# MODELLING OF TURBINE BLADE VIBRATIONS VIA COMPUTATIONAL INTELLIGENCE METHODS

*S. Norton[1], T. Ramsay[1], K. Karatzas[1], J. Fridh[2], P. Petrie-Repar[2]*

[1]Department of Mechanical Engineering, Aristotle University of Thessaloniki, Greece.
snorton@kth.se, ramsay@kth.se, kkara@eng.auth.gr

[2] Division of Heat and Power Technology, KTH Royal Institute of Technology, Sweden.
jens.fridh@energy.kth.se, paul.petrierepar@energy.kth.se

## ABSTRACT

**A method for modelling turbomachine blade vibration events is proposed, based on computational intelligence algorithms. The method utilises steady thermodynamic data and blade tip-timing data to identify high amplitude vibration events and to draw underlying relationships between steady-thermodynamic input channels and resultant blade motion characteristics. Several computational studies probe specific process aspects in order to improve model prediction accuracy and several methods of data-feature reduction are established to further enhance vibration predictions.**

**Overall, the study shows promise of what prediction capabilities can be achieved with seemingly limited instrumentation. Drawbacks in matters of tip-timing interpretation, quality/quantity of data and process limitations are discussed. Consequential future objectives are outlined to envisage onward predictive accuracy.**

## KEYWORDS

Blade Vibration Prediction, Blade Tip Timing, Computational Intelligence, Health Monitoring.

## NOMENCLATURE

| | |
|---|---|
| BTT | Blade Tip Timing |
| CI | Computational Intelligence |
| EO | Engine Order |
| HCF | High Cycle Fatigue |
| KTH | Kungliga Tekniska Hogskolan (Royal Institute of Technology, Stockholm) |
| RPM | Revs per Minute |
| SOM | Self Organizing Maps |
| t-SNE | t-distributed Stochastic Neighbour Embedding |

## INTRODUCTION

Blade vibration has always been a problematic aspect of turbomachinery, leading to machine downtime and high maintenance costs. Statistics on industrial gas turbines show that 28% of the primary causes of failure are related to turbine blades and rotor components (Bloch, 1982). Furthermore, over an 8-year period, Dundas (1994) showed for gas turbines that high cycle fatigue (HCF) of turbine blades accounted for 12% of damage costs and an additional 12% of damage costs were accrued by HCF of compressor blades; similar studies extended into the steam turbine sector (Mazur et al. 2006, Stange, 2012, Ziegler et al. 2013).

Although various characteristic data (thermodynamic and otherwise) is collected in operation through machine monitoring systems, it is the amplitude and frequency of blade vibrations that are prime factors of incurred problems, as they determine the amplitude of the alternating stress and hence

OPEN ACCESS
Downloaded from www.euroturbo.eu
Copyright © by the Authors

blade damage. Of course this data is not commonly collected, but both parameters can otherwise be extracted with techniques such as blade tip-timing (BTT). This paper focuses on characterising vibrational measurements of turbine blades as a function of steady-state thermodynamic data via computational intelligence (CI) techniques in order to develop a blade vibration prediction model. A simple metric of 'measured blade vibration amplitude' is utilised to identify vibration events, but this can be extended and potentially annexed to further robust fatigue models, to give real-time conditional monitoring of blade health.

Initially, BTT data is pre-processed to clearly identify the vibration events. Then, discretisation into practical data points is required to aid computation. As a next step, data dimensionality reduction techniques were applied to aid in visualisation of the synchronised dataset, and a procedure is suggested to prioritise the most important features. The resultant dataset is used to model recorded blade vibration levels by employing CI algorithms, using selected thermodynamic data as model inputs. Cross-validated modelling (Kohavi, 1995) allows for models to be trained and tested on the limited amount of data available: 10 equally sized subsets of data were produced via random sampling, each set containing 10% of all data. Then, 9/10 sets are used for training, while the remaining one is left for testing. This is repeated a total of ten times, each time using a different subset as the test set, leading to 10 classifiers. These 10 classifiers are averaged to give the final model. The predicted values are then cross-checked by means of correlation coefficient and error measurements. In pursuit of better vibration predictions, several computational studies into variations of initial data conditioning and reduction methods are then conducted to explore their effects. To note, this paper is built upon the research conducted in parallel by the two primary authors as part of the THRUST Erasmus Mundus joint Master's programme. The idea is to convey the potential usefulness of the method by providing enough detail to demonstrate its function. More detailed discussion of methods used can be found in the representative thesis documents (Norton 2016, Ramsay 2016).

## ORIGINAL EXPERIMENT AND DATA

Modelling is based on experimental data made available by the TURBOPOWER project (sponsored by the Swedish Energy Agency). Using the KTH test turbine, the original experiment was performed on a newly commissioned blisk developed by Siemens. The goal was to detect numerically predicted synchronous resonances and identify safe operating regions free of high blade vibration.

At six operational stages (time ranges, viewable in Figure 1) of the experimental run, blade vibration measurements were acquired using the Agilis optical blade tip-timing (BTT) measurement system (Agilis 2016), which identified BTT amplitudes for various increasing and decreasing sweeps in RPM and load types. These sets of blade amplitudes were used to train the CI algorithms in order to create the predictive ability of the proposed models. Naturally, sampling rate relies directly on the speed of operation but is of the order of 100 Hz. A total of 111 individual probes of the 'Steady Operational Monitoring System' recorded varying performance data of the test turbine at a rate of 0.5 Hz. Note that the discrepancy in sampling rates is addressed later by means of interpolation. The experiment was successful in identifying several synchronous resonance crossings; most notably the Mode 1 and EO42 crossing, shown in the Campbell diagram in Figure 2 that produced excessive levels of vibration with potential to cause serious damage at extensive operation.

Additionally, a source of uncategorised asynchronous vibration was detected throughout the testing. Overall, distinct vibrational events are evident within the collected data that are able to be identified and predicted by the methods detailed in this work. Note that Figure 1 shows data from a single blade. For simplicity, all studies conducted herein are regarding this. However, the authors acknowledge that with consideration to multiple blades, averaging or even blade differences could add further value to predictive capabilities.

