A comparative study of principal component analysis and kernel principal component analysis for photogrammetric shape-based turbine blade damage analysis

Benjamin Katerere Gwashavanhu[a]

Abrie J Oberholster[b]

Stephan P Heyns[c]

[a]Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa. benji.gwashavanhu@hotmail.com

[b]Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa. abrie.oberholster@up.ac.za

[c]Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa. stephan.heyns@up.ac.za

Abstract

Photogrammetry is popular as a non-contact full-field data acquisition technique for machine condition monitoring purposes. Two popular implementations thereof are Digital Image Correlation (DIC) and 3D Point Tracking (3DPT). Since these two approaches require surface preparation in the form of applying speckle patterns or discrete markers, their use is negated in several industrial applications, such as the online condition monitoring and/or assessment of horizontal axis wind turbines. A shape-based photogrammetric approach that focuses on analysing changes in the form of a rotating blade's boundary contour is a viable alternative that does not require any surface preparation. 3D shapes of a blade at a particular instance can be extracted from a pair of stereoscopic images of the blade. Shape Principal Component Descriptors (SPCDs) determined from applying Principal Component Analysis (PCA) to Fourier coefficients of chain-code shape signatures of the blade contour can be determined. These can be regarded as indicators of the form of the shape, and as the blade rotates, variations in the SPCDs can be investigated to better understand the dynamics of the blades.

This paper focuses on the application and performance evaluation of data reduction and classification techniques for a novel shape-based photogrammetry condition monitoring technique. Investigations

are conducted to show that applying data reduction methods to raw time domain SPCDs can result in successful classification of differently damaged blades. Post-processing strategies are developed to ensure a more robust shape based condition monitoring tool for analysing turbine blades. Kernel Principal Component Analysis (KPCA) is implemented and it is shown that it out-performs the conventional PCA in terms of classifying different blade damage modes. The feasibility of using multi-domain statistical features as feature vectors to which PCA or KPCA is applied for classification purposes is also investigated. Results indicating how well differently damaged blades can be distinguished are provided, and it is clearly illustrated that multi-domain statistical features are more robust to noise contamination in the signals compared to using the raw SPCDs data.

Keywords: Photogrammetry, shape principal component analysis, condition monitoring, principal component analysis

1. Introduction

Vibration based dynamic analysis constitutes a major component of Structural Health Monitoring (SHM), more reliable and efficient than other approaches such as structural visual inspections. To apply this technique to rotating machinery, machine dynamic responses have to be captured using more specialized equipment since conventional contact transducers such as accelerometers are less than ideal. These are point-wise in nature, can potentially mass load the specimen under investigation, and they require telemetry systems for data transfer from the accelerometers to the data logging systems. In addition to the system easily becoming very complex if data has to be captured at multiple locations on a structure, telemetry systems can introduce a significant amount of noise in the measurements. Also, a system will have to be shut down for the intrusive sensor installation before measurements can be taken when using contact sensors. Furthermore, installed sensors do not last for extended periods of time if the machine is operated in a harsh environment. Considering the size of the structure under investigation, the issues associated with contact transducers are even more problematic when analysing large structures such as wind turbine blades.

With the huge drive to develop clean alternative sustainable renewable energy sources, the sizes of wind turbines being built are continuously increasing to harness more power from the wind. According to the 2023 Global Wind Energy Council report, 680 GW of wind capacity are expected to be added globally between the years 2023 and 2027 [1]. This type of growth creates a need for measurement approaches that are better suited to easily capture operational data that can be used for the condition monitoring of operational wind turbines. Owing to the harsh environments in which these structures operate, failures typically occur due to strong winds, fatigue, moisture lightning and fire [2]. Not only does wind turbine blade failure result in revenue losses during downtime and costly repairs ranging from single blade failure effects to total structural collapse scenarios, but incidents have also been recorded in which lives have been threatened or even lost as a result of these failures [3]. Optical non-contact measurement techniques have been receiving a lot of attention as alternative approaches to contact transducers for SHM, especially photogrammetry.

1.1. Photogrammetry

Photogrammetry utilizes cameras to capture a sequence of images of a structure under excitation, and then extract displacements or strains which can be analysed to extract the dynamic behaviour of the structure. A single camera can be used to capture in-plane structural motion, or multiple cameras can be synchronized to capture more degrees of freedom. After the system has been set up such that the structure of interest is in the Field-of-View (FOV) of all cameras involved, system calibration is conducted to extract the camera intrinsic and extrinsic parameters that are used in the determination of real-world displacements or strains. The approach is more attractive in that it does not interfere with the structure under investigation, meaning that it will not alter the aerodynamic behavioural properties of the blades in any way. No telemetry systems are required, and the approach is full-field in nature (measurements can be taken for entire sections on a wind turbine that are in the FOV of the cameras).

Two popular forms of photogrammetry are Digital Image Correlation (DIC) and 3D Point Tracking (3DPT). DIC refers to photogrammetric strain measurements in deforming structures. Since this approach relies heavily on tracking subsets of a speckle patterned image, DIC is usually employed for the analysis of structures that do not have significant rigid body motions between subsequent image pairs. The technique is especially popular for measuring strains in statically loaded specimens, and crack growth analysis in applications involving fatigue life investigations. Whilst a speckle pattern is applied on a structure for strain measurements using DIC, 3DPT focuses on tracking markers that are attached to a dynamic structure to provide structural displacements.

Najafi et al. [4] developed a robust 3DPT approach for analysing a 1.64-m horizontal axis wind turbine, at a stand-off distance of 7.5 m from two high-speed cameras. In the investigation, the authors demonstrated this as an alternative measurement technique to the conventional accelerometers. For stereoscopic camera setup calibration, a 3D calibration approach is suggested. A calibration grid size of least 120×120 m^2 is required for a modern-day wind turbines. The 3D calibration uses exact point locations on a structure as the calibration coordinates. These points are captured using a Leica surveillance device with a laser range finder [4]. For calibration conducted via rotation of the blades, the entire rotor area can be defined through rotating the blades by specific angle steps so that collections of known coordinates can be defined. The investigations done illustrated that this type of calibration proved to be more accurate. Blade out-of-plane displacements were then successfully captured using this approach.

In terms of wind turbine analysis using a pair of high-speed cameras, Winstroth et al. [5] applied DIC on a full-scale 3.2 MW wind turbine. Surface preparations were done on four radial locations on the blades, and captured data was compared to aeroelastic simulations. The obtained measurements illustrate the applicability of using a photogrammetric approach to analyse a full-scale wind turbine in which the cameras are separated by 105 m and situated at a distance of about 205 m in front of the turbine blades. In addition to demonstrating the feasibility of the approach, in-plane, out-of-plane and

torsional blade responses were also measured whilst the turbine was in operation. For blade pitch angle step inputs, the small variations in the torsion, in-plane and out-of-plane blade responses detected illustrated that the optical system had a high spatial accuracy. The comparison done by the authors between DIC and aeroelastic simulations indicated good agreement of measurements, demonstrating that DIC can be used for aeroelastic simulation code validation.

Ozbek et al. [6] also analysed a full-scale wind turbine (2.5 MW, 80-m diameter) through point tracking using four Charged Coupled Device (CCD) cameras. At a stand-off distance of 220 m, operational deflection shapes of the blades could be captured and deformations with an accuracy of ±25 mm identified. Whilst the maximum overall error for markers at the tip of the blade could reach 35 mm, the authors suggested that these systematic errors could be reduced by using more sophisticated camera calibration systems, and also by utilizing hardware and data processing methods more specialized for analysing wind turbines.

Since this approach does not require extra cable installations and focuses only on the markers attached on the surface of the blade, it is generally more suited for analysing large structures when compared to the conventional fiber optics strain gages and accelerometers. When considering horizontal-axis wind turbines for instance, accelerometers are not only intrusive, but they are also not capable of capturing large structures' very low frequencies in the ranges of 0.2 Hz to 0.5 Hz [4]. Ozbek et al. [6] highlighted the pros of adopting this measurement approach, which include the fact that a single system can be used to monitor several turbines. Data capturing systems which are obviously located on the ground can be easily accessed and analysed to detect any faults. The measurements can also be stored to develop proper condition monitoring archives.

Shortcomings in terms of the applicability of photogrammetry for analysing rotating structures do exist. For an experiment of magnitude in the range of a 3.2 MW wind turbine, the preparations that have to be done are quite elaborate [5]. Self-adhesive foils have to be used to apply the speckle

pattern for tracking on the blade whilst it is still on the ground before mounting. This limits/restricts the practicality of this technique, something which the work presented in this paper aims to address.

The use of Unmanned Aerial Vehicles (UAVs) to capture footage (image sequences) of blades has been investigated. Khadka et al. [7] employed a DIC system installed on a drone to measure deflections of a six bladed 113 cm rotor with a hub height of 154 cm. Using a dynamic image stitching technique of the DIC measurements, the authors were able to characterize the dynamics of the entire blade with high accuracy. In addition to complicated speckle pattern application, the requirement that the drone flight mode path be programmed for each specific turbine, and the windy conditions of the locations where wind turbines are installed were highlighted. Ensuring that patterns are applied during manufacturing and equipping the drones with Light Detection and Ranging (LIDAR) to improve drone the control were proposed possible solutions to these issues. Guan et al. [8] proposed a similar UAV based technique for defect detection of a 94 m diameter wind turbine. In the investigation, a combination of ground cameras at a standoff distance of 248 m and stereo camera system mounted on an UAV was used. Successfully captured were the turbine global dynamic measurements and local blade deformations.

A shape-based approach such as boundary edge extraction from a captured image, followed by shape variation analysis similar to what Gwashavanhu [9] proposed, has the potential of being more industrially applicable. In that study, 3D boundary contours of blades captured using a set of stereoscopic high-speed cameras were decomposed into six uncorrelated shape descriptors called Shape Principal Component Descriptors (SPCDs) using Principal Component Analysis (PCA). Subsequently, data reduction and classification techniques are applied to the determined SPCDs. Variations of these descriptors from one image to the next were successfully analysed to distinguish differently damaged rotating blades. Further post-processing strategies are investigated to better utilize this technique for SHM of turbine blades.

1.2. Data reduction approaches

Captured vibrational data in the form of strains or displacements are typically analysed through the use of data reduction statistical methods such as PCA and Independent Component Analysis (ICA). This is done to extract features of reduced dimensionality that retain most of the variations present in the original data. Data reduction techniques are widely considered in condition monitoring applications for classifying dynamic behaviours of machines that have different forms or levels of damage. Elsamanty [10] employed PCA to illustrate how conventional vibration dimensionality reduction can be accomplished whilst preserving the variations in the data. Through analysis wind turbine data captured using a Supervisory Control and Data Acquisition System (SCADA), a number of authors have successfully predicted abnormalities in the operation of a wind turbine using a combination of Convolution Neural Network (CNN) and KPCA data processing approach [11, 12, 13]. The following sections take a deeper dive into the implementation of such data reduction approaches.

1.2.1. Principal component analysis

Using PCA in a machine learning application, dimensionality reduction and visualization of measurements are achieved by projecting the possibly correlated data into a lower-dimensional feature space of uncorrelated data using a linear transformation [14]. Given a multivariate data set $\boldsymbol{X}$, the problem of PCA essentially reduces to solving an eigenvalue problem stated in Equation (1).

$$\boldsymbol{X}^{new} = \boldsymbol{V}^T \boldsymbol{X} \tag{1}$$

A covariance matrix of the original correlated feature vectors is calculated to determine the eigenvectors ($\boldsymbol{V}$). Each column of the eigenvector matrix represents a single principal component, whose scores are given by the corresponding row of matrix $\boldsymbol{X}^{new}$.

Data reduction aims at representing the original data with principal components much fewer than the size of the original data. This is however not the case in Shape Principal Component Analysis (SPCA). In SPCA, the ability of PCA to decompose potentially correlated data into components, each of which corresponds to a distinct feature, is exploited. Given a shape, resulting principal components can be associated with distinct shape variations, which means that several identifiers (SPCDs) can be monitored or analysed to investigate how the form of a shape changes from one stage to the next.

