

A Sequential Bayesian Approach to Online Power Quality Anomaly Segmentation

Xu Jiang, *Student Member IEEE*, Bruce Stephen, *Senior Member IEEE*, Stephen McArthur, *Fellow IEEE*

Abstract— Increased observability on power distribution networks can reveal signs of incipient faults which can develop into costly and unexpected plant failures. While low-cost sensing and communications infrastructure is facilitating this, it is also highlighting the complex nature of fault signals, a challenge which entails precisely extracting anomalous regions from continuous data streams before classifying the underlying fault signature. Doing this incorrectly will result in capture of uninformative data. Extraction processes can be confounded by operational noise on the network including harmonics produced by embedded generation. In this paper, an online model is proposed. Our Bayesian Changepoint Power Quality anomaly Segmentation allows automated segmentation of anomalies from continuous current waveforms, irrespective of noise. Demonstration of the effectiveness of the proposed technique is carried out with operational field data as well as a challenging simulated network, highlighting the ability to accommodate noise from typical network penetration levels of power electronic devices.

Index Terms— Power Distribution, High Resolution Power Quality Monitoring, Power System Anomaly Detection, Changepoint Analysis

I. INTRODUCTION

Utilities have moved to forestall operational problems by enhancing the visibility of their network with high frequency Power Quality (PQ) monitoring [1]. This is motivated by the premise that faults have early stage signatures of their incipience and these manifest themselves in waveform artefacts [1]. There have been two key areas of focus. The first is on specific pieces of equipment and assets and identifying their anomalous behavior from PQ monitoring data [2] [3]. The second, and less common, is analysis of network monitoring data to segment abnormal behavior from normal signals [1] [4] [5]. Both areas are key problems within smart grids, and both would benefit significantly from enhanced online anomaly detection and segmentation – online in the sense that data must be processed as it is received to avoid needless storage and offer early warning of operational problems. The challenge of this is that a large amount of noise from complex distribution networks and heterogeneity of anomalies hampers accurate detection, and conventional metrics are not able to measure the performance of segmentation tools in a meaningful way.

PQ monitoring implemented at substation level has the

potential to observe the whole downstream distribution network. As fault locations get further away from the substation, the signals captured can be weak and hidden in environmental noise and harmonics. This means waveform detection at substation level will require a robust approach to its detection method in order not to miss faults or parts of fault signatures which would in turn hamper their classification. Conventional high-resolution anomaly detection has previously utilized the differential waveform RMS method [2], Mean Absolute Variation in Squared Amplitude (MAVSA) [2] and Kalman Filtering (KF) [6] as examples. However, the former two approaches are prone to false alarms because they are built on local statistical characteristics, such as waveform RMS and mean value, rather than the entire waveform. In other works, [4] proposed a generic anomaly detection method using hypothesis testing. None of these works discussed anomaly detection performance in a non-linear noise environment or were tested against streams of data in an online manner. With the increasing use of Low Carbon Technologies (LCT) in distribution networks, the harmonics and noise produced by convertors/inverters can degrade the power quality of the grid, and can be challenging to robust anomaly detection [7]. Furthermore, previous works have tended to require adjustment of the threshold of anomalous behavior to accommodate dynamic noise levels, which can also obstruct practical implementation [2].

This paper develops and demonstrates Bayesian Changepoint PQ anomaly Segmentation (BCPQS), an online and generic PQ event segmentation method for distribution network level faults. The proposed model tackles the complex problem of segmentation of high-resolution PQ signals and automatically learns the bounds of periods of anomalous behavior in an online model with minimal historical data. This overcomes key barriers and problems outlined in the literature reviewed, and allows a move to effective online anomaly and fault detection. The contributions of the proposed methodology are: 1) the changepoint based anomaly segmentation model is more robust to non-linear noise from electronic devices on distribution networks than conventional detection; 2) The false positive rate and false negative rate for anomalous data regions are lower than conventional methods; and, 3) a new means of performance evaluation measurement is proposed that is more appropriate for continuous data feeds. The paper is organized

as follows: Section II outlines how the Bayesian Changepoint detection method [10] is implemented for online segmentation of PQ disturbances. Section III provides an overview of the fundamental elements of power quality disturbances and how these can be simulated for testing classifiers and anomaly detectors, which is considered in the practical context of distribution network monitoring. Section IV discusses the data used for validation, parameter initialization, the proposed benchmarks and a new performance measurement for models segmenting streamed PQ data. Section V demonstrates the practical applicability with case studies on both operational and simulated data with comparison against a number of established benchmark models.

II. BAYESIAN CHANGEPOINT PQ ANOMALY SEGMENTATION

This section will detail the approach BCPQS uses to extract PQ events from a continuous stream of single phase current sampled at a high resolution in an online manner. Plant failure on power distribution networks usually starts with power quality distortion, which manifests as either immediate interruption (e.g. equipment damage, animal contact and lightning strikes) or periodic anomalies (e.g. arcing [8]) of changing magnitudes. Any approach to capturing instances of these from normal operating data faces the following practical considerations:

- **Temporal dependency:** PQ data is non-stationary over time during disturbances and faults which requires the subsequent predictions to depend on the previous observations rather than assuming every observation is independent

- **Weak delineation from normal behavior:** Faults may cause anomalies to be dispersed across time which dispels the assumption that a new underlying regime will delineate a fault with a single structural break in stochastic behavior.
- **Quantification of Uncertainty:** power utilities need to know how confident they can be in automated decisions made so any approach must offer transparency as to the certainty that a segment of data is anomalous.
- **Online Operation:** The rate at which data is acquired means that online analysis must be conducted as data storage requirements will not be scalable to a network wide solution. Furthermore, online learning of anomalous behavior is preferable owing to the individual operating characteristics of networks.
- **Adaptive thresholds:** The heterogeneity of anomalies in distribution networks means fixed thresholds will not be sufficient for detection – changes in loading, which occur over longer timescales than faults, render these ineffective.

Based on these expectations, a changepoint analysis approach can be seen as an appropriate solution. Changepoint analysis finds abrupt changes [9][10] in time series data, which can in practice distinguish between different operating regimes, such as identifying industrial plant faults [11]. The conventional changepoint approaches include Hidden Markov Model [12], state space models such as the Kalman Filter [6] and the sequential application of hypothesis testing [13]. The approaches are designed to identify a particular abrupt change; however, individually they do not meet the practical constraints given above. Thus, the approach proposed here is the application of Bayesian Changepoint detection[9], sequentially updating it to model the uncertainty of anomaly occurrence across a detection window.

A. Processing Design

As Fig. 1 shows, raw data is broadly segmented in a sampling window, which ensures that the abnormal condition is learned only from local transients rather than long-term load variation. Then an abnormal component extraction algorithm is applied to transform the sinusoid in the window into a transient signal. Following this, the resulting signals are input to a sequentially updated changepoint detection model with initial conditions that can be either randomly selected or set using prior knowledge. These stages will now be described in detail.

