MULTI-PCA DRIVEN APPROACH for FAULT DETECTION and ROOT CAUSE ANALYSIS of PROCESS EQUIPMENT

Jinendra K Gugaliya
ABB Ability Innovation Centre
Bangalore, India
jinendra.gugaliya@in.abb.com

Rahul Kumar Vij
ABB Ability Innovation Centre
Bangalore, India
Rahul.Kumar-vij@in.abb.com

Srini Ramaswamy
ABB, Power Generation
Cleveland, USA
srini.ramaswamy@us.abb.com

Lakshwin Shreesha M K
BMS College of Engineering
Bangalore, India
shreeshalakshwin@gmail.com

Abstract

Principal Component Analysis (PCA) is quite popular for fault detection and diagnosis in industrial applications. PCA assumes linear relationships among the features and serves to represent them as a linear combination. However, a typical industrial application can have non-linearity due to operation at multiple operating regions or inherent non-linear relationships among the features. This paper proposes a novel clustering based Multi-PCA approach which can divide the overall non-linearity into simpler linearity's which can subsequently be modelled by multiple PCA models. The clustering is done with the use of domain knowledge where the fact that an operation of an asset at different operating points can lead to multimodal distribution of the variables. The proposed approach is structured systematically with the following steps 1) Feature set selection 2) Hierarchical Density Based Spatial Clustering (HDBSCAN) and 3) Fitting a PCA model in each cluster. The proposed approach retains the computational simplicity of the PCA compared to models based on other non-linear modelling approaches such as neural network based autoencoders. Finally the paper also proposes a simplified Root Cause Analysis (RCA) algorithm for identifying the cause of the fault.

1 Introduction

Industrial assets such as motors, pumps, fans, turbines etc. are subject to faults and failures due to operation at excess load conditions or due to aging effects. Identifying that an industrial asset is drifting towards an abnormal condition is the key to avoid unplanned downtime of an industry due to asset failure. In literature, there are two important approaches to tackle this challenge of detecting abnormal asset health. The first approach is based on detailed know-how and physics of the asset and second approach is black box. The first approach works well for simpler assets such as a motor, as the underlying physics is well established. However, this approach is not easily scalable and it requires one

to develop physics based models for every asset. Additionally, as industrial assets become complex, such an approach is difficult to implement. Hence, there is a significant shift to apply data driven approaches for asset health monitoring. Due to availability of low cost sensors and digital technology, lots of data can be collected from an industrial asset and approaches based on machine learning principles can be applied to learn the model for the asset, in a semi-automated manner. Such an approach is easily scalable and can be applied to a variety of machines.

Some of the earliest fault detection techniques were model based. One such popular algorithm was based around analytic redundancy, wherein a comparison between the inputs of the monitored system and the output obtained from an analytical mathematical model was carried out to detect the presence of a fault (M. Frank 1990). However this comparison was a naive estimate and failed to capture faulty conditions in high dimensional spaces. Following which several approaches based on multivariate-statistical process control methods (MacGregor and Kourti 1995); (Kresta and Marlin 1991); (Macgregor 1994) were presented for the diagnosis of complex physical processes. The usage of state observers by modelling faults as state variable changes (Isermann 2005) provided a better strategy for aberration detection compared to statistical processes albeit at higher computational cost.

A pressing need to capture and localize abnormalities at reduced computational rates brought about the usage of Principal Component Analysis (PCA). PCA defines a new outlook to the data and aims to capture hidden structure underneath data redundancy and noise (Pearson 1901). An abundant list of algorithms based around the PCA is evident in literature. One such approach involves using the $\mathbf{Q}$ and $\mathbf{T^2}$ statistic (Villegas, Fuente, and Rodríguez 2010) for fault detection. This methodology was subsequently simulated for fault detection in a waste water treatment plant (Garcia-Alvarez 2009) wherein the authors showcase results which capture local linear structure only. An improvement to the conventional PCA model was brought about by introducing the dynamic PCA (DPCA) (Russell, Chiang, and Braatz 2000) which is established by considering the dependency of current observations on previous time instances as well. A non-linear modification to the PCA involved the combination of using the Kronecker product, wavelet decomposition
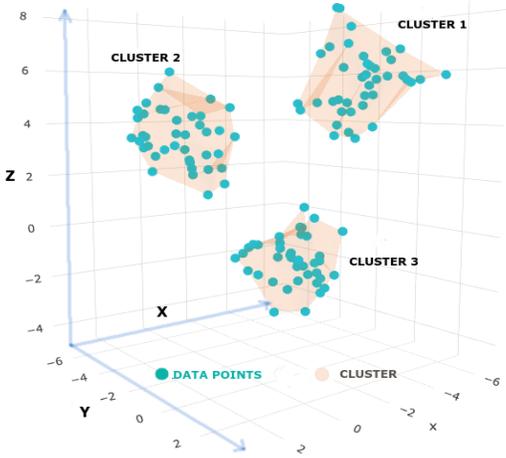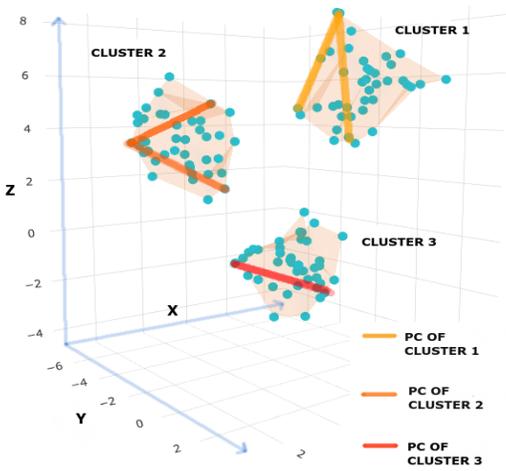
(a) Clustered Data

