

# Signal Features for Classification of Power System Disturbances using PMU Data

Milan Biswal\*, Yifan Hao<sup>†</sup>, Phillip Chen\*, Sukumar Brahma\*, Huiping Cao<sup>†</sup> and Phillip De Leon\*

\* Klipsch School of Electrical and Computer Engineering  
New Mexico State University, Las Cruces, NM, 88003, USA.

<sup>†</sup> Computer Science Department  
New Mexico State University, Las Cruces, NM, 88003, USA.

**Abstract**—Event identification is one among numerous applications being researched for PMU data. This application is intended to increase visualization of power system events, as well as for protection and control, including verification of relay operation to detect any misoperations. This paper uses data from field as well as from simulation to test a large variety of features using two well-known classifiers on a common dataset to find the most suitable features for disturbance data recorded by PMUs. The approach also uses data from only one PMU instead of data from multiple PMUs used by researchers so far, thus significantly reducing the data to be processed. It is shown that simple observation-based features capturing shape and statistics of disturbance waveforms work better than some well-known features derived from domain transformations. Classification accuracy and speed achieved with these features are shown to be satisfactory and suitable for the intended applications.

**Index Terms**—Classification, feature extraction, phasor measurement unit, power system disturbance

## I. INTRODUCTION

As Phase Measurement Units (PMUs) proliferate throughout power systems, researchers have begun to investigate automatic methods for the detection and identification of disturbance events using PMU data in an effort to increase visualization of power systems [1]–[3]. It is expected that near real-time detection and identification of disturbances will be used for power system protection and control [4]. The current state of the art of protection and control applications can benefit from detection of relay misoperations, a problem that has remained elusive to engineers.

*Misoperation* is a term that indicates that a relay has operated when there was no fault in its operating zone. Certain misoperations have historically occurred due to defective relay components which have also been called *hidden faults* [5]. In addition there are many instances where a misoperation due to a hidden fault either initiated or contributed to system instability, resulting in a blackout [5], [6]. Another type of misoperation that routinely occurs during early stages of a blackout is *load encroachment*, where a distance relay misinterprets overload as a system fault. With modern relays the

hidden fault may also manifest as a cyber attack, where the intruder can misoperate a relay at a time calculated to result in instability. In light of these issues, event detection and identification using PMU data can play an important role in verifying a relay operation. If a relay operated, but no fault was detected by the event identification tool, this operation can be flagged as misoperation.

Previous research work elaborates this idea to outline the concept of supervisory protection, where event identification using PMU data can form a *supervisory layer* that verifies every relay operation, and informs the operators if a misoperation is detected [6]. Depending on the state of the system (moderately or heavily loaded), a judicious response can be initiated, either manual or automated. This application, however, requires speed - the faster the detection, the better the chances of successful countermeasures. All papers describing visualization of system events using PMU data, extract patterns from disturbance files generated by multiple PMUs and use a classifier to map the files to a disturbance event [1]–[3], [7]. However, this requires waiting for the disturbance files to be generated, which are typically 3 min long [8], before even starting the process. In addition, the most commonly used feature extraction technique has been Minimum Volume Enclosing Ellipsoid (MVEE), which relies on the measurements from *all* PMUs reporting to a Phasor Data Concentrator (PDC) to formulate an ellipsoid [2]. Properties of the ellipsoid such as volume, major and minor axes, rate of change of volume are taken as the features of the data. Classification of the disturbance event is then based on feature vectors built from these features. Although effective, this method is computationally intensive, and ignores the *spatial correlation* between data recorded by different PMUs for a particular disturbance. It is intuitive that the *nature* of a disturbance pattern in different PMU measurements for the same disturbance would be the same, although the strength of the pattern may vary depending on the vicinity of a particular PMU from the location of the disturbance.

This discussion brings out the need for the following improvements in the application of event identification using PMU data:

- 1) The need to accelerate the process - instead of waiting

---

This work was supported by the National Science Foundation, CREST Grant HRD-1345232.

for PMUs to generate disturbance files, place the event detection mechanism inside PDC, and start the classification process right after an event is detected by PDC.

- 2) The need to reduce dimensionality of data - implement a method within PDC that can identify the strongest disturbance signal.
- 3) The need to develop signal features and classification methods that quickly and accurately identify the disturbance from a single PMU-data and requires only a small length of disturbance data.

Obviously accuracy and speed are the two most important metrics to evaluate the classifier for this application.

In this paper we describe a method to identify the strongest signal, called the Strongest Signal Selection (SSS). Then, we investigate signal features with complexity lower than that of MVEE in order to identify the type of disturbance. The features under consideration are based on

- classical methods such as the Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), and Principal Component Analysis (PCA),
- time-frequency transforms such as the S-Transform (ST), where the mother wavelet is replaced by a Gaussian window and a phase term, as well as a new Hilbert Spectrum which decomposes signals into AM-FM components,
- shapelets that capture the most significant part of the raw waveform,
- deterministic and statistical features in the data such as slope, variance, etc.