B. Sampling Window Design

Based on the findings of past work [2], the length of the sampling window is specified as 8 cycles. To ensure the detection result is not affected by the initialization stage at every window, the detection avoids this and double counting anomalies by sliding over the raw data at 7.8 cycles (i.e. less than window length) to roughly segment the continuous input signal.

C. Fundamental Power Signal Separation

As described in [4], the current signal can be thought of as a sinusoid (the AC fundamental and its harmonics) with abnormalities (noise and transient components) superimposed. The conventional approach to decoupling the sinusoidal

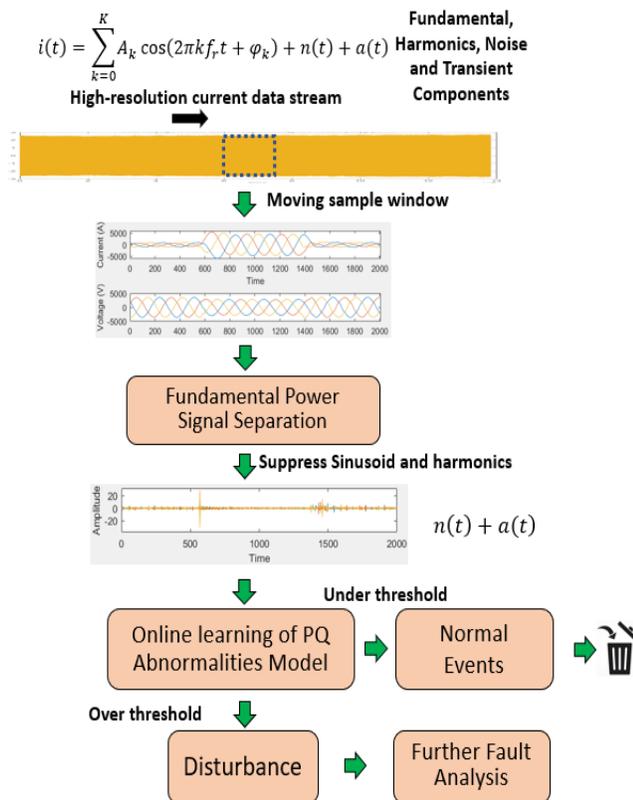


Fig. 1. The high-level process proposed for online segmentation of PQ data

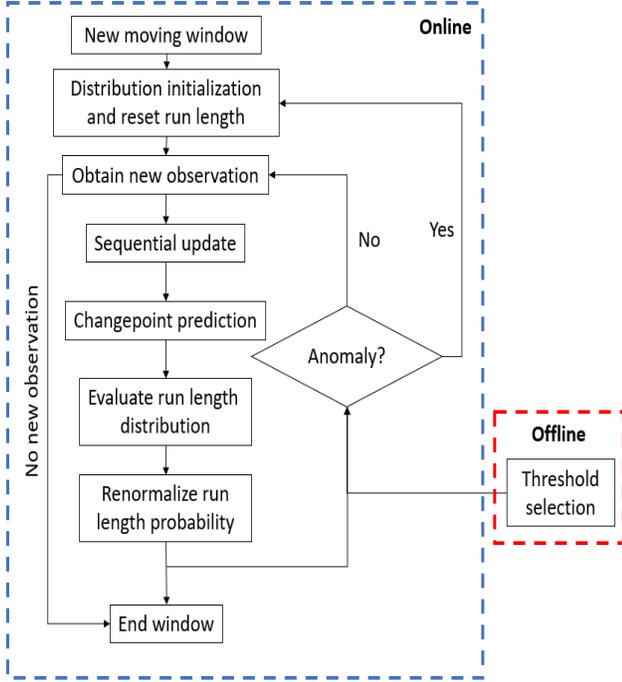


Fig. 2. Bayesian Changepoint anomaly detection process

components from the abnormal components of the signal are to superimpose faults over assumed normal signals using a Fast Fourier Transform (FFT) [4] or superimpose the fault over the previous cycle [2]. The latter method is adopted here to reduce the computational complexity and accommodate the supposed absence of knowledge of the form of faults. The extracted component residual has been validated as Gaussian distributed for normal operation [4].

D. Online Learning of PQ Abnormalities

After noise extraction, the residual noise signal from the single phase current is then run through a changepoint detection model [9] to segment abnormal events through identifying the start point and the end point as changepoints. To identify changepoints within the sampling window (Section IIB), the problem can be formulated as $P(S_i | x_{1:i})$, where i is the time step, $x_1 \cdots x_i \cdots x_k$ are a sequence of PQ current observations, x_k is the last observation in the sampling window and S_i is the binary state (healthy or abnormal) at time i . Thus, the formula represents a probability of the normal or abnormal state at time i . If the probability strongly indicates an anomaly existed at time i , then the waveform at that time could be saved and put in a fault analysis model for further processing. The process of online segmentation of a PQ abnormality for a single detection window is shown in Fig. 2. Online learning of the bounds of an anomalous region works on the principle of run length estimation [9], that is, how long a model can be sequentially updated with new PQ observations before a random model of behavior is considered to be more likely than the current one. To achieve this in this application context, BCPQS assumes the parameters of the start point in the sampling window have been reset to represent a normal event – this is undertaken offline using historical normal events. Each new observation updates model hyper-parameters, which are then utilized to evaluate the changepoint probability and run length distribution. Finally, the

run length probability is normalized and used to detect anomalies by comparison with probability thresholds learned offline. When a changepoint has been encountered, the prior distribution is reset, and parameters are sequentially calculated as before until another changepoint is found. After a start point is found, the fault segmentation function is triggered to look for the corresponding end point of the abnormal event and segment it out. The end point is determined when three consecutive cycles are free of changepoints [2]. This section will now detail: the initialization of the BCPQS model parameters, how these are sequentially updated as more data is observed and how the detector model works in the presence of PQ noise.

1) Noise Distribution Initialization

As discussed in [4], the non-periodic component of a normal signal's behavior can be modeled by Gaussian distribution with the presence of any abnormalities resulting in the violation of this assumption. A key part of BCPQS arises from adopting the exponential family form of a Gaussian distribution with conjugate priors over its parameters. The Bayesian formulation of this results in a Student-t distribution being used to track abnormal performance as a probability rather than a Gaussian distribution which has the additional benefit of being more robust to outliers [14]. The distribution is parameterized by the prior hyperparameters which in turn can be used to formulate expressions for the mean, variance and degree of freedom for the distribution by integrating over all possible values of the prior distribution[9].

