4th Asia Pacific Conference of the Prognostics and Health Management,
Tokyo, Japan, September 11 – 14, 2023 R02-11

Event detection in a noisy time series data using static smoothening and gradient variation analytics

Saravanaram T[1], Unnat Mankad[2], Simi Madhavan Karatha[3], Arati Halaki[4], and Carmine Allegorico[5]

[1,2,3] *Baker Hughes, Bengaluru, 560037, India*
saravanaram.t@bakerhughes.com
unnat.mankad@bakerhughes.com
simimadhavan.karatha@bakerhughes.com

[4] *Wissen Infotech, Bengaluru, 560100, India*
arati.halaki@wisseninfotech.com

[5] *Baker Hughes, Florence, 50127, Italy*
carmine.allegorico@bakerhughes.com

Abstract

Data-driven prognostic industrial asset health management is essential to improve reliability and availability of industrial machineries. Industrial sensors capture multiple phenomena of underlying assets like environment impact, system degradation, process variation, instrument noise, control system response or user induced actions. Some of these phenomena have distinct signature and have impact on component life and remaining useful life. Capturing these events' signature help to apply advanced AI algorithm to categorize various failure modes and early detection. For Ex: Some of transient events can contribute to thermal cycling of component and in turn reduces life of high temperature and low thermal mass components. Due to random and nonstandard nature of these events, it is extremely challenging to detect and extract these events. Existing change point detection algorithms have limitations to detect sudden variations, which is common due to process or control actions. The noisy signal adds additional challenge to differentiate between important event and noise. In this work of time series analysis, we propose a new approach for consistent estimation of numbers and locations of the change-points. With this tunable algorithm combined with event labelling and pattern search, we can detect events of our functional needs and use them as a feature for our prediction models. This methodology has opened exciting opportunity to further analyse these events with development of classification system and time to fail prediction models and also apply large language models for time series data.

1. INTRODUCTION

Rotating equipments are often among the most critical items in industrial domain including Oil and Gas/Energy plants serving various applications, such as Liquefied Natural Gas (LNG), pipeline, refinery, petrochemicals and power generation. One of the most important aspects for the operators of these turbo machines are continuous availability and reduced downtime covering the entire life cycle. Unscheduled shutdown of the turbo machines can have impact on the whole plant downtime with associated significant loss of production Iannitelli et al. (2018)

Baker Hughes have developed monitoring capabilities which are offered as a service, applied to a broad installed fleet of rotating equipment including gas turbines. Baker Hughes' iCenter ecosystem continuously acquire different sensor parameters of its deployed assets at customer premises. These large number of operational data from the everyday operation of turbo machine is usually collected and analyzed as soon as new data sets arrive in the monitoring center. Anomaly detection rules and models are designed to scan through the data and notify the monitoring and diagnostic engineers, if any anomalies or emerging problems are detected. These analytics help to maximize monitored units availability and uptime. Allegorico & Mantini (2014)

Sensor time series data which indicates intrinsic behaviour of the underlying system, typically are non-stationary due to influence of dependent and independent variables interactions, noise of the instruments, system interaction, etc. The given system could also be going through cyclic variation due to dependent and independent interactions. Understanding these cyclical variations could be of area of interest to study its impact on the underlying system.

There could be many fields where identifying this cyclical behaviour is desired. In the field of Turbo machines, it could be sudden variation of ambient, system degradation, process upsets, input conditions variations, control response or user induced actions. These could lead to thermal cycling of internal components and in case of high temperature exposure and low thermal mass components could lead to reduced life Szczepankowski & Przysowa (2022)

Typically, these cyclical variations are non-standard in nature. Additionally, noise in the system adds additional challenge to distinguish between event and noise. Accurately identifying and quantifying them poses a great challenge in the field of time series data analysis.

2. MINOR CYCLE

Typically, these cyclical variations are non standard in nature and their pattern could very depend on the trigger. This marks a real challenge in identifying and evaluating them. One of the possible solutions could be classification system, where a supervised learning classifier is able to detect and quantify the cyclical variation. However, as the pattern is not standardized, it would lead to exhaustive list of training samples and this restricts the application of the same.