Gwashavanhu et al. [15] gave a detailed 2D description and illustration of this concept. Definining a shape boundary contour using a shape signature is a common practice. The shape signatures are essentially 1D functions such as the chain-code, shape centroid distance, complex coordinates, the curvature function or the cumulative angular function [16]. Application of PCA to Fourier coefficients of a chain-code shape signature results in components each of which represents a unique shape form variation that can be linked to a specific machine dynamic behaviour [15]. A similar type of shape analysis was conducted by Iwata et al. [17], in which different variations of citrus leaf shape were investigated and quantified using PCA. Variations in the shapes of cereal grains were also successfully investigated using PCA. Mebatsion et al. [18] illustrated that principal components of standardized elliptic Fourier descriptors can be used to quantitatively capture grain shape variability. They concluded that this type of analysis is more objective and precise when compared to classification via human visual inspection.

For a 3D shape, the proposed procedure for determining the SPCDs is illustrated in Figure 1. The covariance matrix of the 6×N matrix of Fourier coefficients ($a_n, b_n, c_n, d_n, e_n, f_n$) is calculated using the mean ($\overline{\boldsymbol{X}}$), and ith column entries ($\vec{X}_i$) of the matrix $\boldsymbol{X}$**.** Eigenvector $\boldsymbol{V}$ and eigenvalues $\boldsymbol{I}$ that satisfy the eigenvalue problem can then be used to determine a new uncorrelated multivariate vector $\boldsymbol{X^{new}}$. Each row of matrix $\boldsymbol{X}^{new}$ (which contain the principal component scores and denoted as $\vec{X}_i^{\,new}$ for row i) represents the uncorrelated linear combinations of the original data, accounting for the variance in the data. Each column of matrix $\boldsymbol{V}$ (i.e., $\vec{V}_i$) represents a specific shape feature, termed the Shape Principal Component (SPC). $\boldsymbol{V}$ can be considered to be composed of geometric shape modes. The physical relationships between different $\vec{V}_i$ vectors and the form of a typical shape were investigated in detail by Gwashavanhu et al. [15].

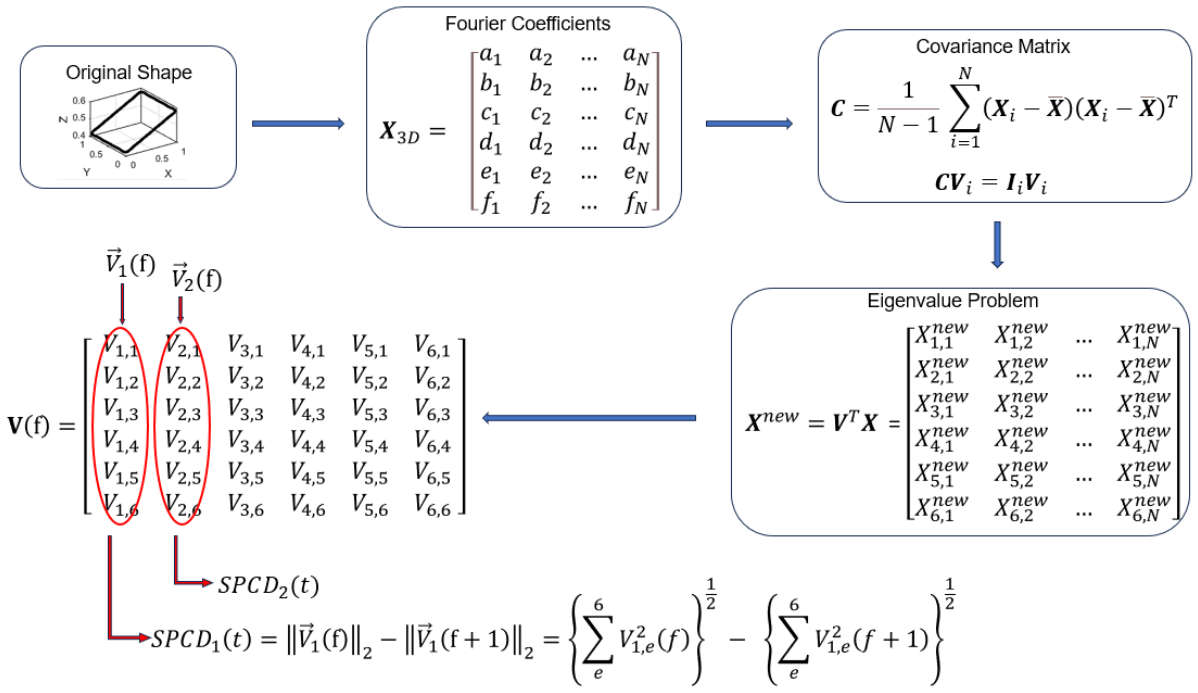

Figure 1: Procedure for determining Shape Principal Component Descriptors

The uncorrelated $\vec{V}_i$ s represent unique shape features whose variations can be monitored separately to get an in-depth insight on how the overall shape changes in a dynamic situation. Each $\vec{V}_i$ represents an independent shape feature which allow a quantitative analysis of the shape to be conducted by use of $\boldsymbol{X}^{new}$ as the ordinary quantitative characters [18]. For two consecutive image contours from two frames f and $f+1$, each described by a 6×6 multivariate matrix $\boldsymbol{V}$, the i^{th} measure of variation between two consecutive frames (Shape Principal Component Descriptor, $SPCD_i$) can be considered as the difference between the consecutive $\vec{V}_i$ Euclidean norm values (Figure 1). The difference between the two frames was considered for an improved SNR in this case.

In this study, the application of conventional data reduction techniques (PCA and KPCA) to calculated time-domain signals (SPCDs in this case) is investigated to introduce new robust data processing techniques that can be adopted for accurate and reliable condition monitoring of rotating structures [9].

1.2.2. Kernel Principal Component Analysis (KPCA)

Since PCA is based on a linear transformation approach, the feature vectors considered are assumed to be extracted from data that has a Gaussian distribution. In investigations where nonlinear relationships may exist among variables or time intervals, the use of PCA may result in a loss of the nonlinear features [9]. This nonlinearity is quite common in complicated industrial systems which have responses that are especially nonstationary. For SHM purposes, it is worth noting that some forms of damage in machines can cause a linearly behaving structure to have nonlinear dynamics associated with its vibration responses.

KPCA addresses this issue by first nonlinearly transforming the original data into a different feature space through the use of various kernels. The kernels used are typically polynomial, sigmoid and Gaussian kernels. PCA is then conducted in the new feature space on the data with nonlinear dependences between variables. As described by Nguyen & Golinval [19], the concept of KPCA boils down to the nonlinear mapping $\boldsymbol{x}_k \mapsto \emptyset(\boldsymbol{x}_k)$ with $\boldsymbol{x}_k \in \Re^n, (k = 1, \ldots, M)$, which represents a higher dimensional space. For centered data $\left(\sum_{i=1}^{M} \emptyset(\boldsymbol{x}_i) = 0\right)$, the covariance matrix used in the PCA eigenvalue problem can then be calculated using Equation (2).

$$C = \frac{1}{M}\sum_{i=1}^{M} \emptyset(\boldsymbol{x}_i)\emptyset(\boldsymbol{x}_i)^T \tag{2}$$

The Gaussian kernel function, defined by Equation (3), for which $2\sigma^2 = w$ is the width of the kernel, is most frequently used for KPCA investigations.

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(\frac{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}\right) \tag{3}$$

Also considered in this study is the polynomial kernel given in Equation (4), where d is a positive integer.

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j + 1)^d \tag{4}$$

Compared to the linear PCA, KPCA can essentially extract the higher-order nonlinear relationships between data, retaining a greater degree of the complex information between the data [20]. With the

right or appropriate choice of parameters, different kernel functions produce very similar results [19]. In classification applications such as optical character recognition, KPCA has been shown to outperform PCA in terms of feature extraction [21].

1.3. Statistical features

According to He et al. [21], mechanical systems are generally complex owing to the stochastic processes that essentially contribute to the measured structural responses. This makes deterministic time function based investigations insufficient to analyse the machines. It is often difficult to capture the behaviour of a machine from the raw data time series directly. Feature extraction from the raw time data has to be conducted and the 'hidden' patterns identified from these features.

1.3.1. Time domain features

Typical time domain statistical features that can be extracted from the data and used as the feature vectors for either PCA or KPCA are given in Table 1.

Table 1: Time domain statistical features

Feature	Formula	Feature	Formula
Absolute mean value, $\lvert\bar{x}\rvert$	$\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$	Variance, D_x	$\frac{1}{N-1}\sum_{i-1}^{N}(x_i-\bar{x})^2$
Maximum peak value, x_p	$max\lvert x_i\rvert$	Kurtosis, β	$\frac{1}{N}\sum_{i=1}^{N}x_i^4$
Root mean square, x_{rms}	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$	Crest factor, C_f	$\frac{x_p}{x_{rms}}$
Square root value, x_r	$\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i\rvert}\right)^2$	Shape factor, S_f	$\frac{x_{rms}}{\lvert\bar{x}\rvert}$

1.3.2. Circular domain features

When analysing rotating structures, a set of circular domain features can be extracted. In some applications, these features can contain information more descriptive of the machine behaviour compared to their time domain counterparts. These features include circular mean, circular variance, circular skewness and circular kurtosis.

Caesarendra et al. [22] proposed a circular domain analysis approach based on a Piecewise Aggregate Approximation (PAA) data reduction process, in which the following listed steps are followed:

(i) reduction of the vibration data using PAA and construction of neighbourhood correlation plot of the reduced data

(ii) neighbourhood correlation plot ellipse least-square fitting for pattern classification

(iii) plotting the distribution of the ellipse shape in the circular domain, and calculation of the circular domain features

PAA is a data reduction technique applied to large time series data in which a sequence of sampled vibration data $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$ is divided into w windows of equal size, referred to as frames [22]. For each frame, the reduced data representation is then taken as the mean value of the data in that frame. The equation below is used for this calculation, in which the vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ represents the PAA reduced data set.

$$x_n = \frac{1}{w} \sum_{j=w_{n-1}+1}^{w_n} y_i \tag{5}$$

A neighbourhood correlation plot is then simply x_{n+1} vs x_n. Caesarendra et al. [22] illustrated how the orientation of an ellipse fitted onto the neighbourhood correlation plot data points is frequency-dependent. This becomes very useful when analysing non-stationary signals of a machine captured over an extended period. The onset of damage in a structure is typically associated with a change in the frequency of the vibrational responses. Thus if neighbourhood correlation plots are developed for different time segments in measurement sequence, a change in the ellipse orientation will be an indication of dynamic behavioural changes in the machine system. Properties of the fitted ellipse such

as its orientation, size and aspect ratio can be considered as circular domain features for the time signal of interest.

Other circular domain features can be obtained by representing signals with features extracted from their random circular variables. Hilbert transformation of time signals can be utilized to determine the instantaneous phase of the time series, allowing complex signal representations of real-valued signals [23]. For a time domain signal $x(t)$, an analytical signal $x_a(t)$ can be determined using Equation (6).

$$x_a(t) = x(t) + j\mathcal{H}\{t\}(t) \tag{6}$$

In the equation above, $\mathcal{H}\{t\}$ is the Hilbert transform of $x(t)$. The analytical signal can then be expressed as:

$$x_a(t) = \psi(t)e^{j\theta(t)} \tag{7}$$

$\psi(t) = |x_a(t)|$ is the amplitude envelope of the analytical signal, and $\theta(t) = \arg(x_a(t))$ the instantaneous phase with $\theta(t) \in [0, 2\pi]$.

Using the instantaneous phase, the circular domain features can be determined using equations given in Table 2, [23].