The conjugate prior of the Student-t distribution is the Normal-Gamma distribution[14], $\mu, \Sigma \in NG(u, k, \alpha, \beta)$ which is expressed as:

$$NG(\mu, \Sigma | u, k, \alpha, \beta) = Z \cdot \sum^{\alpha - \frac{1}{2}} \cdot e^{-\frac{\Sigma}{2}(k(\mu - u)^2 + 2\beta)} \quad (1)$$

$$Z(u, k, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{2\pi}{k}\right)^{-\frac{1}{2}} \quad (2)$$

Where u, k, α, β are hyper-parameters and Γ denotes the Gamma function. Gamma function is a well-known factorial generalization function amounting to an approximation of a continuous factorial. The initial hyper-parameters of the observation distribution can change the shape of the distribution which affects the sensitivity of changepoint detection [14]. The initial values of μ_0 and k_0 can be determined by the mean and variance of the mean value of normal behavior data. According to [14], the variance of the sample is determined by the initial parameters α_0 and β_0 for the Gamma distribution.

2) Sequential Update of Noise Distribution

The parameters of the predictive distribution need to be sequentially updated based on prior observation. To achieve this, a sequential Bayesian estimation is utilized to derive parameter updates [14] and the corresponding Bayesian posterior [14] is:

$$u_{i+1} = \frac{k_i u_i + \bar{x}}{k_i + 1} \quad (3)$$

$$k_{i+1} = k_i + 1 \quad (4)$$

$$\alpha_{i+1} = \alpha_i + \frac{1}{2} \quad (5)$$

$$\beta_{i+1} = \beta_i + \frac{1}{2}(x_j - \bar{x})^2 + \frac{k_i(\bar{x} - u_i)^2}{2(k_i + 1)} \quad (6)$$

This permits a sequential update to be carried out using just the sufficient statistics of the distribution – in the case of the Gaussian distribution, this amounts to a running sum for the mean.

The new parameters are estimated based on the previous distribution and the current observation. Then the updated new parameters can produce a new Student-t distribution to evaluate the next changepoint probability.

3) Changepoint Detection

The BCPQS model is used to evaluate the run length probability of PQ noise over a window. Fig. 3 shows an example of how the exponential family distribution works on the PQ noise and how the model updates for new observations. For each run length probability calculation, the hyper-parameters can be

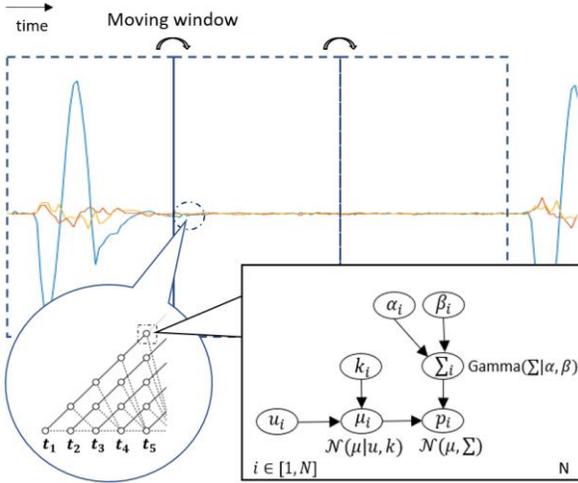


Fig. 3. Window based tracking of changes in noise distribution

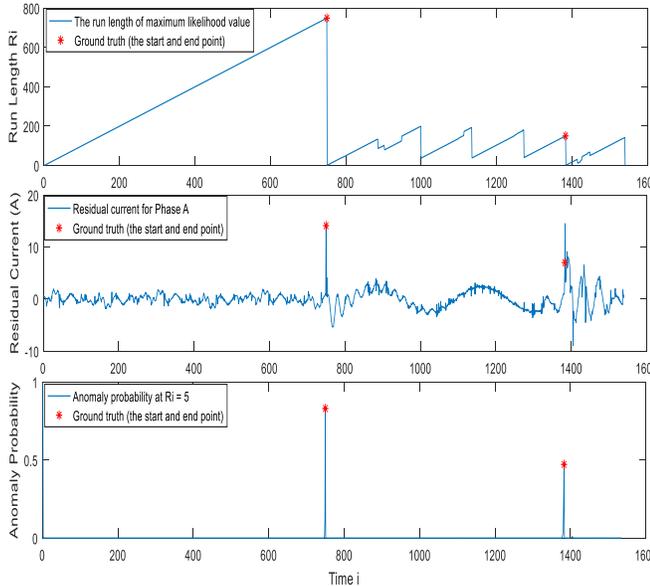


Fig. 4. Comparison of indicators for changepoint result, (a - top) how the run length of the maximum likelihood varies with time, (b - centre) how the residual current varies with time, (c - bottom) how the anomaly probability at $r_i = 5$ varies with time

updated with a new observation injected, then put in Normal-Gamma distribution to calculate distribution prediction and run length probability. The run length probability is expressed as:

$$P(r_i, x_{1:i}) = \sum_{r_{i-1}} P(r_{i-1}, x_{1:i-1}) P(r_i | r_{i-1}) P(x_i | u_i, k_i, \alpha_i, \beta_i) \quad (7)$$

Where r_i is run length at time i , $r_i = r_{i-1} + 1$ for normal event and $r_i = 0$ for anomaly. The run length probability can be calculated for every possible scenario and can form a trellis of probable run lengths over the sampling window.

E. Offline Threshold Selection

An appropriate threshold and associated approach are required to pick up the abnormal segmentation from continuous current signals. The run length specifies the point in time that an anomaly occurs by its associated likelihood value – if this is lower than that for a shorter run length, it is indicative that it should be reset. Owing to the time period required for a fault to develop, $r_i = 5$ has been observed experimentally to identify the minimum plausible run length at a given time. This is because the purpose of the model is to segment significant disturbances, such as incipient faults, rather than subtle deviations from normal operation, such as demand fluctuation; if r_i is too small, the model can be too sensitive to the outliers; large r_i makes it possible to pick up changepoints with small variance. An example is given in Fig. 4. The run length is sensitive to waveform distortion. As shown in Fig. 4a and 4b, the small signal inference around the time 900, 1000, 1140 and 1280 can result in changepoint reset but these are not expected to detect. In Fig. 4c, $r_i = 5$ makes start point and end point more apparent and easily detectable through setting an appropriate probability threshold.

Threshold of anomaly probability is chosen through analysis of historical events and how they affect the false alarm rate [4], and a buffer of three extra cycles on either side of the anomalous region is retained to capture subtle transients as noted in [2].

III. MODELING CONSIDERATIONS FOR PQ ANOMALIES

As noted in [4], PQ measurements under normal conditions on an operational network can be decomposed into noise components and normal load components. Transient events contain components in addition which can indicate the operational context of the distribution network [15]. In this waveform, the voltage and current can usually be expressed in the following terms

$$S(t) = \sum_{k=0}^K A_k \cos(2\pi k f_r t + \varphi_k) + n(t) + a(t) \quad (8)$$

Where t is the time, $S(t)$ is the signal measured at substation, A_k and φ_k represent amplitude and phase of k th harmonic component, K is the highest harmonic order, f_r is the fundamental operation frequency, $n(t)$ is high-frequency random noise and $a(t)$ represent the transient components. In order to rigorously test the capabilities of an anomaly detector,

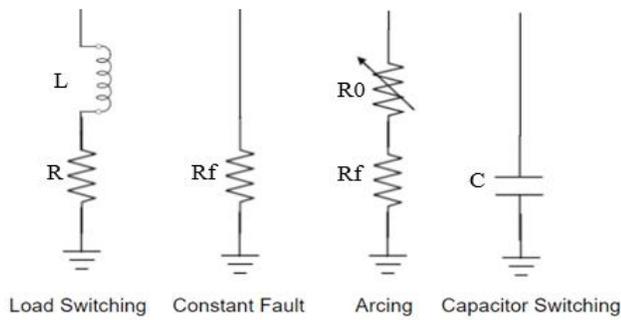


Fig. 5. Definition of transient models in terms of RLC

it is necessary to identify the features of PQ disturbances that it may encounter. This section covers the individual aspects of this and discusses how these can be realistically modelled.