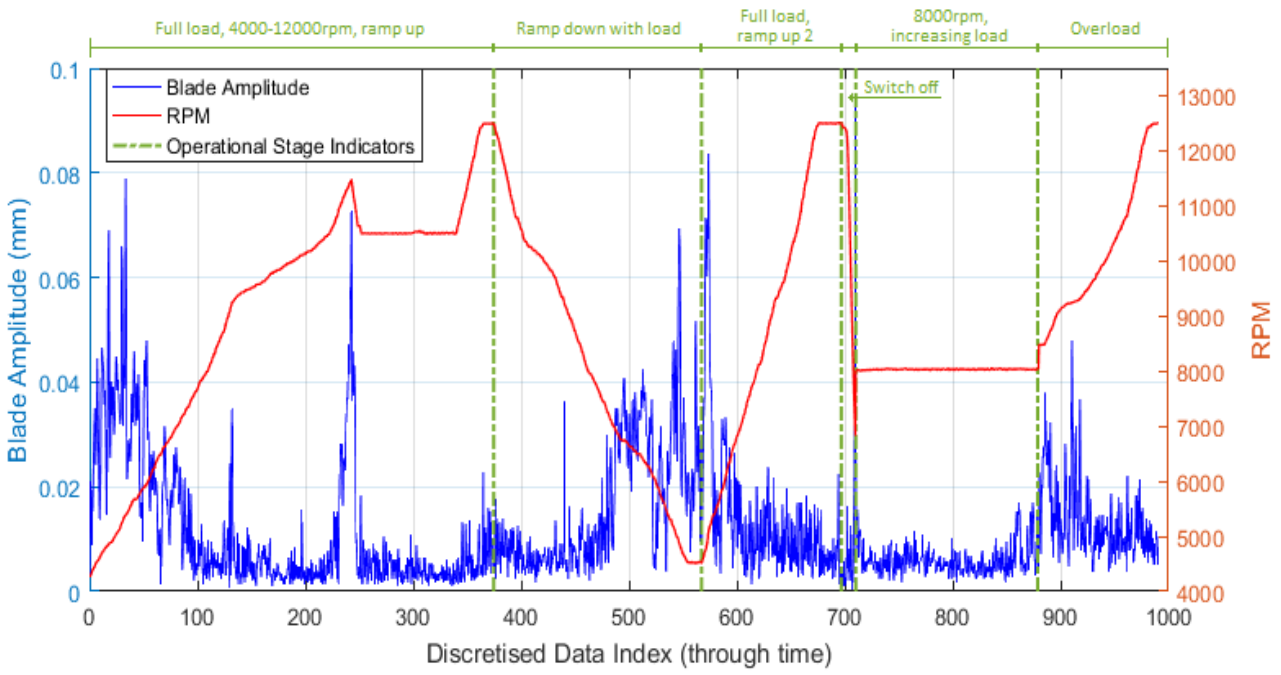



**Figure 1: Superposition of machine RPM and measured blade amplitude through time (blade #2, probe averaged, envelope discretisation, window size 50).**

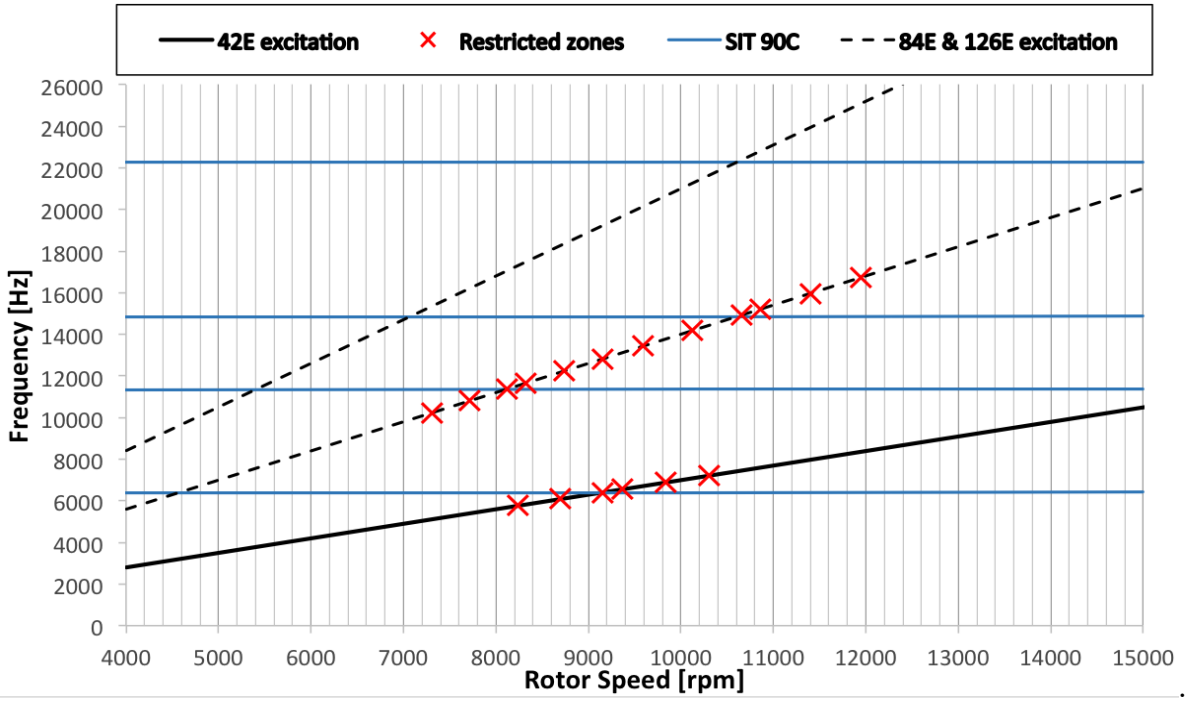



**Figure 2: – Campdell diagram showing crossing of interest at 42E excitation.**

## METHODS

### Data Pre-Processing

Synchronisation of the BTT and thermodynamic data series was not a priority during the original experiment: data was recorded on two computers with differing internal clocks, resulting in the inability to carry out a time-based synchronisation. Fortunately, due to the nature of the experimental

test runs, the intermittent increase and decrease of rotational speed was captured across all data recording systems. Consequently, a methodology was developed to match the time-series of data points based on matching RPM patterns in order to achieve synchronisation of the BTT and thermodynamic data (Norton, 2016).

Different data sampling rates was another issue, resulting in large differences in the number of data points between the two datasets. To rectify this, the sampling rate of the thermodynamic data was artificially increased by spline-interpolation (Hazewinkel et al., 2001) between the synchronised points. Some obvious levels of noise in some of the thermodynamic data channels were dealt with via a 5-point triangular sliding average function a priori (Smith, 1999). This did not affect larger scale values and trends. Finally, Butterworth (1930) signal filtering was performed over the BTT data series to reduce noise while preserving resonant vibration events as best as possible (Norton, 2016).