Table 2: Circular domain statistical features

Circular statistical feature	Formula
p^{th}-order sample trigonometric moment, μ_p	$\frac{1}{N}\sum_{n=1}^{N}\exp(jp\theta_n)$
Variance, σ^2	$1 - \lvert\mu_1\rvert$
Skewness, γ	$\frac{\lvert\mu_2\rvert \sin(\arg(\mu_2) - 2\arg(\mu_1))}{(\sigma^2)^{\frac{3}{2}}}$
Kurtosis, κ	$\frac{\lvert\mu_2\rvert \cos(\arg(\mu_2) - 2\arg(\mu_1)) - \lvert\mu_1\rvert^4}{(\sigma^2)^2}$

Combining the time domain statistical features and the circular domain features, one can calculate a set of multi-domain statistical features (MSFs) that are representative of the properties of a signal. Data reduction methods such as PCA and KPCA can then be used to analyse these features for dynamic behavioural classification purposes.

2. Experimental setup description and data acquisition

A pair of IDT NX8-S2 high-speed cameras were set up at a standoff distance and angle such that the entire specimen was in the FOV of both cameras. The cameras were used at their maximum resolution of 1600×1200, and images were captured at a frame rate of 4000 frames per second (FPS). 25 mm lenses were employed, and the exposure was set at 248 µs. A camera standoff distance, d = 140 cm, and the baseline distance, b = 11.5 cm were used (Figure 2).

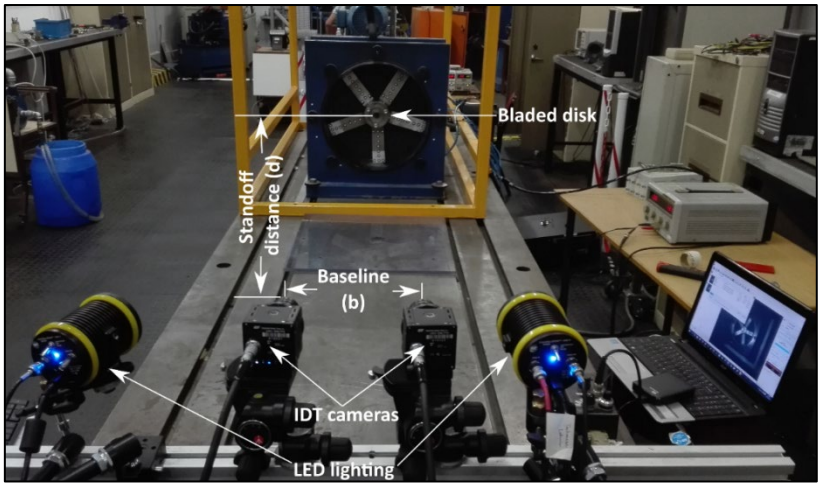

Figure 2: Stereophotography system setup

The calibration parameters for this setup are given in Table 3.

Table 3: Calibration quality parameters

Maximum projection error	Calibration quality	Maximum rectification error	Rectification quality
0.4075	0.9817	0.3910	0.9898

Calibration is considered successful if the quality is greater than 0.7, making 0.9817 acceptable [9]. The maximum rectification error (indicating how well the rectified images are aligned) was 0.391,

significantly less than the threshold maximum acceptable value of 1.5. The maximum projection error is computed as the error between left image grid points projected onto the right image and the original points in the right image. A value of 0.9890 for the rectification quality was obtained, which is very close to 1 (perfect rectification). Overall, the system was sufficiently calibrated to capture reliable stereo-photogrammetric measurements.

Excitation of the blades was achieved through the use of two air jets (Figure 3). Detailed information on the fitting of contours on the blade edges can be found in the work by Gwashavanhu [9].

The 5 bladed system analysed was composed of 3 damaged blades and two healthy blades. Of the damaged blades, one of them had two cracks (referred to as blade 1), and the other two just a single crack (blades 3 and 5). The aluminium blade geometries in millimetres are given in Figure 4, and their vibrational characteristics can be obtained from the work by Church [24].

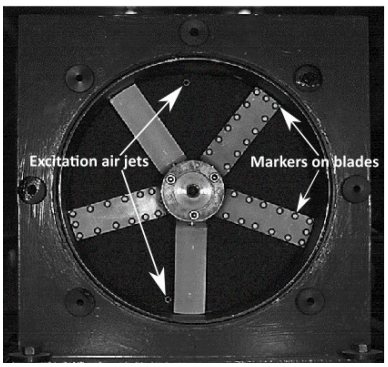

Figure 3: Excitation and blade markers setup

Three rotational speeds of 650 rpm, 1000rpm and 1460 rpm were considered for this investigation. The harmonics for 650 rpm and 1000 rpm almost coincided with the bending mode shape frequencies of the blades, which are 108 Hz, 127 Hz and 119 Hz for blades 1, 2 and 3 respectively [24]. Therefore, the blade out-of-plane responses were quite pronounced due to resonance, and the shape variations more distinct at these rotational speeds. To evaluate how the approach performs for a rotational speed that has potentially hidden shape variation characteristics, the 1460 rpm rotational speed was therefore considered.

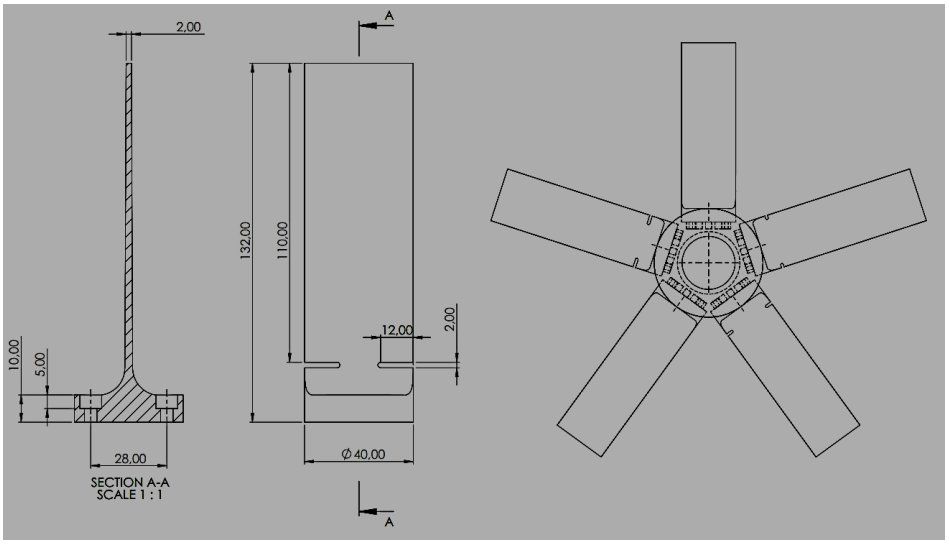

Figure 4: Single blade and bladed disk models

Due to the manner in which the blades were arranged, and owing to the small crack sizes and the fact that the blades are of a low density metal (aluminium), unbalance effects were not considered. For the rest of the paper, blade 1 corresponds to a blade with two cracks, blade 2 a healthy blade, and blade 3 a blade with a single crack.

3. Data analysis and results evaluation

SPCDs were determined for each image pair captured during testing. Since the descriptors are determined for a multivariate vector composed of six Fourier coefficients calculated for the 3D (x-, y- and z-directions) blade boundary contours, six SPCDs are obtained for each image pair. The first four SPCDs are associated with in-plane shape variations (x- and y-directions), whilst the last two are linked to the out-of-plane variations (z-direction). Crack damage at the root of the blades is expected to have more influence on the out-of-plane responses of axial blades, owing to the more significant decrease in blade out-of-plane root stiffness that is introduced with the damage. Thus for this investigation, $SPCD_5$ and $SPCD_6$ were the ones analysed.

Owing to the low Signal-to-Noise-Ratio (SNR) of the captured data due to minimal amplitude out-of-plane blade responses, basic frequency domain analysis of the SPCDs cannot be used to distinctively differentiate the blades [9]. The use of PCA to project the SPCDs data into a different feature space was therefore employed. Time segments of the SPCDs corresponding to each rotation of a blade were

considered as identifiers associated with the behaviour of that particular blade. Standard PCA was applied to reduce the data size and then project the data in a 2D feature space composed of PCs 1 and 2. A similar approach is usually used for spike signal processing, [24 - 26], in which spikes generated by various neurons (nerve cells) are detected, sorted and aligned to differentiate and identify the neurons generating them.

$SPCD_5$ and $SPCD_6$ for the three rotational speeds are illustrated in Figure 5. Whilst all SPCDs clearly show isolated clusters for the three blades for 650 rpm and 1000 rpm [9], the proposed PCA approach cannot classify the blades for $SPCD_6$ at 1460 rpm. In the case of 650 rpm $SPCD_5$, poor classification can be attributed to possible poor correlation between the fifth geometric mode and the blade response shape variation at 650 rpm.

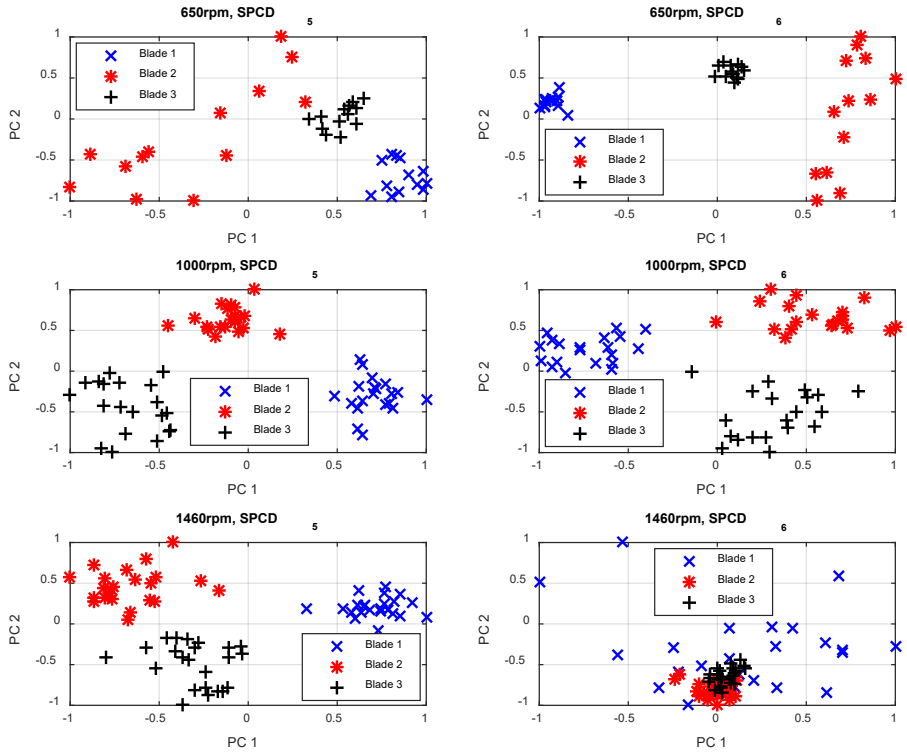

Figure 5: PCA $SPCD_5$ and $SPCD_6$ projected in a 2D feature space for 650 rpm, 1000 rpm and 1460 rpm

Further investigations conducted and presented in the following sections of this paper illustrate that a linear transformation of the data using PCA may be the reason why classification is not possible for

some rotational speeds. Non-linear relationships between different SPCDs time segments (per rotation) are suspected to be more dominant for a rotational speed (1460 rpm) whose harmonics do not almost coincide with some of the out-of-plane bending natural frequencies of the blades [9, 23].

Three different investigations were conducted to develop post-processing approaches that are more robust and more suitable for processing shape analysis based data. These are listed below.