A. PQ Transients

Many faults have precursors that are observed long before they evolve into a serious outage. Precursors, such as effects from electrical arcing and mechanical aging, can repeat and generally increase in severity resulting in outright failure of components. The duration of these transients varies, but incipient faults can occur more frequently as they develop further. The duration of the incipient fault development offers a possibility for utilities to forestall serious faults or outages in advance. However, not all transient components are attributable to faults, some of them are caused by normal operation of the network, such as load switching [16]. These regular events can result in small changes to signals which might confuse an anomaly detector into a false alarm. To ensure the proposed segmentation technique is robust to this, this paper simulates different transient components to test the robustness of anomaly segmentation methods. Some transients are related to faults, such as Kizilcay's arcing model [16] [17], capacitor switching model and three-phase fault model [4]. Others are not, such as the load switching model [4]. The underlying electrical characteristics of these models are shown in Fig. 5.

Kizilcay's arcing model has previously been validated [17] and used to test distribution network PQ anomaly detection [4][16]. This model is representative of arcing in air in a resonant-earthed medium voltage system [17] and also emulates high impedance faults (HIF) in overhead lines on distribution networks [17]. These arise in the operational context in two situations: broken overhead lines and faults that earth through arcing suppression coils. In both of these cases the high impedance nature of the fault results in a very small fault current which can be challenging to detect as an anomaly. As Fig. 5 shows, load switching in the paper represents the inductive load switching, which is common in daily load schedules to achieve system balance on a distribution network. Load switching usually results in small transient components in the PQ waveform which could confuse an anomaly detector: capacitor switching is used to correct voltage and power factor, in doing so though, it can bring high magnitude and high frequency transient overvoltage which may damage customer side equipment, resulting from oversensitive tripping in response to these momentary excursions [18]. It can significantly affect the network health compared with load switching [18]. Although capacitor switching transients can be mitigated by limiters in the operational context [19], it is

worthwhile to monitor capacitor switching to ensure it is functioning correctly. Therefore, for anomaly detection, capacitor switching should be taken into account for fault type, but load switching should be treated as a normal event.

B. PQ Noise

PQ noise challenges the correct identification of transient components using intelligent detectors [20]. Noise is commonly modelled as high-frequency white noise which is usually caused by electromagnetic interference (EMI). EMI can damage plant and degrade quality of service. Thus, standards with respect to Electromagnetic Compatibility (EMC) [20] are published to mitigate resulting threats. Although noise such as EMI can be limited, PQ noise is inevitable, and they can exist on any device, especially for electronic devices [20]. Therefore, it is necessary to verify the robustness of an anomaly segmentation in the presence of such forms of noise. [4] has verified that PQ noise under normal operation almost obeys a Gaussian distribution. However, the noise can be time-variate under different environments. Thus, existing high-resolution detectors can make mistakes on diagnosis if the detector is sensitive to the small variance [4].

C. Harmonic Content

PQ harmonics are usually produced by industrial plant and power electronic devices in commercial buildings [20]. This facet of normal operation could affect anomaly segmentation performance. Although a lot of work has already proposed schemes to reduce or eliminate harmonics [20], high-frequency harmonics can be difficult to completely filter out. Therefore, anomaly segmentation on the network must be robust to the harmonics from realistic penetrations of power electronic devices. Furthermore, with a large number of residential customers producing harmonics through domestic PV inverters, which can affect network condition in the long-term, and these can be represented using non-linear load models [21]. To mimic harmonics from a non-linear load, a simple generic power electronic model [22] is utilized and shown in Fig. 6. This model utilizes diode switching along with an impedance model to emulate the combined effect of a variety of non-linear loads seen from a utility connection point. Diode switching can produce the same ripples which dominate the harmonics generated by power electronics which makes it an ideal noise emulator.

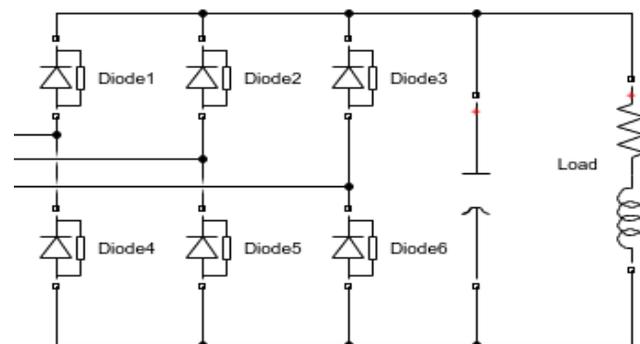


Fig. 6. Circuit representation of the harmonic model used

TABLE I
FAULT SETUP IN IEEE 13-BUS SIMULATION

Fault type	Fault Location	Duration (cycle)	Impedance
High Current Arcing Fault	633,634,645	0.25 to 1.25	$u_0=300\sim 4000$ V $r_0=0\sim 0.015$ ohm $\tau=5e-5\sim 4e-4$ s
Low Current Arcing Fault	680,652,692	0.25 to 1.25	$u_0=1000\sim 6000$ V $r_0=0.1\sim 100$ ohm $\tau=5e-5\sim 4e-4$ s
Constant Impedance Fault	680,652,692	2 to 3.25	$R_g=0.1\sim 1500$ ohm $R_f=0.1\sim 10$ ohm
Capacitor Switching Load Switching	680,652,692	N/A	$C=1e-6\sim 4e-6$ F
Load Switching	675,611,646	N/A	$R_g=0.3\sim 10$ ohm

TABLE II
NOISE CONFIGURATION IN IEEE 13-BUS SIMULATION

Noise type	Fault Location	Duration (cycle)	Configuration
White Noise	N/A	N/A	SNR = 70,80,90 dB
Harmonics	671	N/A	$R=0.1\sim 100$ ohm $L=0.0005$ H $C=60e-6$ F

IV. MODEL TESTING CASE STUDIES

To test the model capability, simulated and operational distribution network monitoring data are used. The purpose of using both is to ensure sufficient operational realism, in the case of the operational data, and the full spectrum of fault possibilities, in the case of the simulation. The BCPQS model is compared against several existing works to demonstrate its effectiveness and practical advantages.

