrating how each step added meaning while reducing the dataset size.

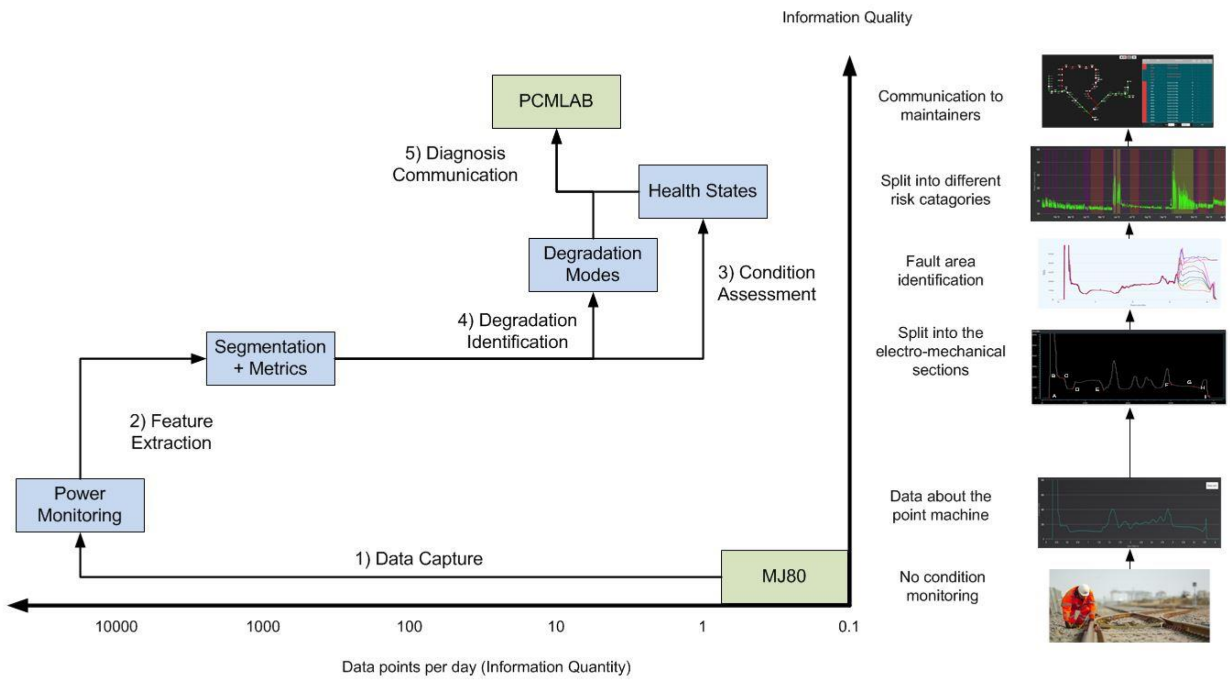

Figure 1: The Framework and its application to the MJ80 point machine

One point of interest is the choice to run the degradation identification and condition assessment independently and in parallel. It may seem more intuitive to build up a dictionary of known degradation modes that can be identified and used to assign the health states. However, this comes with the risk that combinations of modes or unknown modes might not be identified which might lead to the mis-assignment of health states. In the framework above, our choice is to assign the health state independently of the degradation identification. In doing this we've prioritised maximising the number of 'at risk' points identified above being able to assign a clear root cause for these. This prioritisation is driven by a strategic decision that the maintainer is much more likely to be able to identify the issue on site than to be able to assess the risk level of the machine remotely.

3 THE FRAMEWORK

3.1 Data Capture

3.1.1 Aim

The raw data captured is the foundation for all further analysis and therefore its precision, quality and focus dictates how easily areas of the machine's behaviour can be identified. Choosing the right technique is a trade-off between maximising the usefulness of the data measured against the cost and risk of implementation.

3.1.2 Implementation

On the MJ80, we monitor the three phases of voltage, three phases of current and the detection relays, all at a sampling rate of 100Hz. From this, the power of the motor and direction of the point move can be calculated. Figure 2 shows the power data for two point moves (note that the inrush has been clipped to make the rest of the detail clearer). These profiles are for sequential point moves, yet show a completely different character and can already provide useful information about the machine.

Voltage, current and detection monitoring can be implemented in the local control room where the points are being powered from. The detection monitoring informs when the machine is safely locked in each direction and therefore the direction of each move. The measurements are made using current clamps or isolated voltage connections. This setup minimises the safety risk and exposure to the harsh railway environment, making it a relatively cost effective and low maintenance solution. The power of the motor directly relates to its torque and speed and therefore the amount of stress the mechanisms are under, thereby capturing a large amount of the machine behaviour.