We consider the  $K$  Nearest Neighbor (KNN) and Support Vector Machine (SVM) classifiers to classify these features because of the former's simplicity and low test-stage computational cost [9] and the latter's superior performance in our previous study [6].

In this paper, we focus on fault and generation loss events—the two most significant disturbance events. The bulk of the PMU data required in this study is generated by dynamic simulation of the Western Area Coordinating Council (WECC) system in the USA using General Electric's Positive Sequence Load Flow (PSLF) software. We also have actual field data. Since field data are limited, we use simulation data for training the classifiers and then both simulation and actual PMU data (separately) for evaluation.

## II. IDENTIFICATION OF STRONGEST SIGNAL

The input data arrives from multiple PMUs to the PDC in real time. PMUs have built in thresholds for voltage and frequency to detect disturbances [8], which can also be easily implemented in PDC to detect a disturbance-trigger. These PMU data are then fed to the SSS module. This module first preprocesses the signals to remove bad data [8], and uses the processed signal to identify the disturbance data stream with the highest strength. Although PMUs record multiple variables such as voltage phasors, current phasors, frequency and rate of change of frequency, we have shown in [8] that the voltage magnitude and frequency data carry sufficient information for

event recognition. We use the voltage magnitude alone to find out the strongest signal and the associated PMU.

Let  $V_i(n)$  and  $f_i(n)$  denote the measured voltage magnitude and frequency signals recorded by the  $i^{\text{th}}$  PMU at the  $n^{\text{th}}$  sample index. Also suppose that an event is triggered at  $n_0$ .

The positive and negative deviations of the voltage signal,  $n$  samples after triggering, for the  $i^{\text{th}}$  PMU, are calculated according to (1) and (2) respectively, which define the positive,  $\Delta_i^+(n)$  and negative,  $\Delta_i^-(n)$  deviations of the  $n^{\text{th}}$  sample after the triggering point from the base voltage  $V_i^{\text{rms}}$  of the  $i^{\text{th}}$  PMU:

$$\Delta_i^+(n) = \begin{cases} V_i(n_0 + n) - V_i^{\text{rms}}, & \text{if } V_i(n_0 + n) \geq V_i^{\text{rms}} \\ 0, & \text{else} \end{cases} \quad (1)$$

$$\Delta_i^-(n) = \begin{cases} V_i^{\text{rms}} - V_i(n_0 + n), & \text{if } V_i(n_0 + n) < V_i^{\text{rms}} \\ 0, & \text{else} \end{cases} \quad (2)$$

$V_i^{\text{rms}}$  is calculated as the averaged rms value of the measured voltage magnitude by the  $i^{\text{th}}$  PMU, over the ten samples prior to  $n_0$ . Ten cycles are generally used in measurement aggregation for power signal analysis [10].

Thus,  $V_i^{\text{rms}}$  is defined as

$$V_i^{\text{rms}} = \frac{1}{10} \sqrt{\sum_{n=n_0-10}^{n_0-1} V_i(n)^2}. \quad (3)$$

A new metric based on energy deviation, referred to as Cumulative deviation in Energy (CE), is defined for quantifying the strength of the signal recorded by a PMU. The CE for the  $i^{\text{th}}$  PMU,  $m$  samples after triggering is expressed as:

$$\gamma_i(m) = \sum_{k=0}^{m-1} [\Delta_i^+(n+k)]^2 + \sum_{k=0}^{m-1} [\Delta_i^-(n+k)]^2 \quad (4)$$

From experimental verification, we observed that a value of  $m=15$  was found to be sufficient to detect the strongest signal. Even at 30 frames per second (fps) used by older PMUs, this amounts to 0.5 s.

## III. FEATURE EXTRACTION TECHNIQUES

Though many feature extraction methods exist in literature, there is no qualitative approach to determine which method will work the best for a given dataset. In this paper, we applied a number of existing classical feature extraction techniques. We also designed several new features based on the waveform statistics and the waveform shapes. The features were extracted from voltage waveforms and frequency waveforms. Before extracting the features, we clipped the waveforms to only include samples 0.5 s before the triggering point and 1.5 s after (and including) the triggering point. This is because most critical changes in disturbance waveforms are captured by this window. This can be verified by the sample disturbance waveforms shown later in Section IV in Fig. 1. We describe the feature extraction methods in the following subsections.

## A. Classical Methods

1) *Discrete Fourier Transform*: The well-known DFT [11] converts time-domain data to the frequency-domain through an expansion into a basis of complex exponentials

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \quad (5)$$

where  $X(k)$  are the DFT coefficients.