**Data Discretisation**

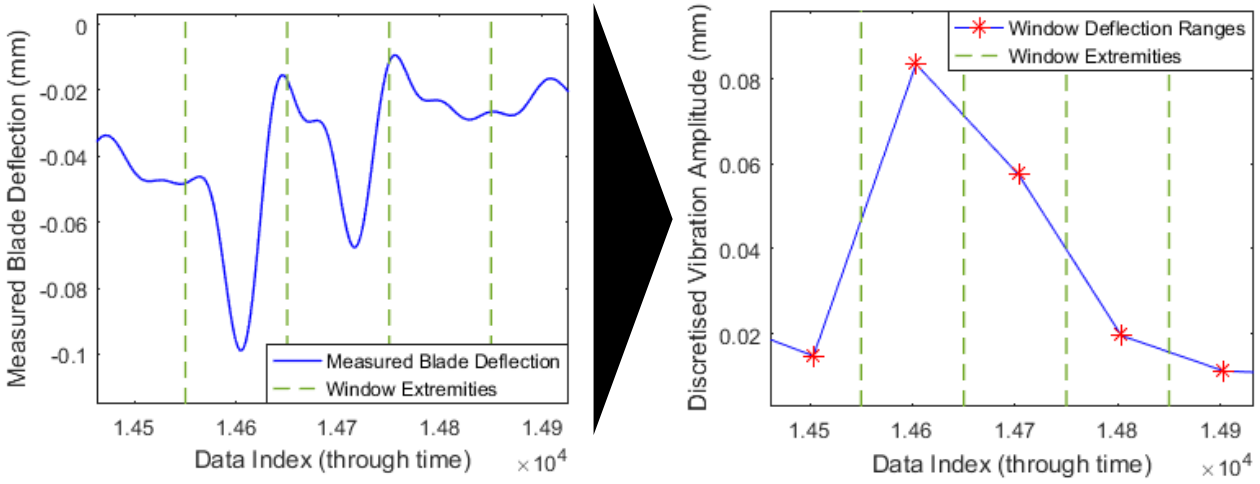
Two separate techniques were developed to extract the measured blade amplitudes within a user defined time window. Importantly, discretisation is necessary due to the cyclic nature of the TT signal: no individual TT data point contains an accurate description of vibration amplitudes at that time. This discretisation must also be applied to the associated steady operational data channels.

'Windowing' refers to the approach of selecting a number of data points in time, then calculating parameters of interest within that to be appended as a single instance into a new refined data series. The displacement range was determined within each window for the BTT data via the "Envelope" method and the "Direct Range" method.

The Direct Range method calculates the range within a given window by taking the difference between the minimum and maximum displacement. In comparison, the Envelope method creates two spline curves around the upper and lower peaks in the data series. Within a given window the difference between the two curves is determined, resulting in a set of ranges for that window (window size being "manually" selected by the user). The final amplitude range value for that window is taken as the average of the calculated set. The use of differences results to the fact that (absolute) negative vibration displacements may lead to positive discretised vibration amplitudes. Basic illustrations of both methods are shown in Figure 3 and Figure 4. It should be mentioned that the tip-timing method does not have a sampling ratio high enough to capture physical blade vibrations - rather it is only an asynchronous 'shadow' of them. As Diamond et al., 2014, point out: "based on the ToA difference between a vibrating and non-vibrating blade, the tip displacement is sampled at each proximity probe. This sampling rate is completely dependent on the shaft speed as well as the number and spacing of tip-timing probes (i.e. if the rotor is turning at 25 Hz and one has 4 proximity probes, the effective sampling rate is 100 Hz). As a result, the extracted signals are inherently sub-sampled and conventional signal processing techniques such as the Fast Fourier Transform (FFT) have limited use. Thus, alternative algorithms need to be developed to extract blade vibration information".

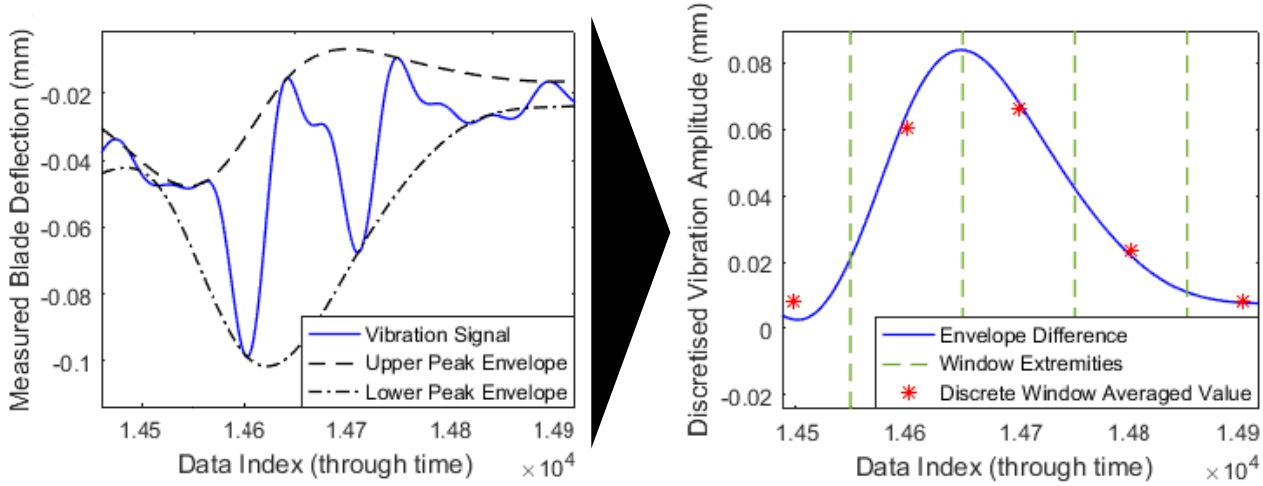
**Visualisation via Dimensionality Reduction Techniques**

Data visualization is a powerful tool for analysing and understanding the structure of a multidimensional dataset, such as the one addressed in our study. However, due to the commonly high dimensionality of experimental data it is often impossible to directly visualize a dataset. Thus, dimensionality reduction methods were employed in order to represent the data in a lower dimensional space, ideally two or three dimensions. Three unique dimensional reduction methods were investigated in this thesis: Self Organizing Maps (SOM), Sammon Mapping, and t-distributed Stochastic Neighbour Embedding (t-SNE). At a top level, all of the techniques are nonlinear, but each one clusters the data by minimizing a different cost function, as discussed in detail by Ramsay (2016).




1. Measured vibration signal is discretised and local peak values identified.
2. Ranges between local peak values are calculated within each window.
3. Local ranges are then appended as the BTT discretised series.

**Figure 3: Direct-Range discretisation procedure for BTT data.**




1. Spline envelope functions calculated above and below vibration signal.
2. Difference between two envelopes calculated.
3. Specified window discretisation.
4. Local mean difference value appended as the BTT discretised series.

**Figure 4: Envelope method discretisation procedure for BTT data.**

SOM are neural networks composed of a two-dimensional array of randomly weighted neurons. As data points are passed through the neural network, they are matched with a winning neuron, causing the network topology to adjust and eventually form clusters of similar attributes. (Kohonen, 1997). The unified distance matrix (U-matrix) is used as the basic visualisation method for SOM. It represents the Euclidean distance between neighbouring neurons (data) and thus depicts the relations between neighbouring data.

Sammon mapping is a nonlinear variant of a traditional Principle Component Analysis (PCA) (van der Maaten et al. 2009). Euclidean pairwise distances are determined in the high dimensional space then mapped to a lower dimensional space while preserving the high dimensional topology as best as possible (Sammon, 1969).