3.1.3 Review

Although the power monitoring gives a lot of detail about the machine behaviour, from further analysis it quickly became clear that there were some gaps in what it could detect. The most critical of these was track movement, as there could be rapid changes between point moves as a train passes over the point. If more attention had been paid to the machine context, in this case that the machine was on ballast, more appropriate monitoring might have been chosen. For example, vibration or position monitoring might help close out some of these gaps. These, however, are expensive to implement.

3.2 Feature Extraction

3.2.1 Aim

Feature extraction should identify and quantify interesting variation in the data. This output should be a much smaller amount of information which is of a higher quality.

3.2.2 Implementation

On the MJ80, it was decided to use temporal segmentation to identify key features. The segmentation points (A to I), as labelled in Figure 2, split up the point move into its different electro-mechanical sections. These segmentation points are identified automatically, isolating each feature for further analysis. For example, in Figure 2 it can be seen that the unlock section, between points D and E, is longer and requires more power for the green profile.

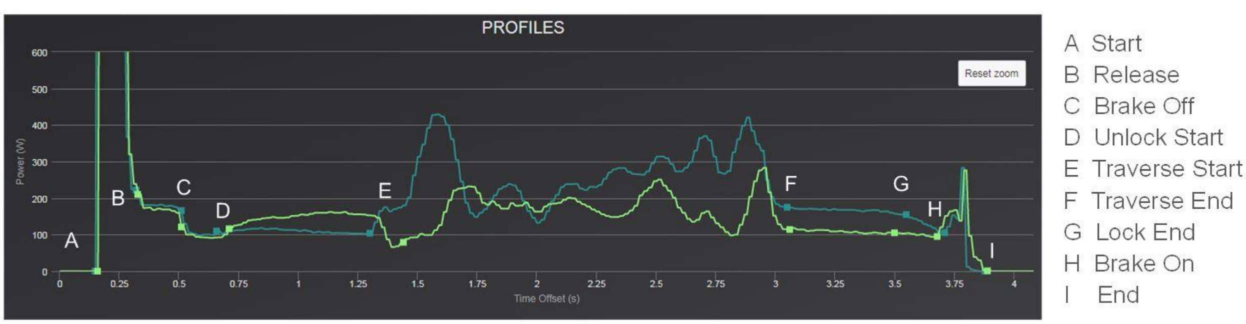

Figure 2: Segmentation of two sequential profiles (blue and green), on a graph of power (W) against time (s)

In order to identify the segmentation points, a genetic program with a tree structure was used.

Genetic programs use a principle similar to Darwin's 'survival of the fittest' to evolve the best model to solve the problem. Figure 3 shows a typical genetic program format. An initial population of random models is created to become the first generation. These models are assessed and ranked according to how good they are, using the fitness definition and training data. The best models are preserved through 'elitism' while new models are made by mutating or combining members (crossover) of the population to form the next generation. This loop is stopped after the number of generations goes over a set limit or when an acceptable level of 'fitness' is achieved.

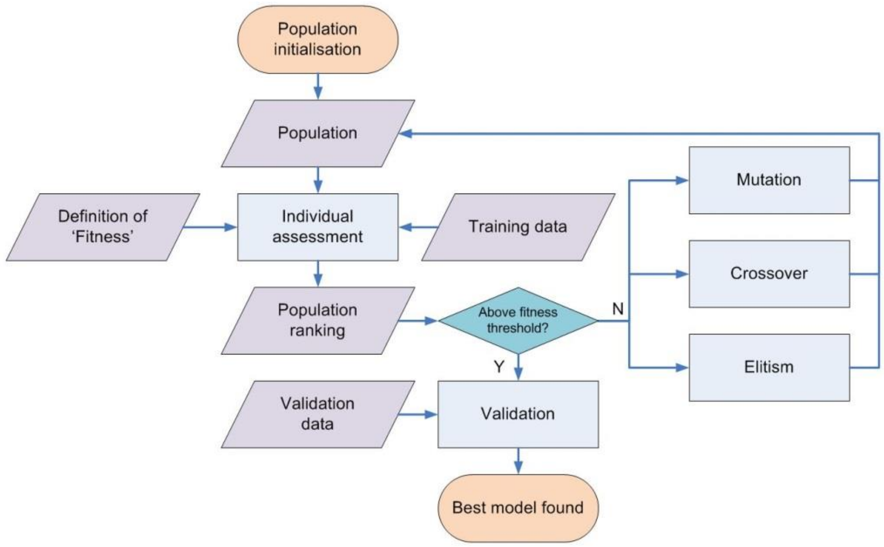

Figure 3: A typical genetic program

The tree structure shown in Figure 4 is used to create the models' individual features. Filtering removes any noise from the data; a search area improves the focus and speed of the model; while the template is used to search for the segmentation point. The search configuration can then be used to add any other parameters of interest. The model works by finding the best fit between the template shape and a portion of the profile.

