Corruption-Resilient Detection of Event-Induced Outliers in PMU Data: A Kernel PCA Approach

Kaustav Chatterjee and Nilanjan Ray Chaudhuri
School of Electrical Engineering and Computer Science
The Pennsylvania State University
University Park, PA 16802
Email: kuc760@psu.edu, nuc88@psu.edu

***Abstract*—Bad data outliers and malicious corruption in Phasor Measurement Unit (PMU) data having signature similar to that of a highly nonlinear event-induced oulier can challenge reliable event detection when linear principal component analysis (PCA)-based metrics are used. This paper presents a moving window based kernel PCA approach for accurately detecting event-induced outliers in presence of such corruptions in data. It is demonstrated that with appropriate tuning of kernel parameters, the change in the square of the norm of principal component score between successive windows along the direction of maximum variance in feature space can be used as a metric for corruption-resilient detection of event-induced outliers. Analytical justification for the same is provided along with a bound on this change. The performance of the proposed metric is validated on both synthetic data and field measurements.**

***Index Terms*—Kernel PCA, PMU, Event Detection, Bad Data Outlier, Cyber Attack.**

NOTATIONS

$\mathbf{x}_i$	$\in \mathbb{R}^m$	Measurement vector at i^{th} time instant
$\mathbf{X}^{(k)}$	$\in \mathbb{R}^{m\times N}$	k^{th} window of measurement data
$\phi(\mathbf{x_i})$	$\in \mathbb{R}^F$	Mapping of $\mathbf{x}_i$ in feature space
$\mathbf{\Phi}^{(k)}$	$\in \mathbb{R}^{F\times N}$	k^{th} window of feature space samples
$\mathbf{C}$	$\in \mathbb{R}^{F\times F}$	Feature space covariance matrix
$\mathbf{v}_j$	$\in \mathbb{R}^F$	j^{th} eigenvector of covariance matrix $\mathbf{C}$
$\mathbf{K}^{(k)}$	$\in \mathbb{R}^{N\times N}$	Kernel matrix for k^{th} window
$\boldsymbol{\alpha}_j$	$\in \mathbb{R}^N$	j^{th} eigenvector of kernel matrix $\mathbf{K}$
ζ_j	$\in \mathbb{R}$	Norm of the PC score along $\mathbf{v}_j$

I. INTRODUCTION

The advent of Phasor Measurement Units (PMUs) capable of reporting high fidelity time synchronized measurements across wide geographies, have significantly improved the situational awareness in control rooms [1], [2]. Powered by advanced data analytics these synchronized phasor measurements have found many applications in power system monitoring including real time event detection and classification [3].

Utilities in North America- BPA, PJM, TVA, and NYPA among others, have an expanding network of PMUs reporting data to respective control centers, where the samples are time-aligned by Phasor Data Concentrators (PDCs) and are fed to modules of dedicated algorithms looking for specific event signatures in the data. In this context, an event signature is defined as characteristic patterns in system response following a disturbance like fault, line tripping, generator outage, etc.

The performance of an event detector is not only determined by how accurately it can differentiate between the onset of an event and ambient condition, but also by its ability to do so when samples are corrupted with bad data outliers with signatures similar to that of an event outlier. To understand this better, consider a spurious outlier in voltage measurement arising out of an unreliable sensor or loss of a data packet in transit, padded as zero. A robust detector exploiting the correlation in data should not interpret this transient dip in voltage as a fault and trigger the alarm for an event. The origin of bad data however is not limited to noisy and erroneous measurement systems alone- it can creep in through malicious intruders exploiting the security breaches in the cyber layer [4]. Under scenarios where a sensor or a pool of sensors have been compromised to replay pre-recorded disturbance data, it becomes challenging for the event detector to distinguish attacks from events. With threats looming on our grids in the era of post-Ukrainian grid attack [5], it is imperative to instill resilience into monitoring and control operations so that it becomes difficult for an attacker to fool an operator into taking wrong decisions. Thus motivated, the goal of this paper is to design a robust event detection scheme resilient to such corruptions in PMU data streams.

There could be two broad approaches to address the challenge of bad data jeopardizing the reliability of event detection. Firstly, one can design a data preprocessor to act as an anomaly detection [6] and correction engine [7], which would clean the incoming signals of any measurement inconsistencies. Authors in [7]–[9] have used low-rank matrix completion methods to reconstruct signals from grossly corrupted and attacked data samples. The output of these could then be used for a variety of monitoring applications including event detection. However, techniques like robust PCA face challenges in distinguishing outliers at the onset of an event from corruption. The other approach is to ensure that the event detector is itself robust to bad data. This paper discusses one such strategy of circumventing the issues of bad data by exploiting the degree of correlation in data channels and the low rank property of a measurement window.

