

# Optimising remedial outcomes for Gas Turbines through large scale data analysis

Jason McGinty\*, Stefanos Kollias\*\*, Georgios Leontidis\*\*\*

*Machine Learning Research Group, School of Computer Science, University of Lincoln, Brayford Pool, LN6 7TS, Lincoln, UK, \* [16654927@students.lincoln.ac.uk](mailto:16654927@students.lincoln.ac.uk) / [Jason.McGinty@siemens.com](mailto:Jason.McGinty@siemens.com), \*\* [skollias@lincoln.ac.uk](mailto:skollias@lincoln.ac.uk), \*\*\* [gleontidis@lincoln.ac.uk](mailto:gleontidis@lincoln.ac.uk)*

**Abstract.** An investigation into approaches to model and predict the costs, risks and outcomes, relating to a common failure mode within a large population of remotely monitored engines will be presented. This investigation will cover the relevant aspects of lifecycle management, customer operating cycle, visibility of classified issues remotely using low sample rate remote monitoring systems and identification of the most appropriate repair regime. Unlike previously identified studies, the datasets used are larger comprising sensor readings from over a 120 gas turbine units (distributed across the world) which total 100 million observations taken across 117 parameters. These are cross referenced with relevant service events (preventative, corrective, fault investigation and observational reports) and service execution data (spares and manpower required) gathered over a period of up-to 11 years. We have now succeeded in identifying behaviours in a particular problem domain using techniques such as K-means, that indicate non-optimal decisions may have been made (with the benefit of hindsight). This indicates that the intended outcome of the overall research to produce of quantitative models of the different strategies and applying them to optimize for the best outcome both in terms of and customer and supplier objectives in a justifiable form is possible.

## 1. Introduction

In the world of industrial power generation (and other industrial sectors) there has been a shift in business models from Original Equipment Manufacturers (OEM) with the ongoing service, corrective and upgrades portions of the business being seen as the biggest drivers of growth rather than the traditional staple of new unit sales (Drugugbo, 2016).

This has required some re-alignment by affected OEM's to change their product mix and find ways to competitively service their customer base. One of the ways OEM's have hoped to achieve this is through the use of remote diagnostic systems and services in order to have visibility of issues that occur at customers sites. Traditionally these services have been usually simply for post-mortem or for the use identification of individual issues.

Remote monitoring of Gas Turbines is still in many respects an art form, rather than a true science. Due to the complexity of the system, relatively high cost (in capital, maintenance, running and potential loss of revenue) and serious consequences of a wrong decision (such as loss of life), is still very manually controlled.

We have however a large quantity of information that has been collected over the years, and now with the gradual digitalization of other processes it become possible to automatically add context to much of the information that we have been collecting. This leads to us now being able to review historic data in a new light that was not possible potentially at the time, and analysis the effectiveness of the repair strategies used the processes of labelling and annotating a large dataset (comprising a total of 11 years of operating data from up to 120 engines, not all engines we in operation or connected for the duration of that period), this is taken) is still ongoing, and involved. The data primarily contains information provided by customers who are under the various long-term agreements, this means there is considerably more contextual information available in the form of site visit reports, parts orders and maintenance history to annotate the collected sensor data with.

A number of papers (such as those looking at machine fault detection (Zhang , 2014)), or signal reconstruction to improve availability (Yang, 2013) have been produced by various research groups in relation to the same core dataset, however their purview has been quite narrow in terms of the scope of the project regarding analysis of individual failures modes in a very specific (and often novel way), without looking at the context of the surrounding remedial actions that would need to be taking place.

This paper presents an overview of the direction in which the ongoing research is taking, illustrated by a sample of the data currently available. This for reasons of commercial sensitivity the data and the user, product and its technical features has been obfuscated.

## 2. Research Goal

With the rapidly increasing availability of usable information would it be possible to create a self-optimizing decision support system that would improve the consistency and quality of our support in major events domain of faults which are addressing are the areas where there are two (potential competing) disruptive and relatively costly remedies that could result in a fix for an issue, but which the chances of success are not guaranteed. These issues generally occur when there are major subcomponents that can be changed out in lieu of a full engine exchange, but could trade off longer term reliability or disruption with the customer if they were to fail.