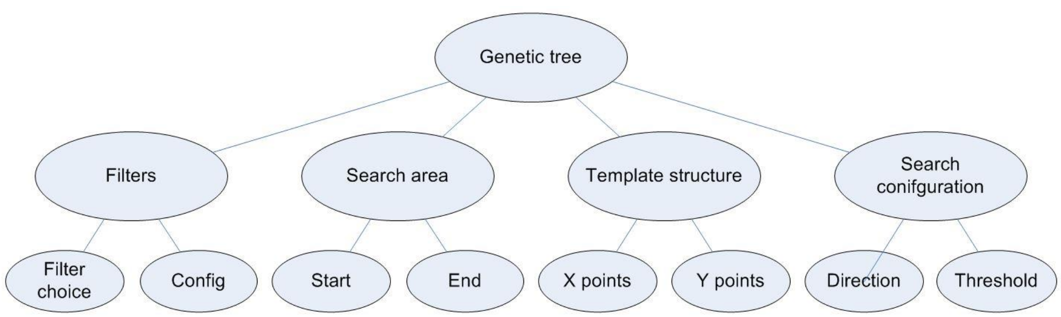

Figure 4: The genetic tree structure used for segmentation

In order to assess the fitness of each individual model a training set is required. Our training set was made up of a diverse range of profiles with a specific segmentation point labelled by an expert. During the assessment stage of the genetic program, the models predict where they think the segmentation point should lie on each profile and are ranked according to how closely they agree with the labels. Trained candidates can then be validated by running them over a different selection of labelled data.

Once the profile is segmented, each section can then be quantified using three metrics, chosen to capture a wide range of behaviours. These metrics can be combined to provide additional meaning. For example, a section with a high peak power but relatively low electrical energy indicates a high level of power variation within the section.

Metric	Significance
Time	Time to complete
Electrical Energy	Amount of effort exerted
Peak Power	Maximum amount of torque/stress experienced

Below is an example of the effectiveness of these new metrics. In Figure 5, the top graph shows the peak power for the section relating to the internal locking of the motor for each point move. The data is clearly clustered into two different power bands: around 160W for moves in one direction (green arrow) and around 240W for the other direction (blue arrow). A typical value of peak power is below 200W so the higher power moves are immediately suspicious. Comparing the two subsequent moves in the bottom graph, there is a clear bulge at the end of the blue profile. From its location we can tell there is a fault in the internal locking mechanism. Without these specific metrics, significant issues like this can be masked by larger variations in other sections that might carry much less interest.