Signal features can be extracted directly from the DFT coefficients or their magnitudes. The major limitation of DFT is its inability for multi-resolution analysis, and thus it cannot capture temporal variation of a time series.

2) *Discrete Wavelet Transform*: The DWT decomposes time-domain data into multiple levels of resolution through the use of basis functions derived from a mother wavelet [12]. The resulting DWT coefficients can also serve as signal features. The choice of the mother wavelet is a critical factor in deciding the performance of DWT and is often chosen with a goal of obtaining a sparse representation, i.e. the least number of non-zero DWT coefficients. The data can be visualized from a different resolution at each level of the decomposition, thus capturing temporal variations.

3) *Fast variant of discrete S transform features*: Fast variant of discrete S transform (FDST) is a linear time-frequency transform which is an extension from the wavelet class of algorithms, with specific merits such as fast computation, absolute phase reference and Fourier basis modeling [13]. The FDST is capable of segregating a disturbance waveform in terms of oscillatory modes which correspond to a range of Fourier frequencies. The frequency domain expression for FDST of a time-series  $V_i$  of length  $m$  samples is expressed as:

$$S_{i,k} = \begin{cases} \frac{1}{m} \sum_{k'=0}^{m-1} \bar{V}_{k'} \bar{W}_{(k'-i),k} e^{j2\pi k k' / m} & , \text{if } k' = k \\ & \text{and } \bar{V}_{k'} \geq \alpha \\ 0 & \text{Otherwise.} \end{cases} \quad (6)$$

The meaning of each component in this equation is explained as follows. First,  $i = 0, 1, \dots, (m-1)$  denote the index of the original sample And  $k = 0, 1, \dots, m/2$  denote the index of the sample in the frequency domain. Second,  $\bar{V}_{k'}$  is the DFT of  $V_i$  Third,  $\bar{W}_{k',k}$  is a discretized frequency domain normalized Gaussian window function which is expressed as:

$$\bar{W}_{k',k} = e^{-\frac{1}{2} \left( \frac{2\pi k'}{k} \right)^2}, \quad k' = 0, 1, \dots, (m-1) \quad (7)$$

The parameter  $\alpha$  is a threshold, whose value is zero in our analysis to preserve all the frequency components.

Six features that gave superior performance were selected through trials. The features quantified in various ways the time-frequency resolution (TFR) energy and dominant oscillatory modes. Since results documented in Section IV show that this approach did not come up among the top performing methods, we do not describe these features in detail here due to space constraints.

4) *Principal Component Analysis*: PCA transforms the coordinates of data in  $\mathbb{R}^D$  such that the data are expressed in terms of  $D$  principal axes, arranged in order of decreasing variance [14]. PCA has the ability to reveal hidden dynamics underlying a complex dataset by expressing the data as a linear combination of new basis vectors or Principal Components (PCs). The first PC of a given set of correlated variables encodes the maximum variance of the data, and is a linear combination of the variables having the maximum variability among all linear combinations. The second PC defines the next largest amount of variation not accommodated by the first PC and is orthogonal to the first PC and so on for subsequent PCs. PCA can also be used to project the data onto a lower dimensional subspace by selecting the  $M < D$  PCs associated with the maximum variance of the data [14].

## B. Methods Designed from Waveform Observation

We also decided to create features based on observation of data. Fig. 1 presented and explained later in Section IV-A shows that the different disturbance waveforms are characterized by changes that are fairly distinctive to the naked eye. Post-disturbance values can change quickly, or slowly, increase, or decrease, deviate more or less from pre-disturbance values, and continue to go through such changes for different durations. Motivated by this observation, we designed features described in the following subsections.

1) *Slope Sequence*: Based on the slope, which would capture both the direction and steepness of a line, we propose a new feature, Slope Sequence ( $S^2$ ), that can capture the trend of *changes* in the data. From the entire data sequence, we capture a sequence of slopes, i.e.  $S^2$ , described now.

For a data point  $x(n)$ , its  $K$ -step slope value is calculated as

$$\lambda(n) = \frac{x(n+K) - x(n-K)}{2K}. \quad (8)$$

For  $K = 1$ , given a waveform with  $N$  sampled points,  $S^2$  is a sequence  $\lambda(2), \lambda(3), \dots, \lambda(N-1)$ .

2) *Domain specific shapelets*: In the past several years, *shapelets* have been introduced as a new feature for time series data [15]. Shapelets are time series “snippets” (or sub-sequences). One major advantage of using shapelets as time series features is its easy interpretability because it is part of the original time series. Ever since the initial introduction of shapelets by Ye and Keogh [15], shapelets discovery has generated significant interest from independent research groups (e.g., [15]–[19]). Since brute-force algorithm to compute shapelets is not feasible because of the exponential number of shapelet candidates, more efficient algorithms [17], [20] have been proposed.