Extensive literature exists on use of multivariate statistical methods for event detection and localization in power systems including [10]–[13] and references therein. In [10] authors have combined the classical principal component analysis (PCA) with T^2 and Q statistics to capture the occurrence of an event and to separate it from normal operating conditions. Subsequently in [11] this idea was extended to the geometric interpretation of T^2 and Q showing that they can be separately used to detect generation mismatch and islanding events respectively. Although the issue of a single PMU malfunctioning with a fixed bias has been discussed in [10], it does not comment on the performance of the detection algorithm with larger outliers comparable to magnitude of faults or under fault replay attacks in one or more data streams. Another limitation of PCA-based methods is in the assumption of linear relationship between the measured variables. With the nonstationary nature of system dynamics due to change in load, generation and operating conditions, and inherent nonlinearity associated with an event such assumptions may lead to inaccurate results.

Improvements on these have been suggested in [12] and [13] using a kernel PCA (K-PCA) based nonlinear technique. Instead of computing the principal components of the input data, this method maps the data to a higher dimensional space where the assumptions of linearity hold and then use the 'kernel trick' to solve a linear PCA problem. However, a few important questions still remain: 1) Since K-PCA is computationally intensive over linear PCA, can the window size of computation be reduced for real-time monitoring without compromising on the detector accuracy? 2) How sensitive are T^2 and Q to data anomalies and noise in a reduced window? and finally 3) Can better indices be developed with higher selectivity to events and lower sensitivity to data anomalies while working on sub-second windows?

To address these questions, this paper presents a K-PCA-based detection index derived from the norm of the principal component (PC) score of a moving window of data mapped in feature space. It is demonstrataed that with an appropriate choice of a kernel function the change in the largest eigenvalue of the kernel matrix can be used as an detector for occurrence of events. Unlike the T^2 and Q statistics presented in literature, the proposed metric works fine with small windows and is insensitive to noise, corruption and spurious outliers. The other contribution of the paper is in the derivation of an upper bound on the change in the largest eigenvalue of kernel between two successive windows in terms of the data samples without explicitly performing an eigen decomposition. Further it is shown that the bound can itself act as a detector, thereby significantly reducing the computation in real time.

The remaining paper is organized as follows. In Section II, a brief theoretical overview of kernel PCA is presented, an outline of the detection logic is discussed and an upper bound on the change in kernel eigenvalues is derived. In Section III, results from case studies on both simulated data and actual field measurements are presented to validate the claims of the event detection. Also, comparisons are drawn with existing T^2 and Q statistics based methods. Concluding remarks and discussions are summarized in Section V.

II. Kernel PCA for Detecting Event Outliers

In multivariate statistics, PCA [14] is a powerful tool for linear dimensionality reduction and feature extraction. However, as discussed before, the nonlinear and nonstationary nature of power system response impedes the success of PCA in event detection. Kernel principal component analysis (K-PCA) [15] proposed by Scholkoph et. al in 1998, is a generalization of PCA to nonlinear dimensionality reduction which has found a wide range of applications in process control, spectrum sensing, image processing, and other fields.

Kernel PCA constitute a two step process: 1) Mapping the input data to a higher dimensional feature space where the relation between the mapped variables is linear, and 2) Using PCA in feature space for dimensionality reduction. Being a kernel method, K-PCA does not explicitly compute the mapping to the higher dimensional feature space, but uses a function of the measured data to encode the mapping information. The choice of this function (called kernel) thus decides the efficacy of the mapping and the analysis that follows in the mapped domain.

Let $\mathbf{x}_i$ be a vector of detrended measurements from m PMU channels obtained at instant i, and $\phi(\mathbf{x}_i)$ be the mapping of $\mathbf{x}_i$ in feature space, $\phi : \mathbf{X} \rightarrow \Phi$. Without loss of generality let us assume that the data in feature space is centered.

We consider a window of size N sliding in time, with the vector of latest observations augmented to the last column and the oldest observation discarded from the first.