(b) PCA applied to each cluster

Figure 1: Multiple PCA applied on clustered data

and sliding median filter for determination of a fault in non-linear data-sets (Zhang, Li, and Hu 2012).

All of these methods suffer from the inability of the Hotelling's T^2 to identify and isolate the responsible feature (The Hotelling's T^2 test is simply a multivariate counterpart of the T-test). Further the fault detection index used is extremely sensitive to anomalies, making them susceptible to false positives. In order to fix this problem, a new fault detection index based on the sum of the squares of the last few principal components weighted by the inverse of their variances was developed which yielded a good detection rate in the dependent as well as independent variables (Benaicha et al. 2010). This work also points out that hierarchical contribution plots provide sufficient partitioning to localize any anomaly, provided the stochastic nature of bloc size is evaluated by a definite formula. A latest development pertinent to the PCA involves decomposing variables using the Empirical Mode Decomposition (EMD) (Du and Du 2018). Fault

detection is subsequently carried out by applying PCA to the decomposed variables and detecting small shifts in data using a Cumulative Sum control cart (CUMSUM).

Kernel PCA (Schölkopf, Smola, and Müller 1996) extends the idea to the non-linear case wherein the kernel trick is used to learn a linear representation in non-linear space. While these models have been deployed with significant success to capture non-linearity's their success is correlated with the assumption of the data generating process. Most often, an RBF kernel is used so as to encode maximum uncertainty about the data generating distribution and justifiably so because the central limit theorem points to the behavior of several random variables to be Gaussian distributed. However, this may not be the case for several real world processes. Therefore selecting the correct kernel may prove to be an exhaustive process which scales exponentially with more data and even after, may produce poor results. Our method circumvents both these problems while remaining inexpensive.

Artificial Neural Nets (ANN) have also been employed for aberration detection. One such work carries out anomaly detection and root cause analysis using a Bayesian network (Amin 2018). A similar analysis which borrows the technique of decomposing the T^2 statistic proved to be extremely effective for non-linear fault diagnosis (Verron, Li, and Tiplica 2010).

While most of the proposals in literature are promising, the need to maintain a balanced stance on computational simplicity while simultaneously having to achieve significant sensitivity for fault detection is still a problem that remains unsolved. A simple yet robust solution that the PCA offers is limited in scope due to it's linear nature (Villegas, Fuente, and Rodríguez 2010), (Garcia-Alvarez 2009), (Russell, Chiang, and Braatz 2000).

Our main contribution is an extension of the classical PCA framework for non-linear systems. We propose a systematic approach to capture non-linearity through several linear models (by applying several PCA models on chunks of localized data) while retaining the computational simplicity of the single PCA. This paper successfully demonstrates the proposed approach on an industrial asset having several years of historical data.

The Multi-PCA (Fig.1) offers a simple solution of breaking the non linearity through clustering and then building a PCA model for each cluster. This approach results in a framework that can detect the faults with reasonable accuracy. The concept proposed by Liling Ma et al. (Ma et al. 2004) presents a similar idea of using multiple PCA models. In this case however, process monitoring is achieved by weighing each of the sub-PCA models using the K-means clustering technique and creating a decision boundary based on Hotelling's T^2 statistic. K-means clustering is biased to choose local data points because it splits the space into Voronoi cells. Moreover, it performs poorly when tasked with finding clusters in data inherent with varying densities and is acutely affected by the choice of K. K-means clustering does not in any way identify noise prevalent in the data and assigns them to a cluster regardless of its influence. As opposed to the naive clustering processes adopted

in (Ma et al. 2004), the Multi-PCA approach proposed in this paper, employs the Hierarchical density based Spacial clustering (HDBSCAN) algorithm (McInnes, Healy, and Astels 2017), which is radically inexpensive and robust. The elbow method developed about the mean square error serves as a sufficient statistic for setting the hyper parameters of the HDBSCAN, following which the Multi-PCA modelling approach is applied to the clustered space. While in (Ma et al. 2004) the SOFM (self-organizing feature map) neural network (Kohonen and Honkela 2007) calculates fault thresholds using the multiple PCA components, we showcase that determining thresholds from reconstructions of projected data provides similar results at a fraction of the computational cost.

We also present a novel feature selection strategy to select essential features for clustering. It is to be noted that our model can only detect known fault states that the Multi-PCA model encounters during training. Hence it is necessary to provide a wide array of fault cases to the model.