Motivated by the recent advances in shapelets discovery, we extracted shapelets features by leveraging domain knowledge from PMU data. A shapelet is *any subsequence* that differentiates one type of sequences from other types. To design our domain specific shapelets, we need to figure out what is the *intrinsic nature* that makes the disturbance waveforms different. For the PMU data, we observe that the disturbances

caused by the same physics, or in other words, corresponding to the same disturbance event, show similar waveforms. We also observe that the critical changes occur only in a short time after the a disturbance is triggered. Therefore, we only need to extract shapelets from a subset of the waveform close to the trigger point. We denote the domain specific shapelets as *Dshapelet*.

The first step of extraction *Dshapelet* finds the global extreme points, including global maximum and global minimum, from the chosen subset. The anchor extreme point is chosen as the first extreme point. This anchor extreme point is used for the extraction of the *Dshapelet* feature. Next, our algorithm extracts one subsequence that is centered at the anchor extreme point and has length  $2L+1$ . If the anchor point is at index  $n_1$ , then the extracted subsequence (or *Dshapelet*) is  $V_{n_1-L}, \dots, V_{n_1-1}, V_{n_1}, V_{n_1+1}, \dots, V_{n_1+L}$ . For instance, if the anchor extreme point is 20, which can be the global maximum point, for length  $L=3$ , the *Dshapelet* that we extract is the subsequence from samples 17 to 23.

3) *Slope Sequence for Dshapelet* ( $S^3$ ): Both the  $S^2$  feature and the *Dshapelet* feature have their own advantages: the  $S^2$  feature captures the trend of a sequence by calculating the changes of consecutive sequence values, while the *Dshapelets* feature captures the shape of the most dramatic change. We then design a new feature by combining these two features to combine their strengths. The new feature of a waveform is the slope sequence of its *Dshapelet*, and is denoted as  $S^3$ .

4) *Statistics-based features*: The research in [21] developed features, based on the statistics of the data, specifically for fault and generation loss classes. These features are the elements of the statistics-based feature vector reported in this paper. We briefly describe these features.

- 1) Median vs. mean: The disturbance waveform is normalized as

$$V_{\text{norm}}(n) = |V(n) - \bar{V}| \quad (9)$$

and the feature is calculated as

$$F_1 = \frac{\bar{V}_{\text{norm}} - \tilde{V}_{\text{norm}}}{\sigma_V} \quad (10)$$

where  $\bar{V}_{\text{norm}}$ ,  $\tilde{V}_{\text{norm}}$ ,  $\sigma_V$  is the mean, median, standard deviation respectively of  $V_{\text{norm}}$ .

- 2) Difference median and mean difference: The backward difference magnitude of the disturbance waveform is

$$\dot{V}(n) = |V(n+1) - V(n)| \quad (11)$$

and the feature is difference of the mean and median of  $\dot{V}$

$$F_2 = \bar{\dot{V}} - \tilde{\dot{V}}. \quad (12)$$

- 3) Variance distribution: The disturbance variance distribution feature first calculates the variance due to each point in the waveform as

$$V_{\text{var}}(n) = [V(n) - \bar{V}]^2. \quad (13)$$

Elements of (13) are then sorted in descending order yielding  $\hat{V}_{\text{var}}(n)$  and the minimum number of points

(minus one),  $n_{\text{var}}$  that could account for 60% of the total variance is found, i.e.

$$\sum_{n=0}^{n_{\text{var}}} \hat{V}_{\text{var}}(n) < 0.6 \sum_{n=0}^{N-1} \hat{V}_{\text{var}}(n) \quad (14)$$

where  $N$  is the total number of elements in  $\hat{V}_{\text{var}}$ . This percentage was optimized experimentally. Finally,  $n_{\text{var}}$  is normalized yielding the feature

$$F_3 = \frac{n_{\text{var}}}{N}. \quad (15)$$

- 4) Change distribution: This feature is calculated similarly to  $F_3$ , but rather than using the element-wise variance,  $\dot{V}(n)$  is used and sorted in descending order, yielding  $\hat{\dot{V}}(n)$ . Then the minimum number of points (minus one),  $n_{\text{ch}}$  that could account for 80% of the total change is found, i.e.

$$\sum_{n=0}^{n_{\text{ch}}} \hat{\dot{V}}(n) < 0.8 \sum_{n=0}^{N-2} \hat{\dot{V}}(n). \quad (16)$$