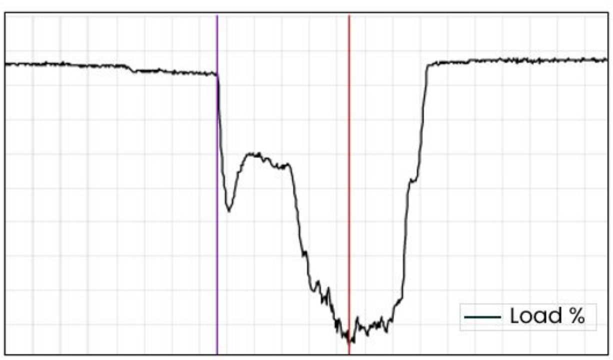

Figure 1. Typical cyclical variations in turbo machines

Main objective of our novelty is to accurately predict the start and end point of the cycle, reducing the probability of false alarm and miss detection, mainly to capture the abrupt changes in the cycles. Time series sensor data of a Gas Turbine was considered for the analysis. As shown in the Fig.1, sensor data shows a cyclical behaviour, due to some of the factors highlighted already. Accurate identification and quantification of these cycles could be of importance as it would help understand the impact severity on the system. This could also be a strong feature in developing an intelligent system where one of the key feature could be type of cycles the unit has undergone. Similar events could be used in clustering providing additional details.

3. CHANGE POINT DETECTION METHODOLOGIES

In order to capture the cycle, it is important to identify the change points of the cycles. Change points are the points where the data changes its pattern in terms of direction. If change points are identified accurately, this could mark the beginning or end of the cycle.

In change points identification the most basic detection is that of hypothesis testing Neyman & Pearson (1933), Johnson et al. (2017) where the goal to detect which of two distinct hypotheses has better statistical support. A cumulative sum chart (CUSUM) is the sum of the z-standardized realizations of the time-series widely used control chart to detect the deviation of the individual values or subgroup mean from the adjusted target value. If a time-series has constant zero mean, the cumulative sum of its realizations converges to a zero-mean Normal distribution. Thus, if the cumulative sum diverges from a zero-mean Normal distribution, a change-point in the underlying time-series might have occurred. The key factor in this method is how you should set the change point threshold values in CUSUM. After all, setting the threshold too loose will lead to undetected change points. On the other hand, narrow thresholds can easily lead to frequent false alarms. To overcome false detection, Cross Entropy method to a sequential change-point problem is applied to obtain estimates for thresholds in the Shiryaev-Roberts procedure Sofronov et al. (2012) and the CUSUM procedure, considered the generated sequences to illustrate the effectiveness of the approach to find threshold value.

The Shewhart control chart is a common control chart methods applied in SPC (Statistical Process Control) Cheremisinoff (2001), however they are unable to detect small shifts in the process. To overcome this issue, Roberts (2000)introduced a control chart using the exponential weighted moving average (EWMA) statistic, which is more efficient to detect a small shift in the process. EWMA method is a control chart for the data which are quantitative and continuous in measurement in time order. EWMA chart plots exponentially weighted moving average values, and a weighting factor is found in order to estimate the influence of older data points on the average value compared to more recent one. EWMA chart uses all information that a given sample provides. EWMA control charts are used to monitor processes over time.

3.1. Leverage Standard Libraries

To identify the change points, rupture library from Python was used. “Ruptures is a Python library for off-line change point detection. *Rupture Python Package* (n.d.) provides methods for the analysis and segmentation of non-stationary signals. Implemented algorithms include exact and approximate detection for various parametric and non-parametric models”. Rupture uses different methods to identify Change

Points on the non-stationary data. PELT method (Killick et al., 2012) was used with different sensitivity on "Penalty value" which is a tunable parameter of this method. It was observed that with lower penalty value, method identified higher change points and with higher penalty value, method identified lower change points. This is due to the consideration that Penalty value is the cost of defining a change point in the data set.