(i) PCA vs KPCA evaluation for time domain SPCDs

(ii) raw data (time domain SPCDs) analysis vs multi-domain feature-based analysis

(iii) noise sensitivity evaluation for PCA and KPCA

3.1. PCA vs KPCA

The time domain SPCDs determined by Gwashavanhu [9] were analysed per shaft revolution and were directly used as feature vectors for PCA. As mentioned in the introduction, the linear transformation based PCA can result in loss of the non-linear properties of a signal. This is not the case with KPCA. In efforts to try and retain most of the signal characteristics when employing data reduction techniques for SHM, KPCA was therefore considered for the determined $SPCD_5$ and $SPCD_6$.

The choice of the kernel parameter does influence KPCA results with regards to the contribution of the first principal component. This contribution can be in terms of the percentage of the first eigenvalue variance for instance. According to Nguyen & Golinval [19], the Gaussian kernel function introduced in Equation (3) is known to be flexible when it comes to the choice of the kernel function width. As a comparison, a polynomial kernel defined by Equation (4) is also considered.

To evaluate how well a technique can distinguish different behavioural states of a machine through cluster performance analysis, two parameters namely the intra-class scatter S_w, and inter-class scatter S_b can be considered. These are calculated using Equations (8) and (9), [28].

$$S_b = \sum_{i=1}^{c} (\mu_i - \mu)(\mu_i - \mu)^T \tag{8}$$

$$S_w = \sum_{i=1}^{c} \sum_{f_k \in C_i} (f_k - \mu_i)(f_k - \mu_i)^T \tag{9}$$

In the equations above, μ_i is the average value of a class $C_i = (1, 2, \ldots, c)$, μ the total average and f_k the features extracted by PCA or KPCA. Essentially, a smaller S_w and a larger S_b represents better clustering classification performance. The quality of the clustering for an investigation can then be defined as the ratio S_b/S_w.

Illustrated in Figure 6 is a plot comparing the performance of a polynomial kernel to a Gaussian kernel for varying kernel parameters d and w, (Equations 3 and 4). As can be noted, the polynomial kernel S_b/S_w magnitudes are ten times less than the Gaussian kernel (0.015 vs 0.15). For the Gaussian kernel, a kernel width of 3 results in the best blade classification performance of the 1460 rpm, $SPCD_6$. Therefore in this study, the Gaussian kernel with a width of 3 was considered.

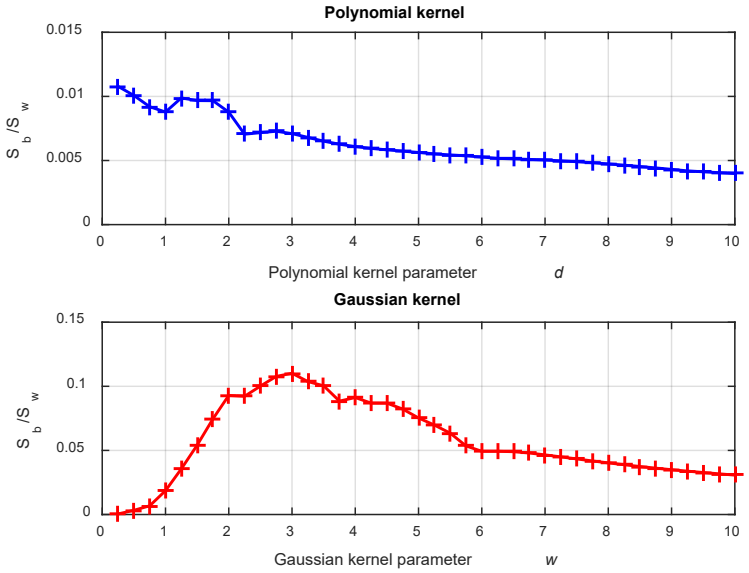

Figure 6: Comparison of polynomial and Gaussian kernels for various kernel parameters

The results obtained for projections in the 2D and 3D feature spaces are given in Figure 7 and Figure 8 for 650 rpm rotational speed.

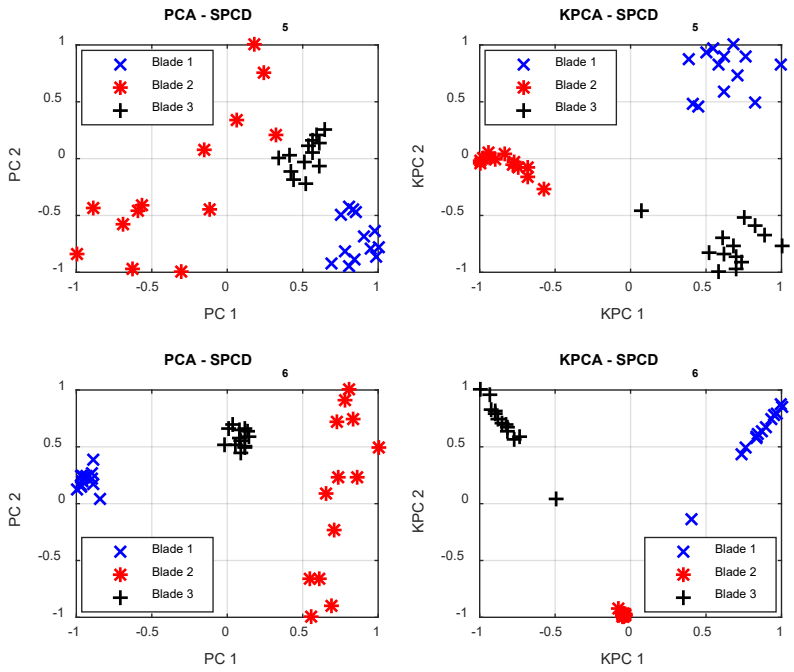

Figure 7: 650 rpm PCA and KPCA SPCDs 5 and 6 projected in a 2D feature space

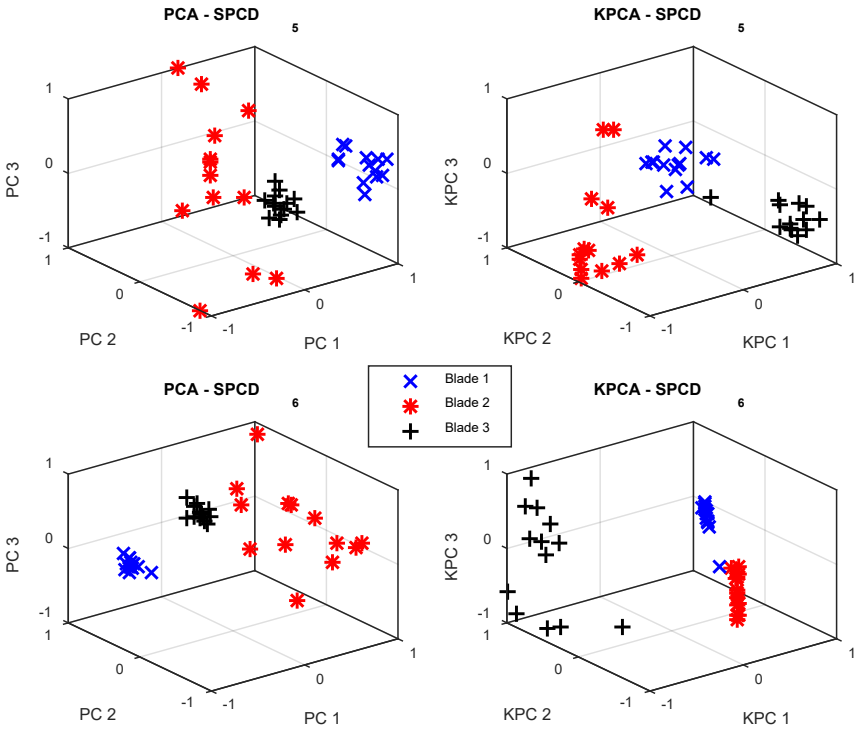

Figure 8: 650 rpm PCA and KPCA $SPCD_5$ and $SPCD_6$ projected in a 3D feature space

For 650 rpm, using the non-linear kernel transformation of the SPCDs resulted in more compact clusters in which the data points are arranged in a linear manner ($SPCD_5$). The improved clustering is better visualized in the 2D feature space (Figure 7).

Results for the 1000 rpm rotational speed are given in Figure 9 and Figure 10. It is observed that the 1000 rpm KPCA does not seem to improve the clustering in the measurements visually. A quantitative analysis presented later on will give a better insight into the performance of PCA and KPCA at 1000 rpm.

Illustrated in Figure 11 and Figure 12 are results for 1460 rpm. It can be observed that KPCA significantly improved blade distinction in comparison to PCA. For $SPCD_6$, PCA is unable to properly cluster the data for blade visual identification. This can be attributed to the low SNR of SPCD 6 at that speed. Unlike the initially considered rotational speeds (650 rpm and 1000 rpm) that have harmonics that nearly coincide with the out-of-plane blade bending natural frequencies stated above, at 1460 rpm the blade responses are not that well defined. Thus from the results presented, it is evident that the blade behaviour is better characterized by an approach that captures the nonlinear relationships between the different shaft revolutions.

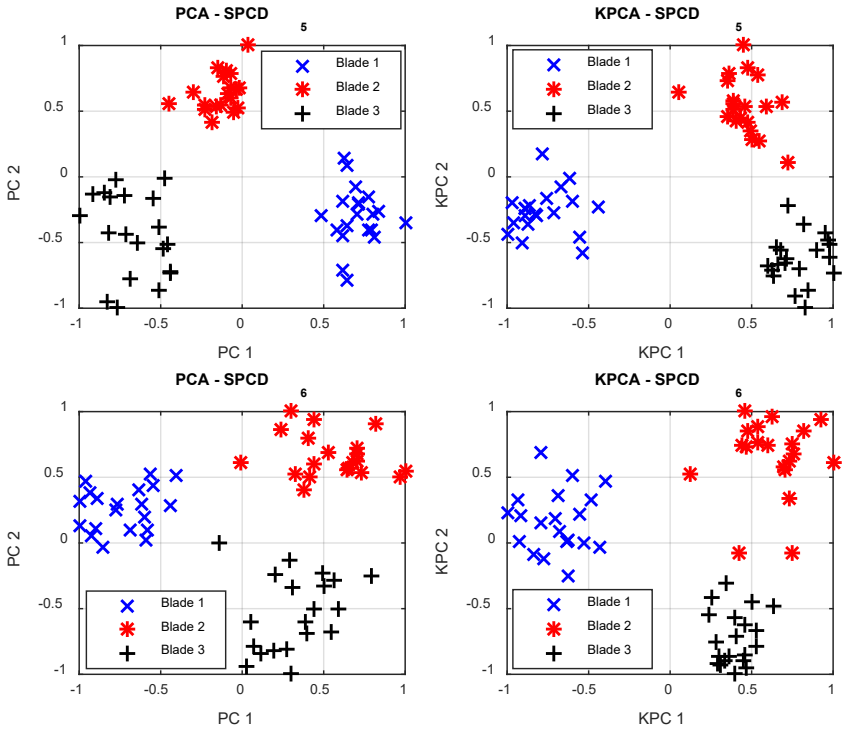

Figure 9: 1000 rpm PCA and KPCA $SPCD_5$ and $SPCD_6$ projected in a 2D feature space