This percentage is also optimized experimentally. Finally,  $n_{\text{ch}}$  is normalized yielding the feature

$$F_4 = \frac{n_{\text{ch}}}{N}. \quad (17)$$

- 5) IMF Correlation: The Hilbert Spectrum, described in [22], is computed for the voltage waveform resulting in typically four Intrinsic Mode Functions (IMFs). The correlations between each IMF and the voltage waveform are found. The feature is the ratio of the cross-correlation between the first IMF and the original waveform and the maximum correlation between the original waveform and all other components. One is added to both correlations so that anticorrelations do not cancel each other out. Details on the calculation of the IMF Correlation feature,  $F_5$  can be found in [21].
- 6) Frequency falling ratio: This feature quantifies how steady the frequency decrease is, i.e. the frequency decrease from the maximum to the minimum values and the time interval supporting this decrease. Our observations revealed a steadily decreasing frequency is typically associated with generation loss, which reflects the physics of the event. Details on the calculation of the frequency falling ratio feature,  $F_6$  can be found in [21].

## IV. EXPERIMENT DATA AND RESULTS

### A. Data

Both actual and simulated disturbance data were utilized in our evaluation. Field data were gathered from a PDC serving four PMUs sampling at 30 frames per second (fps) located in the 345-kV network of a WECC member utility in the United States. Each PMU is set to trigger at  $\pm 10\%$  change in the base voltage, or  $\pm 0.05$  Hz change in the base frequency (60 Hz), or  $\pm 5$  Hz/second change in the rate of change of frequency. As mentioned in Section III, the disturbance files collected

from the PDC were pre-processed to construct data sequences with 60 samples recorded in 2 seconds—0.5 s (15 samples) before the triggering point and 1.5 s (45 samples) after (and including) the triggering point. The field data set contain 23 waveforms corresponding to generation loss and 58 waveforms corresponding to faults.

The number of events captured by field data is small and hence poses challenges for training and evaluating classifiers. Therefore, the PSLF dynamic simulation tool from GE was used for creating a dataset of simulated PMU events belonging to different event classes. In order to simulate actual operating conditions of the grid, we took a load flow base case for the WECC system, tagged as 2008 Heavy Summer, which can be loaded into PSLF. Voltage and frequency probes were placed at the four buses where the actual PMUs are located, and similar triggering criteria was used as used by the actual PMUs. Though data were sampled at 240 fps, they were downsampled to 30 fps to be consistent with the field data. They were also sized the same as the field data files.

Seven disturbance types, based on their importance in power systems, were simulated: faults (FLT), generation loss (GL), load switching (LS) on/off, reactive power switched in/out, and synchronous motor switching off. For these disturbance types, we generated 3495 events in total, and the breakdown of the events is given in Table I. Examples disturbance data are shown in Fig. 1. Notice the field waveforms for fault and generation loss contain noise, whereas all the other simulated waveforms are noise-free.

TABLE I: Number of simulated data events for seven disturbance classes.

class	events
Fault (FLT)	1260
Generation loss (GL)	558
Load switching (LS) off	348
Load switching (LS) on	354
Reactive power switched out	497
Reactive power switched in	442
Synchronous motor switching off	36
<b>Total</b>	<b>3495</b>

## B. Results

The features described in Section III were extracted from the data and fed to both the KNN and SVM classifiers. In addition, we also fed raw data directly to the classifiers, so the effectiveness of different feature extraction techniques can be compared with a baseline. The accuracy of our classifiers is determined by the number of correct classifications as a percentage of total number of classifications.

All coefficients were selected for DFT and and DWT. For the PCA-based features the largest 20 principal components were selected based on trials. Least Asymmetric wavelet was chosen as the mother wavelet for DWT; however other mother wavelets like Daubechies and Coiflet also gave comparable results.  $k=1$  was set for the  $S^2$  method, and  $L=4$  was chosen for Dshapelet. For  $S^3$ , we generate the Dshapelet with  $L=5$ , and then  $k=1$  for slope. The larger  $L$  ( $=5$ ) is used for  $S^3$