A. IEEE 13-Bus Simulated Fault Data

The IEEE 13-bus model [23] is an unbalanced radial distribution network at medium voltage level which consists of underground cables, overhead lines, shunt capacitors and an in-line transformer. The network and disturbance configuration are shown in Fig. 7. According to [4][16], the initial parameter values for the IEEE 13-bus simulation are given in Table I.

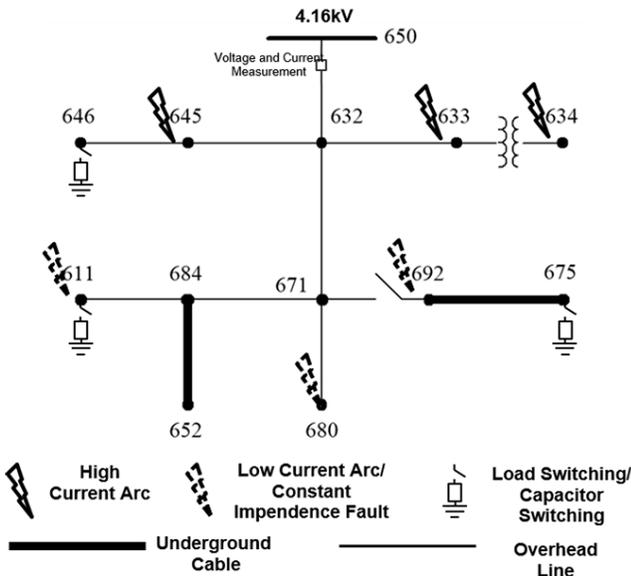


Fig. 7. IEEE 13-bus simulation with fault injection locations shown

Using this model, simulation is used to generate events to cover all possible parameter values: 1632 load switching events, 1278 high current arcing faults, 928 low current arcing faults, 480 capacitor switching operation and 560 constant impedance faults. To add additional operational realism, simulation of 10-50A low current arcing is also included which aligns with high impedance faults seen in practice [24]. Moreover, the fault duration is randomly assigned within a realistic range. Furthermore, the system short-circuit level can affect the disturbance PQ waveform characteristic obtained at the substation. A more robust network will obscure or minimize the changes to PQ waveform. The total load of the network is approximately 3.8MVA. Accordingly, the short-circuit level of the network can be set to 20MVA, 60MVA and 100MVA to test accuracy under different short-circuit levels [25]. Furthermore, some subtle disturbances, such as high-impedance faults, are difficult to detect because the noise and harmonics conceal the transient components. Therefore, an appropriate level of white noise and harmonics are added to the simulated signals to validate robustness to noise [26] [27]. Through testing 2000 normal condition signal segmentations from field data [27], 30dB signal-to-noise ratio (SNR) for current and 40dB SNR for voltage are commonly used for power delivery in distribution networks. According to IEEE 519-1992 [28], THD(%) limits of voltage and current are 5% and 15% respectively. This range determines the configuration of the harmonics generated by the impedance model described in Section III, which is given in Table II.

B. Operational Substation Waveform Data

To augment the tests against simulation data, two days of continuous operational 25kV substation data [27] are used to test the capability of the proposed method. The topology of this network implies an absence of power electronic devices which may present less of a challenge than the simulation cases. Fault occurrences have not been marked up on this data set, so prior to usage anomalous regions are manually identified so that the effectiveness of the changepoint can be determined. The RMS value for every cycle is extracted and shown in Fig. 8. According to these RMS values, only a single anomalous region can be found approximately 41 hours into the data where four abnormal events are picked up. In this region, four significant waveform distortions can be seen and are given in Fig. 9. The magnitude of PQ waveforms in Fig. 9 is far beyond the rating current [2] and harmonics limits in standards [28][29] - these abnormal events are capable of degrading network performance over time. Therefore, they will be used as the ground truth to test the effectiveness of the proposed anomaly detection. Additionally, as Section IVA discussed, the noise and harmonic levels of the simulated signals are learnt from the field data.

C. Model Initialization

Section IID.1 noted that normal operational data is required for the initial values of the prior distributions. In order to achieve this online, the first 200 moving windows, each comprising 8 cycles of normal operation [2], are used to initialize the prior distributions by fitting a Student-t distribution by Maximum Likelihood Estimation. Using the resulting location, shape and degree of freedom estimates for the sample Student-t, the initial values of u , k , α and β can be recovered [30] from:

$$v = 2\alpha \quad (9)$$

$$\mu = u \quad (10)$$

$$\lambda = \sqrt{\frac{\beta}{\alpha} \times \frac{1+k}{k}} \quad (11)$$

Where v is degree of freedom, λ is the scale factor and μ is the location factor. Sensitivity analysis reveals that the mean of the current almost remains stable at 0 which indicates the initial Gaussian mean can (intuitively for normal operation) be set to 0.

D. Benchmark Changepoint Detection Models

The Kalman Filter [6], Differential RMS [4][2] and MAVSA [4] detectors are used as benchmarks. Differential RMS detects anomalies through observing whether the differential waveform

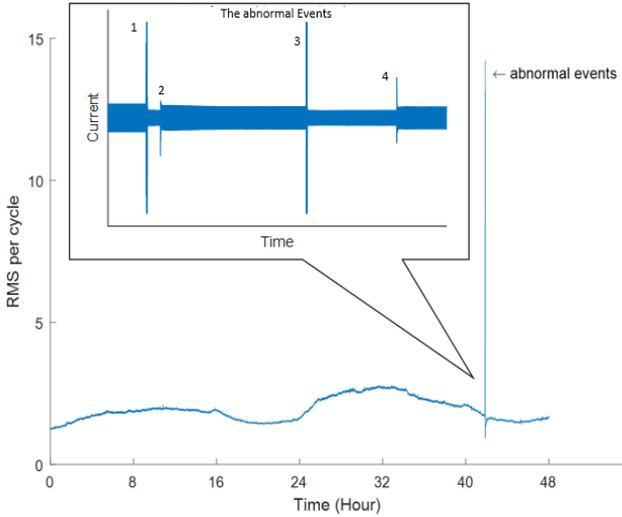


Fig. 8. Anomalous regions for the 48 hour long industrially sourced operational data set used for ground truth purposes

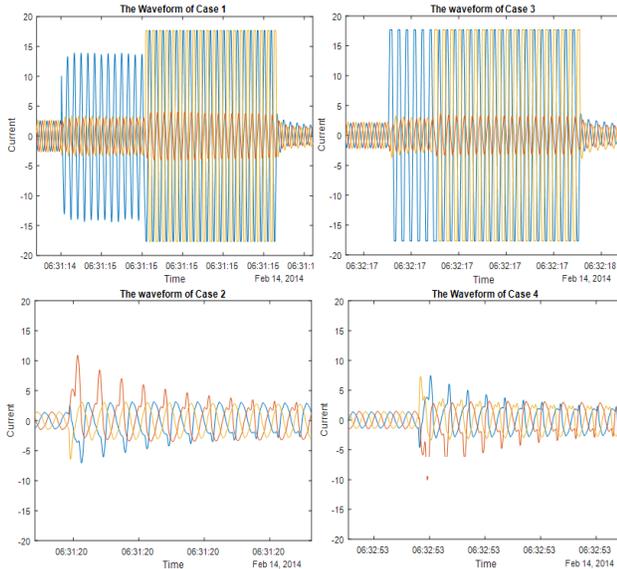


Fig. 9. The three-phase current waveform of the four abnormal events identified in Fig. 8.

between consecutive cycles is out with a predefined threshold [4]. MAVSA compares the difference of the squared value between two consecutive cycles with predefined thresholds to detect anomalies [4]. Anomaly detection by Kalman Filter utilizes the sequential estimation of a non-stationary Gaussian distribution from a dynamic signal; the signal is segmented through comparing the log likelihood of the observation with a predefined threshold [6].