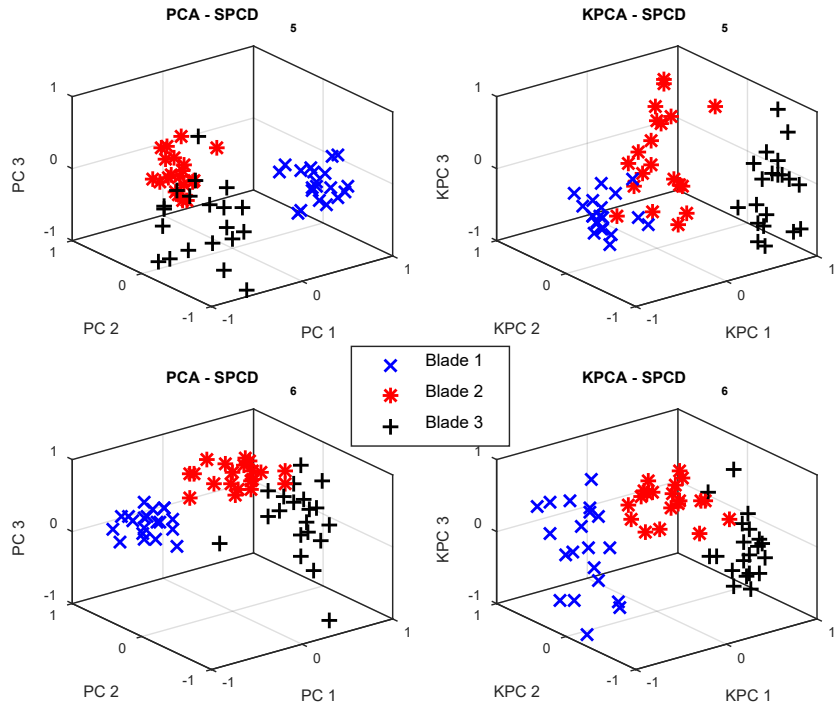

Figure 10: 1000 rpm PCA and KPCA $SPCD_5$ and $SPCD_6$ projected in a 3D feature space

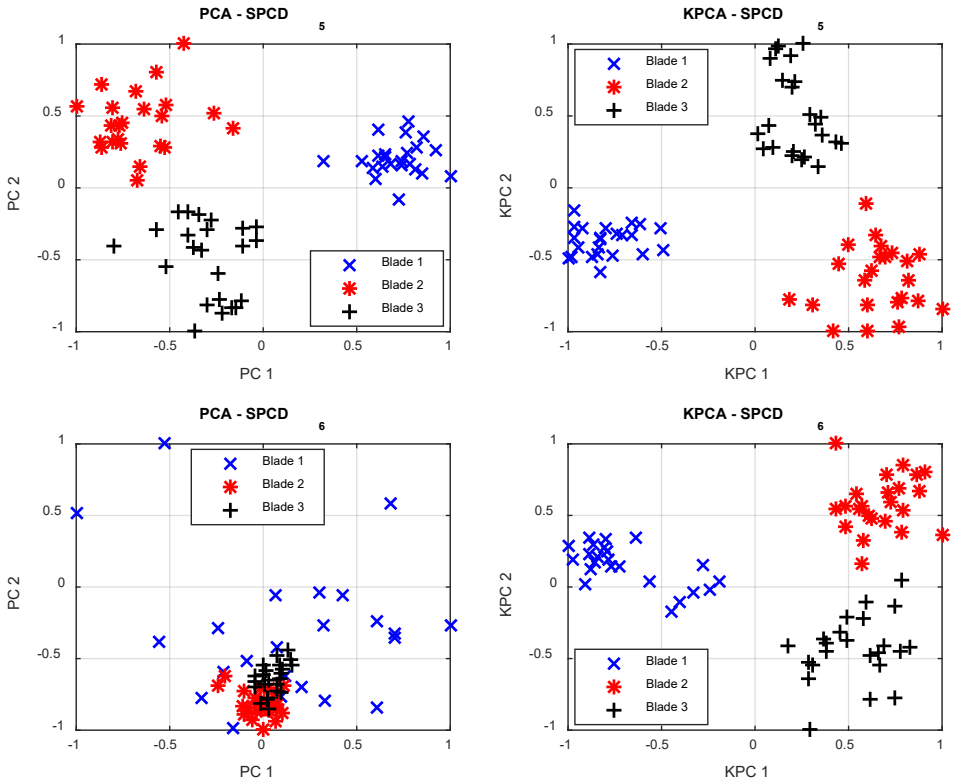

Figure 11: 1460 rpm PCA and KPCA $SPCD_5$ and $SPCD_6$ projected in a 2D feature space

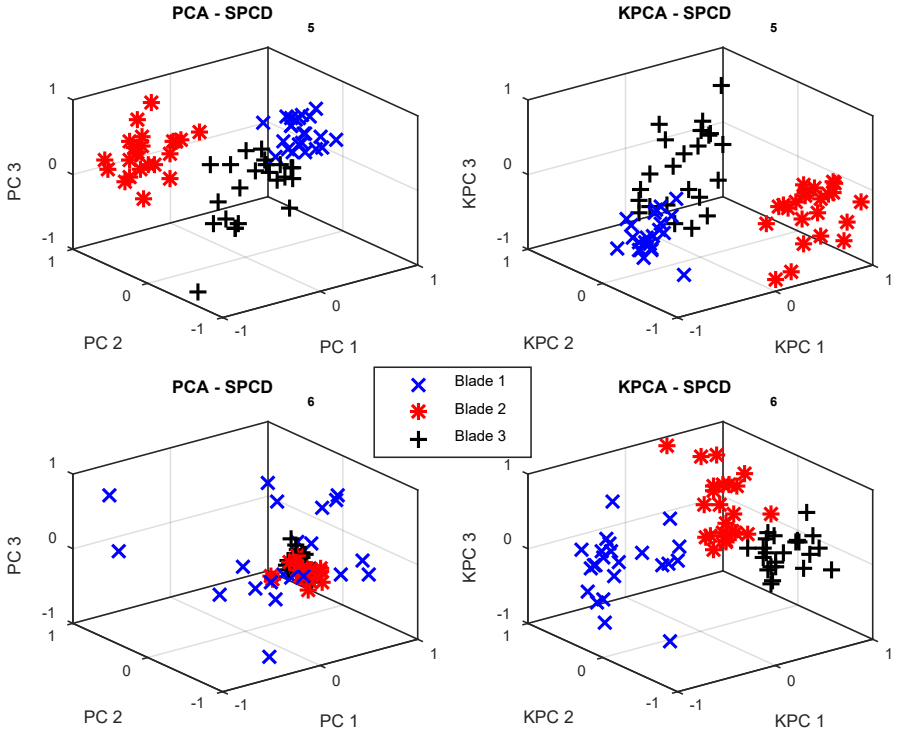

Figure 12: 1460 rpm PCA and KPCA $SPCD_5$ and $SPCD_6$ projected in a 3D feature space

In evaluating the performance of PCA vs KPCA, the clustering parameters ratio S_b/S_w was determined for the three rotational speeds investigated as shown in Figure 13.

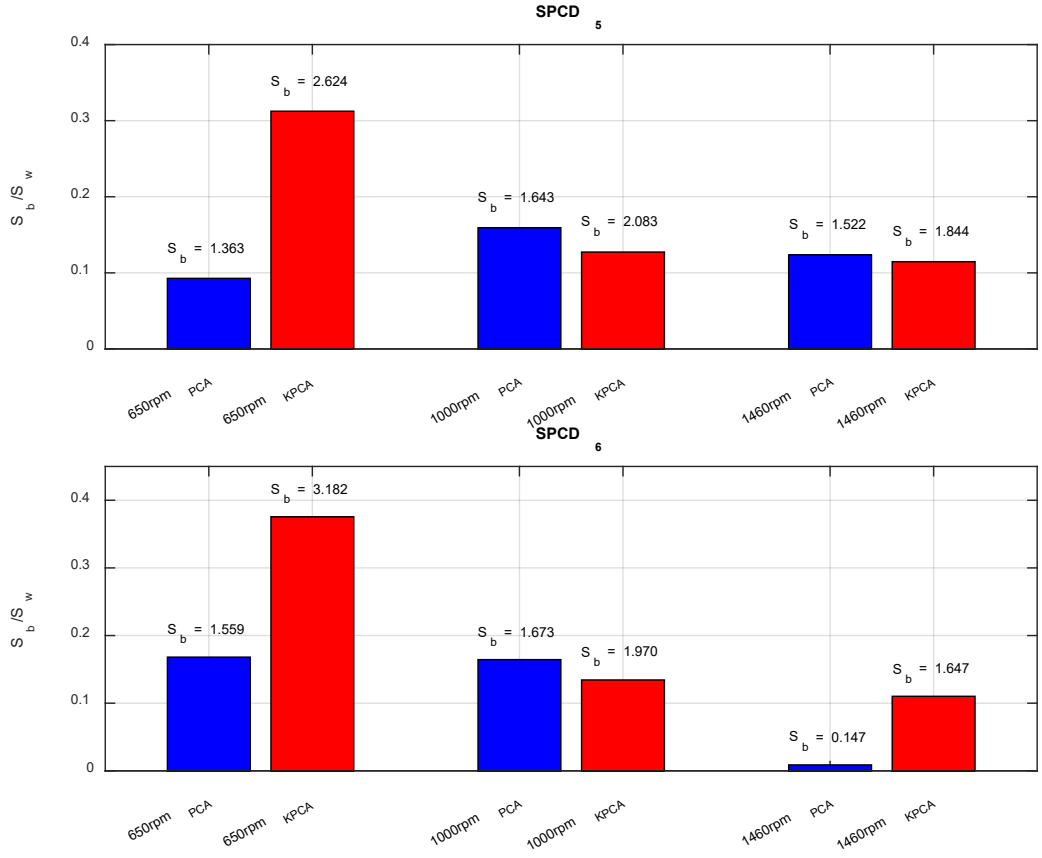

Figure 13: PCA and KPCA clustering performance for time domain SPCDs raw data

As can be noted from Figure 13, KPCA outperforms PCA for 650 rpm $SPCD_5$ and $SPCD_6$, and also 1460 rpm $SPCD_6$ in terms of overall clustering. The overall clustering performance takes into account both the cluster compactness (intra-class parameter S_w) and the cluster separation (inter-class parameter S_b) by way of the ratio S_b/S_w. However, in terms of the separation of clusters for the accurate distinction of different blades with varying damage, the inter-class cluster values can be argued to be of more importance. Whilst S_w values are not always larger for KPCA, it can be noted that the S_b values are larger for KPCA compared to PCA for all the rotational speeds. This applies to both $SPCD_5$ and $SPCD_6$.

3.2. Multi-domain statistical features based evaluation

Eight time-domain statistical features were calculated for each shaft rotation (Table 1). Circular domain features (Table 2) were also determined. After combining these features, a multi-domain feature set of 12 features was obtained, with PCA and KPCA applied for the classification of the different damaged blades.

For 650 rpm, the resulting Normalized Feature Values (NFVs) are illustrated in Figure 14. It can be observed that the three blades have different statistical features.

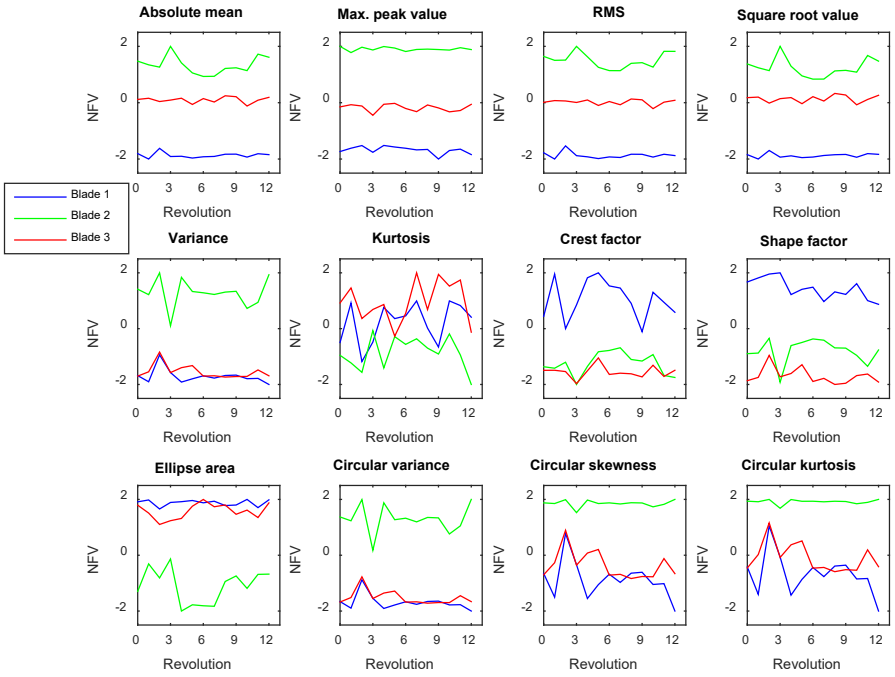

Figure 14: Averaged MSFs, $SPCD_6$, 650 rpm

MSFs of the combined time domain statistical features and the circular domain features were then considered to investigate if blade identification can be improved. In the following sections, clustering results for PCA and KPCA applied to the raw SPCDs time-domain data were being compared to the results for PCA and KPCA applied to the MSFs.