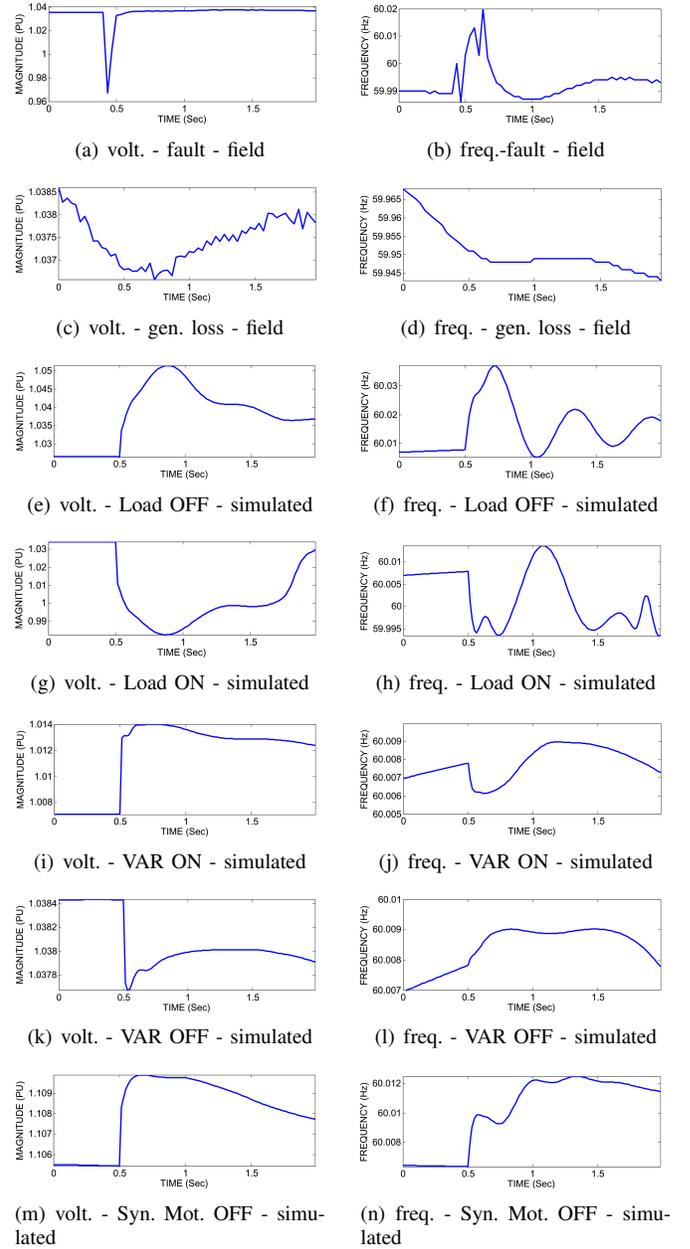


Figure 1: PMU recordings of different events.

because the slope sequence calculation reduces the length of shapelets. Finally for the statistics-based features, we use a feature vector consisting of the features  $[F_1, \dots, F_5]$  when only the voltage waveform is used and  $[F_1, \dots, F_6]$  when both voltage and frequency waveforms are used. For the KNN classifier, we obtained good results with  $k = 1$  and for the SVM, a radial basis function with Gaussian kernel gave the best results.

Every disturbance event affects the voltage as well as frequency of the power system to varying extent. We carried out classification exercise using 1) voltage data only, 2) frequency data only, and 3) both voltage and frequency data. For the third case, the features extracted from voltage and frequency

data were concatenated.

As explained in Section IV-A, we have ample datafiles for seven simulated events, but only 81 field datafiles corresponding to two events - faults and generation loss. For robust evaluation of the performance of feature extraction methods and classifiers *given these constraints*, we created four types of experiments:

- 1) use simulated data corresponding to Generation Loss and Fault for training a two-class classifier, and use field data for testing (results in Table II),
- 2) use simulated data corresponding to all types of disturbances for training a seven-class classifier, and use field data (two classes) for testing (results in Table III),
- 3) use simulated data corresponding to Generation Loss and Fault for training a two-class classifier *and* testing, using 10-fold cross validation (results in Table IV),
- 4) use simulated data corresponding to all types of disturbances for training a seven-class classifier *and* testing, using 10-fold cross validation (results in Table V).

Tables do not show results with frequency data only, because the concatenated data yielded better results in all cases. Since the last two cases involve only simulation data, tenfold cross-validation was used [23] and the dataset was partitioned 90% for training and 10% for testing. For each training/testing configuration, *ten* different tests were performed to guarantee that each fold is used at least once as training data and used at least once as testing data. Reference [6] provides more details of this approach.

TABLE II: Training (1818 simulated events with faults and generation loss) and testing (81 actual events)

	Voltage		Voltage and Frequency	
	1NN	SVM	1NN	SVM
Raw data	77.7	93.8	64.2	71.6
DFT	88.8	72.8	80.2	71.6
DWT	80.2	81.4	62.9	71.6
FDST	86.4	69.1	88.9	74.1
PCA	71.6	71.6	71.6	71.6
$S^2$	81.5	95.1	88.9	93.8
Dshapelet	87.6	<b>100</b>	69.1	71.6
$S^3$	95.1	97.5	88.9	<b>100</b>
Statistics	<b>97.5</b>	93.8	<b>97.5</b>	84.0

Observing the results in these tables, where highest accuracy numbers are boldfaced, it is clear that concatenating the features of voltage and frequency is desirable, although the voltage waveform alone can also give good results. For such features,  $S^3$  is able to classify faults and generation loss with near perfect accuracy, when combined with SVM (Tables II, III, IV). For the case in Table V, where all 7 classes are being classified, the accuracy with  $S^2$  is above 90%, but nowhere near perfect. Another important observation is that raw data, and feature extraction methods that choose subsets of raw data, as well as those that quantify properties of raw data (DShapelet,  $S^2$ ,  $S^3$ , *Statistics*) perform better than the classical methods that decompose or transform the data