Eigen decomposition of the covariance matrix $\mathbf{C}$ in feature space can then be expressed as,

$$\mathbf{C}\mathbf{v}_j = \lambda_j \mathbf{v}_j \implies \frac{1}{N}\sum_{i=1}^{N} \phi(\mathbf{x_i})\phi^T(\mathbf{x_i})\mathbf{v}_j = \lambda_j \mathbf{v}_j \quad (1)$$

$$\implies \mathbf{v}_j = \sum_{i=1}^{N} \phi(\mathbf{x_i})\{\frac{\phi^T(\mathbf{x_i})\mathbf{v}_j}{N\lambda_j}\} = \sum_{i=1}^{N} \alpha_{j_i}\phi(\mathbf{x_i}) \quad (2)$$

Substituting (2) in (1) and pre-multiplying by $\phi^T(\mathbf{x_l})$,

$$\begin{aligned}\sum_{i=1}^{N} \phi^T(\mathbf{x_l})\phi(\mathbf{x_i}) \sum_{n=1}^{N} \alpha_{j_n}\phi^T(\mathbf{x_i})\phi(\mathbf{x_n}) \\ = N\lambda_j \sum_{n=1}^{N} \alpha_{j_n}\phi^T(\mathbf{x_l})\phi(\mathbf{x_n})\end{aligned} \quad (3)$$

Let $\mathbf{\Phi} = \left[\begin{array}{ccc} \phi(\mathbf{x_1}) & \dots & \phi(\mathbf{x_N}) \end{array} \right]$ and let $\mathbf{K}$ be the matrix of inner products in feature space, $\mathbf{K} = \mathbf{\Phi}^T\mathbf{\Phi}$.

$$\sum_{i=1}^{N}\sum_{n=1}^{N} \mathbf{K}_{li}\mathbf{K}_{in}\alpha_{j_n} = N\lambda_j \sum_{n=1}^{N} \mathbf{K}_{ln}\alpha_{j_n} \quad (4)$$

Concatenating the expression in (4) for $l = 1, 2, \dots N$,

$$\mathbf{K}^2\boldsymbol{\alpha}_j = N\lambda_j\mathbf{K}\boldsymbol{\alpha}_j \implies \mathbf{K}\boldsymbol{\alpha}_j = N\lambda_j\boldsymbol{\alpha}_j \quad (5)$$

This is the expression for eigen decomposition of the kernel matrix with eigen pair $(N\lambda_j, \boldsymbol{\alpha}_j)$. To ensure that the eigen vectors of the feature space covariance matrix are orthonormal,

$$\mathbf{v}_j^T\mathbf{v}_j = (\mathbf{\Phi}\boldsymbol{\alpha}_j)^T(\mathbf{\Phi}\boldsymbol{\alpha}_j) = 1 \implies \left\|\boldsymbol{\alpha}_j\right\|_2 = \frac{1}{\sqrt{N\lambda_j}} \quad (6)$$

The eigen vectors of the covariance matrix are the principal directions in the feature space.

If the mapped data in feature space is not centered, one can alternatively center the Kernel matrix as follows.

$$\widetilde{\mathbf{K}} = [\mathbf{\Phi} - \{\frac{1}{N}\sum_{i=1}^{N}\phi(\mathbf{x_i})\}\mathbf{J_N}^T]^T \;\; [\mathbf{\Phi} - \{\frac{1}{N}\sum_{i=1}^{N}\phi(\mathbf{x_i})\}\mathbf{J_N}^T]$$
$$= \mathbf{K} - \frac{1}{N}\mathbf{K}\mathbf{J_{N\times N}} - \frac{1}{N}\mathbf{J_{N\times N}}\mathbf{K} + \frac{1}{N^2}\mathbf{J_{N\times N}}\mathbf{K}\mathbf{J_{N\times N}} \tag{7}$$

where, $\mathbf{J_N} \in \mathbb{R}^N$ and $\mathbf{J_{N\times N}} \in \mathbb{R}^{N\times N}$ with all entries as 1.