2 Preliminaries

This section explains Multiple Principal component analysis (Multi-PCA), describes it's algorithmic flow chart for fault detection and provides an overview of the Hierarchical Density Based Spacial Clustering algorithm along with illustration of the feature selection methodology for the process of clustering.

2.1 Multiple Principal Component Analysis

The Principal Component Analysis (PCA) (Pearson 1901) is an orthogonal transformation that carries out dimensionality reduction by converting a multivariate space into a subspace which preserves maximum variance of the original space in minimum number of dimensions. PCA can be thought of as looking at data from it's most informative viewpoint in the transformed space.

The Multi-PCA borrows this characteristic and extends it to non-linear data by clustering the data space into several clusters and applying an independent PCA on each (Fig.1). The essence of clustering the data space is to account for several operating regions prevalent in the steady state data of the plant.

To formally describe the process of fault detection using the Multi-PCA, consider a standardized (zero mean and unit variance) data matrix $\mathbf{X} \in \mathbf{R^{nxm}}$ (representing the steady state model of the plant), where n indicates the number of samples and m denotes the number of feature variables. Clustering the data space $\mathbf{X}$ leads to clusters $\mathbf{x_1, x_2, x_3, ..., x_q} \in \mathbf{X}$ where $\mathbf{q}$ is equal to the number of clusters. Assuming each component ($\mathbf{x_i} \parallel \mathbf{i = 1, 2, .., q}$) to be independent of each other, the co-variance matrices $\mathbf{C_i}$ of $\mathbf{x_i} \ \forall \ i = 1, 2, .., q$ describing the variance between the features can be constructed as:

$$\mathbf{C_i} = \frac{1}{1 - \mathbf{n_i}} \mathbf{x_i^t x_i} \quad \forall \ i = 1, 2, \ldots, q \tag{1}$$

the singular value decomposition of $\mathbf{C_i} \in \mathbf{R^{mxm}}$ is given as:

$$\mathbf{C_i} = \mathbf{V_i} * \mathbf{\Lambda_i} * \mathbf{V_i^t} \quad \forall \ i = 1, 2, \ldots, q \tag{2}$$

where $\mathbf{\Lambda_i}$ is a diagonal matrix containing the eigenvalues of C_i in descending order($\lambda_1 > \lambda_2 > \ldots > \lambda_m$). Λ is obtained by solving the characteristic equation:

$$\mathbf{W_i} - \lambda \mathbf{I} = \mathbf{0} \tag{3}$$

where $\mathbf{W_i} = \mathbf{C_i} * \mathbf{C_i^t}$, $\forall \ i = 1, 2, .., q$ and the columns of $\mathbf{V_i}$ are the eigenvectors of $\mathbf{C_i}$. The transformation matrix for each cluster $\mathbf{P_i}$ is formulated by choosing a eigenvectors (columns of $\mathbf{C_i}$) corresponding with a eigenvalues.

$$\mathbf{T_i} = \mathbf{x_i} \ \mathbf{P_i} \quad \forall \ i = 1, 2, .., q \tag{4}$$

equation 4 describes the transformation of each cluster to a reduced dimension and $\mathbf{P_i}$ denotes the transformation matrix for it's respective cluster. A standard measure used to calculate a i.e. number of principal components, based on desired variance is specified by the cumulative percent variance (CPV) formulation:

$$\mathbf{CPV(a)} = \frac{\sum_{\mathbf{j=1}}^{\mathbf{a}} \lambda^{\mathbf{j}}}{\mathbf{trace(C_i)}} * \mathbf{100} \tag{5}$$

2.2 Sequence of events for Multi-PCA fault detection

We use the trends of *Hotelling's* T^2 and Q statistics to analyze abnormality in the data and tailor it to cater for fault detection in the case of Multi-PCA.
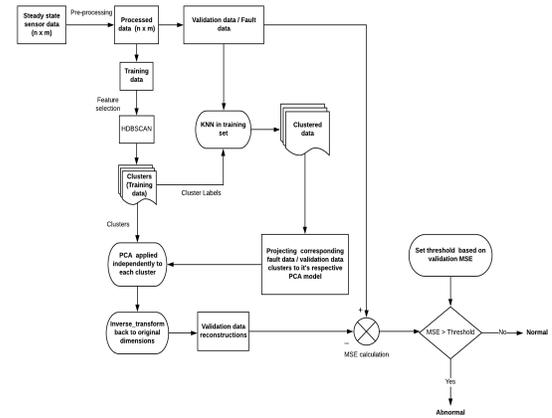
Figure 2: Multi-PCA fault detection

Steady state sensor data collected from an industrial asset is normalized and treated as training data. Fault data recorded during the malfunction of the industrial asset serves as test case to detect anomalies using the Multi-PCA model.

The model is formulated based on the training data alone, consider $\mathbf{X}$ to be the training data and $\mathbf{X_f}$ to be the fault data. $\mathbf{X}$ is clustered into different operating regions by the Hierarchical Density based Spacial Clustering Algorithm (subsection 2.3) yielding clusters with unique cluster Id's. The K-Nearest Neighbors (KNN) classifier is now employed to classify test data points into one of the cluster Id's based on

majority voting of the data points by considering it's K nearest neighbor's. KNN is based on the Eucledian distance and provides a simple, inexpensive and robust solution to designate data points into different clusters.