TABLE III: Training (3495 simulated events from 7 disturbance types), testing (81 actual events with FLT and GL)

	Voltage		Voltage and Frequency	
	1NN	SVM	1NN	SVM
Raw data	49.4	81.5	61.7	71.6
DFT	55.6	71.6	72.8	71.6
DWT	50.6	75.3	60.5	71.6
FDST	<b>86.4</b>	69.1	<b>88.9</b>	74.1
PCA	71.6	71.6	71.6	71.6
$S^2$	50.6	<b>87.7</b>	56.8	85.2
Dshapelet	56.9	74.1	61.7	71.6
$S^3$	77.8	82.7	72.8	<b>100</b>
Statistics	32.1	81.5	54.3	84.0

TABLE IV: Training (90% of 1818 simulated events with FLT and GL), testing (10% of 1818 simulated events with FLT and GL); ten-fold cross validation

	Voltage		Voltage and Frequency	
	1NN	SVM	1NN	SVM
Raw data	99.1	92.7	98.9	89.5
DFT	98.5	87.3	98.3	83.9
DWT	99.1	90.4	98.9	88.1
FDST	<b>99.5</b>	96.9	98.5	98.6
PCA	69.8	69.3	69.3	69.3
$S^2$	<b>99.9</b>	99.4	<b>99.9</b>	<b>99.2</b>
Dshapelet	<b>99.4</b>	97.9	<b>99.7</b>	96.0
$S^3$	<b>99.9</b>	99.3	<b>99.6</b>	<b>99.3</b>
Statistics	<b>99.9</b>	<b>99.7</b>	<b>99.8</b>	<b>99.5</b>

in to other domains. This is not surprising looking at the disturbance waveforms in Fig 1, where the changes in voltage and frequency waveforms are *visually different* for faults and generation loss. The last observation is about the robustness of the better performing method against noise in field data. Simulation data used for training has no noise, but the field data has noise as seen from Fig. 1 (a) through (d). With the chosen parameters for  $S^2$  and  $S^3$ , the noise does not interfere with classification. It should be mentioned here that the features for the *Statistic* approach were designed for only fault and generation loss events. Therefore, its performance for cases where all 7 classes are used for training (Tables III and V) is inferior.

## V. DISCUSSION AND FUTURE WORK

As explained in Section I, event identification has two purposes: 1) system visualization, which is a SCADA-like tool, and 2) supervisory protection, which verifies relay operation. In this research, we have identified candidate features which can lead to accurate event classification. It has been shown that simple features based on slope and statistics work uniformly better than classical methods. This is an important finding. We have also been able to reduce the data to be processed significantly by detecting the strongest PMU signal. We are able to work with 1.5 s of post-disturbance data, which, even with delays in communication and computing (slope/statistic calculations impose very low calculation burden), is much

TABLE V: Training (90% of all 3495 simulated events with 7 disturbance types), testing (10% of all 3495 simulated events with 7 classes); ten-fold cross validation

	Voltage		Voltage and Frequency	
	1NN	SVM	1NN	SVM
Raw data	80.6	75.2	86.9	77.8
DFT	74.8	66.5	81.8	64.3
DWT	80.5	74.1	86.9	74.9
FDST	73.9	70	82.9	77.4
PCA	36.7	36.1	36.1	36.1
$S^2$	<b>87.5</b>	<b>80.5</b>	<b>93.3</b>	<b>91.2</b>
Dshapelet	78.2	69.9	86.3	79.6
$S^3$	79.8	71.7	92.5	88.5
Statistics	77.1	69.2	82.8	76.2

faster than any methods reported for event identification in literature so far.

For supervisory protection, faults need to be classified with 100% accuracy, because it is supposed to verify relay operations, detecting any misoperations. This has been achieved with acceptable speed. Visualization is supposed to enrich SCADA data, which is received once every 2-4 seconds. The proposed application provides better speeds; however, accuracy is not 100% when we consider all classes. This needs to be addressed in future work. It was observed that most misclassifications involved load/reactive power switching with generation loss, which makes sense due to the similarity in physics underlying these events. It is probable that more PMUs in the system will alleviate this problem, since the PMU nearer to the currently misclassified event may produce stronger signatures and hence better features.