E. Performance Evaluation Criteria

Continuous streams of monitoring data will mostly represent normal network operation. This resulting high prevalence can result in misleading evaluation metrics which can indicate a very high accuracy even when no detection takes place. However, abnormal events usually should be more important than normal events. This necessitates metrics to reward different scores for True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) rates and use these to reflect detector performance. Furthermore, a good performing anomaly detector requires correct detection of the abnormal signal as early as possible in order to segment out the entire fault trace. Therefore, an evaluation window is required to score every instance. Receiver Operating Characteristic (ROC) is a conventional metric used to evaluate the performance of anomaly detection [4]. This metric can reflect the impact of the probability of the detection and the probability of false alarm. However, ROC cannot reward early detection.

The Numenta Anomaly Benchmark (NAB) score [31] has been used to evaluate online anomaly detection against its ground truth. NAB labels ranges of data as anomalous, therefore it can penalize premature detection, reward early detection and penalize late detection outside the scoring window as Fig. 10 shows. It is formulated as follows:

$$S = \frac{2}{1 + e^{c(y-n)}} - 1 \quad (12)$$

where S represents the performance score, parameter c can control the rate of the drop, parameter n can be used to adjust the zero-cross point and y is the output of anomaly detection. Accordingly, if a disturbance lasts for three consecutive cycles, the segmentation can be identified as an abnormal event [2]. Therefore, missing detection for more than three cycles should

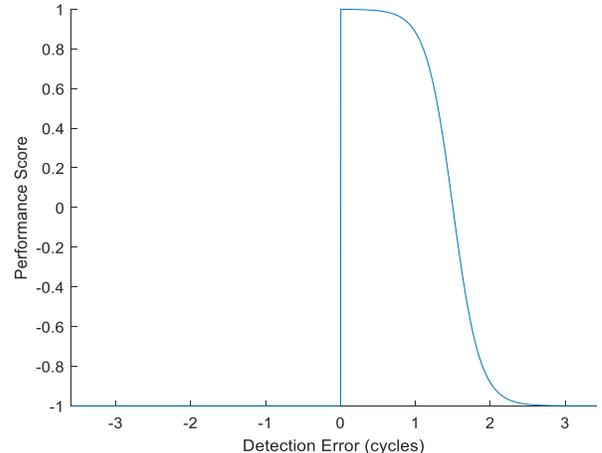


Fig. 10. The NAB scoring window shape to visualize the reward zone and penalization zone

be heavily punished by evaluation metrics as Fig. 10 shows. To reward a small detection error and heavily penalize missed detection, a narrow reward range is required, and the curve is expected to decrease more gently at the beginning. Therefore, $c=0.02$, $n=N*1.5$ where N is the number of samples in one cycle in the paper. For multiple detection within a single detection window, only the earliest result is used to contribute to the score. The overall score is:

$$S_t = A_{TP}S_{TP} - A_{FP}S_{FP} - A_{FN}S_{FN} \quad (13)$$

$$NAB = \frac{S_t - S_t^{null}}{S_t^{perfect} - S_t^{null}} \quad (14)$$

where A are weights for the confusion matrix elements to raise or lower the importance of true or false positives, and NAB is the ultimate score considered with all instances. $S_t^{perfect}$ represents all the instances are detected with best performance and S_t^{null} means the worst performance for all the instances. True Positive (TP) rates should be much more important than True Negative (TN) rates, so the reward for TP should be higher than that of TN, meaning $A_{TP} \gg A_{TN}$. Therefore, 0 is selected for A_{TN} with all others set to 1. The NAB score can obtain 0 for worst case and 100 for perfect detection. Accordingly, a NAB score of over 65 can be considered acceptable [31].

V. RESULTS

The BCPQS model is implemented in Python and executed on an Intel i7-6700HQ 2.6 GHz CPU. The model complexity is linear in time [10] which results in it taking approximately 0.12s to process a moving window (approximately 8 cycles for 3840Hz three-phase current data). This section will discuss the performance against the different detectors described in Section IVD, using both the simulated and operational data described in Sections IVA and IVB.

A. Segmentation Performance Using Simulated Network Data

To test the capability of segmentation, the detection accuracy taking account of the false alarm rate is informed by ROC

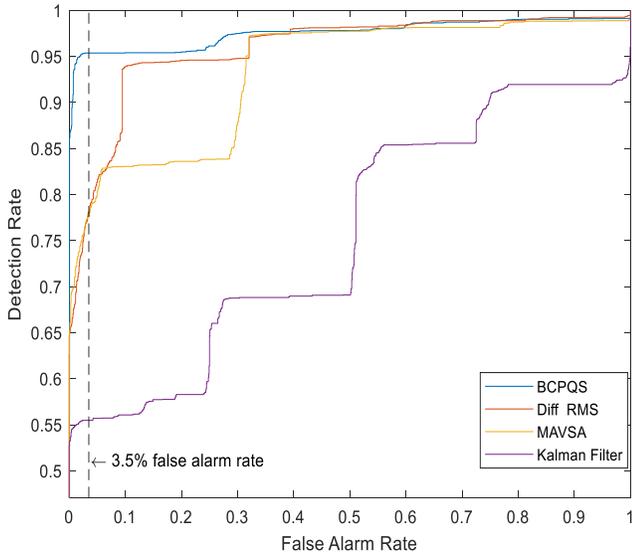


Fig. 11. ROC curve of comparing the proposed methodology against benchmarks

curve. The fault data from Section IVA is utilized to estimate ROC curve shown in Fig. 11.

The proposed BCPQS approach takes a larger area of the ROC curve which means it outperforms other detectors on the data set with non-linear load noise. Furthermore, the threshold of 3.5% specified by [4] at the same false alarm rate can be calculated for the following performance comparison, and the thresholds used are given in Table III.