The third nonlinear dimensional reduction technique used is the t-SNE method, whose metric for data point similarity is based on probability rather than Euclidean distance. A Student t-distribution is defined for each data point and an optimal mapping is determined based on the conditional probability that one data point would pick another data point as its "neighbour" (van der Maaten et al. 2008).

**Prioritisation of Data**

In the context of this paper "prioritisation" refers to a selection process for determining which thermodynamic sensors are deemed to be the most important among the ones used in the monitoring process. Thermodynamic sensors are located at various axial, radial, and circumferential locations, and many of their measurements are highly correlated to one another and share similar correlations to the amplitude of blade vibration. This suggests that a limited set of data, coming from specific sensors, may be considered as representative for the data set under investigation and can thus be used without hindrance to computational modelling. The prioritisation procedures were developed taking into account the visualizations produced by each dimensional reduction technique. Using these procedures (Ramsay, 2016), a set of prioritised datasets was created for each dimensional reduction technique, then numerically evaluated using CI-based algorithms.

**Numerical Modelling**

A computational study was conducted to detail a sensitivity analysis of 11 different computational intelligence algorithms considered for this study, by their ability to make vibration predictions from the overall (synchronised and refined) data set. The amplitude values were the target of the predictions, and they were produced exclusively by the trained prediction algorithm which had no access to the measured values. In this way, vibration amplitudes are being "predicted" purely by the algorithm looking at the thermodynamic data series.

Each one of the algorithms was used to construct a model that was then tested in terms of its ability to forecast the vibration data by using as inputs all other data that were available in this study. The study was conducted using the machine learning environment WEKA (Hall et al. 2009). Among algorithms tested , KSTAR (Cleary et al., 1995), Random Committee (Hall et al., 2009) and Random Forest (Ho, 1995) were found to be the most successful in this case (Norton, 2016) and their results are presented in this paper.

To monitor and compare success of the models, two key performance metrics are considered: Correlation Coefficient (Equation 1) is calculated to represent the linear interdependence of the measured and predicted vibration variables; Relative absolute error (Equation 2) is the average of the numeric difference between predicted and measured data points, divided by the magnitude of the expected (measured) value. Although not greatly quantified, the authors also make references to "time to build numerical models" as well as "total time of prediction process" or "sub-processes".

$$r = \frac{\sum_i (p_i - \bar{p})(a_i - \bar{a})}{\sqrt{\sum_i (p_i - \bar{p})^2} \sqrt{\sum_i (a_i - \bar{a})^2}} \qquad [1]$$

$$RAE = \frac{\sum_i |p_i - a_i|}{\sum_i |a_i - \bar{a}|} \qquad [2]$$

**Flowchart of Process**

To aid understanding, a flowchart of the entire prediction and feedback process including the methods, the algorithms and the calculated values (vibration attributes based on thermodynamic data) imperative to each step, is visualised in Figure 5. The loop closes with the comparison of predicted to the measured vibration, via the "Modelling Performance" parameters (correlation coefficient and mean absolute error).

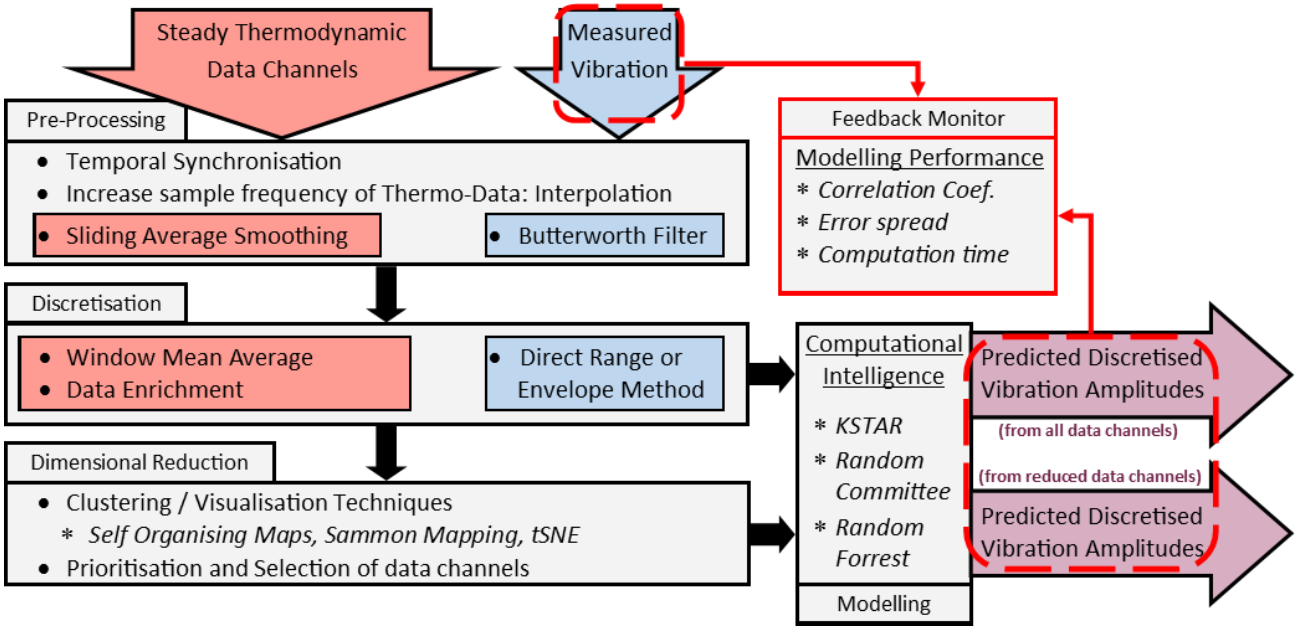



**Figure 5: Flowchart of Prediction Process.**

## COMPUTATIONAL STUDIES

A number of studies were conducted in order to investigate the influence of the various preprocessing, dimensionality reduction and prioritization methods on the forecasting ability of the models that were developed. This will help in the further development of the methods being presented. Indicative results are presented in the next subchapters.

### Effect of Window Size on Numerical Modelling

An investigation was conducted into the effects of discretisation window size on the computational modelling of vibration events, searching for trends in correlation coefficient and relative absolute error. The intention was to find the best window size setting for improving vibration predictions. Both the direct-range and envelope methods were considered (Norton, 2016).

### Effect of Discretisation, Dimensional Reduction and Prioritisation on Numerical Modelling

An investigation was conducted into the effects of discretisation and dimensional reduction-prioritisation of the combined BTT/thermodynamic dataset on vibration predictions. Considering both the Direct Range and Envelope discretisation methods, the study compares the modelling performance of three CI-based algorithms (KSTAR, Random Committee and Random Forest) in combination with SOM, Sammon mapping, and t-SNE dimensional reduction techniques. Data visualisation and prioritisation are performed for all test cases, resulting in a set of eighteen prioritised models (Ramsay, 2016).