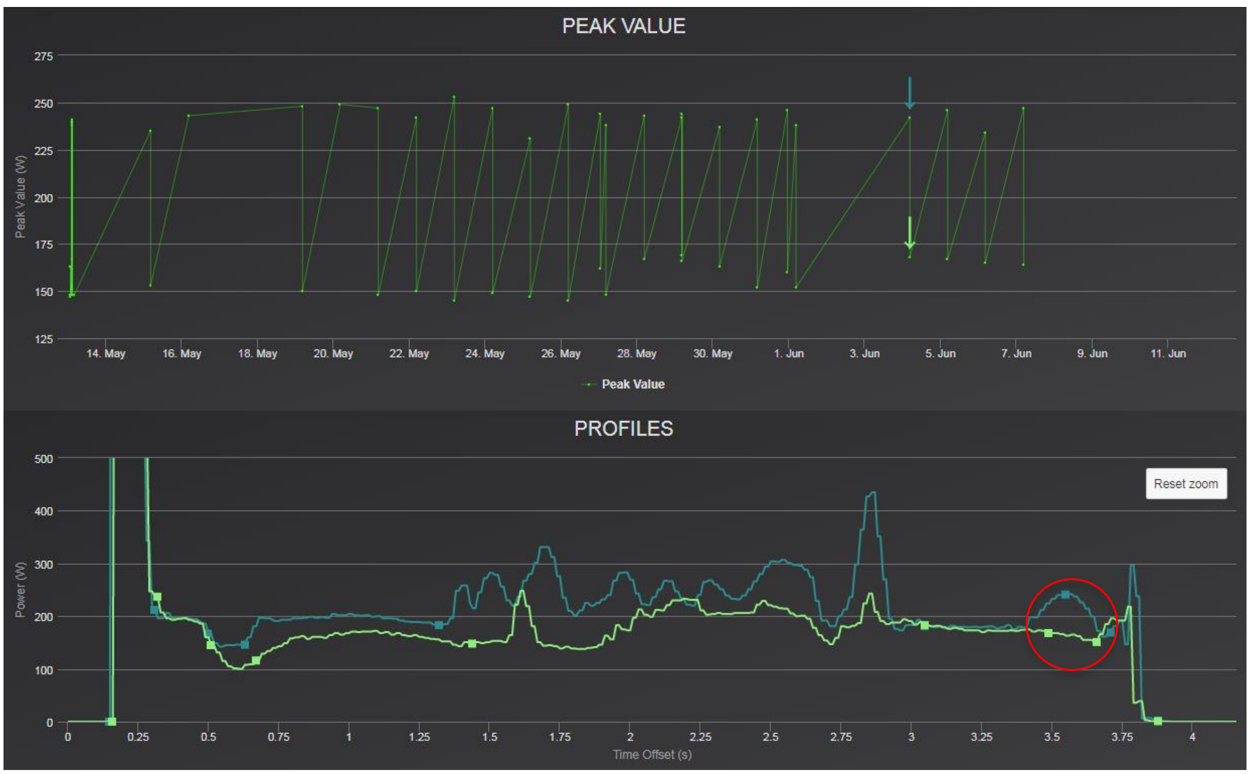

Figure 5: (Top) A graph of peak power (W) between segment points G and H over time (Days). (Bottom) The segmented profile with the area of interest highlighted

3.2.3 Review

Temporal segmentation was chosen as the feature to extract for three reasons:

- Simple to represent graphically
- Relatable to existing maintainer knowledge
- Delivers a large improvement in information quality

Genetic programs were used because they are:

- Understandable - it is possible to create a good model manually
- Fast to train - training the model usually takes under 30 minutes
- Highly robust to noise - the output model works on over 2 million point moves
- Flexible to a wide range of profile shapes as shown above
- Transferable - the program could also segment other point machines, for example clamp lock and HWs.

However, segmentation isn't the only feature on offer. Classification, using clustering, is an alternative approach. Figures 2 and 5 already demonstrate the wide range of section positions and profile shapes the segmentation algorithm must be robust to. In some scenarios, classification might be an easier approach to apply, with the drawback that it often requires more detailed interpretation afterwards.

One good case for clustering is a different point machine, known as the Dimetronic. As shown in Figure 7, this machine has a relatively featureless profile, for which the genetic program struggles to find segmentation points. Clustering the profiles into subcategories using an algorithm such as K-means, as shown in Figure 6, can help reduce the variation. If the variance for each cluster can be sufficiently reduced, respective segmentation points can be reliably identified.