BCPQS outperforms other detectors for all but one fault type as Table IV shows. As a benchmark, KF is good at identifying constant impedance and load switching, but poor for detection of arcing faults. Differential RMS and MAVSA get similar performance, as Table IV shows, and are good at identifying significant over-current change. However, both Differential RMS and MAVSA are built on local statistical characteristics, such as RMS and Variance, which can result in a false positive rate that is too high for practical implementation. Finally, summarizing performance for all event types, the start point and the end point of the event are used to validate the capability of segmentation using NAB, as Table V shows. Generally, the start point of the abnormal events are easier for detectors to capture which may be caused by the way faults gradually end rather than abruptly change back to normal operation.

B. Segmentation Performance Using Operational Data

Using the set of continuous operational data described in Section IVB, the proposed BCPQS method is compared against the Differential RMS method with results shown in Table VI. Because there are only 4 anomalous episodes in this data, there are too few cases to produce a ROC Curve. This also inhibits the selection of a threshold for the benchmark Differential RMS model: if the same thresholds as Section VA are used to pick up

TABLE III
THRESHOLDS OF DETECTORS AT 3.5% FALSE ALARM RATE

Detector	BCPQS	KF	Differential RMS	MAVSA
Threshold	0.0747	91.4024	0.0426	20963.88

TABLE IV
NAB SCORE OF FAULT DETECTOR PERFORMANCE, AT FALSE ALARM RATE 3.5%, STARTING TIME DETECTION

Fault type	BCPQS	KF	Differential RMS	MAVSA
High Current Arcing Fault	98.1	61.66	87.38	85.71
Low Current Arcing Fault	92.15	13.50	64.38	65.63
Constant Impedance Fault	92.66	85.86	83.84	85.43
Capacitor Switching	86.68	74.22	68.99	66.60
Load Switching	96.65	97.74	96.52	96.51

TABLE V
ANOMALY START AND END POINT TEST, NAB SCORE AT FALSE ALARM RATE 3.5%

Reference Point	BCPQS	KF	Differential RMS	MAVSA
Start Point	94.46	53.86	74	73.79
End Point	86.65	34.45	66.24	64.23

the abnormal events, the proposed method will continue to work well but the benchmark method produces thousands of false alarms – a NAB score of 0.00525. This is because the benchmark methods work on the absolute magnitude of signal which can be sensitive to different voltage and current levels. Therefore, in order to continue to make a comparison the threshold used for the simulation data in Section VA is supplanted with the fixed threshold proposed in [2] for use with Differential RMS – which raises its NAB score to 72.84. Under these conditions, both methods are able to detect the four abnormal events with no false alarms. However, the BCPQS model detects the abnormal event earlier than the Differential RMS method as Fig. 12 shows. The resulting NAB scores are shown in Table VI where BCPQS is demonstrated as the significantly better performer. This can be attributed to the BCPQS model building a distribution of plausible values of the next value [10] in the PQ current using the values observed since the last changepoint rather than a single prediction being compared to a single threshold.

VI. CONCLUSIONS AND FURTHER WORK

The contribution of this paper has been to develop online sequential Bayesian methods to automatically detect and segment operational fault and anomaly data collected from distribution networks under typical operation and realistically challenging simulated conditions. Against comparable approaches, the proposed method, Bayesian Changepoint Power Quality anomaly Segmentation, offers improved false alarm rates, earlier detection of disturbances and also provides associated severity metrics, interpretable as fault stage development. In addition, this paper has introduced a new

metric for evaluating the effectiveness of detecting anomalies online from continuous streams of distribution network monitoring data. Both operational data from a typical distribution network and simulations of various realistic challenges including high power electronic penetrations have been used to demonstrate its effectiveness against state of the art approaches.

The consequences of faults on distribution networks can be wide ranging with regulatory penalties have been formulated to reflect this and promote higher standards of operation. The effects of equipment degradation have been historically difficult to detect owing to the coverage and extent of distribution networks; however, works such as [1] [4] have shown that remote detection can be possible. High sampling frequency Power Quality data has been shown to capture fault signatures at the waveform level but only when intensively monitoring the network in a continual manner. In order to make such technologies cost effective to scale to operational, the data they produce must be intelligently managed in order to permit it to be captured and processed on a ‘grid edge’ device with minimal computing power. The segmentation algorithm contributed by this paper supports such advances by accurately identifying only the periods of interest within continuously monitored data and thus efficiently using storage or communication resource to report the data associated with a disturbance. Unifying these outputs with pattern recognition techniques has the potential to apportion situational awareness to these automatically segmented network disturbances which can then be acted on according to company specific procedure.

VII. REFERENCES

- [1] J. A. Wischkaemper, C. L. Benner, B. D. Russell, and K. Manivannan, “Application of Waveform Analytics for Improved Situational Awareness of Electric Distribution Feeders,” *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2041–2049, 2015.
- [2] Transmission & Distribution Committee, Power Quality Subcommittee, and IEEE Working Group on Power Quality Data Analytics, “Electric Signatures of Power Equipment Failures,” 2018. [Online]. Available: http://grouper.ieee.org/groups/td/pq/data/downloads/Signatures_Equipment_Failures_V2018Dec.pdf. [Accessed: 21-Feb-2019]
- [3] T. S. Sidhu and Z. Xu, “Detection of incipient faults in distribution underground cables,” *IEEE Trans. Power Deliv.*, vol. 25, no. 3, pp. 1363–1371, 2010.
- [4] B. Li, Y. Jing, and W. Xu, “A Generic Waveform Abnormality Detection Method for Utility Equipment Condition Monitoring,” *IEEE Trans. Power Deliv.*, vol. 32, no. 1, pp. 162–171, 2017.
- [5] B. D. Russell and C. L. Benner, “Intelligent Systems for Improved Reliability and Failure Diagnosis in Distribution Systems,” *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 48–56, 2010.
- [6] W. D. Penny and S. J. Roberts, “Dynamic Models for Nonstationary Signal Segmentation,” *Comput. Biomed. Res.*, vol. 502, pp. 483–502, 1999.
- [7] G. D. Antona, C. Muscas, and S. Sulis, “Localization of Nonlinear Loads in Electric Systems Through Harmonic Source Estimation,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3423–3430, 2011.
- [8] R. D. Telford, S. Galloway, B. Stephen, and I. Elders, “Diagnosis of Series DC Arc Faults—A Machine Learning Approach,” *IEEE Trans. Ind. Informatics*, vol. 13, no. 4, pp. 1598–1609, 2017.
- [9] R. P. Adams and D. J. C. MacKay, “Bayesian Online Changepoint Detection,” 2007. [Online]. Available: <https://arxiv.org/pdf/0710.3742.pdf> [Accessed: 21-Feb-2019]
- [10] P. Fearnhead, “Exact and efficient Bayesian inference for multiple changepoint problems,” *Stat. Comput.*, vol. 16, no. 2, pp. 203–213, 2006.
- [11] S. Maleki, C. Bingham, and Y. Zhang, “Development and

TABLE VI
NAB SCORE FOR OPERATIONAL DATA ANOMALY DETECTION

NAB	BCPQS	Differential RMS
Using the threshold from simulated data	95.07	0.00525
Using the threshold in [2]	95.07	72.84