## COMPUTATIONAL RESULTS

### Effect of Window Size on Numerical Modelling

Due to the requirement of the Direct Range method having to consider at least one full cycle of measured vibratory oscillation (~90-100 points), data points of particularly small window sizes must be disregarded. Figures 6 & 7 show a common trend for both discretisation methods: as discretisation window size is reduced, prediction correlation coefficient increases while relative absolute error decreases. Two main contributing factors can be attributed to this: Firstly, as window size is decreased, more discretised data points are available to train the model, thus reducing uncertainty; secondly, contemplating Figures 3 & 4, reducing window size also increases the resolution of the “max amplitude” of a vibratory oscillation that is captured with both methods.

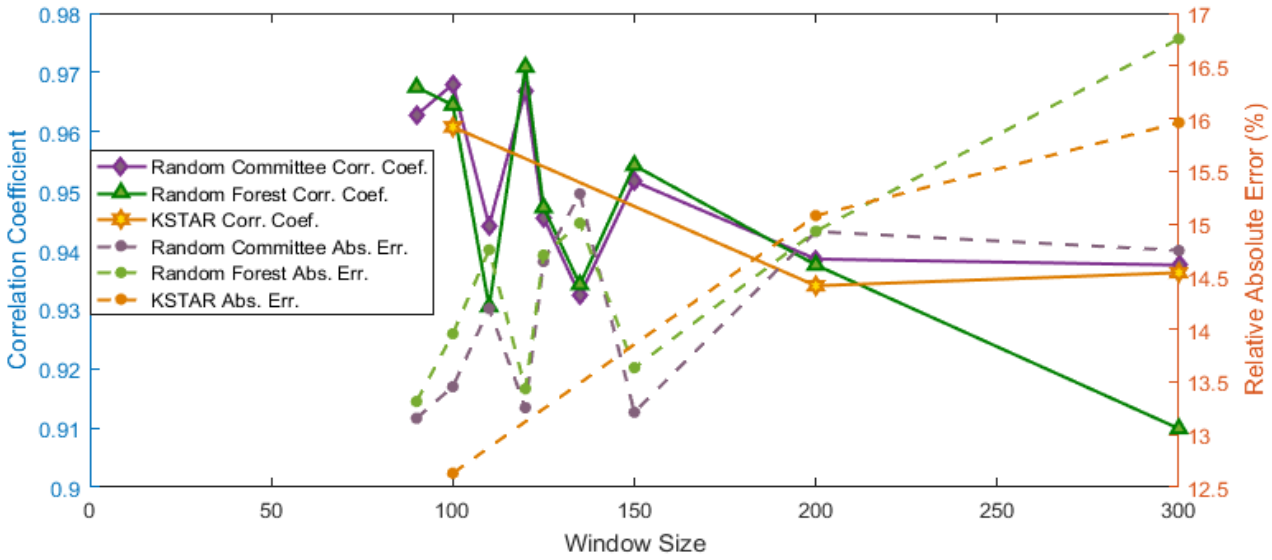



**Figure 6: Effect of discretisation window size, on vibration prediction accuracy of the three computational intelligence methods, for the "Direct-Range" discretisation method.**

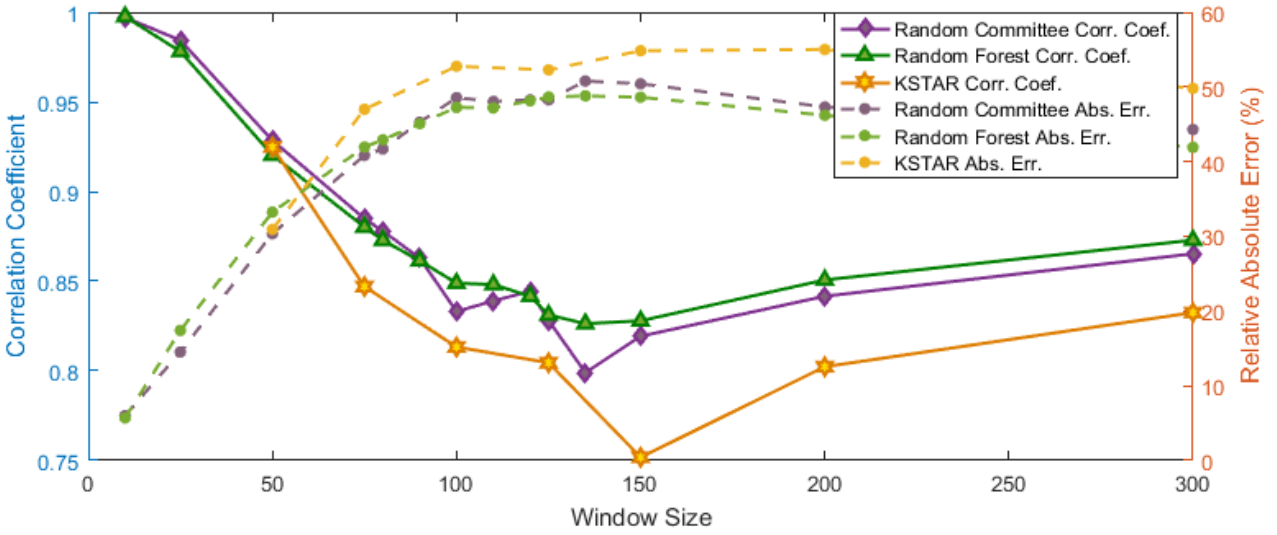



**Figure 7: Effect of discretisation window size, on vibration prediction accuracy of the three computational intelligence methods, for the "Envelope" discretisation method.**

The Envelope method (Figure 6) seems to converge more steeply (correlation coefficient to one and error to zero) when approaching a window size of zero, whilst displaying poorer performance values than the Direct Range method at higher discretisation window values. This is attributed to how the Envelope method requires a higher window resolution to sharply define vibration events: this can be seen by comparing figures 3 & 4. The Envelope method, however, is ultimately able to lead to better vibration prediction performance as it can reduce the discretisation window size adequately in order to effectively depict vibration events. Instability in the region of window size 100-200 in Figure 6 is speculated to arise from partial capturing of blade vibratory oscillations by the Direct-Range method (please note that figures 6 and 7 demonstrate different y-axis scaling to allow for the optimum visualisation of data values and their trends rather than homogenize relevant scales).

**Effect of Discretisation, Dimensional Reduction and Prioritization on Numerical Modelling**

The clustering results shown in Figure 8 are organised into four main groups based on thermodynamic data channel type and location: a temperature sensor group, two pressure sensor groups, and a group for other operational parameters (i.e. rotational speed, power, etc). The first pressure group (Pres Group1) includes midstream pressure sensors, while the second pressure group (Pres Group2) includes pressure sensors mounted in the casing of the test turbine. To better interpret these results it is important to understand that a data cluster forms when the weight vectors of each data channel are similar to one another, suggesting correlation between those data channels. A point that does not lie within a cluster suggests that the information from that data channel is unique, and potentially valuable (or an outlier). It is also important to note that the axes in the plots of Figure 8 represent arbitrary weight vector coefficients used to compare data channels. The cause for disparity between scales used for t-SNE and Sammon mapping is due to the difference in which each method minimizes its cost function when comparing weight vectors (Ramsay, 2016). Overall, the axes of Figure 8 should be interpreted as representing differences (dissimilarities) between n-dimensional data points and their clusters that are repulsed from dissimilar clusters: the highest the distance in t-SNE, the highest the dissimilarity between data points.