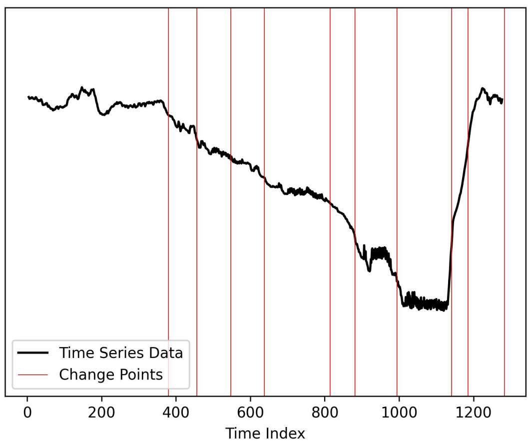

Figure 2. Change point detection with penality 5

As it can be observed from Fig.2, the method could find out change points, but they were not accurate. At the same time one big limitation of this method is sensitivity to Penalty value and for non standard pattern of cycles happening in the time series, this method may not be a most suitable one for the application.

3.2. Evaluating Energy of Signals

Silva et al. (2020) demonstrated successful application of Continuous Wavelet Transform (CWT) which could be successfully applied for early fault detection in gas turbines. To extract the features of time series data in frequency domain, well known FFT and wavelet transform were studied. Wavelet transform (WT) is robust to noise variance and has higher time-frequency resolution than the Fourier transform (FT). CWT with Mexican hat and appropriate scaling values are considered to extract energy of the signal (Cheremisinoff, 2001). Higher variations in the signal leads to higher energy in CWT and this was used as a key parameter to handle the signal.

Wavelet transform of the signal is defined

$$W_{\psi}(s,\tau) = \int_{-\infty}^{\infty} f(x)\psi_{s,\tau}(x)\, dx \tag{1}$$

Energy can be defined as

$$E_{\tau} = \frac{1}{S_{max}} \int_{0}^{\infty} |W_{\psi}(s,\tau)|^2 ds \tag{2}$$

Attempts were made to filter out low energy portion of the signal for the change point analysis. It was assumed that cyclical portion of the signal would mean higher variation of the signal in time domain and hence higher energy in frequency domain. Fig.3 shows the extracted Energy by CWT on the signal.

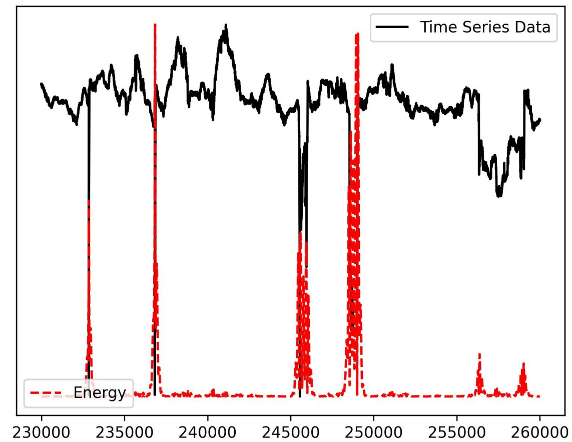

Figure 3. Energy representation of data extracted from CWT

Attempts were made to apply PELT methodology on the signal filtering out lower energy values in frequency domain and sensitivity was carried out with respect to different Penalty. It was observed that this has greatly improved the speed of execution of the method.

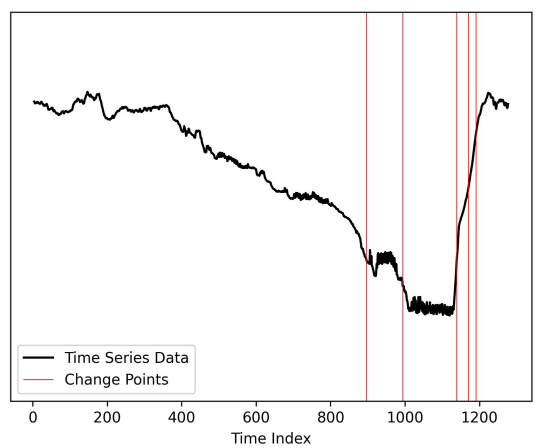

Figure 4. Change point detection on energy with penality 3

However this did not lead to fulfilling the criteria of accuracy of identification. Fig.4 shows the outcome, where results were still function of Penalty value. It is observed that some of detected points have either delay in detection or the early detection which are considered as the probability of false alarm or probability of miss detection, which effect the overall performance of probability of detection of the algorithm.