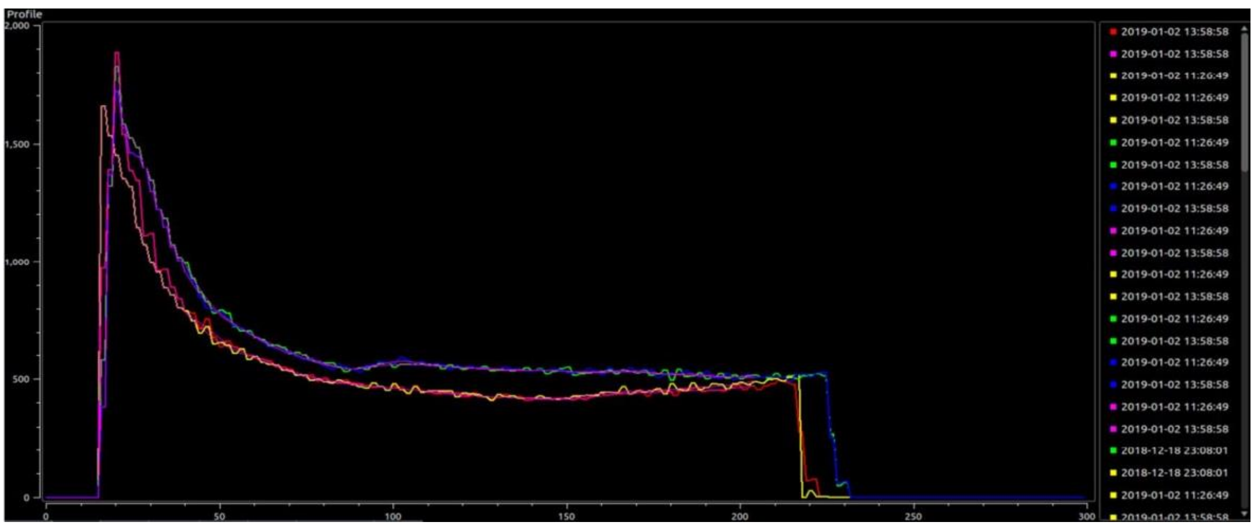

Figure 6: Two Dimetronic clusters each made up of 10 profiles

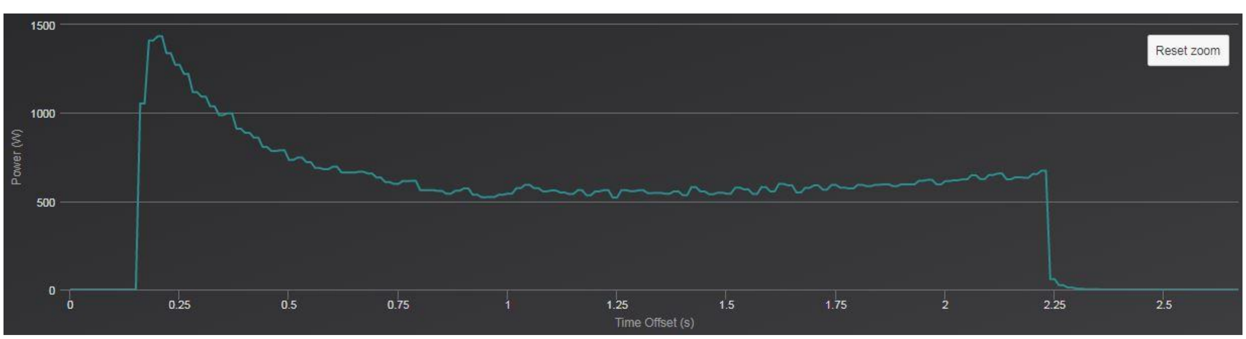

Figure 7: A typical Dimetronic profile

3.3 Condition Assessment

3.3.1 Aim

Defining a machine's condition can be achieved by assigning a risk category of failure to the point machine. Using risk categories recognises that all machines are at risk and that failures are also influenced by additional factors that are much more difficult to measure or predict. This risk category can then inform the urgency of response required by the maintainer.

3.3.2 Implementation

For the MJ80, a risk level system purely focused on failure risk was used. This step builds upon the feature extraction step using the three metrics for each of the 8 sections of the profile as an input. A genetic program uses this to generate an output, assigning the point machine a red or neutral health state. By running the algorithm over a large amount of historic data, statistics about these health states could be produced as shown below. These KPIs were chosen to try and best capture the behaviour of the health states.

No	Training KPI	Result	Impact	Aim
1	Percentage failures captured whilst in a red health state	>70%	Failure prediction rate	Maximise
2	Relative likelihood of a failure in a red health state compared to a neutral health state	>25	Significance of red state	Maximise
3	Percentage likelihood that a red health state contains at least one failure	>30%	Red health state risk level	Maximise
4	Average number of red states seen by a point machine in a year (Historic)	<1	Minimises number of false alerts (empirical)	Minimise
5	Percentage time spent in red health states (Historic)	<15%	Minimises effect of false alerts	Minimise

To understand the choice of these KPIs further, Figure 8 shows an example of an algorithm in action. This graph shows 2 years' worth of historic data tracking the duration of the point move, with red health states coloured red and failures represented by red crosses. This shows how effective the genetic program was at achieving strong results against its training KPIs:

1) 100% of the failures occur in a region of red health in the sample below
2) N/A – no failure occur in neutral, a good result
3) 100% of the red stats contain failures
4) 5 red health states were seen over the 2 year period. This is higher than the target but is minimal alerting considering the number of actual failures realised on this machine
5) < 20% of the time was spent in red health. This is a good result considering the high number of failures.

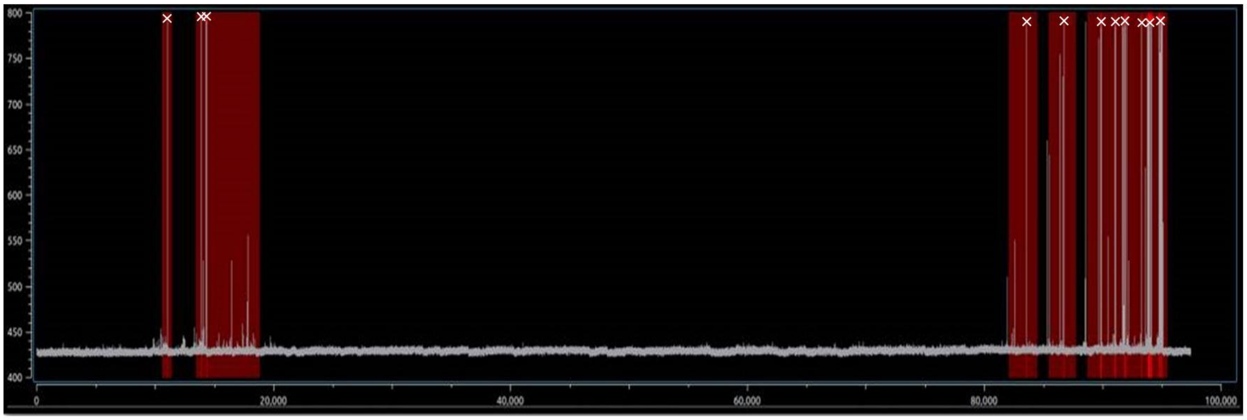

Figure 8: A graph of point move duration (number or samples at 100Hz sample rate) over time (point move event number). The areas highlighted in red show red health states and the white crosses show failures

A two state based health system was chosen as it is simple, understandable and stable, with minimal changes of health state seen. This has additional benefits such as robustness to the noisy environment, easy adjustability to suit maintainer's requirements and a clear message to maintainers.