## VI. CONCLUSION

This paper reports on the discriminating ability of a wide range of features when applied to disturbance data recorded by PMUs. Many features are proposed for the first time. Testing is performed under a common dataset and using the same classifiers. This exercise allowed us to investigate and balance computational complexity with classifier accuracy which is critical in real-time applications using PMU data. The use of noisy field data tests the robustness of the features. Results show that simple features based on slope and statistics, applied to a rather small 1.5 s data-window on data from only one PMU, offer the best discrimination, and outperform traditional methods for both simulated as well as field data. The candidate PMU is found based on detecting the strongest signal. The  $S^3$  feature performs the best overall. Accuracy and speed of event identification are excellent, and adequate for the intended applications of better visualization and supervisory protection.

## REFERENCES

[1] O. Dahal, S. Brahma, and H. Cao, "Comprehensive clustering of disturbance events recorded by phasor measurement units," *IEEE Transactions on Power Delivery*, vol. 29, no. 3, pp. 1390–1397, June 2014.

[2] J. Ma, Y. V. Makarov, C. H. Miller, and T. B. Nguyen, "Use multi-dimensional ellipsoid to monitor dynamic behavior of power systems based on PMU measurement," in *2008 IEEE Power and Energy Society General Meeting*, July 2008, pp. 1–8.

[3] L. Fan, R. Kavasseri, Z. Miao, D. Osborn, and T. Bilke, "Identification of system wide disturbances using synchronized phasor data and ellipsoid method," in *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, July 2008, pp. 1–10.

[4] "Use of synchrophasor measurements in protective relaying applications," Report of PSRC Working Group, 2012. [Online]. Available: [http://www.pes-psrc.org/Reports/Use\\_of\\_Synchrophasor\\_Measurements\\_in\\_Protective\\_Relaying\\_Applications\\_final.pdf](http://www.pes-psrc.org/Reports/Use_of_Synchrophasor_Measurements_in_Protective_Relaying_Applications_final.pdf)

[5] D. C. E. de la Garza, *Masters Thesis - Hidden Failures in Protection Systems and its Impact on Power System Wide-area Disturbances*. Virginia Polytechnic Institute and State University, 2000.

[6] O. P. Dahal, H. Cao, S. Brahma, and R. Kavasseri, "Evaluating performance of classifiers for supervisory protection using disturbance data from phasor measurement units," in *2014 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Oct 2014, pp. 1–6.

[7] J. Ma *et al.*, "The characteristic ellipsoid methodology and its application in power systems," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2206–2214, Nov 2012.

[8] O. P. Dahal and S. M. Brahma, "Preliminary work to classify the disturbance events recorded by phasor measurement units," in *2012 IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–8.

[9] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison Wesley, 2005.

[10] "Electromagnetic compatibility part 4: 30: Testing and measurement techniques—power quality measurement methods," *IEC 61000-4-30 Std*, 2003.

[11] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson/Prentice Hall, 2008.

[12] A. Jensen and A. la Cour-Harbo, *Ripples in mathematics: the discrete wavelet transform*. Springer Science & Business Media, 2001.

[13] M. Biswal and P. Dash, "Detection and characterization of multiple power quality disturbances with a fast s-transform and decision tree based classifier," *Digital Signal Processing*, vol. 23, no. 4, pp. 1071–1083, 2013.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[15] L. Ye and E. J. Keogh, "Time series shapelets: a new primitive for data mining," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 947–956. [Online]. Available: <http://doi.acm.org/10.1145/1557019.1557122>

[16] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme, "Learning time-series shapelets," in *SIGKDD*, 2014, pp. 392–401. [Online]. Available: <http://doi.acm.org/10.1145/2623330.2623613>

[17] E. J. Keogh and T. Rakthanmanon, "Fast shapelets: A scalable algorithm for discovering time series shapelets," in *SDM*, 2013, pp. 668–676. [Online]. Available: <http://dx.doi.org/10.1137/1.9781611972832.74>

[18] K. Chang, B. Deka, W. W. Hwu, and D. Roth, "Efficient pattern-based time series classification on GPU," in *ICDM*, 2012, pp. 131–140. [Online]. Available: <http://dx.doi.org/10.1109/ICDM.2012.132>

[19] J. Lines, L. M. Davis, J. Hills, and A. Bagnall, "A shapelet transform for time series classification," in *SIGKDD*, 2012, pp. 289–297. [Online]. Available: <http://doi.acm.org/10.1145/2339530.2339579>

[20] A. Mueen, E. J. Keogh, and N. Young, "Logical-shapelets: an expressive primitive for time series classification," in *SIGKDD*, 2011, pp. 1154–1162. [Online]. Available: <http://doi.acm.org/10.1145/2020408.2020587>

[21] P. Chen, "Automatic Classification of Disturbance Events," Master's thesis, New Mexico State University, Las Cruces, NM, 2015.

[22] S. Sandoval and P. L. De Leon, "Theory of the hilbert spectrum," *arXiv*, Oct. 2015, math.cv/1504.07554.

[23] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.