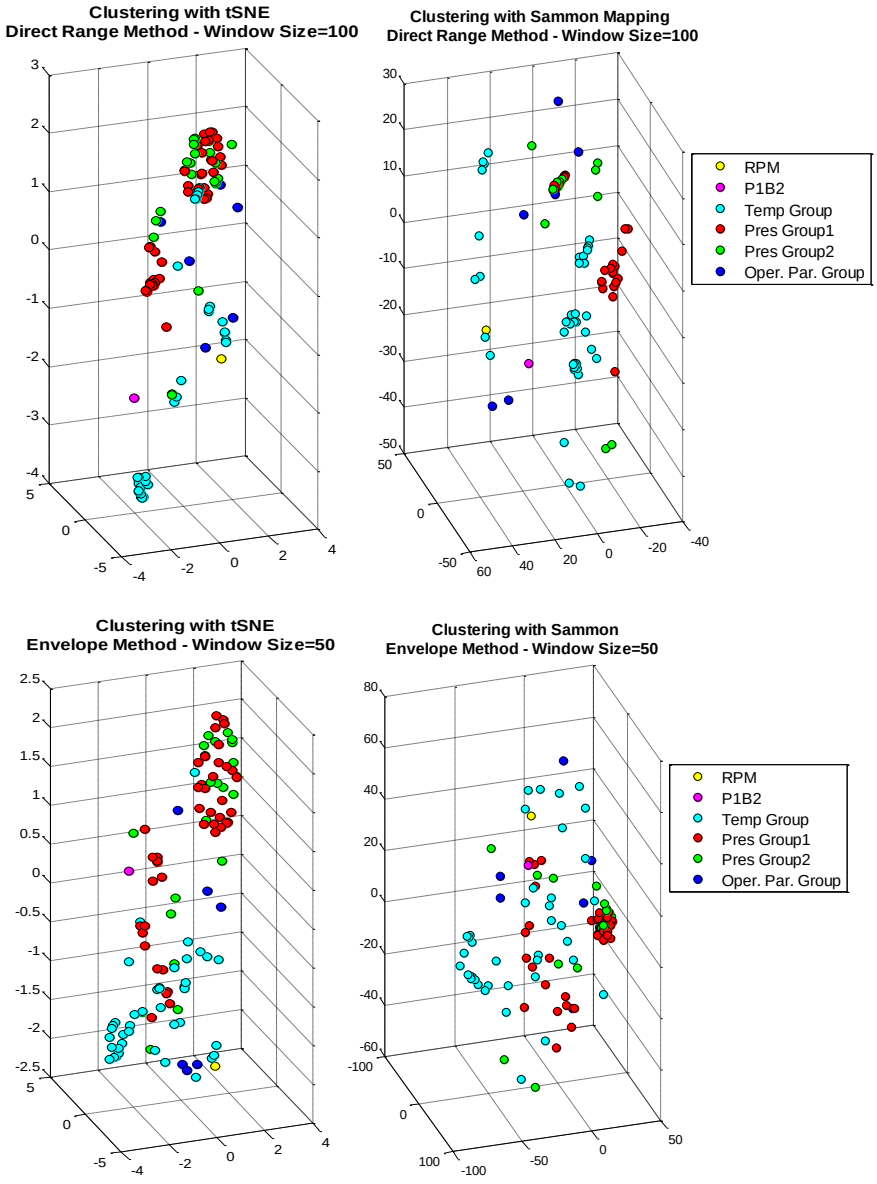



**Figure 8: Direct Range v. Envelope method: Clustering for t-SNE and Sammon mapping.**

For both discretization methods the results showed that the t-SNE method was superior to Sammon mapping and SOM for visualizing data as it results in more well structured clusters. The only downside to this method is that feature prioritization for datasets with many variables can be quite cumbersome due to the requirement of manually defining clusters after visualization.

The U-matrices from Figure 9 presents with the distances between neighbouring data points: the highest the distance, the more well-structured the cluster of points (thus the U-matrix may be regarded as a visualization of the "fence height" between data, and thus "well fenced" areas can be considered as clusters of similar data points) . On this basis, Figure 9 clearly depicts the difference between the Envelope and Direct Range discretisation methods in regards to data visualisation and clustering: low values (commonly visualised with the aid of dark blue colours) represent tight clusters of data, and high values (commonly visualised with yellow-red colours) represent a clear separation between neighbouring neurons. SOM clustering results showed that the Envelope discretisation method was less effective in producing meaningful data visualisation as the resulting clusters are less clear in comparisons to the ones produced by the Direct Range method. For both discretisation methods it's possible to identify three main clusters in the dataset. However, each method results in a unique topology for the three main clusters. Arguably, within the large central cluster, the Direct Range method better defines potential sub-clusters, whereas any sub-clusters are nearly indistinguishable with the Envelope method.

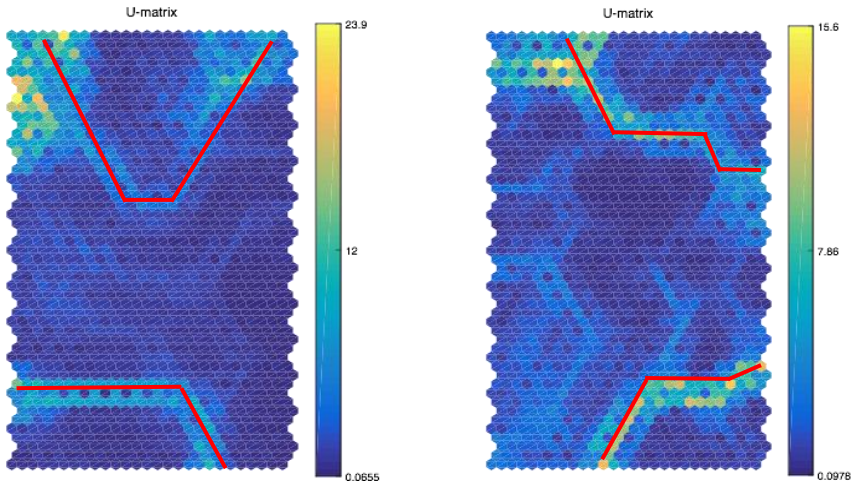



**Figure 9: SOM U-matrix comparison of Envelope (left) and Direct Range (right) methods. Three main clusters are identified with the red-coloured line. Axes are analogous to the number of neurons used for the SOM transformation and have no physical meaning.**

CI-based modelling was applied to each of the prioritised datasets (SOM, Sammon, and t-SNE), results of which are reported in Figure 10 as percentage difference from a non-prioritised dataset that includes all thermodynamic attributes. In most cases, the correlation coefficient diminished and the relative error increased as a result of prioritisation depending on the dimensional reduction technique. However, the largest reduction in correlation coefficient was by only 1.2%, with a corresponding increase of 9.5% in relative error. Most notably, some of the prioritised datasets slightly outperformed the non-prioritised dataset. Of those that performed better than the non-prioritised dataset, the relative error increased by only 3% on average. This implies that a subset of parameters can effectively be used for predicting vibration behaviour with acceptable accuracy. The reason this is possible is due to the fact that the reduced data set only retains the attributes deemed to be most important (i.e. those with highest determined correlation to the discretised blade amplitudes), while uncorrelated attributes are eliminated, thereby eliminating unwanted "noise" that negatively affects the modelling performance.