Maintenance carried out on the machine can cause large changes of machine behaviour. It was found that this had a large influence on the output of the machine learning algorithm. In order to accommodate these changes, the sensitivity of the algorithm was turned down after maintenance while the system gathered information about the new behaviour.

A genetic program was used to develop the health state model. Figure 9 shows the tree structure for this problem. This tree allows the model to choose which metadata it would like to use, how much historic data it needs and the weighting and threshold level for red health.

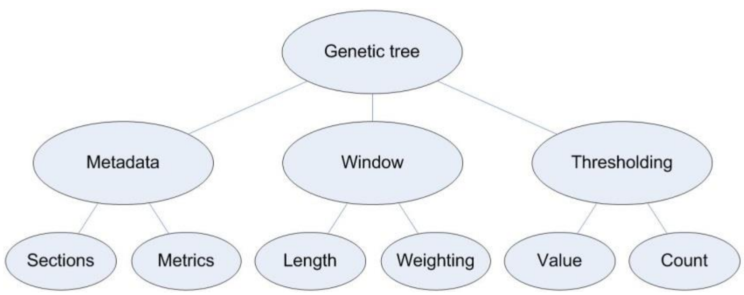

Figure 9: Genetic tree for the health state algorithm

The models are then run over a large amount of data and their fitness assessed against the 5 KPIs proposed earlier, with weightings applied to indicate the importance of each statistic.

3.3.3 Review

The red and neutral health states provide a clear separation of assets risk of failure, reducing the amount of interpretation required by an end user. However, using only two health states could make the problem too black and white. Finding a way to include an intermediate health state while retaining the clarity of the current system could augment the benefits to the end user.

3.4 Degradation identification

3.4.1 Aim

The aim of this section is to identify characteristics in the data that correlate with known degradation modes of the machine. In these scenarios, specific and precise information can be given about the action required to rectify the issue.

3.4.2 Implementation

Point machines are rugged mechanisms and can tolerate significant amounts of wear and misalignment and can operate quite far from the ideal setup without any immediate increased risk of failure.

We have already seen how a red health state is assigned to a point machine with a high risk of failure, agnostic to root cause. The remaining lower risk states can then be categorised further. A yellow health state is used when any of the known wear patterns or misalignment types are recognised, but the machine's overall risk is not high enough to merit a red health state.

To understand typical degradation modes, training data was created. For example, to assess the impact of misalignment of the locking mechanism an experiment was run on a training point machine, to see how it responded. Figure 10 shows the results of these tests: where shims were incrementally added to one of the locking mechanisms. Each addition makes a distinctive change to the profile shape, causing an increase in the total energy required.

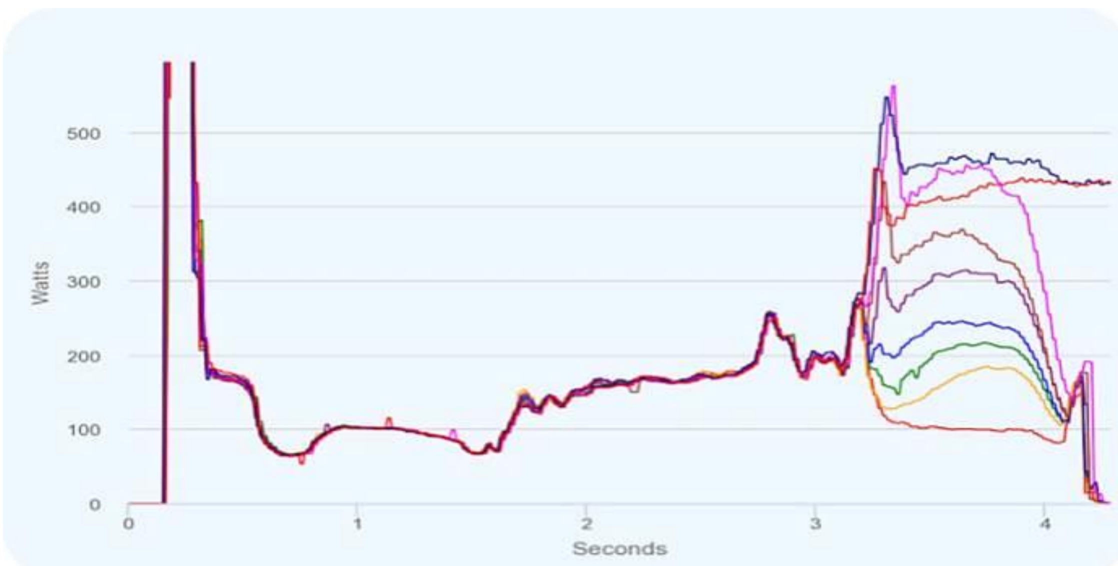

Figure 10: Adjusting the lock setup by incrementally adding shims. Each line represents a different increment.