If the construction of the mapping ϕ is such that the feature space is isometric to the input space, the inner product in the feature space can be expressed as a positive semi-definite symmetric function of pre-images in input space. This is commonly referred to as the '*kernel trick*' in literature and the positive semi-definite function $\mathcal{K}$, is called a *kernel*. Some commonly used kernels include- Gaussian kernel, linear kernel, polynomial kernel, etc. In this paper, we will use the polynomial kernel in (8) for detecting event outliers.

$$\mathbf{K}_{ij} = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) = \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T\mathbf{x}_j)^d \tag{8}$$

The idea behind the detection is to compare between successive time windows, the projection of the mapped data along the direction of maximum variance in feature space. The magnitude of this quantity is captured in the norm of the PC score along the first principal direction. The PC score norm (ζ_1) of a window of mapped data along the direction of maximum variance $\mathbf{v}_1$ can be computed as,

$$\zeta_1 = \left\|\mathbf{v}_1^T\mathbf{\Phi}\right\|_2 = \left\|\boldsymbol{\alpha}_1^T\mathbf{K}\right\|_2 = N\lambda_1\left\|\boldsymbol{\alpha}_1^T\right\|_2 = \sqrt{N\lambda_1} \tag{9}$$

■ *Remark I*: The norm of the projection, as in (9), is also the square root of the largest eigenvalue of the kernel matrix obtained from the measurement window at that instant. With a suitable choice of kernel function and proper tuning of hyper-parameters it can be ensured that eigenvalue of kernel for a window with incoming event outlier is significantly higher compared to the windows containing ambient and anomalous data. This will thus be used as the metric for event detection.

Let the kernel matrix for the i^{th} window of data be $\mathbf{K}^{(i)}$,

$$\left[\begin{array}{c|cccc} (\mathbf{x}_i^T\mathbf{x}_i)^d & (\mathbf{x}_i^T\mathbf{x}_{i+1})^d & \dots & (\mathbf{x}_i^T\mathbf{x}_{i+N-1})^d \\ \hline (\mathbf{x}_{i+1}^T\mathbf{x}_i)^d & (\mathbf{x}_{i+1}^T\mathbf{x}_{i+1})^d & \dots & (\mathbf{x}_{i+1}^T\mathbf{x}_{i+N-1})^d \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{x}_{i+N-1}^T\mathbf{x}_i)^d & (\mathbf{x}_{i+N-1}^T\mathbf{x}_{i+1})^d & \dots & (\mathbf{x}_{i+N-1}^T\mathbf{x}_{i+N-1})^d \end{array}\right]$$

Let, $\mathbf{K}^{(i)} = \left[\begin{array}{c|c} p & \mathbf{c}^T \\ \hline \mathbf{c} & \mathbf{A} \end{array}\right]$ and $\mathbf{P_r}\mathbf{K}^{(i)}\mathbf{P_c} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{c} \\ \hline \mathbf{c}^T & p \end{array}\right] = \mathbf{B}$

where, $\mathbf{P_r} = \left[\begin{array}{c|c} \mathbf{0}_{(N-1)\times 1} & \mathbf{I}_{(N-1)\times(N-1)} \\ \hline 1 & \mathbf{0}_{1\times(N-1)} \end{array}\right]$ and $\mathbf{P_c} = \mathbf{P_r}^{-1}$

From (5) the eigen decomposition of $\mathbf{K^{(i)}}$ can be written as,

$$\begin{aligned} \mathbf{K}^{(i)}\boldsymbol{\alpha}_j &= N\lambda_j^{(i)}\boldsymbol{\alpha}_j^{(i)} \\ \implies \mathbf{P_r}\mathbf{K}^{(i)}\mathbf{P_c}\hat{\boldsymbol{\alpha}}_j^{(i)} &= N\lambda_j^{(i)}\mathbf{P_r}\mathbf{P_c}\hat{\boldsymbol{\alpha}}_j^{(i)} = N\lambda_j^{(i)}\hat{\boldsymbol{\alpha}}_j^{(i)} \end{aligned} \tag{10}$$

Thus, the matrices $\mathbf{K}^{(i)}$ and $\mathbf{P_r}\mathbf{K}^{(i)}\mathbf{P_c}$ have same eigenvalues. The kernel matrix for the $(i+1)^{th}$ window of data be $\mathbf{K}^{(i+1)}$,

$$\left[\begin{array}{ccc|c} (\mathbf{x}_{i+1}^T\mathbf{x}_{(i+1)})^d & \dots & \mathbf{x}_{i+1}^T\mathbf{x}_{i+N-1})^d & (\mathbf{x}_{i+1}^T\mathbf{x}_{i+N})^d \\ \vdots & \ddots & \vdots & \vdots \\ (\mathbf{x}_{i+N-1}^T\mathbf{x}_{i+1})^d & \dots & (\mathbf{x}_{i+N-1}^T\mathbf{x}_{i+N-1})^d & (\mathbf{x}_{i+N-1}^T\mathbf{x}_{i+N})^d \\ \hline (\mathbf{x}_{i+N}^T\mathbf{x}_{i+1})^d & \dots & (\mathbf{x}_{i+N}^T\mathbf{x}_{i+N-1})^d & (\mathbf{x}_{i+N}^T\mathbf{x}_{i+N})^d \end{array}\right]$$