Application of PCA and KPCA on the 650 rpm MSFs resulted in the results indicated in Figure 15 and Figure 16.

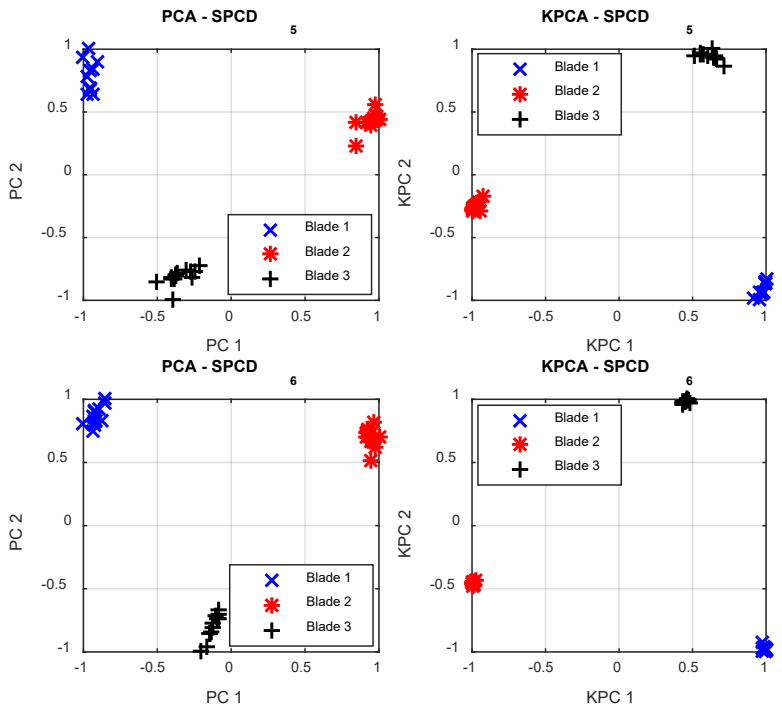

Figure 15: MSFs 650 rpm PCA and KPCA 2D

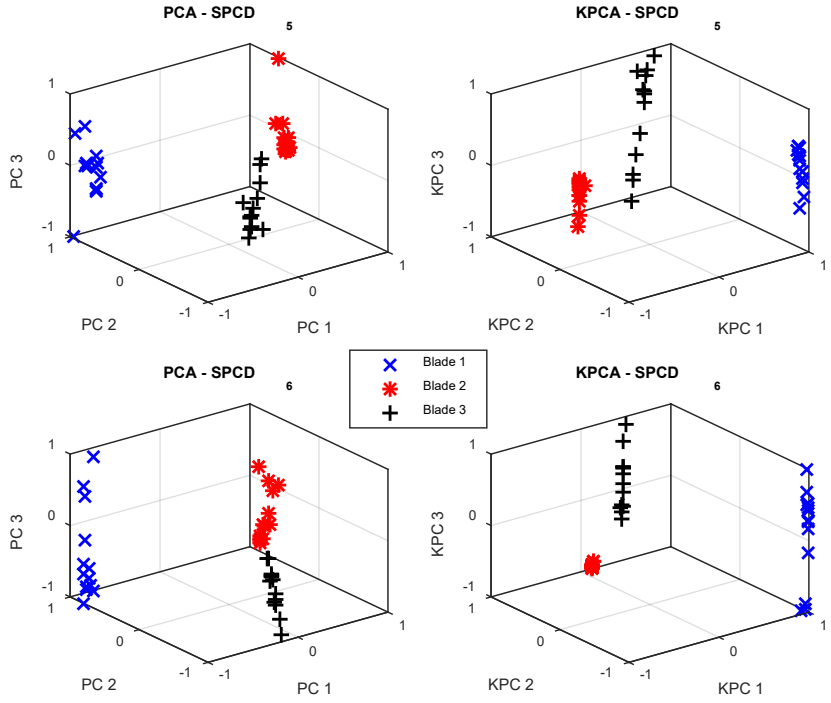

Figure 16: MSFs 650 rpm PCA and KPCA 3D

As can be observed in Figure 15 and Figure 16, MSFs produce clusters that are well separated in comparison to the raw data clusters presented in Figure 7 and Figure 8. The same is concluded for the 1000 rpm and the 1460 rpm results presented in Figure 17 to Figure 20.

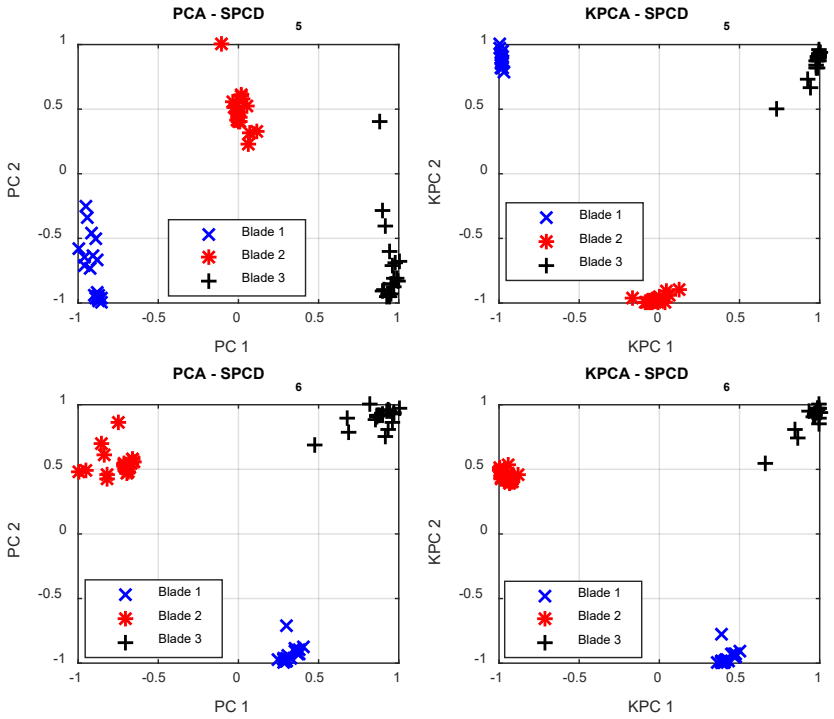

Figure 17: MSFs 1000 rpm PCA and KPCA 2D

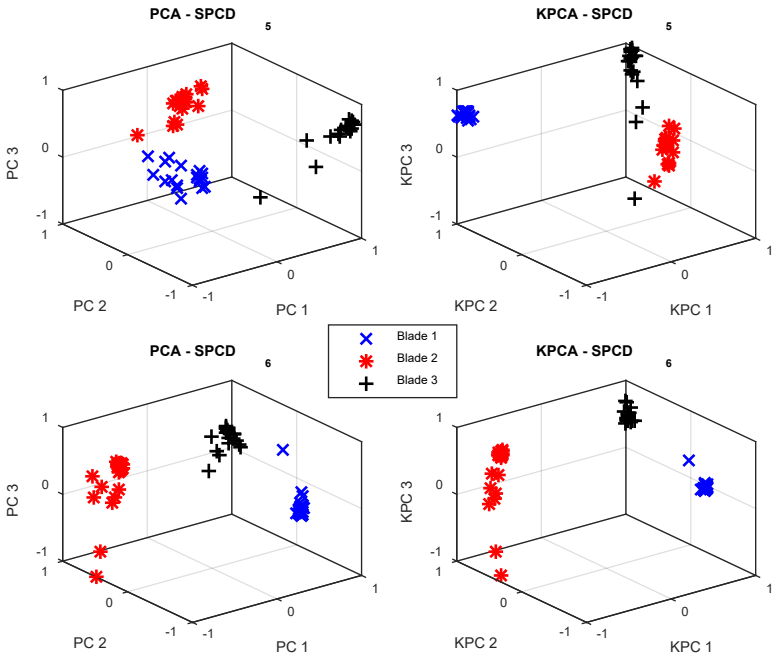

Figure 18: MSFs 1000 rpm PCA and KPCA 3D

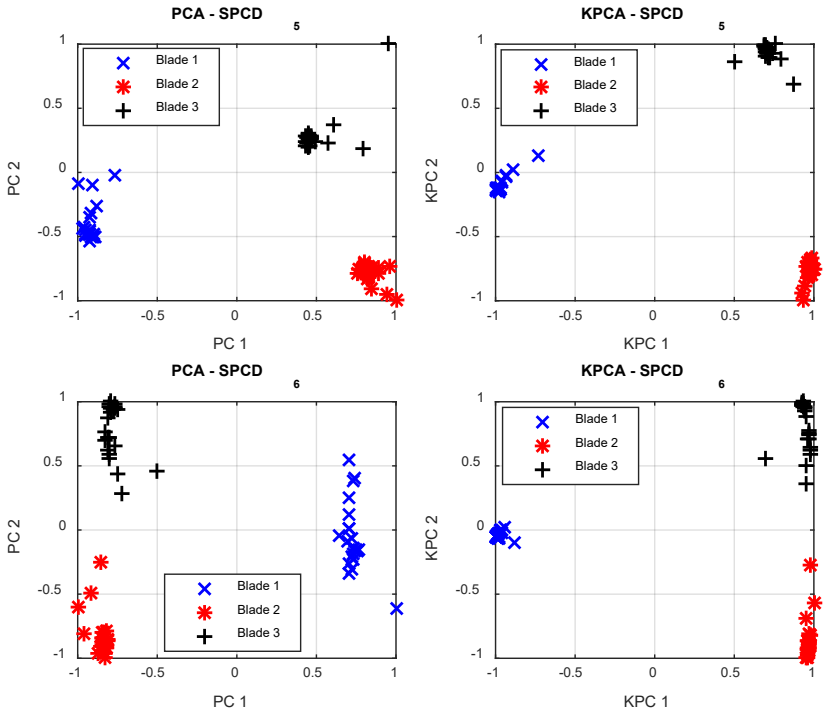

Figure 19: MSFs 1460 rpm PCA and KPCA 2D

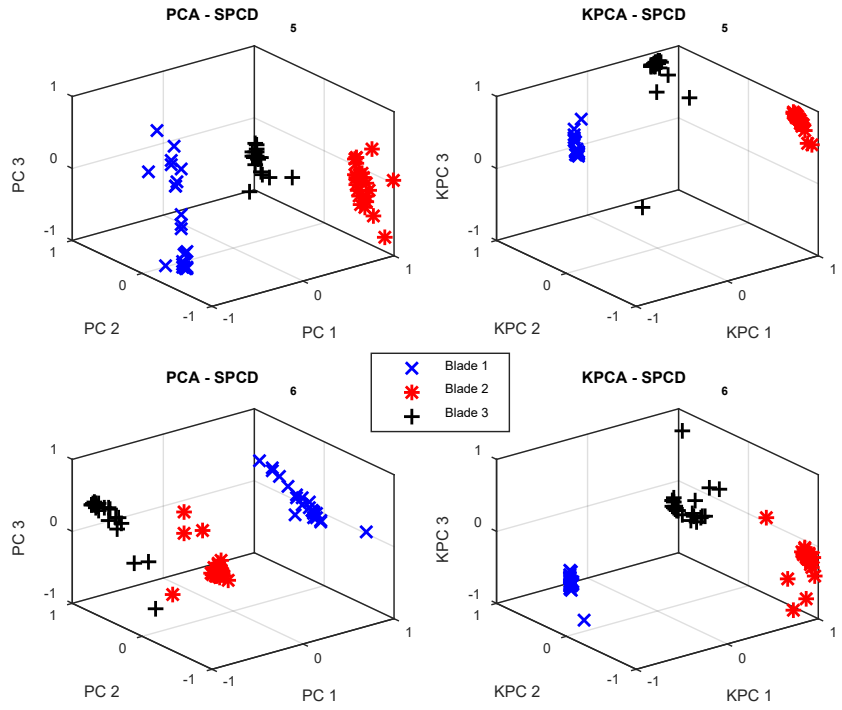

Figure 20: MSFs 1460 rpm PCA and KPCA 3D

Clustering parameter ratio S_b/S_w was also determined for the three rotational speeds investigated, this time for PCA and KPCA data calculated from the MSFs. Results obtained are shown in Figure 21.