This fault mode can then be quantified through the electrical energy of the lock section. Figure 11 shows how tracking this metric over time easily identifies the inappropriate setup of the machine. This can then be flagged to the maintainer using the yellow health state.

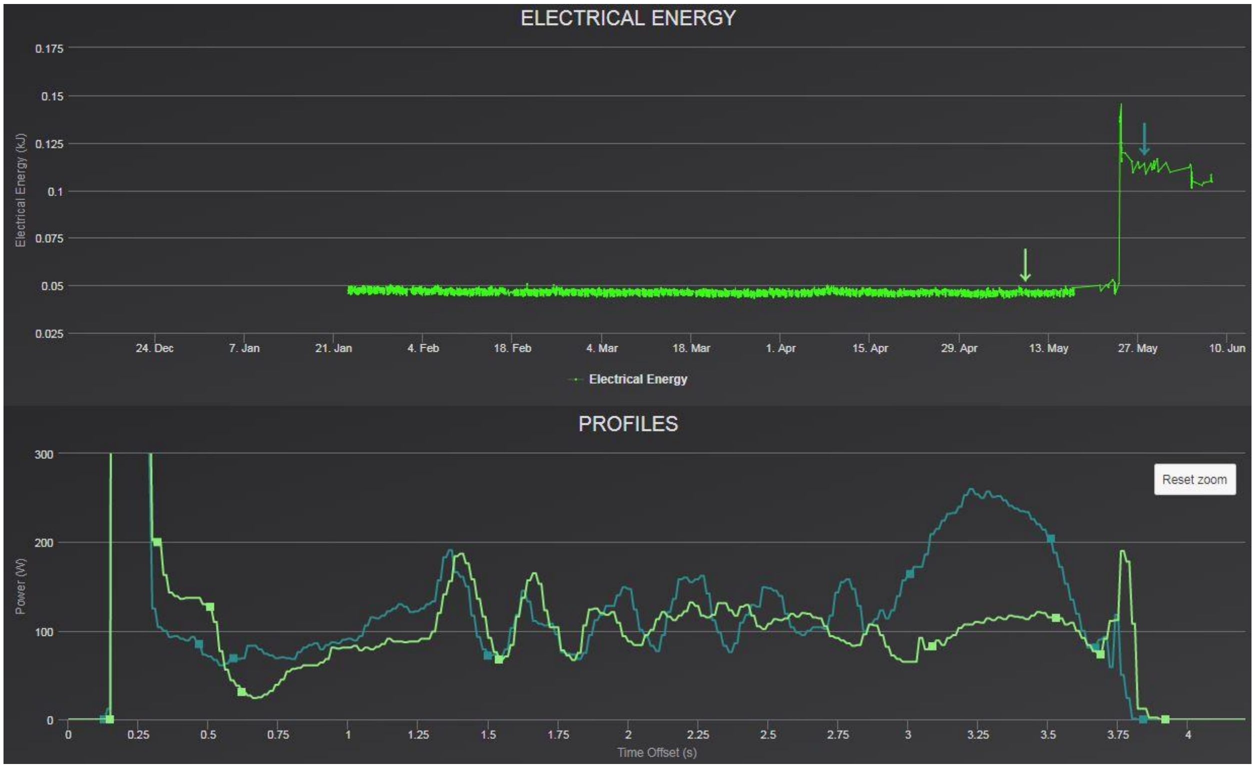

Figure 11: (Top) A graph of electrical energy (J) in the lock section over time (Days). (Bottom) The segmented profile showing two contrasting profiles. The colour of the arrows in the top graph relate to the profile colours.

3.4.2 Review

Simple degradation identifying algorithms work well on simple degradation modes. This gives a lot of benefits for relatively little development work. The maintainer now has a much better idea of what work they will need to carry out on the point machine during their maintenance visit.

A limitation in this case is the ability to identify complex combinations of degradation modes. This might be overcome by augmenting the metrics derived in the feature extraction step. Classifying the sections into common categories based upon behaviour could provide the additional information needed.

3.5 Diagnosis Communication

3.5.1 Aim

Once the most likely meaning has been derived from the data, it is important to communicate this to the maintainer as effectively and efficiently as possible.

3.5.2 Implementation

For the user interface, we developed a live web system for this communication with two clear sections: a high level summary page and an investigation page.

Figure 12 shows the high level analysis page. Here users can explore their assets through the map on the left and on the right view a brief summary of the highest priority assets. Users can then quickly asses what further action is needed, from carrying out further analysis to immediately planning in maintenance.