It is also hoped that the work will allow for the integration of approaches to deal with No Fault Found conditions occurring which are become a major focus of cost reduction across affected industries (Khan, 2016), and in this domain results in large number of resources being used to investigate and replace components potentially unnecessarily.

## 3. Current Approach and Methods

The current core focus of the research at this point is to create the pipeline and infrastructure shown in Figure 1. This illustrates the high-level structure of the system and the data-sources that are available for the decision-making algorithms. It is novel due to the extent of cross-use of information between the parties. This is analogous to the system that was presented at the preceding conference in this series by Animah (Animah, 2016) in their description of a novel system for managing life extension . however it is focused on automatic feature extraction rather than expert driven constraints collection.

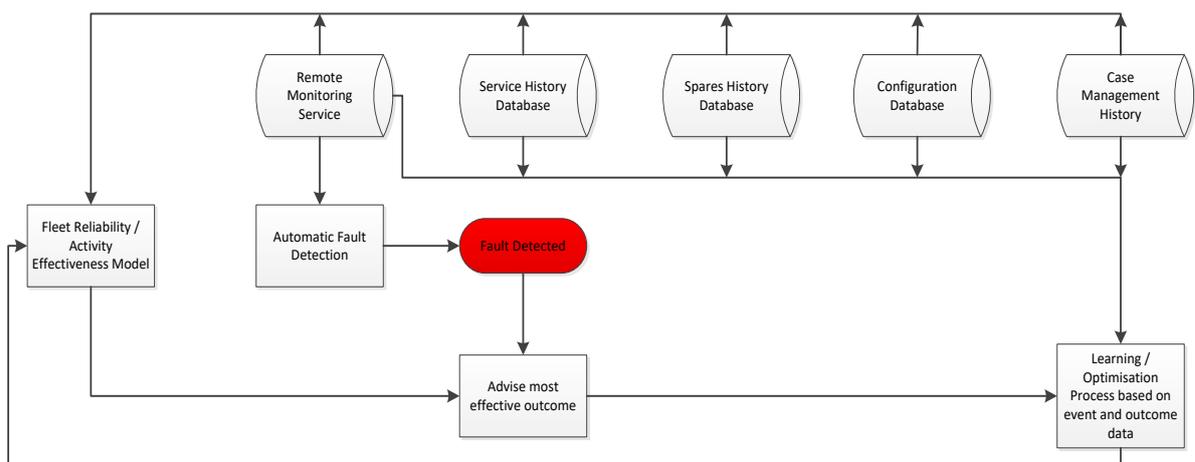


Figure 1: High-level overview of proposed system.

#### 4. The initial sample problem

The fault which has been selected as the basis for the initial study has been selected from a set of common issues that have a high impact upon the availability of the engine and which can be seen remotely via telemetry. For the purposes of this investigation the problem selected is a condition which can occur in which there are two potential options for remediation, the first is that Gas Turbine can be repaired at site for a moderate cost in terms of material and manpower, but more time in operation and with an unknown chance of success (Option 1) or the Gas Turbine can be replaced wholesale at a much greater cost, but with much less downtime for the customer (Option 2). There have been a number of occasions however where option 1 (the parts replacement) has been undertaken only to require a full replacement shortly thereafter and also cases where a full replacement has taken place where a partial replacement may have proved adequately corrective.

#### 5. Assessing the health of the engine

There are many variables that are taken into account when making a decision about an engine. Currently the majority of these are taken in a subjective fashion using a number of heuristics, which are believed to have a better than average chance of success. By having the service visit histories we are now able to identify the engines actual given point in its life including the parts and actions undertaken which may have resulted in a change in operation. Such as those seen in Figure 4 following the engine changes. Due to their complexity and the bespoke manor in which such engines are built they each have a unique character, but should ultimately within a small tolerance provide the output power within an acceptable efficiency. The extent of this variability is shown in Figure 2. Which also on the right hand panel shows the changes that occur have occurred on a single sample engine through its life.