Let, $\mathbf{K}^{(i+1)} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{b}^T & q \end{array}\right] = \mathbf{B} + \mathcal{E}$, where $\mathcal{E}$ is the perturbation in the kernel matrix with respect to the previous window.

$$\mathcal{E} = \left[\begin{array}{c|c} \mathbf{0}_{(N-1)\times(N-1)} & \mathbf{b}-\mathbf{c} \\ \hline \mathbf{b}-\mathbf{c}^T & q-p \end{array}\right] \tag{11}$$

Using the classical perturbation bound [16], between two successive windows, the difference in the squared PC score norms along $\mathbf{v}_1$ can be expressed as,

$$\begin{aligned} \Delta(\zeta_1)^2 &= |(\zeta_1^{(i+1)})^2 - (\zeta_1^{(i)})^2| = |N\lambda_1^{(i+1)} - N\lambda_1^{(i)}| \\ &= |\lambda_1(\mathbf{K}^{(i+1)}) - \lambda_1(\mathbf{K}^{(i)})| = |\lambda_1(\mathbf{B}+\mathcal{E}) - \lambda_1(\mathbf{B})| \\ &\leq \|\mathcal{E}\|_2 = \lambda_1(\mathcal{E}) \end{aligned} \tag{12}$$

where, $\lambda_1(\mathcal{E})$ corresponds to the largest eigenvalue of $\mathcal{E}$.

$$\lambda_1(\mathcal{E}) = \frac{(q-p) + \sqrt{((q-p)^2 + 4(\mathbf{b}-\mathbf{c})^T(\mathbf{b}-\mathbf{c})}}{2} \tag{13}$$

■ *Remark II*: Between two successive windows i and $i+1$, the term $q-p$ equals to $\|\mathbf{x}_{i+N}\|^{2d} - \|\mathbf{x}_i\|^{2d}$ and each entry in the vector $\mathbf{b}-\mathbf{c}$ equals to $(\mathbf{x}_{j+1}^T\mathbf{x}_{i+N})^d - (\mathbf{x}_{j+1}^T\mathbf{x}_i)^d$, for $j = i, (i+1), \dots (i+N-1)$. If both the windows are from ambient condition, then the difference between the outgoing data vector $\mathbf{x}_i$ and the incoming data point $\mathbf{x}_{i+N}$ is small, only source of difference being the ambient noise and minor variations about equilibrium. However, if $\mathbf{x}_{i+N}$ corresponds to the onset of an event, the difference would be significantly larger compared to the previous case. Further, with an appropriate choice of d (> 1) the difference can be amplified. Thus, both $q-p$ and $\mathbf{b}-\mathbf{c}$ would be significantly larger contributing to a large $\lambda_1(\mathcal{E})$ for an ambient-to-event transition compared to ambient-to-ambient transition of data windows.

If $\mathbf{x}_{i+N}$ corresponds to a bad data or malicious corruption, depending on the number of PMU channels corrupted the difference with $\mathbf{x}_i$ would vary. In most cases, only a small fraction of PMUs can be assumed to be affected by bad data or malicious corruption. Thus, the difference would always be less than that compared to an event. With high values of the parameter d, the event condition would be amplified more compared to an anomaly condition, and clear separation between them can be obtained. Thus, the upper bound on the deviation in PC score norm, as expressed in (13) can be used as a metric for detecting the onset of an event. A major advantage of using this is that the closed form expression in (13) can be computed using inner products and completely eliminates the need for an eigen decomposition of the kernel matrix, as required in the PC score based detector in (9).

■ *Remark III*: Event-induced outliers are important in the detection process described above. Onset of events that do not produce an outlier signature may not be efficiently detected. The determination of an appropriate kernel function is another aspect of research.

III. Case Studies

A. Event-Outlier Detection from Simulated Data

The 16-machine 5-area New England - New York system [17], with PMUs at all inter tie buses was used to generate the synthetic data for simulations. Voltage magnitudes of buses-18, 27, 41, 42, 49, 53, 54, 60, and 61 were considered for analysis after appropriate detrending of the data. The reporting rate of the PMUs was assumed to be 50Hz and a moving window of 25 samples, equivalently of 0.5s duration was studied for detecting the onset of an event.