3.3. Change Point Detection Using Statistical Features

Statistical features such as mean and standard deviation gives the properties of the distribution of the data. Variation in the mean shifts the distribution and standard deviation will spread out the distribution. (Roberts, 2000), defined controls charts utilizing moving averages to detect small shifts in the process.

An attempt was made to extract additional statistical features of the signal and processing same in order to identify the change points. Signal when undergoing a higher variations would lead to higher value of standard deviation and the same is shown in Fig.5 where rolling standard deviation is high-

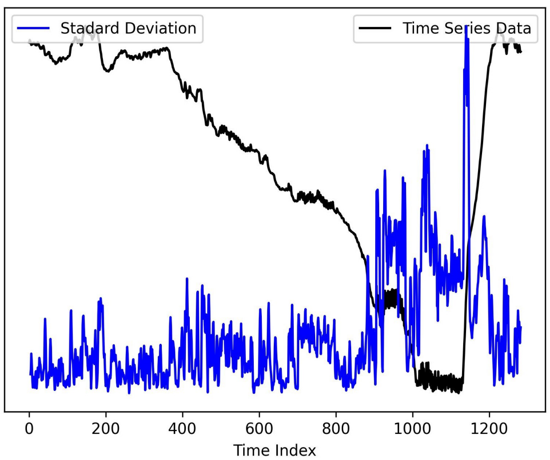

Figure 5. Rolling standard deviation

A methodology was defined in which rolling average of standard deviation and rolling average of mean was calculated. The choice of filter width has a great impact on the final results. The basic formula (for filter width L=M+1) is stated as follows.

$$\hat{x}[n] = \frac{1}{M+1}\sum_{k=0}^{M} x[n+k] \tag{3}$$

$x[n]$ and $\hat{x}[n]$ are raw data and process data respectively.

Rate of change of these properties were analysed with differential of the same calculated as

$$\text{Differential @ given time interval} = \frac{dx}{dt} \tag{4}$$

For the cyclic variations, the standard deviation of differential of these parameters are expected to be higher than other operating conditions. Change points were then identified based on the hypothesis testing, when standard deviation of these properties was higher than threshold, null hypothesis is rejected. An iterative approach was carried out for different rolling window and different threshold on standard deviation differential to identify the change points.

Fig.6 provides the outcome of this method for multiple thresholds for identification of change points. It is noted that the method is able to identify the change points, however it is not accurate due to below mentioned points
• Shift generated due to rolling average
• Sensitivity on Threshold definition criteria

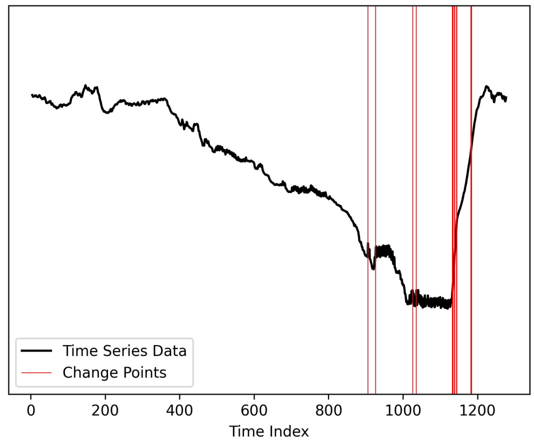

Figure 6. Threshold calculated by standard deviation of differential of rolling standard deviation and differential of rolling standard mean

4. Novel methodology to identify accurate change points

Due to the limitations of previous developed/available approach the problem was analysed with focus on each of the subsections.