Following this, principal components are determined independently for each cluster in the training data (Eq. 4). Let $\mathbf{j}$ denote the output clusters from KNN for the test data, $\mathbf{j} \in (\subseteq \mathbf{q}) \mid (\mathbf{q})$ itself i.e, j can either contain a few or all the clusters of the training data set $\mathbf{X}$. Equation 4 can now be extended as:

$$\mathbf{T_{jf}} = \mathbf{y_{jf}} \; \mathbf{P_j} \quad \forall \, \mathbf{j} = 1, 2, \ldots, \mathbf{q} \; or \; < \mathbf{q} \qquad (6)$$

where, $\mathbf{y_{jf}}$ are the test data clusters classified by KNN and $\mathbf{P_j} = \mathbf{P_i} \parallel \forall \, \mathbf{j} \in \mathbf{i}$. Equation 6 transforms the data onto the new space based on the steady state transformation matrices $\mathbf{P_i}$. Inverse transformations are applied to revert the training and the fault data back to 'm' dimensional space.

$$\widehat{\mathbf{y_{jf}}} = \mathbf{T_{jf}} \; \mathbf{P_j^t} \quad \forall \, \mathbf{j} = 1, 2, \ldots, \mathbf{q} \; or \; < \mathbf{q} \qquad (7)$$

the inverse transforms of eq. 7 carries with it the error of projection. This error is expected to be large for the faulty case and is computed for each cluster as:

$$\mathbf{E_f} = \mathbf{Y_{jf}} - \widehat{\mathbf{Y_{jf}}} \quad \forall \, j = 1, 2, .., q \qquad (8)$$

The threshold is set based on the validation set. KNN is used to classify the validation data points into clusters, each of these clusters are projected onto their respective steady state PCA model and are reconstructed back. The reconstructed data is compared with the original validation data set to produce a sample error. A threshold of 3 standard deviations from the mean of the sample error is set, which serves as the decision boundary to detect anomalies. Fig.2 depicts the algorithm.

2.3 Hierarchical Density Based Spacial Clustering Algorithm (HDBSCAN)

The concept of Multi-PCA requires clustering the data into different operating zones (Fig. 1). Literature presents several algorithms for clustering, however each of these have a trade off to account for in terms of computation cost and data size. Figure 3 presented in (McInnes, Healy, and Astels 2017) depicts the superior performance of the HDBSCAN over the current state of art algorithms.

HDBSCAN transforms a N-dimensional space according to the density of the data by defining a new distance metric. It's hyper-parameters include *minimum cluster size* and *minimum samples* which were decided through the elbow technique. Using the distance matrix thus obtained it constructs a minimum spanning tree based on Prim's algorithm (Prim 1957). A dendrogram is formed by arranging the edges of the spanning tree in the increasing order of their distance and thereby creating clusters for each edge group. The important clusters are retained by a measure of $\lambda = \frac{1}{distance}$ giving an indication of *"how long"* the clusters retain themselves.

HDBSCAN scales well to large datasets and is effective at global clustering. The algorithm also detects outliers in the data and classifies them as noise. These outliers are data
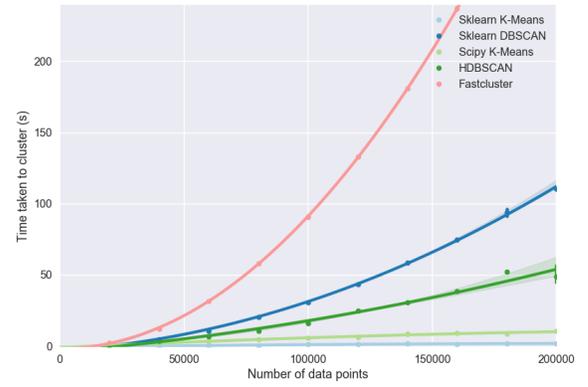
Figure 3: Comparison of clustering algorithms

points that are obtained as a result of sensor faults and signify noise in the dataset. For example: A faulty tachometer may output a negative value of speed or a very large value that is improbable. HDBSCAN was found to identify such stray data points and these points were removed. HDBSCAN thus provided a way to account for sensor related noise and drift.

2.4 Feature selection for clustering

Clustering of a multidimensional dataset requires feature selection. Correlated features do not aid the clustering methodology. They increase computational time without improving *cluster quality*. A need to present only important features to the clustering methodology has led to several algorithms developed in literature. Michael Fop and Thomas Brendan (Fop and Murphy 2017) present several approaches involving Gaussian mixture models and latent class analysis models. Dirichlet process mixture models were also proposed for variable selection (Kim, Tadesse, and Vannucci 2006). As opposed to finding a common feature subset that is relevant to all clusters, Yuanhong Li et al. (Li, Dong, and Hua 2008) developed a localized feature selection method for clustering.

The proposed method of feature selection is based on inspecting the density plots of each variable and looking for features exhibiting distinct operating regions (multi-modal distributions). This method provided sufficient simplification and served as a robust criterion for selecting distinct variables for clustering. The idea behind such a feature selection method is that, if an asset operates at $\mathbf{N}$ distinct operating regimes, then one can expect $\mathbf{N}$ distinct peaks in its density plot.