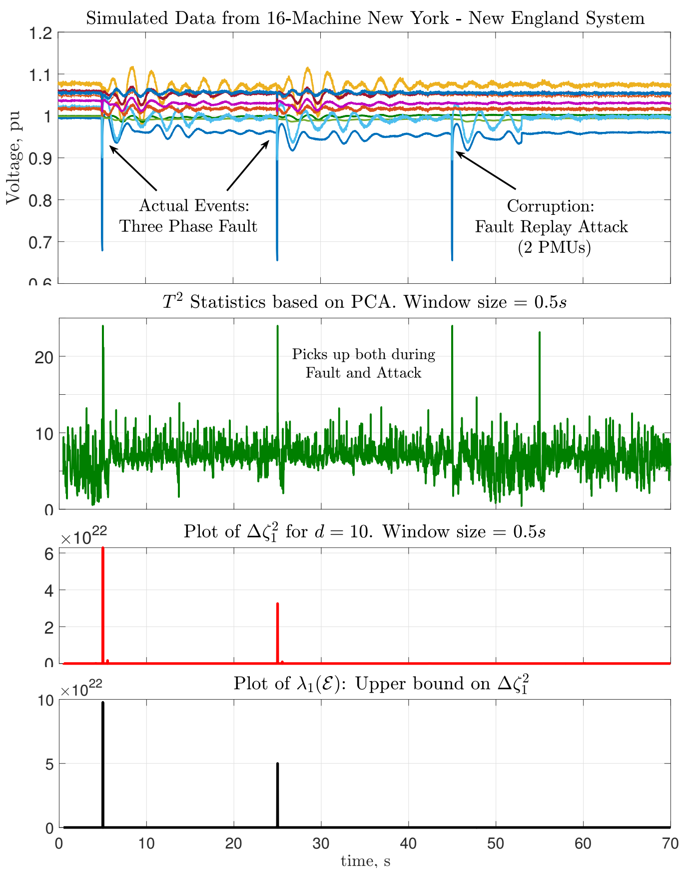

Fig. 1: Fault outlier detection in 16-machine NE-NY system.

The performance of the detection principle was studied for multiple event scenarios. One such case is reported in Fig. 1 with two instances of three phase fault and a fault replay attack on the voltage measurements of buses 18 and 42. The fault at t = 5s at bus 18 is cleared by opening the line 18-42. This is followed by a self-clearing fault at t = 25s near bus 18. The recorded fault data from t = 25s is replayed at t = 45s at the mentioned buses.

It can be seen from the figure that a T^2 based event detection metric fails to differentiate between events (fault a 5s and 25s) and attack (at 45s). However, when $\Delta\zeta_1^2$ is used as a metric for detection, it clearly identifies the event and does not have a false triggering during the attack. Also, the relative magnitude of the index in a window capturing the onset of an event is significantly higher compared to ambient and attack cases. Thus, any clustering algorithm using $\Delta\zeta_1^2$ as a feature would identify these instants as onset of events. The plot for the bound on the change in $\Delta\zeta_1^2$ is also reported in Fig. 1. It validates the claim that $\lambda_1(\mathcal{E})$ can also be used as a metric.

B. Event-Outlier Detection from Field Measurements

Synchronized frequency measurements [18] from four locations in the Indian grid were obtained for analysis. Multiple disturbances in the grid have been captured in the frequency data from these locations. Two of these cases have been reported in this paper. The reporting rate of the PMUs were 50 Hz, and the analysis was carried on a moving window of 25 samples.