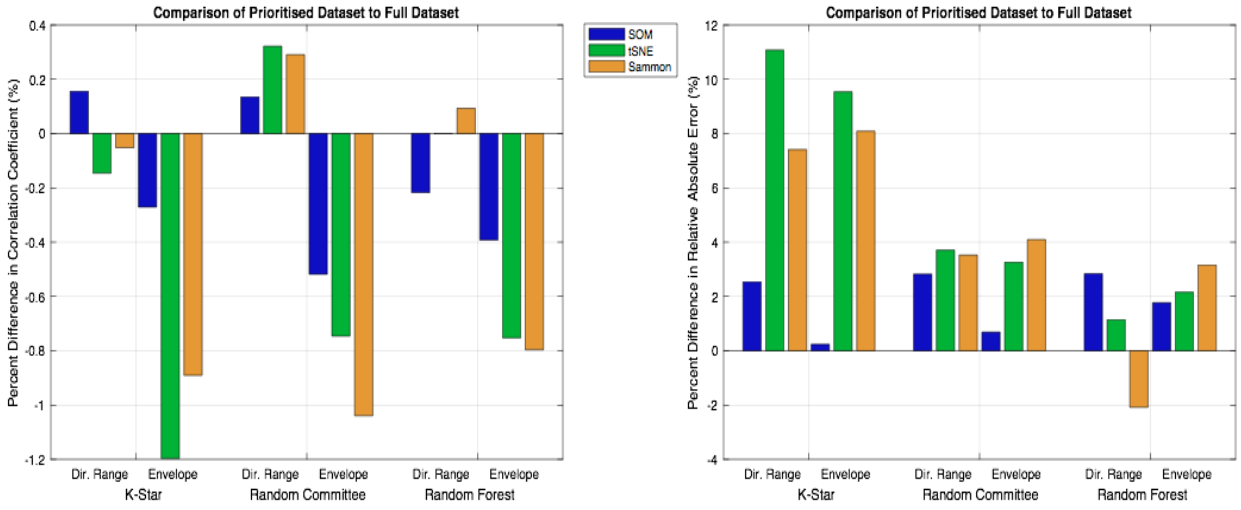



**Figure 10: Comparison of model performance of reduced datasets represented as a percent difference from the full dataset for different CI algorithms and discretisation methods. Correlation Coefficient (left); Relative Absolute Error (right)**

Of the two methods, the datasets discretised with the Direct Range method consistently outperformed those discretised by the Envelope method; however, both methods resulted in similar changes in relative error.

In regards to which dimensional reduction technique produced the best performing dataset, the results are mixed. The optimal prioritised dataset differs depending on which computational intelligence algorithm is applied. The optimal combinations appear to be SOM analysis with the KStar algorithm; t-SNE with the Random Committee algorithm; and Sammon mapping with the Random Forest algorithm. However, for all dimensional reduction techniques, the Random Committee algorithm produced the highest correlation coefficients with minor increase in relative error. Based on these results, the authors assumed the combination that generates the "best case" model to be the Direct Range method with a window size of 100 data points, dimensionally reduced with the t-SNE method, and trained using the Random Committee classification algorithm. This model was used to create a set of predicted blade amplitudes for verification against the measured blade amplitudes, results of which are shown in Figure 11.

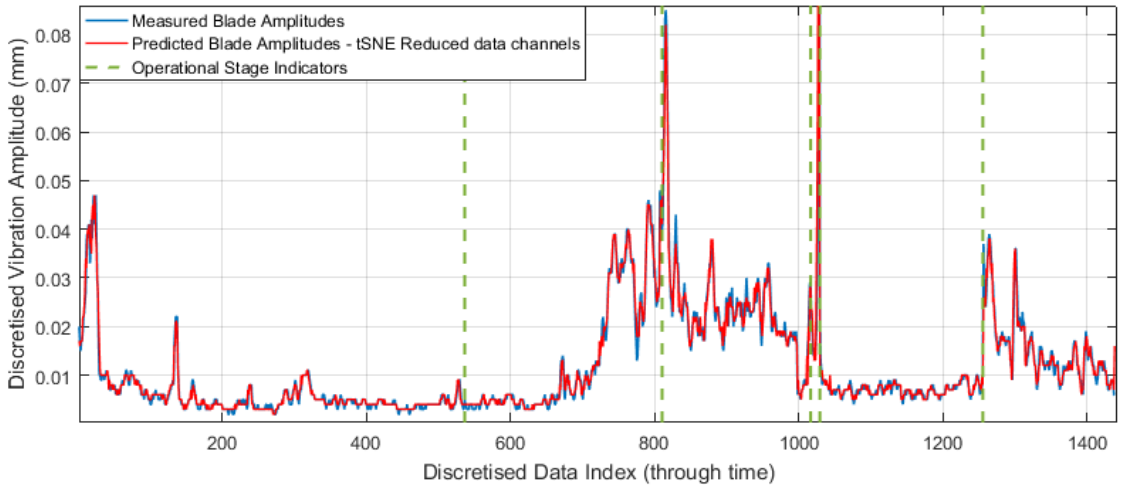



**Figure 11: Comparison of "best case" model with measured blade amplitudes.**

The model produces a signal with a correlation coefficient of 0.965 and relative absolute error of 13.9%, resulting in strong agreement between the predicted and measured values. Overall, the

predicted blade amplitudes are slightly lower than the measured blade amplitudes, suggesting that the model would be slightly underconservative in practice. However, safety factors can be applied in order to err on the side of conservatism. One important thing to note is the absence of a peak at a discretized data index of 250, when compared to the probe-averaged signal in Figure 1. The absense of this peak in Figure 11 is likely a result of using a single probe and a single blade to perform the predictions. When using a single probe and blade to train the model it is quite possible the blade's maximum deflection could go undetected due to phasing; i.e. the blade is always in a quasi-neutral position each time it passes by that probe. However, the probe spacing is optimised such that the blade will not be in the same position (i.e. at a different phase) when passing a subsequent probe. The authors were aware of this phenomenon, but for the sake of simplicity to prove the feasibility of such a model, only the data from one probe and blade was used for model training and testing.