4.1. Preprocessing the Signal

In this method, primarily the denoising of the signal is done. Rolling mean is one of the popular statistical methods to smoothen the signal, considering its drawback as discussed in section 3.3, here forward and backward moving averages on a variable rolling window was carried out to overcome the shift generated by uni-directional moving average

A novel methodology was defined which considers multi pass forward and backward moving averages on a variable rolling window to overcome the shift generated by uni-directional moving average and to smoothen the signal or to remove the noise in the time series data. Forward Moving Average algorithm replaces each original value by the average over its neighbour values. Refer eq.3

Backward Moving Average algorithm replaces each original value by the average over its neighbour values. The basic formula (for filter width L=M+1) is stated as follows.

$$\hat{x}[n] = \frac{1}{M+1} \sum_{k=-M}^{0} x[n+k] \tag{5}$$

The process is repeated multiple time to denoise the signal such that the originality of the signal is retained.

From the denoised signal, 2-point slope of the curve is calculated by using linear regression method

$$m = \frac{y_2 - y_1}{x_2 - x1} \tag{6}$$

and change points are defined when slope changes the direction from -ve to +ve and vice versa as shown in Fig.7

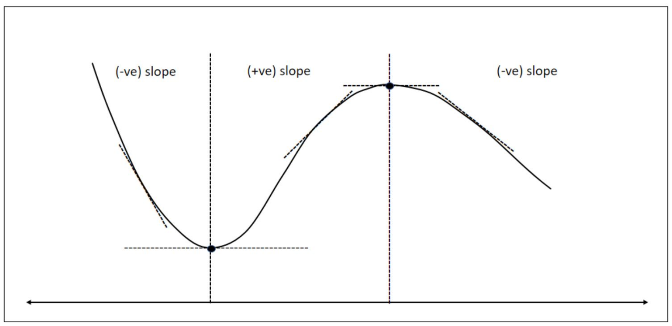

Figure 7. Positive negative change of slope

4.2. Filtering Change Points with Higher Severity

Once the change points have been identified, threshold analysis is required to filter out less severe change points. A broader scale of threshold would lead to many false positives and a narrow threshold would lead to missed detection.

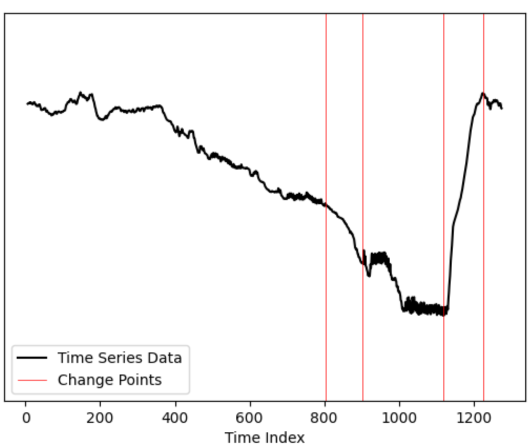

Figure 8. Detected change point using adaptive threshold

Also defining absolute value of threshold may not serve the purpose due to non standard nature of the cycle. An adaptive threshold dynamically analysing the slope variation was defined as a tunable parameter in the algorithm which adapts itself based on the nature of the cycle. Iterative analysis was carried to fine tune the range of adaptive threshold that would ensure proper detection capability.