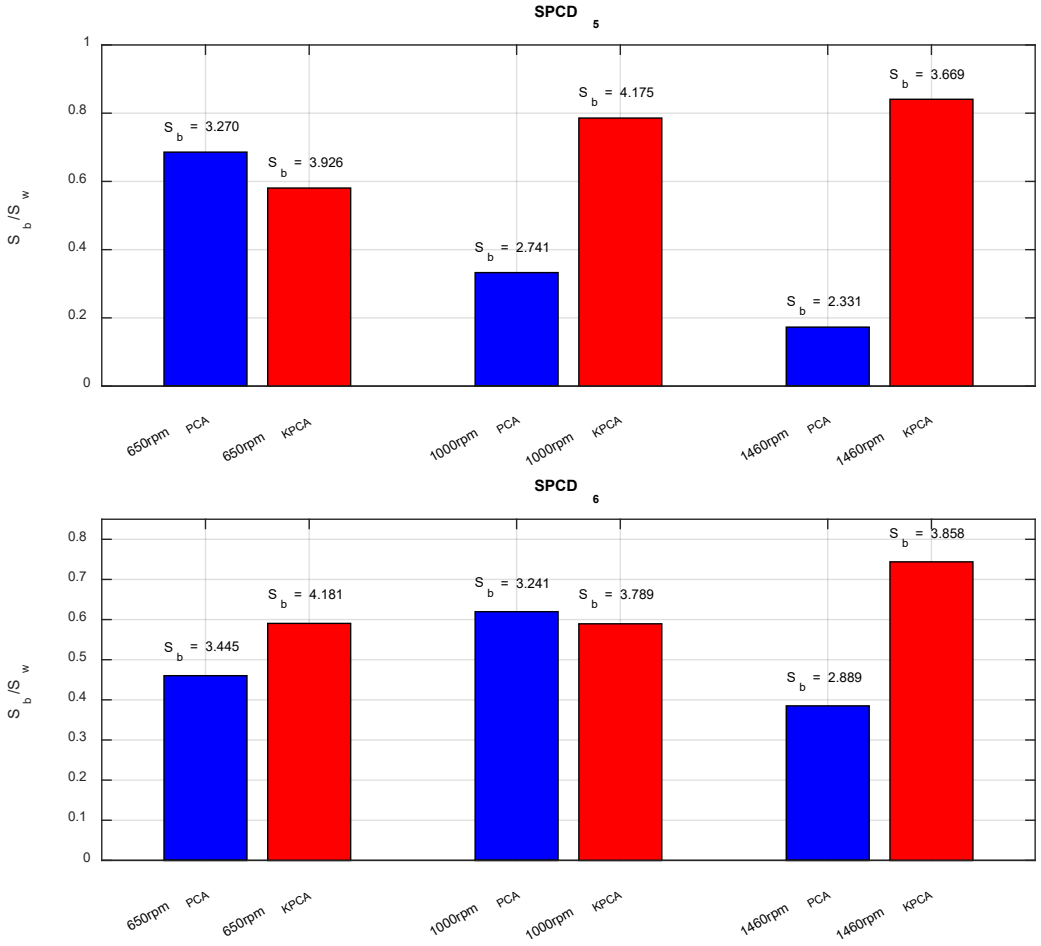

Figure 21: PCA and KPCA clustering performance for MSFs

Figure 21 illustrates how KPCA also tends to produce better clusters than the PCA approach, in terms of the higher S_b determined for each case analysed. Comparing the range of S_b/S_w for the raw SPCDs data and the multi-domain statistical features calculations (Figure 13 and Figure 21), it can be concluded that using features extracted from the raw data provides results that make distinguishing the blades much easier. MSF S_b/S_w parameters are in the range of 0 to 1. They are only between 0 and 0.4 for the raw SPCDs data.

3.3. Robustness to noise

An investigation was conducted to analyse the sensitivity of the proposed approaches to signal noise. The signals were contaminated with pseudorandom Gaussian white noise of zero mean and standard deviation values (σ) ranging from 0.1 to 1.5 at 0.1 increments. This is demonstrated in Figure 22 for the blade 1 SPCD 6 signals at 1000 rpm.

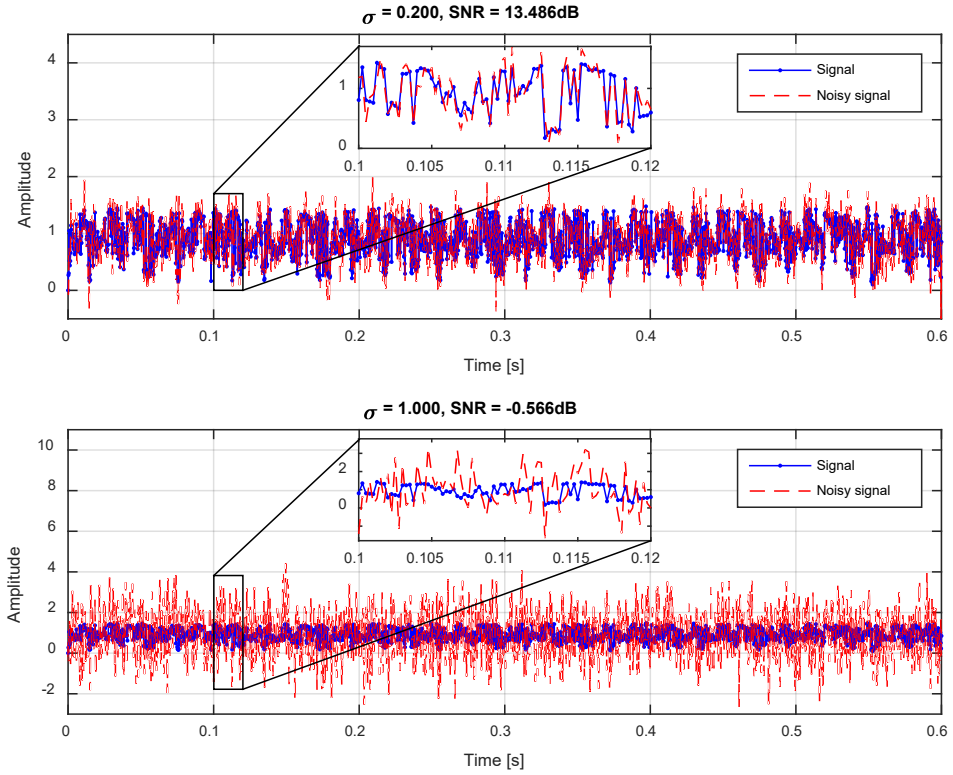

Figure 22: SPCD contamination with Gaussian white noise

At 1000 rpm, it is observed in Figure 23 that the SNR is nearly 0 dB for $\sigma = 1$, indicating that the signal and the noise almost have the same energy at that instant.

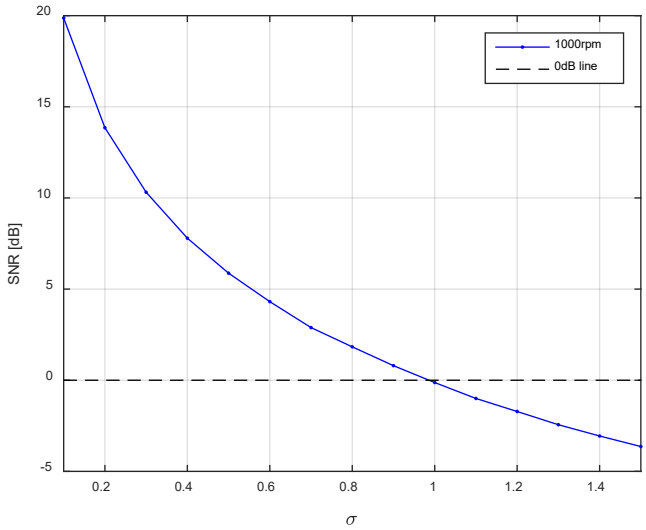

Figure 23: SNR variation with σ

Illustrated in Figure 24 is the clustering performance of PCA and KPCA applied to the MSFs and the order-based time-domain data for the $SPCD_6$ at the three rotational speeds considered.