## DISCUSSION

The overriding aim of the presented work is to develop a method of conditional monitoring that directly forecasts vibration related turbomachine blade parameters in order to assist in monitoring and maintenance scheduling. Initial blade deflections captured via blade tip timing are synchronized with thermodynamic data measurements and are then discretised into amplitudes. Then, thermodynamic data are used to model and predict the aforementioned amplitudes, reaching a correlation coefficient of 0.965 in the best achieved case. Alternative discretisation methods and parameters are discussed, while the use of a subset of the initial thermodynamic data is deemed to be appropriate as it leads to a satisfactory modelling performance on the basis of a smaller input data set.

In full, such a method would require additional implementation of fatigue models that consider the effects of vibration events at the material level. In turn, this would require the input of physical blade motion characteristics, including accurate blade amplitude as well as frequency content. In its current stage of development, the method presented is limited only to identification of vibration events. However, it should be regarded as first step, providing a framework for a comprehensive damage analysis.

Throughout this work, CI-based models were trained on BTT data that had been conditioned to identify clear vibration events. However, secondary analyses depending on the type of vibration are required to extract the physical parameters of amplitude and frequency; such as the "least-squares fitting method" (Przysowa, 2014) for single-synchronous crossings and autoregressive based frameworks (Gallego-Garrido et al. 2007) for asynchronous and multi-synchronous events.

## CONCLUSIONS

This work has described a framework, in which, dissimilar data sets describing seemingly unrelated system parameters can be associated temporally with consideration to differences in measurement and sampling quality. The intended end-result was to be able to predict blade vibration events (measured by a BTT system) by considering only steady operational metrics of a test turbine via CI methods.

A study into the ideal window size for each discretisation method was conducted and results contrasted. It was determined overall that smaller window sizes were better as more data points became available for consequent modelling and that a finer refinement gave more physical detail to vibration events. The two discretisation methods were also compared and strengths and weaknesses identified.

Three unique dimensional reduction techniques were studied to investigate their effectiveness at visualizing the underlying structure of the combined dataset. The datasets were then prioritised based on importance. Lastly, CI-based algorithms were applied to study the effectiveness of the prioritised datasets in predicting vibration events from a single blade of the experimental blisk. Results surpassed a correlation coefficient of 0.95, suggesting a very good model performance that may support operational conditional modelling.

Finally, considering future works, umbrella questions over the physical application of the methodology described in this paper must also be addressed. For example, investigations into the

amount of computational training required for a model's results to be considered optimal; how models change their predictions with increasing data input; and integration of the prediction method to multiple machines accounting for additional attributes such as mass and geometry. However, the generalisation potential of related models needs to be further examined.

**ACKNOWLEDGEMENTS**

The authors acknowledge Siemens Industrial Turbomachinery AB for providing access to the data used in the present paper.

**REFERENCES**

Agilis, (2016). Agilis Measurement Systems Homepage. Available at: http://agilismeasurement systems.com/

Bloch, H.P., (1982). Improving Machinery Reliability in Process Plants, Volume I. Houston, Texas: Gulf Publishing.

Butterworth, S., (1930). *On the Theory of Filter Amplifiers.* In Wireless Engineer (also called Experimental Wireless and the Wireless Engineer), Vol. 7, Pages 536–541.

Cleary, J.G. and Trigg, L.E., (1995). *K*: An Instance-based Learner Using an Entropic Distance Measure*. Dept. of Computer Science, University of Waikato, NZ.

Diamond, D.H., Heyns, P.S., Oberholster, A.J., (2015). A Comparison Between Three Blade Tip Timing Algorithms for Estimating Synchronous Turbomachine Blade Vibration. In: Amadi-Echendu J., Hoohlo C., Mathew J. (eds) 9th WCEAM Research Papers. Lecture Notes in Mechanical Engineering. Springer, Cham

Dundas, R.E., (1994). *A Statistical Study of Gas Turbine Losses and Analysis of Causes and Optimum Methods of Prevention.* ASME International Gas Turbine and Aero-engine Congress, The Hague, The Netherlands. ASME Paper Number: 94-GT-279.

Gallego-Garrido, Dimitriadis, Wright, (2007). *A Class of Methods for the Analysis of Blade Tip Timing Data from Bladed Assemblies Undergoing Simultaneous Resonances—Part I: Theoretical Development.* International Journal of Rotating Machinery Volume 2007, Article ID 27247.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, R., Ian, H., (2009). *The WEKA Data Mining Software: An Update*. SIGKDD Explorations, Volume 11(1), Pages 10-18.

Hazewinkel, M., (2001). *Spline interpolation*, Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4.

Ho T.K. (1995). *Random Decision Forests.* Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. Pages 278–282.

Kohavi, R., (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection.* Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann. 2 (12): 1137–1143.

Kohonen, T., (1997). *Self-Organizing Maps, 2nd edition*. Springer-Verlag Berlin Heidelberg, Germany.

van der Maaten, L.J.P., Postma, E.O., van den Herik, H.J., (2009). *Dimensionality Reduction: A Comparative Review.* Tilburg University Technical Report, TiCC-TR 2009-005, 2009.

van der Maaten, L.J.P., Hinton, G.E., (2008). *Visualizing High-Dimensional Data Using t-SNE.* Journal of Machine Learning Research 9, ISSN: 2579-2605.

Mazur, Z., Garcia-Illescas, R., Aguirre-Romano, J., Perez-Rodriguez, N., (November 2006). *Steam Turbine Blade Failure Analysis.* Elsevier. Engineering Failure Analysis 15, Pages 129-141.

Norton, S.E., (2016). *Analysis and prediction of blade vibration events on the basis of steady operational-monitoring and tip-timing data via computationally intelligent methods.* MSc Thesis. Aristotle University of Thessaloniki. Greece.

Przysowa, R., (2014). *The Analysis of Synchronous Blade Vibration Using Linear Sine Fitting.* Journal of KONBiN. Volume 30, Issue 1, Pages 5–19, ISSN 2083-4608.

Ramsay, T., (2016). *Investigation of Data Reduction Methods for the Analysis of Blade Tip-Timing Vibration Data.* MSc Thesis. Aristotle University of Thessaloniki. Greece.

Sammon, J.W., (1969). *A Nonlinear Mapping for Data Structure Analysis.* IEEE Transactions on Computers, vol. C-18, no. 5, Pages 401-409.

Smith, S.W., (1999). *The Scientist and Engineer's Guide to Digital Signal Processing, Second Edition, Chapter 15.* California Technical Publishing, USA. ISBN: 0-9660176-6-8.

Ziegler, D., Puccinelli, M., Bergallo, B., Picasso, A., (July 2013*). Investigation of Turbine Blade Failure in a Thermal Power Plant.* Case Studies in Engineering Failure Analysis, Volume 1, Issue 3, Pages 192-199.