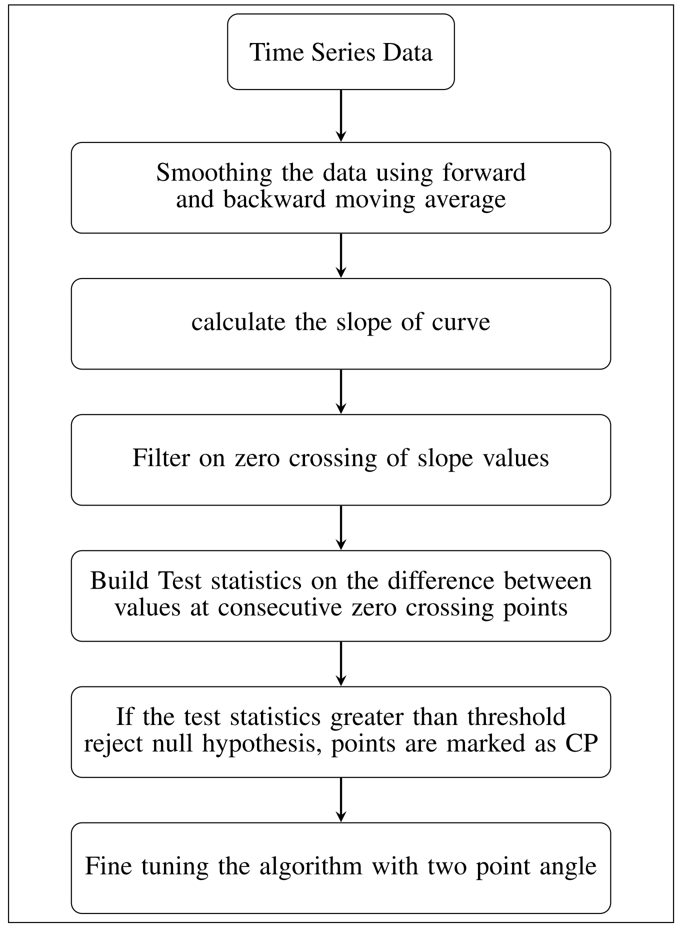

Figure 9. Flow chart for the proposed change point detection methodology

4.3. Fine tuning the algorithm with two point angle

During various internal validations, a minor shift in change point detection was observed. This was improved by evaluating the angle between consecutive points.

$$\theta = arctan \left(\frac{y_2 - y_1}{x_2 - x_1} \right) \tag{7}$$

This angle for each sample for the given cycle was further analyzed and a dynamic threshold was set to overcome the remaining minor shift. This lead to accurate identification of change points with negligible shift as seen in Fig.10

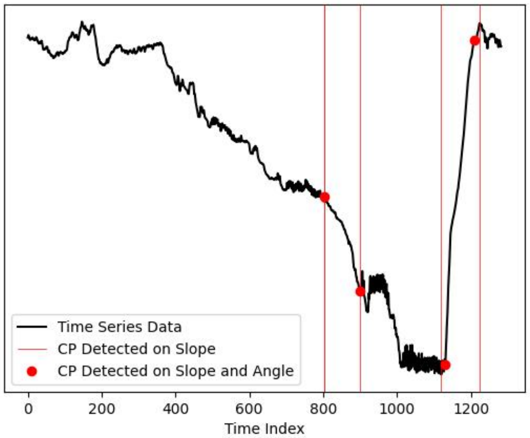

Figure 10. Change point detected on slope and angle

4.4. Symbolic Annotation

The time series signal shows an overall trend of either upward, downward or remain stable at different time interval. The nature of curve is based on the movement of the signal. This task is easily described visually but can become a time consuming, tedious, and repetitive task to search for the desired cycle. Symbolic representation of time series allows discretization of data into symbolic strings.

Here symbolic annotation is generated on the detected change points based on a tuneable parameter which considers differential values between two change points.
1. Positive differential indicates acceleration cycle (A)
2. Negative differential indicates deceleration cycle (D)
3. Differential value closer to zero threshold indicates cycle running stable (R)

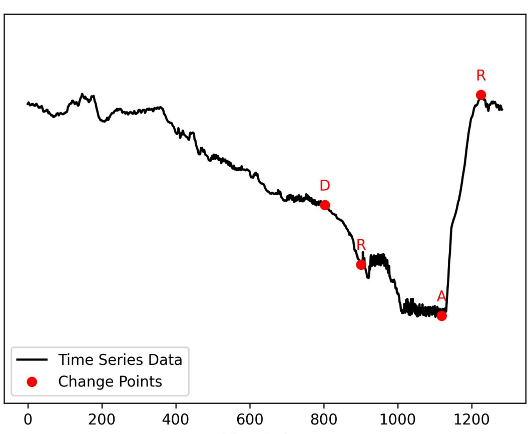

Figure 11. Symbolic annotation based on the event

The string that results from this step is a symbolic representation of attributes of the signal "DRAR". The frequency of the symbols can evidence repetitive patterns in the data.