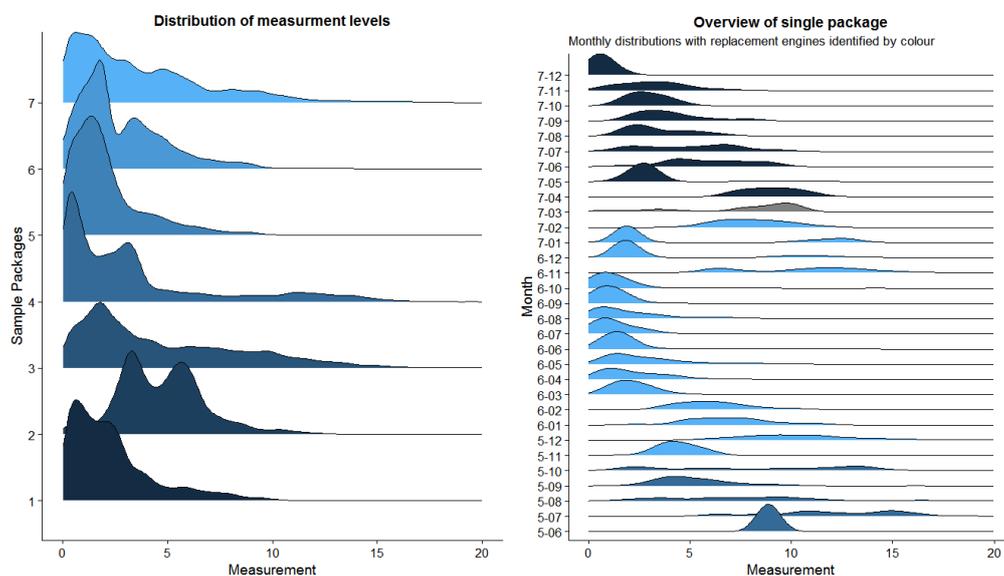


Figure 2: Relative distribution of an observable measurement across sample engines related to the sample problem.

An initial K-Means analysis was undertaken in the R environment (R Core Team, 2017) (which provides the workplace in which this research is currently being undertaken). The clustering was conducted using the common Hartigan-Wong (Hartigan, 1979) method and involved a number of parameters (such as distance between measured variables and the life of the engine at the point of measurement).

As can be seen from Figures 3, The K-means clustering appears to contradict what has been the traditional heuristic levels imposed for that measurements plotted (and as defined by the staged boxes on the plot). This indicates that there are other features which have been included within the clustering that are not generally considered which have an effect on the likelihood of the engine being replaced.