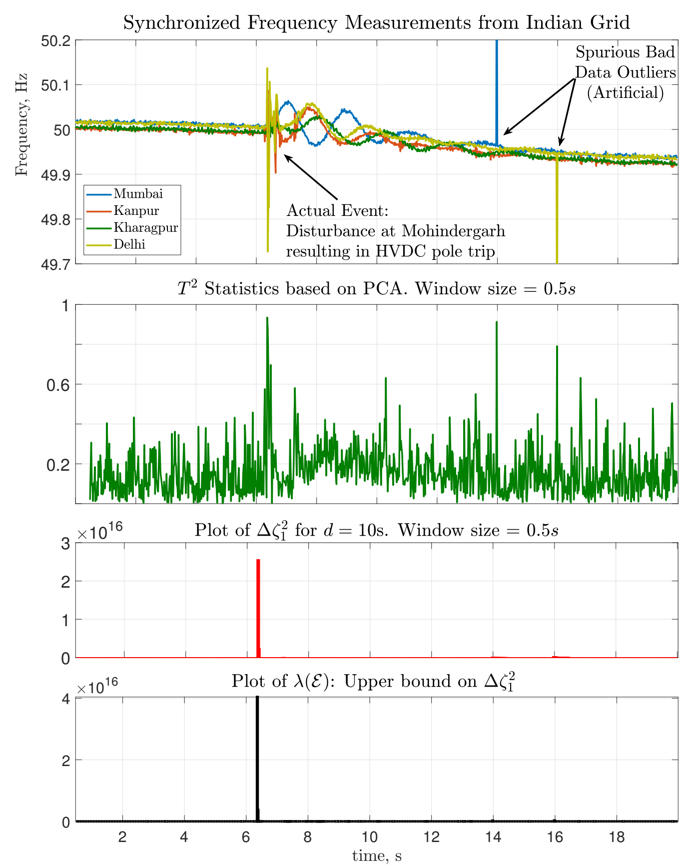

Fig. 2: Detection of Mundra-Mohindergarh HVDC pole trip.

■ *Mundra−Mohindergarh HVDC Pole Trip*: Frequency data from June 2013 capturing the disturbance following the tripping of Pole-1 of the ± 500 kV Mundra-Mohindergarh bipolar HVDC link is considered. In addition, two spurious bad data outliers in magnitude comparable to the frequency dip/rise have been injected artificially into the data. The frequency variation corresponding to the event along with the performance of the detector is shown in Fig. 2. It can be seen that the usual T^2 metric is equally selective for the disturbance and the outliers, also it is sensitive to measurement noise. Thus, any detection approach using T^2 as a feature is prone to false positives. In contrast, the robustness of the $\Delta\zeta_1^2$ metric is illustrated in Fig. 2.

■ *Multiple Trippings at Samaypur substation*: Frequency data is from April 2013 capturing the grid disturbance following a series of failures (including a current transformer and three phase circuit breakers) at Samaypur substation has been considered. This ultimately lead to tripping of multiple lines. In addition, for validating the performance of the event detection scheme, we have artificially injected a disturbance in one of the data channels, as shown in Fig. 3. The plot of $\Delta\zeta_1^2$ clearly shows that the index is robust to bad data and is more selective towards detecting events as compared to the T^2 metric.

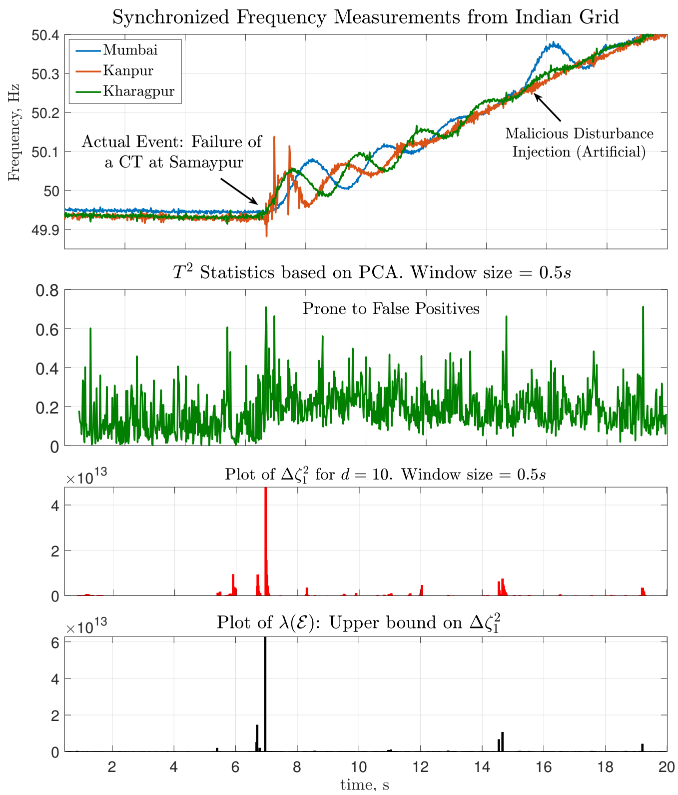

Fig. 3: Detection of mutiple trippings at Samaypur Substation.