4.5. Search Pattern

Once symbolic annotation of signal is done, the next step in the process involves searching for the desired pattern in the

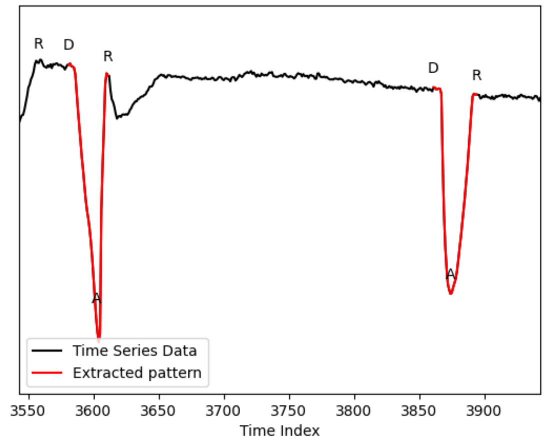

Figure 12. Search pattern

cedure returns the stamps at which positive matches occurred. Here in Fig.12, the intended pattern searched is "DAR" highlighted with 'red' in the signal.

5. Result Analysis

To summarize, the input signal is processed using forward and backward moving average window, change points are detected based on change in direction of slope of the curve, further improved with angle methodology, a symbolic annotation is applied, and a search is performed to retrieve the intended pattern

Actual indexes of change points	**Index identified by rupture package**	**Index identified by Novel approach**
803	380, 456, 547, 638, 815	803
901	881, 994	901
1131	1141	1131
1212	1185	1210

Table 1. Comparison of change point detected on Rupture package and novel approach

In order to validate this methodology on a variety of pattern, 100+ time series patterns of turbo machines including process upsets, start up, shutdown, trip, instrument anomalous behaviour etc. were considered. Table 1 shows the comparison of actual index of change points with respect to rupture package and our propose novel approach on a pattern highlighted in Fig10. As observed, rupture package was not accurate in identifying the exact change points and also had multiple false positives. Our proposed novel approach was quite accurate with minor shift in time index of change points. In order to deploy this methodology on an industrial application, various tests were carried out. Average computation time for 1 Year of operational data for a signal with 1 minute sampling interval was 70 seconds. For an offline / investigative analysis, this is not severe computational efforts.

In order to get real time predictive insights of turbo machinery operation, this methodology could be deployed on an IoT infrastructure with data processed in a batch of certain hours as time interval. Out tests on 12 hours batch of a signal with 1 minute sampling, required 5 seconds of computational efforts. For current state of the art industrial IoT infrastructure,this should not act as a severe constraint.

One of the application of this methodology includes identifying and comparing start cycles of turbo machines across different time period as shown in Fig.13

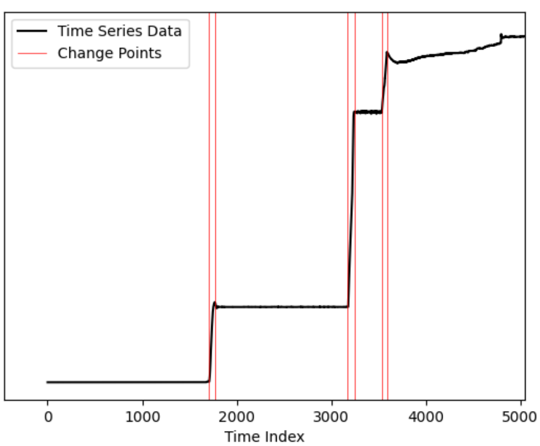

Figure 13. Startup cycle of Turbomachinery

6. Conclusion

In this paper, we discussed about importance of critical events in machinery life and build an algorithm to extract from transient operation data. This will enable new opportunity of physics based event extraction from time series data analytics. On top of it, the time series data is transformed into a more concise symbolic sequence and , a series of similar patterns are extracted and used for event extractions and identifying the characteristics of time series data which will have broad applications in the development of classification models and cycle diagnostics in machineries. This approach also open a new area of research to apply evolving technology like large language model.