Variables with two or more operating regions are chosen for clustering, while the rest are rendered redundant in this particular analysis. An illustration is shown in Fig.4

3 Case Study

We assess the performance of the proposed Multi-PCA through a comparison with the Single PCA approach employing the methodology aforementioned.
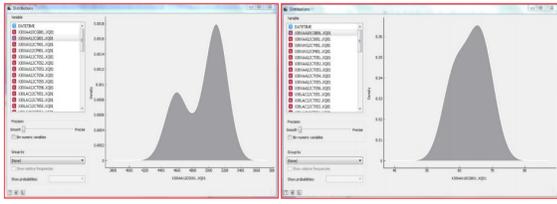
Figure 4: Density plots of two variables, one with two operating regions (L) and one with only a single operating zone (R)

3.1 Data Description

Gas turbine data of a power generation plant comprising of forty five features sampled over one minute intervals was the dataset used for this case study. The training data used to develop the steady state model $\mathbf{X}$, is a matrix of dimensions 37200x45 with total 37200 samples each having 45 variables. The data also has information about the dates on which the faults are reported. Six test files are prepared as test cases to detect anomalies corresponding to the 6 faults. Each fault file contains data 24 hours prior to the fault reporting time. Hence each fault file acts as a test case for the proposed algorithm and ideally should detect possible anomalies.

We demonstrate the effectiveness of the proposed Multi-PCA algorithm over a single PCA algorithm. Table: 1 summarizes the various test case results for both algorithms. In some test cases, both single PCA and Multi PCA algorithms detect faults. Whereas in other test cases (Cases: 1, 3, 5, and 6) only Multi-PCA approach was able to detect the fault. Also, in all test cases, Multi-PCA was able to detect faults.

In order to prove the point further, fault case F5 is analyzed in detail. For F5, both the single PCA and the Multi-PCA models are created based on the training data, and the projected test data is reconstructed back. In Fig.8 the actual and reconstructed signal of a variable called turbine speed is depicted. Fig.8a is the original turbine speed signal, whereas Fig.8b indicates the reconstructed signal using the single PCA model and Fig.8c indicates the reconstructed speed signal using the Multi PCA model. It is clear from Fig.8 that the Multi-PCA approach is able to reconstruct the signal very well and closely matches with that of the training data signal. This is a possibility since the Multi-PCA divides the data into multiple regimes whereas a Single PCA fits to the entire data distribution. Also, feature selection (section 2.4) plays the role of a naive correlation detector and in the case of F5, it identifies seven variables out of forty five to be uncorrelated.

The Hierarchical clustering algorithm (section 2.3) uses only these seven variables to cluster the data into two clusters, while simultaneously detecting outliers prevalent in the data; Fig.5 depicts clustering of the **Turbine Flow Speed** variable. We use the uncorrelated features to ensure that redundant features do not interfere with the process of clustering. Once the single PCA and Multi-PCA models are built, F5 data is projected onto its respective principal components, are reconstructed back and MSE per sample is computed. The results are as shown in Fig.6 and Fig.7. As per Fig.6, the mean squared errors (MSE) vary in range of 0-600 indicating that the single PCA is not able to capture all variation in the data. The MSE threshold is set at 100 (based on the validation data) to decide if a data point is normal or not. In the case of F5, using the single PCA model, majority of the sample errors are within the threshold, providing an incorrect indication that the gas turbine is in normal operation. Therefore, the single PCA model is not confident to mark the data set as faulty. Whereas, in the case of the Multi-PCA approach (Fig.7), a good *fit* to the data in both clusters (clustered by HDBSCAN) is achieved. This snug *fit* places the majority of the sample errors above the calculated validation threshold (different from that of the single PCA), conclusively indicating a fault in F5.

Figure 5: Clustered data (Turbine Flow Speed)
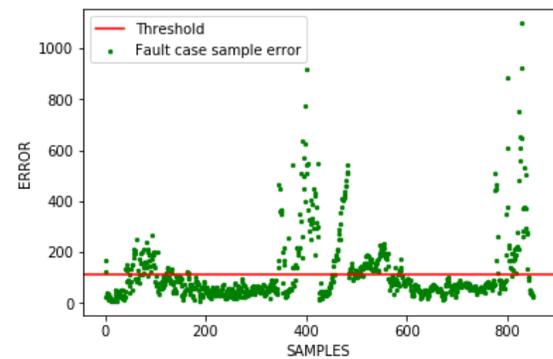
Figure 6: Fault assessment using a single PCA (SPCA)