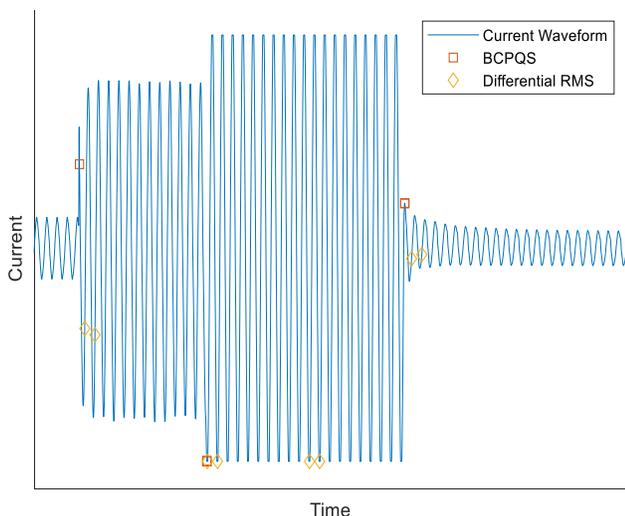


Fig. 12. Anomaly detection for BCPQS benchmarked against Differential RMS for case 1 in the operational data set

- Realization of Change-point Analysis for the Detection of Emerging Faults on Industrial Systems," *IEEE Trans. Ind. Informatics*, vol. 12, no. 3, pp. 1180–1187, 2016.
- [12] P. Fearnhead, "On-line inference for hidden Markov models via particle filters," *J. R. Stat. Soc.*, pp. 887–899, 2003.
- [13] U. Dwivedi and S. Singh, "Denoising Techniques With Change-Point Approach for Wavelet-Based Power-Quality Monitoring," *IEEE Trans. Power Deliv.*, vol. 24, no. 3, pp. 1719–1727, 2009.
- [14] K. P. Murphy, "Conjugate Bayesian Analysis of the Gaussian Distribution," *Def.*, vol. 1, no. 7, pp. 1–29, 2007.
- [15] J. A. Wischkaemper, "Electrical Characterization of Arcing Fault Behaviors on 120/208V Secondary Networks," Texas A&M University, 2011. [Online]. Available: <https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/158892/Wischkaemper.pdf?sequence=1> [Accessed: 21-Feb-2019]
- [16] W. Zhang, Y. Jing, and X. Xiao, "Model-Based General Arcing Fault Detection in Medium-voltage Distribution Lines," *IEEE Trans. Power Deliv.*, vol. 31, no. 5, pp. 2231–2241, 2016.
- [17] M. Kizilcay and P. La Seta, "Digital Simulation Of Fault Arcs In Medium-Voltage Distribution Network," in *15th Power Systems Computation Conference (PSCC'05)*, 2005, no. August, pp. 22–26.
- [18] Y. Tian and W. Xu, "A Method to Calculate Capacitor Switching Transients Using Short-Circuit Programs," in *16th International Conference on Harmonics and Quality of Power (ICHQP)*, 2014, no. 1, pp. 185–189.
- [19] T. Ghanbari, E. Farjah, and F. Naseri, "Three-phase resistive capacitor switching transient limiter for mitigating power capacitor switching transients," *IET Gener. Transm. Distrib.*, vol. 10, pp. 142–153, 2016.
- [20] S. P. Verma and P. Kumar, "Smart Grid, Its Power Quality and Electromagnetic Compatibility," *MIT Int. J. Electr. Instrum. Eng.*, vol. 2, no. 1, pp. 55–64, 2012.
- [21] K. D. Mcbee, M. G. Simões, and S. Member, "Evaluating the Long-Term Impact of a Continuously Increasing Harmonic Demand on Feeder-Level Voltage Distortion," *IEEE Trans. Ind. Appl.*, vol. 50, no. 3, pp. 2142–2149, 2014.
- [22] S. M. Adnaan and M. Tech, "Design and Simulation of a nonlinear Load Model Used to Simulate Voltage Notches and Harmonics Caused by a 6-Pulse Three-Phase Rectifier," *Int. J. Sci. Eng. Technol. Res.*, vol. 6, no. 6, pp. 984–995, 2017.
- [23] K. P. Schneider *et al.*, "Analytic Considerations and Design Basis for the IEEE Distribution Test Feeders Report Prepared by the Test Feeder Working Group of the Distribution System Analysis Subcommittee," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3181–3188, 2018.
- [24] IEEE PSRC Working Group D15, "High Impedance Fault Detection Technology," 1996. [Online]. Available: <http://grouper.ieee.org/groups/td/dist/documents/highz.pdf>
- [25] Russell Bryans, "Calculation of system fault levels," 2017. [Online]. Available: <https://www.spenergynetworks.co.uk/userfiles/file/ESDD-02-006.pdf>. [Accessed: 21-Feb-2019]
- [26] G. Vijayaraghavan, M. Brown, and M. Barnes, *Practical Grounding, Bonding, Shielding and Surge Protection*. 2004.
- [27] IEEE Working Group on Power Quality Data Analytics, "2 day gap-less waveform data." [Online]. Available: <http://grouper.ieee.org/groups/td/pq/data/>. [Accessed: 21-Feb-2019].
- [28] Transmission and Distribution Committee, "IEEE Std 1159™-2009, IEEE Recommended Practice for Monitoring Electric Power Quality," pp. 53, 2009.
- [29] IEEE Power and Energy Society Transmission and Distribution Committee, *IEEE Recommended Practice and Requirements for*

Harmonic Control in Electric Power Systems, in IEEE Std 519-2014 (Revision of IEEE Std 519-1992). 2014.

- [30] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, 2017.
- [31] Christopher Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, pp104, 2006.



Xu Jiang (S' 17) received the BEng., MSc. and PhD degrees in electronic and electrical engineering in 2015 and 2016, and 2020 respectively, from the University of Strathclyde, Glasgow, U.K. He is currently a Research Assistant within the Institute for Energy and Environment, University of Strathclyde. His research interests include application of high-frequency anomaly detection and diagnosis, conditional monitoring, and asset management.



Bruce Stephen (M'09–SM'14) received the BSc degree in aeronautical engineering from the University of Glasgow, UK, in 1997, the MSc degree in computer science from the University of Strathclyde, UK, in 1998 and the PhD degree from the University of Strathclyde in 2005. He currently holds the post of Research Fellow within the Institute for Energy and Environment at the University of Strathclyde, UK. His research interests include power system condition monitoring, renewable integration, characterizing low voltage network behavior and characterization of demand in electrical distribution networks. Dr Stephen is a Chartered Engineer and a Fellow of the Higher Education Academy.



Prof Stephen McArthur (M'93, SM'07, F'15) received his B.Eng. (Hons) and PhD degrees from the University of Strathclyde in 1992 and 1996 respectively. He is a Distinguished Professor of Intelligent Energy Systems in the Institute for Energy and Environment, within the Department of Electronic and Electrical Engineering. His research concerns the application of AI across a broad range of energy systems and smart grid applications.