References

Allegorico, C., & Mantini, V. (2014). A data-driven approach for on-line gas turbine combustion monitoring using classification models. In *Phm society european conference* (Vol. 2).

Cheremisinoff, N. P. (2001). *Condensed encyclopedia of polymer engineering terms*. Butterworth-Heinemann.

Iannitelli, M., Allegorico, C., Garau, F., & Capanni, M. (2018). A hybrid model for on-line detection of gas turbine lean blowout events. In *Phm society european conference* (Vol. 4).

Johnson, P., Moriarty, J., & Peskir, G. (2017). Detecting changes in real-time data: a user's guide to optimal detection. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *375*(2100), 20160298.

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), 1590–1598.

Neyman, J., & Pearson, E. S. (1933). Ix. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *231*(694-706), 289–337.

Roberts, S. (2000). Control chart tests based on geometric moving averages. *Technometrics*, *42*(1), 97–101.

Rupture python package. (n.d.).

Silva, A., Zarzo, A., González, J. M. M., & Munoz-Guijosa, J. M. (2020). Early fault detection of single-point rub in gas turbines with accelerometers on the casing based on continuous wavelet transform. *Journal of sound and vibration*, *487*, 115628.

Sofronov, G., Polushina, T., & Priyadarshana, M. (2012). Sequential change-point detection via the cross-entropy method. In *11th symposium on neural network applications in electrical engineering* (pp. 185–188).

Szczepankowski, A., & Przysowa, R. (2022). Thermal degradation of turbine components in a military turbofan. *Engineering Failure Analysis*, *134*, 106088.

Biographies

Saravanaram T is a Senior Engineering Manager and experienced in new product development, equipment design , services for industrial application at Baker Hughes, Bengaluru, India. He is passionate about solving real world challenges through applied digital. He received his master's degree in thermal engineering from College of Engineering, Guindy, Anna University. In his current role, Saravanaram T lead team of industrial domain experts and digital experts to resolve industrial plant technical issues responsively and proactively to increase uptime of fleet.

Unnat Mankad Unnat Mankad is a Senior engineer and experienced data scientist at Baker Hughes, Bengaluru, India. He received his master's degree in mechanical engineering from Birla Institute of Technology and Science (BITS), Pilani. In his current role, Unnat develops complex analytics for on-line prognostics, diagnostics, and predictive maintenance of Baker Hughes' icenter covered turbo machines. As expert of Gas Turbine and time series sensor data analysis, he also supports data science and machine learning driven algorithm developments focusing on emissions, reliability, and availability improvements.

Simi Madhavan Karatha is a Lead engineer and experienced product service expert at Baker Hughes, Bengaluru, India. She received her bachelor's degree in Mechanical engineering from Sardar Patel College of Engineering, Mumbai University. Simi has 7+ years of experience in Aeroderivative Gas turbine operations, performance and reliability improvement. In her current role, she handles technical industrial site issue resolutions, root cause analysis and drives data driven digital transformation catering to business needs.

Arati Halaki is currently working as a Data Analyst at Wissen Infotech Bengaluru,India and working on the client project at Baker Hughes. She received master's degree in Digital Communication from PES University Bengaluru. She has 2 years of research experience in signal processing for communication, cognitive radio networks and data science. She is also a author of IEEE conference papers and a IET communication journal.

Carmine Allegorico is a Principal engineer and experienced data scientist at Baker Hughes, Firenze, Italy. He received his master's degree in mechanical engineering from University of Napoli Federico II. In his current role, Carmine is a technical point of reference for the analytics discipline providing engineering guidance to other teams, helping to train new engineers and keeping abreast of industry trends and issues. He provides consulting during the development and implementation of advanced solutions for the on-line diagnostic and predictive maintenance, coordinates the creation of internal processes and support the adoption of new platforms and technologies.