In order to test the proposed algorithm's ability to detect normal operation of the gas turbine, a new test file was prepared using 24 hours of the data from normal operation of the gas turbine. This construed data set was not used during training of the Multi-PCA algorithm. Results for this case are as shown in Fig.9, the test file was found to contain three clusters when clustered with a set of ten uncorrelated features and the reconstructed sample errors for each cluster indicated normal operation of the gas turbine as majority of the test samples were well below the threshold values of the corresponding cluster. This result showcases the lack of bias
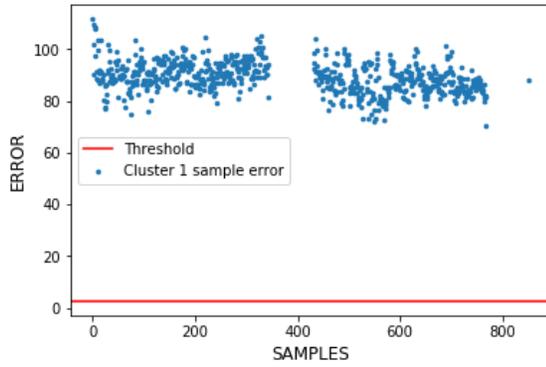
(a) MPCA fault detection in cluster 1 (Fault data)
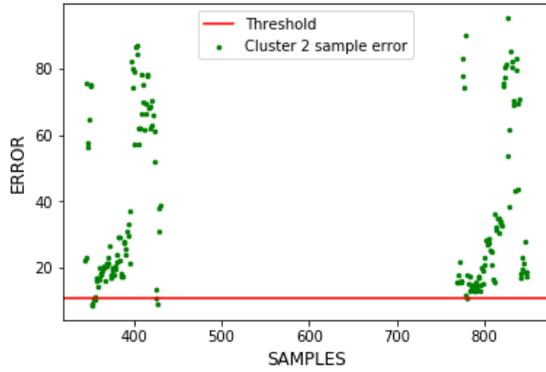
(b) MPCA fault detection in cluster 2 (Fault data)

Figure 7: Fault assessment using Multi PCA

of the Multi-PCA model towards faults while successively demonstrating its ability to classify anomalies well.

Another experiment was conducted in order to test the Multi-PCA against sensor bias faults. A bias was deliberately added to two variables in the steady state dataset which is representative of the normal operation of the gas turbine. The Multi-PCA algorithm was used to detect a fault in such data. The results of this test is shown in Fig.10. The Multi-PCA algorithm indicates abnormal behavior as the MSE per sample is greater than the threshold value for majority of the data points. The threshold at which a bias triggers the fault was found to be approximately three percent above the steady state value of the variables.

Test Case	SPCA	MPCA	Fault Detection
F1.	168.361	119.915	MPCA
F2.	349.667	145.337	Both
F3.	95.498	62.419	MPCA
F4.	1252.578	247.136	Both
F5.	121.629	79.065	MPCA
F6.	71.819	38.1672	MPCA

Table 1: Comparison of errors between Single PCA and Multi-PCA
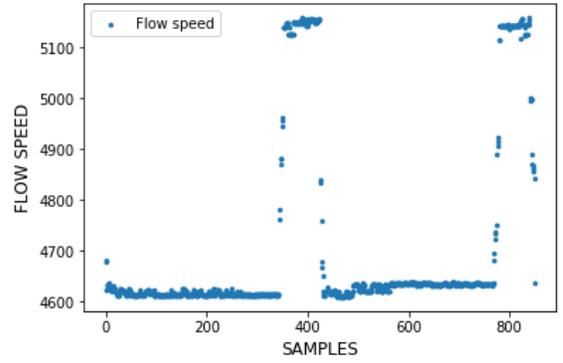
(a) Actual Fault data (Flow Speed variable) (F5)
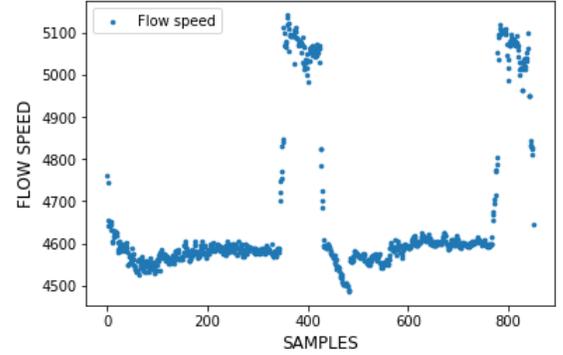
(b) Reconstruction of Flow Speed by Single PCA
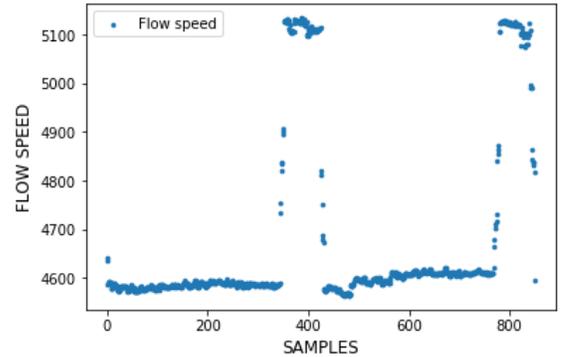
(c) Reconstruction of Flow Speed by Multi-PCA

Figure 8: Comparison of reconstructions by the single PCA and Multi-PCA

3.2 Root cause analysis results

Root cause analysis provides insight into which particular variable is contributing to the anomaly. RCA provides a contribution plot which indicates how each variable is contributing in magnitude to the anomaly. Fault F5 is used to demonstrate the root cause analysis performed using Multi-PCA approach and results for the same are as depicted in Fig.11.