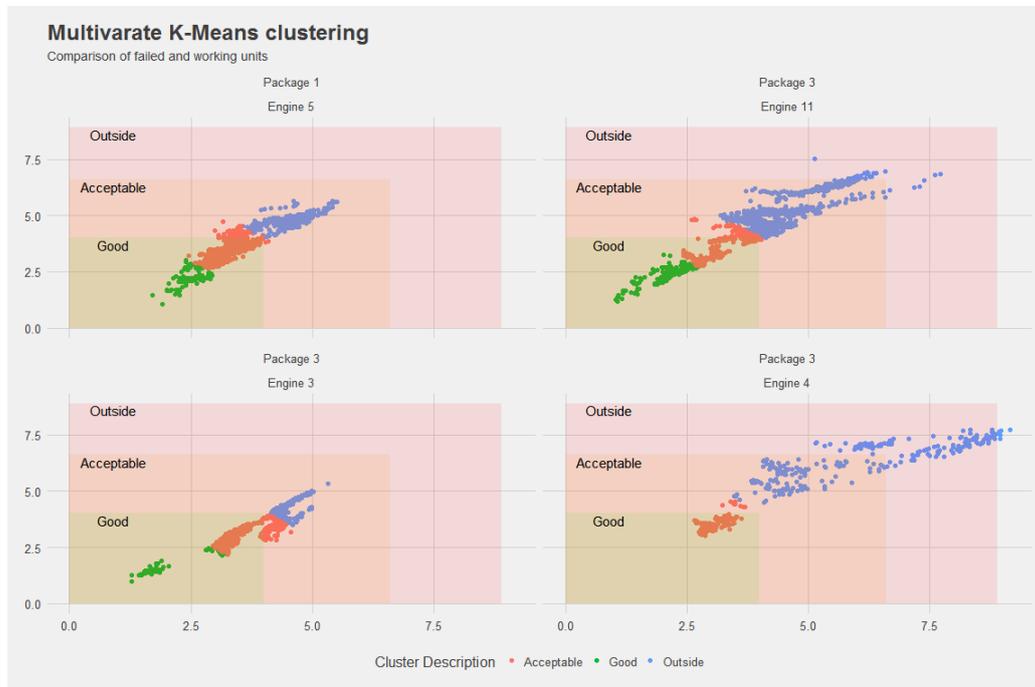


Figure 3: K-means distribution of multiple sites. The left side shows engines where no fault has occurred, right hand side shows two engines which have failed before reaching end of life. Cluster Description describes the predicted membership group using the multivariate approach as shown by the difference in colour. The boxes plotted are the heuristic limits identified for the measurement variables for traditional assessment of health.

As can be seen from the break out of the Package 3 in figure 4, all three engines which have been installed at Package 3 (a package is a full full turbine installation, including supporting equipment and a driven unit) appear to ultimately have issues (although 1 has been identified as not changed early that is simply because it was still in operation at the end of the available data, and had not a that point exceeded its intended life). The option 1 repair which has been attempted three times, appears to have little if any long-term effect. More worrisome is also that even the two engines (in this case the engine is the actual gas turbine itself) that were deemed not to be replaced early in Figure 3, also appear to be showing early signs of issues forming.

In order to identify whether it potentially simply wasteful to undertake the option 1 replacement as you may simply be adding additional costs we have begun considering applying an approach similar to the Number Needed to Treat approach described by Schechtman (Schechtman 2002). There is an assumption that an inconsistent approach has been taken in the previous handling of occurrences of the studied problems which should lead to us being able to identify cases when previous assumptions and heuristics in dealing with them have been misplaced, and through which we can identify a more optimum solution. This is likely to be due to the breadth of considerations that are required to be taken into account outside of the raw operational data from the engine (such as contractual or supply chain issues).