IV. Conclusion and Discussions

A corruption-resilient approach for detection of event-induced outliers was presented in the paper. It was shown that the change in the square of the norm of PC score along the first principal direction in feature space between two successive windows is insensitive to bad data and can be used as a metric for detecting event-induced outliers. It was shown that the PC score-based metric can work with very small window of data. However, this approach may not be successful in detecting events that do not produce outliers at its onset. Moreover, the choice of the kernel function and the kernel parameters play a crucial role in deciding the success of the detection scheme. Our ongoing research is focused on constructing a data-driven kernel matrix without specifying its structure a-priori.

Acknowledgements

The financial support from NSF Award CNS1739206 is acknowledged.

References

[1] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. Springer US, 2008.

[2] V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, J. Fitch, S. Skok, M. M. Begovic, and A. Phadke, "Wide-Area Monitoring, Protection, and Control of Future Electric Power Networks," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 80–93, Jan. 2011.

[3] S. Brahma, R. Kavasseri, H. Cao, N. R. Chaudhuri, T. Alexopoulos, and Y. Cui, "Real-Time Identification of Dynamic Events in Power Systems Using PMU Data, and Potential Applications— Models, Promises, and Challenges," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 294–301, Feb. 2017.

[4] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-Physical System Security for the Electric Power Grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, Jan. 2012.

[5] Analysis of the Cyber Attack on the Ukrainian Power Grid. (), [Online]. Available: https : / / www . nerc . com / pa / CI / ESISAC / Documents/E-ISAC_SANS_Ukraine_DUC_18Mar2016.pdf.

[6] K. Mahapatra, N. R. Chaudhuri, and R. Kavasseri, "Bad data detection in PMU measurements using principal component analysis," in *2016 North American Power Symposium (NAPS)*, Sep. 2016, pp. 1–6.

[7] K. Mahapatra and N. R. Chaudhuri, "Malicious Corruption-Resilient Wide-Area Oscillation Monitoring using Principal Component Pursuit," *IEEE Transactions on Smart Grid*, pp. 1–1, 2017.

[8] P. Gao, M. Wang, J. H. Chow, M. Berger, and L. M. Seversky, "Missing Data Recovery for High-Dimensional Signals With Nonlinear Low-Dimensional Structures," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5421–5436, Oct. 2017.

[9] P. Gao, R. Wang, M. Wang, and J. H. Chow, "Low-Rank Matrix Recovery From Noisy, Quantized, and Erroneous Measurements," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2918–2932, Jun. 2018.

[10] E. Barocio, B. C. Pal, D. Fabozzi, and N. F. Thornhill, "Detection and visualization of power system disturbances using principal component analysis," in *2013 IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid*, Aug. 2013, pp. 1–10.

[11] X. Liu, D. M. Laverty, R. J. Best, K. Li, D. J. Morrow, and S. McLoone, "Principal Component Analysis of Wide-Area Phasor Measurements for Islanding Detection- A Geometric View," *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 976–985, Apr. 2015.

[12] X. Liu, J. M. Kennedy, D. M. Laverty, D. J. Morrow, and S. McLoone, "Wide-Area Phase-Angle Measurements for Islanding Detection- An Adaptive Nonlinear Approach," *IEEE Transactions on Power Delivery*, vol. 31, no. 4, pp. 1901–1911, Aug. 2016.

[13] G. Liu, H. Chen, X. Sun, N. Quan, L. Wan, and R. Chen, "Low-Complexity Nonlinear Analysis of Synchrophasor Measurements for Events Detection and Localization," *IEEE Access*, vol. 6, pp. 4982–4993, 2018.

[14] I. T. Jolliffe, *Principal Component Analysis*. Springer, 1986.

[15] B. Scholkopf, A. Smola, and K. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 51, pp. 1299–1319, 1998.

[16] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The John Hopkins University Press, Baltimore, MD, 1989.

[17] B. C. Pal and B. Chaudhuri, *Robust Control in Power Systems*. Springer, New York, 2005.

[18] K. A. Salunkhe, G. Gajjar, S. A. Soman, and A. M. Kulkarni, "Implementation and applications of a Wide Area Frequency Measurement System synchronized using Network Time Protocol," in *2014 IEEE PES General Meeting — Conference Exposition*, Jul. 2014, pp. 1–5.