The scatter plot (top left corner) is a plot of the sample error for F5. The bar graph indicates the magnitude of contribution of any variables to the fault and weighs them in decreasing order. The top five variables contributing to the fault are identified (Note that the contributing variables are
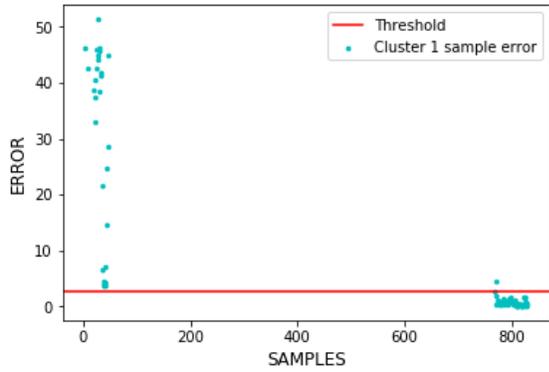
(a) MPCA fault detection in cluster 1
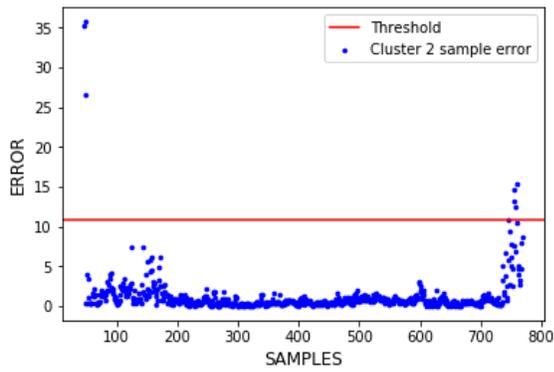
(b) MPCA fault detection in cluster 2
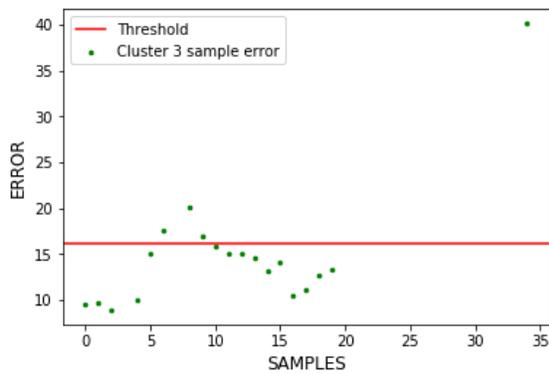
(c) MPCA fault detection in cluster 3

Figure 9: Test for false positives in the Multi-PCA
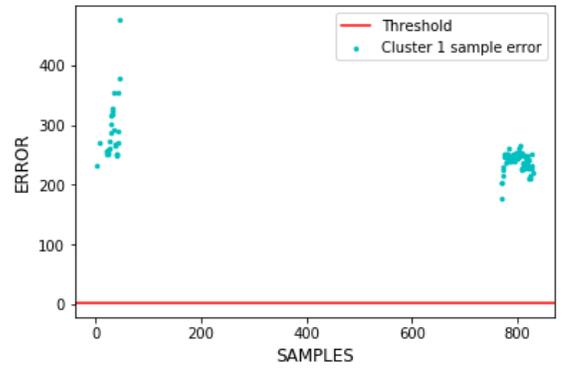
(a) MPCA fault detection in cluster 1
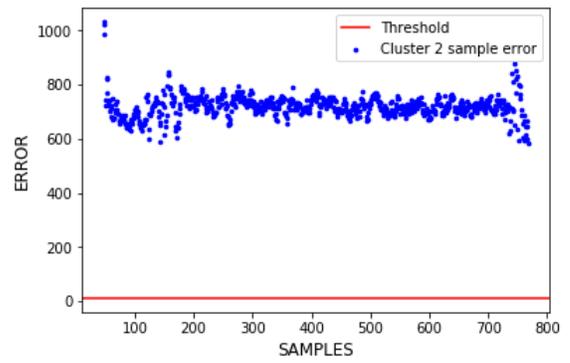
(b) MPCA fault detection in cluster 2
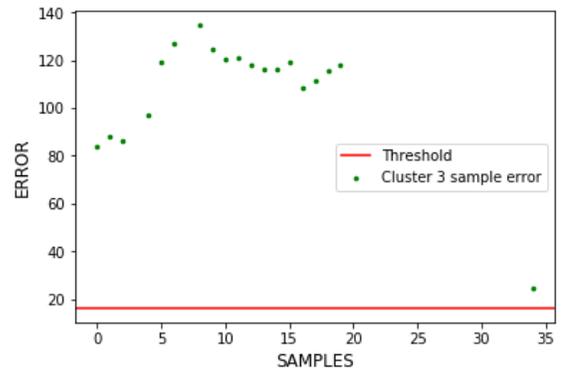
(c) MPCA fault detection in cluster 3

Figure 10: Test case with Deliberate Bias added (Multi-PCA)