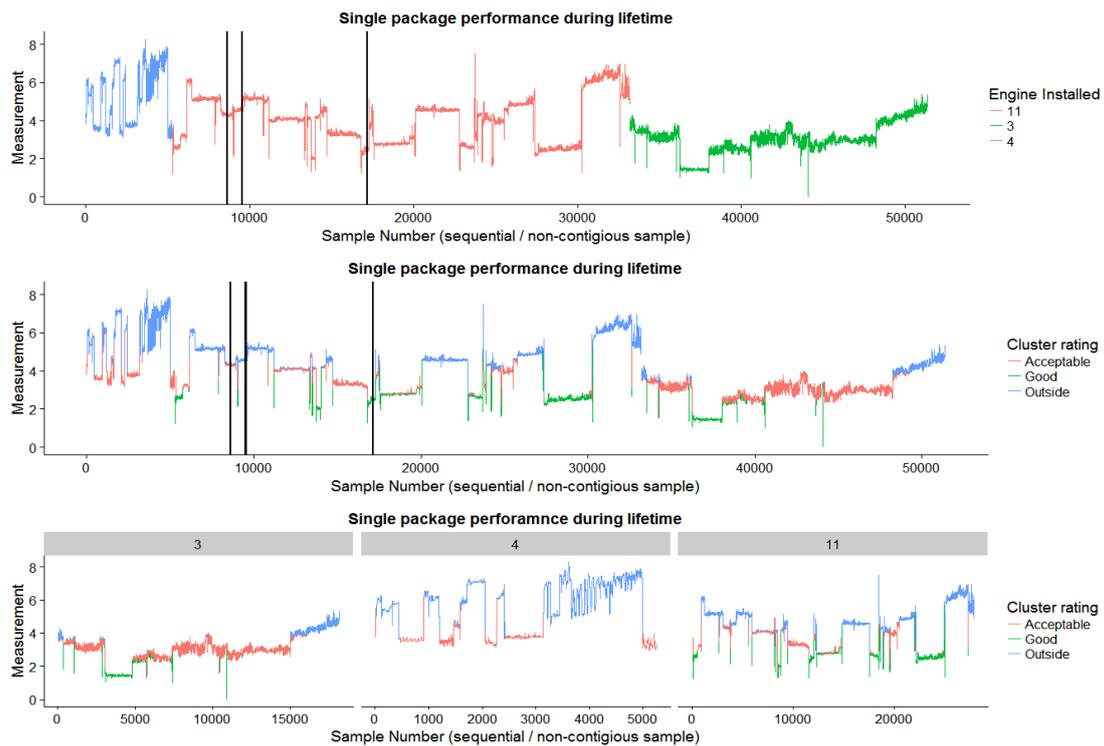


Figure 4: Time series plot of Package 3 (as shown in Figure 4), showing the different Engines installed as different colours (the change in colour indicating and Option 2 repair has been undertaken) are also lined. The point at which Option 1 repairs took place are also depicted by the black vertical lines.

## 6. Future work

Now that the majority of the preparatory work has been undertaken, it is now possible to consider the costs implications of the actions undertaken. Rather than simply consider the physical costs of installation and fitting, it is anticipated that more novel features will be includable within the future model. These issues include but are not limited to:

- Contractual penalties and bonuses,
- Benefit from parts of recovered (i.e. credit for parts)
- Cost of loss of useful life of parts (i.e. those that will not be useable within an overhaul)
- Cost to user of non-availability of Gas Turbine
- Costs or benefits occurred in performance of replacement unit (potentially between +3% and -5% from designated rating), which could lead to increased profit for user.
- Costs incurred due to undertaking of no fault found investigations

Once inclusion of the above features has been completed it is anticipated an accurate working model can be produced and further problem domains can be integrated into the model.

The hope is that this will allow us to provide a labelled dataset from which we can introduce some of the more interesting deep learning approaches (such as that explored by Deutsch (Deutsch, 2017) which is directly applicable to this domain, and Sabour and Hinton's Capsnet architecture (Sabour, 2017)) which show great promise in extracting meaningful information and relationships and providing accurate predictions as to the remaining useful life in-turn provides an opportunity to better

optimize the whole lifecycle of the equipment. It also appears to support the possibility of transfer learning between seemingly dis-similar products but with the information from the surviving processes being applicable across them (i.e. all Gas Turbines obey the same fundamentals, but may be configured slightly differently).

## 7. Conclusion

This paper has presented an overview of a proposed system to capture and optimize the constraints of the decision making process for remedial activities on a Gas Turbine using quantitative data for it. The work conducted to date and documented within this paper, shows that overall goal of automating and optimizing the justification of certain remedial activities within the domain of Gas Turbine maintenance is achievable with the extensive cross-function data sets that are now available .

## Acknowledgement

The work was supported by S.I.T Limited.

## References

- Durugobo, R., Erkoyuncu, J. (2016) *Critical Journal of Operations and Production Management*, Vol. 36 , No.5, pp. 532-571
- Zhang, Y., Bingham, C., Ling, B., Gallimore, M., (2014) Machine fault detection by signal de-noising - with application to industrial gas turbines, *Measurement*, 58, pp. 230-240
- Yang, Z., Ling, B., Bingham C., (2013) Fault detection and signal reconstruction for increasing operational availability of industrial gas turbines, *Measurement* 46, pp. 1938 - 1946
- Animah, I, Shafiee, M. Simms, N. (2016) Optimizing life extension and maintenance policy decisions for safety critical elements, *Proc. Of the 9<sup>th</sup> IMA International conference on Modelling in Industrial Maintenance and Reliability*, pp. 14 -16.
- Khan, S., Farnsworth, M., Erkoyuncu, (2016), A novel approach for No Fault Found decision-making, *CIRP Journal of Manufacturing Science and Technology* 17 ,pp. 18 - 31. Elsevier.
- Hartigan, J. A. and Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistics* 28, pp.100—108
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Deutsch, J. He, D, (2017) Using Deep Learning-Based Approach to Predict Remaining Useful Life of Rotating Components, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol 48, Issue 1 pp. 11-21.
- Sabour, S., Hinton, G., Frosst, N., (2017) Dynamic Routing between capsules [arXiv:1710.09829](https://arxiv.org/abs/1710.09829)
- Schechtman, E. (2002), Odds, Ratio, Relative Risk, Absolute Risk Reduction, and the Number Needed to Treat – Which of These Should We Use?, *Value in Health*, Volume 5, Number 5, pp.431 - 436