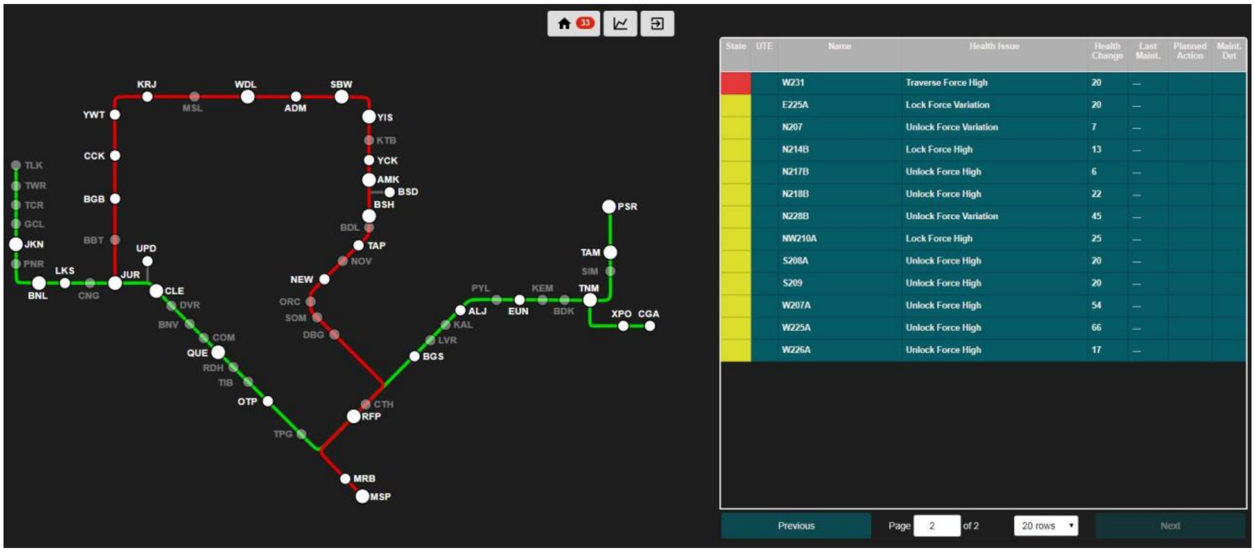

Figure 12: High level asset summary (Screenshot from test system)

The investigation page (Figure 13) is split into 4 sections. On the top left, a time series graph shows the asset's history efficiently through metrics, health states and maintenance events. On the bottom, individual point moves can be compared and analysed using a profile graph. As mentioned earlier, it is important to understand the relationship between maintenance and the point machine. To facilitate this dialogue, maintenance can be inputted into the section at the top right. Finally, in the bottom right, a settings box controls the contents of what is seen in the graphs.

Figure 13: Details and ability to analyse (Screenshot from test system)

The design shown in Figures 12 and 13 ensures easy and natural user flow, using a simple hierarchy between the summary and detail page while maximising the amount of information available on each page.

3.5.3 Review

This new system has received positive feedback from users proving its success. There are plenty of further additions that could be made to the system, from a graphical representation showing which areas of the point machine are degrading to a managerial-oriented page, which gives an overview of the overall performance of their point infrastructure.

4 FURTHER WORK

The framework was applied successfully, as shown with MJ80 case study. Some alternative approaches were presented at the end of each step. Other key learnings that existed across multiple steps are presented below.

4.1 Increased dialogue and partnership with maintainers

There is still a large gap between how traditional maintenance is carried out and what condition monitoring has to offer. There are several benefits gained from working in partnership with maintainers. First, they can help provide extra data about the maintenance which will help reduce variation which is unaccounted for in the models. Secondly, this dialogue will help develop a realistic strategy for maintenance improvement and development that incorporates condition monitoring. Regular communication can be used to help encourage the engagement with the system, for example, getting maintainers to review the data before and after maintenance to further understand the effects of their work. Finally, it will help guide optimising the framework outputs to maximise the benefits to the end user.

4.2 Data integration

Data collection is increasingly becoming cost effective and more prevalent. Building frameworks that can accommodate and make the most of multiple data sources will ensure the acceleration of development in this field. In particular, combining the power data with vibration, weather and track geometry data has huge potential for a holistic understanding of the machine.

4.3 Expand and develop different analysis techniques

There is no right way to analyse the data, but solutions are more likely to be found when a broad range of techniques are tried. In this report, the benefits of using both genetic programs and clustering was demonstrated. There is no doubt that even more benefits that could be gained through the use of Gaussian linear models, Bayesian analysis, neural networks and many other analysis tools.

5 CONCLUSION

This paper presents a new framework to achieve automatic identification of the risk level and fault areas of a point machine from raw sensor data. Techniques for data capture, feature extraction, defining risks levels and fault analysis are explored using the MJ80 point machine as a case study. In particular, the use of machine learning in order to solve complex and variant problems was demonstrated.

Several recommendations were made, highlighting the various benefits of a good feedback loop from the maintainer, the importance of using a wide range of analytic and machine learning techniques and opportunities that arise from combining data from multiple sources.