(a) RCA

(b) Variable 24 in train and test dataset
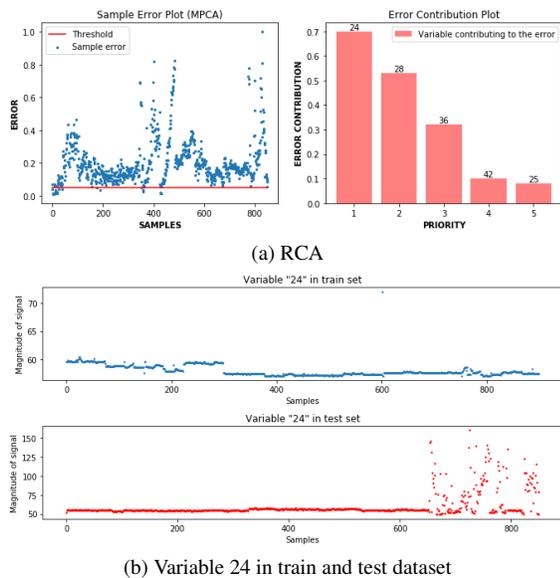
Figure 11: Root Cause Analysis for fault F5

those of the raw dataset (R^m) and **not** of the principal components; All aspects of fault detection are carried out in the raw un-transformed space itself). In order to provide further insight to the subject matter expert (SME), the steady state signal and the aberrant signal are compared as shown in the bottom section of Fig.11. RCA presents the user with a tool to interactively detect the variables contributing to a fault.

4 CONCLUSION

This paper has successfully demonstrated the superiority of the Multi-PCA approach over the single PCA in an industrial case study of a gas turbine. The Multi-PCA approach is able to detect all six faults of the gas turbine. The Multi-PCA approach is also able to detect sensor bias issues in a dataset. This novel approach of feature selection, data clustering followed by PCA model building was found to be quite robust for industrial applications.

References

Amin, M. T. 2018. *Fault Detection and Root Cause Diagnosis using Dynamic Bayesian Network*. Ph.D. Dissertation, Memorial University of Newfoundland.

Benaicha, A.; Guerfel, M.; Bouguila, N.; and Benothman, K. 2010. New Pca-Based Methodology for Sensor Fault Detection and Localization. *In International Conference of Modeling Simulation MOSIM (Vol. 10)*.

Du, Y., and Du, D. 2018. Fault Detection using Empirical Mode Decomposition based PCA and CUSUM with Application to the Tennessee Eastman Process. *IFAC-PapersOnLine* 51(18):488–493.

Fop, M., and Murphy, T. B. 2017. Variable Selection Methods for Model-based Clustering. *arXiv* 12:1–48.

Garcia-Alvarez, D. 2009. Fault detection using Principal Component Analysis (PCA) in a Wastewater Treatment Plant (WWTP). *62-th International Student's Scientific Conference, 2009* 6.

Isermann, R. 2005. Model-based fault-detection and diagnosis - status and applications. *Annual Reviews in Control* 29:71–85.

Kim, S.; Tadesse, M. G.; and Vannucci, M. 2006. Variable selection in clustering via Dirichlet process mixture models. *Biometrika* 93(4):877–893.

Kohonen, T., and Honkela, T. 2007. Kohonen network. *Scholarpedia* 2(1):1568.

Kresta, J. V., and Marlin, T. E. 1991. Multivariate statistical monitoring of process operating performance. *The Canadian Journal of Chemical Engineering* 69(1):35–47.

Li, Y.; Dong, M.; and Hua, J. 2008. Localized feature selection for clustering. *Pattern Recognition Letters* 29(1):10–18.

M. Frank, P. 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. a survey and some new results. *Automatica* 26:459–474.

Ma, L.; Jiang, Y.; Wang, F.; and Gao, F. 2004. Multi-pca models for process monitoring and fault diagnosis. *IFAC Proceedings Volumes* 37:667–672.

MacGregor, J. F., and Kourti, T. 1995. Statistical process control of multivariate processes. *Control Engineering Practice* 3:403–414.

Macgregor, J. 1994. Statistical process control of multivariate processes. *IFAC Postprint Volume* 427 – 437.

McInnes, L.; Healy, J.; and Astels, S. 2017. hdbscan: Hierarchical density based clustering. *J. Open Source Software* 2(11):205.

Pearson, K. 1901. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2(11):559–572.

Prim, R. C. 1957. Shortest connection networks and some generalizations. *Bell System Technical Journal* 36(6):1389–1401.

Russell, E. L.; Chiang, L. H.; and Braatz, R. D. 2000. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 51(1):81–93.

Schölkopf, B.; Smola, A.; and Müller, K.-R. 1996. Nonlinear component analysis as a kernel eigenvalue problem.

Verron, S.; Li, J.; and Tiplica, T. 2010. Fault detection and isolation of faults in a multivariate process with bayesian network. *Journal of Process Control* 20(8):902–911.

Villegas, T.; Fuente, M.; and Rodríguez, M. 2010. Principal component analysis for fault detection and diagnosis, experience with a pilot plant. *Proceedings of the 9th WSEAS International conference on computational intelligence, man-machine systems and cybernetics* 147–152.

Zhang, Y.; Li, S.; and Hu, Z. 2012. Improved multi-scale kernel principal component analysis and its application for fault detection. *Chemical Engineering Research and Design* 90(9):1271–1280.