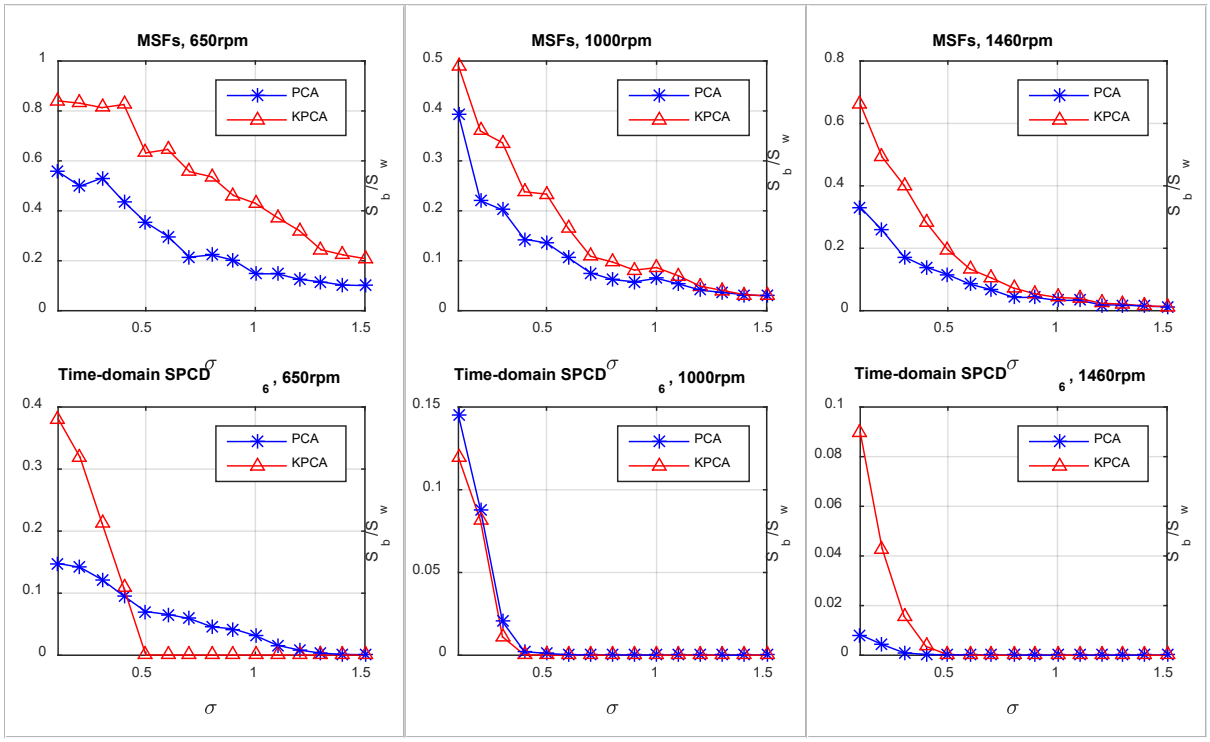

Figure 24: Clustering performance (S_b/S_w) as function of σ

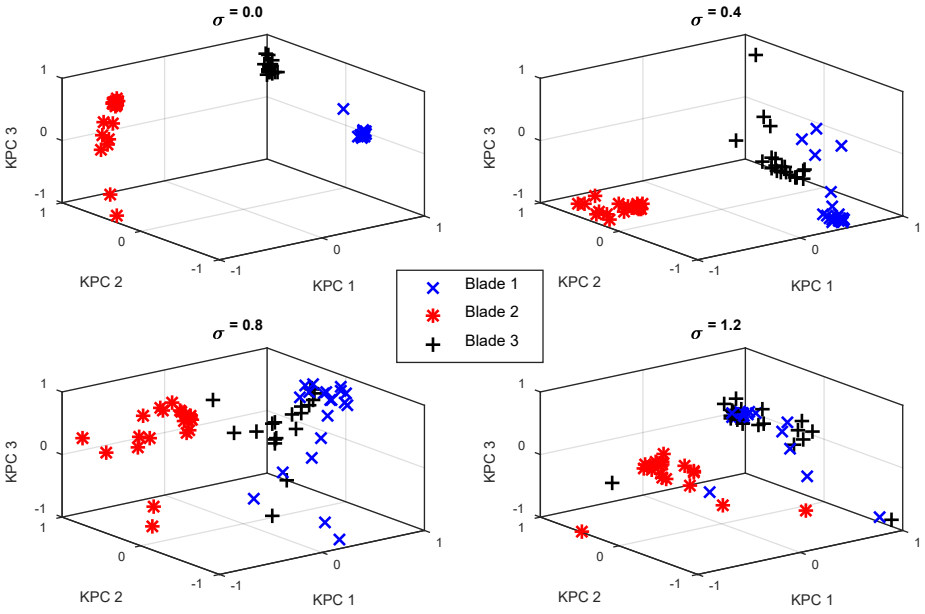

Figure 25: Cluster variation with noise level for KPCA SPCD$_6$ projected in a 3D feature space

Results in Figure 24 indicate that S_b/S_w decreases with decreasing SNR. The clustering performance of PCA and KPCA applied to the raw data degrades significantly with very little noise introduced into the system. For 1000 rpm for instance, S_b/S_w approaches zero at $\sigma = 0.4$, indicating that distinguishing between the different blades' clusters will not be possible. MSF clustering performance

on the other hand is quite robust to noise, and generally degrades substantially when the power of the noise gets closer to $\sigma = 1$. Illustrated in Figure 25 are the results for KPCA projections of MSF values for different levels of noise for $SPCD_6$ at 1000 rpm.

4. Conclusion

In this study, a comparative study of the performance of principal component analysis (PCA) and kernel principal component analysis (KPCA) applied to shape principal component descriptors (SPCDs) for condition monitoring purposes is conducted on a turbomachinery system.

It is illustrated that PCA of the SPCDs associated with the out-of-plane blade responses ($SPCD_5$ and $SPCD_6$) can successfully be used to classify different damaged blades. Projection of these SPCDs into a 3D feature space clearly distinguished between a double-cracked, single-cracked and healthy blades for all speeds except for 1460 rpm. Nonlinear data transformation using Gaussian kernels before performing PCA significantly improved the classification performance, and blades could be distinguished at all speeds. Utilizing a multi-domain statistical feature set composed of eight time-domain and four circular domain statistical features proved to have better blade classification compared to the raw time-domain SPCDs. In terms of robustness to noise, it was successfully illustrated that by using MSFs, successful classification could still be achieved for SPCD signals contaminated with white Gaussian noise with a standard deviation of 1. At higher levels of noise (standard deviation 1.2) the two damaged blades could not be separated but both of them could be distinguished from the healthy blade. This was a significant improvement compared to classification using the raw time-domain data, which was noted to fail at a 0.4 standard deviation noise level.

This study demonstrates that MSFs coupled with KPCA of SPCDs captured using an optical non-contact approach is a viable and very promising condition monitoring strategy expected to perform exceptionally for analysing rotating turbine blades. Future work entails further investigations where more damage screnarios such as delamination and impact damage on other types of blades are considered to effectively investigate the advantages of KPCA in capturing nonlinear blade behaviour.

Full-scale implementation of the approach to analyse industrial wind turbines must also be considered.

5. References

[1] Mark Hutchinson, "GWEC - Global Wind Report 2023," Brussels, Belgium, 2023.

[2] H. F. Zhou, H. Y. Dou, L. Z. Qin, Y. Chen, Y. Q. Ni, and J. M. Ko, "A review of full-scale structural testing of wind turbine blades," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 177–187, 2014, doi: 10.1016/j.rser.2014.01.087.

[3] "Caithness Windfarm Information Forum." Accessed: Oct. 24, 2019. [Online]. Available: http://www.caithnesswindfarms.co.uk/AccidentStatistics.htm

[4] N. Najafi and A. Vesth, "Establishing a robust testing approach for displacement measurement on a rotating horizontal-axis wind turbine," *Wind Energy Science*, vol. 3, no. 1, pp. 301–311, 2018, doi: 10.5194/wes-3-301-2018.

[5] J. Winstroth, L. Schoen, B. Ernst, and J. R. Seume, "Wind turbine rotor blade monitoring using digital image correlation: a comparison to aeroelastic simulations of a multi-megawatt wind turbine," *J Phys Conf Ser*, vol. 524, 2014, doi: 10.1088/1742-6596/524/1/012064.

[6] M. Ozbek, D. J. Rixen, O. Erne, and G. Sanow, "Feasibility of monitoring large wind turbines using photogrammetry," *Energy*, vol. 35, no. 12, pp. 4802–4811, 2010, doi: 10.1016/j.energy.2010.09.008.

[7] A. Khadka, B. Fick, A. Afshar, M. Tavakoli, and J. Baqersad, "Non-contact vibration monitoring of rotating wind turbines using a semi-autonomous UAV," *Mech Syst Signal Process*, vol. 138, p. 106446, Apr. 2020, doi: 10.1016/j.ymssp.2019.106446.

[8] B. Guan *et al.*, "Monitoring the blades of a wind turbine by using videogrammetry," *Opt Lasers Eng*, vol. 152, p. 106901, May 2022, doi: 10.1016/j.optlaseng.2021.106901.

[9] B. Gwashavanhu, "Development and Application of a Photogrammetry Based Statistical Shape Analysis Technique for Condition Monitoring of Rotating Structures," University of Pretoria, Pretoria, 2024.

[10] M. Elsamanty, A. Ibrahim, and W. Saady Salman, "Principal component analysis approach for detecting faults in rotary machines based on vibrational and electrical fused data," *Mech Syst Signal Process*, vol. 200, Oct. 2023, doi: 10.1016/j.ymssp.2023.110559.

[11] A. Zhu, Q. Zhao, T. Yang, L. Zhou, and B. Zeng, "Condition monitoring of wind turbine based on deep learning networks and kernel principal component analysis," *Computers and Electrical Engineering*, vol. 105, Jan. 2023, doi: 10.1016/j.compeleceng.2022.108538.

[12] W. Chen, H. Zhou, L. Cheng, J. Liu, and M. Xia, "Condition monitoring of wind turbine using novel deep learning method and dynamic kernel principal components Mahalanobis distance," *Eng Appl Artif Intell*, vol. 125, Oct. 2023, doi: 10.1016/j.engappai.2023.106757.

[13] G. Hou, J. Wang, and Y. Fan, “Multistep short-term wind power forecasting model based on secondary decomposition, the kernel principal component analysis, an enhanced arithmetic optimization algorithm, and error correction,” *Energy*, vol. 286, Jan. 2024, doi: 10.1016/j.energy.2023.129640.

[14] J. D. Tippmann, X. Zhu, and F. Lanza, “Application of damage detection methods using passive reconstruction of impulse response functions,” 2015, doi: 10.1098/rsta.2014.0070.

[15] B. Gwashavanhu, P. S. Heyns, and A. J. Oberholster, “Shape principal component analysis as a targetless photogrammetric technique for condition monitoring of rotating machines,” *Measurement (Lond)*, vol. 132, pp. 408–422, 2019, doi: 10.1016/j.measurement.2018.09.065.

[16] I. K. Kazmi, L. You, and J. J. Zhang, “A Survey of 2D and 3D Shape Descriptors,” *2013 10th International Conference Computer Graphics, Imaging and Visualization*, pp. 1–10, 2013, doi: 10.1109/CGIV.2013.11.

[17] H. Iwata, H. Nesumi, S. Ninomiya, Y. Takano, and Y. Ukai, “Diallel Analysis of Leaf Shape Variations of Citrus Varieties Based on Elliptic Fourier Descriptors,” *Breed Sci*, vol. 52, pp. 89–94, 2002, doi: 10.1270/jsbbs.52.89.

[18] H. K. Mebatsion, J. Paliwal, and D. S. Jayas, “Evaluation of variations in the shape of grain types using principal components analysis of the elliptic Fourier descriptors,” *Comput Electron Agric*, vol. 80, pp. 63–70, 2012, doi: 10.1016/j.compag.2011.10.016.

[19] V. H. Nguyen and J.-C. Golinval, “Fault detection based on Kernel Principal Component Analysis,” *Eng Struct*, vol. 32, no. 11, pp. 3683–3691, 2010, doi: 10.1016/j.engstruct.2010.08.012.

[20] R. Shao, W. Hu, Y. Wang, and X. Qi, “The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform,” *Measurement*, vol. 54, pp. 118–132, 2014, doi: 10.1016/j.measurement.2014.04.016.

[21] Q. He, F. Kong, and R. Yan, “Subspace-based gearbox condition monitoring by kernel principal component analysis,” *Mech Syst Signal Process*, vol. 21, no. 4, pp. 1755–1772, 2007, doi: 10.1016/j.ymssp.2006.07.014.

[22] W. Caesarendra, B. Kosasih, A. K. Tieu, and C. a S. Moodie, “Circular domain features based condition monitoring for low speed slewing bearing,” *Mech Syst Signal Process*, vol. 45, no. 1, pp. 114–138, 2014, doi: 10.1016/j.ymssp.2013.10.021.

[23] S. C. Martin, G. S. Muster, M. Curilem, G. Fuentealba, C. Gallegos, and C. Melgarejo, “Feature Extraction Using Circular Statistics Applied to Volcano Monitoring,” in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, no. November, 2010, pp. 458–466. doi: 10.1007/978-3-642-16687-7.

[24] C. B. Church, “Turbomachine Internal Pressure and Blade Response Modelling,” University of Pretoria, 2015.

[25] D. Ge and D. Farina, “Spike Sorting,” in *Introduction to Neural Engineering for Motor Rehabilitation*, vol. 22, no. 2, Elsevier, 2013, pp. 155–172. doi: 10.1002/9781118628522.ch8.

[26] C. R. Caro-Martín, J. M. Delgado-García, A. Gruart, and R. Sánchez-Campusano, “Spike sorting based on shape, phase, and distribution features, and K-TOPS clustering with validity and error indices,” *Sci Rep*, vol. 8, no. 1, pp. 33–38, 2018, doi: 10.1038/s41598-018-35491-4.

[27] B. C. Souza, V. Lopes-dos-Santos, J. Bacelo, and A. B. L. Tort, “Spike sorting with Gaussian mixture models,” *Sci Rep*, vol. 9, no. 1, pp. 1–14, 2019, doi: 10.1038/s41598-019-39986-6.

[28] M. H. Gharavian, F. Almas Ganj, A. R. Ohadi, and H. Heidari Bafroui, “Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes,” *Neurocomputing*, vol. 121, pp. 150–159, 2013, doi: 10.1016/j.neucom.2013.04.033.