

# Enabling Grid Stability: Harnessing $\mu$ PMU Data for Data-Driven Analysis of Grid Frequency Events <sup>\*</sup>

Uttam Singh<sup>1</sup>, Maitreyee Dey<sup>1</sup>, and Preeti Patel<sup>1</sup>

Department of Computer Science and Applied Computing, London Metropolitan  
University, 166-220 Holloway Rd, London N7 8DB, UK  
uts0028@my.londonmet.ac.uk, m.dey@londonmet.ac.uk,  
p.patel@londonmet.ac.uk

**Abstract.** The surge in Distributed Energy Resources connected to the power grid has significantly heightened challenges for grid operators. While the integration of Micro-synchrophasor Unit ( $\mu$ PMU) into the power grid infrastructure has facilitated the acquisition of high-resolution data, analyzing such vast datasets presents its own set of challenges. This paper offers insights into fifty days of  $\mu$ PMU data through data analysis techniques focusing on frequency events within the operational and hazardous ranges. Drawing from fifty days of real-world data collected from grid-connected solar farm in England, this paper examines and analyzes multiple frequency events while enhancing comprehension of their diverse patterns. Additionally, this study has explored the Isolation Forest Model for anomalous event detection, with results validated against power quality data.

**Keywords:** Power Grid · Data Analysis · Event Detection · Micro-synchrophasor data · Unsupervised Learning.

## 1 Introduction

Due to the UK government's emphasis on achieving the goal of Net Zero carbon emissions [1] by 2050, the growing number of Distributed Energy Resources (DERs) incorporation has significantly increased the complexity of the electrical network. Previously, there existed a clear distinction between electricity producers and consumers, making it easier for grid operators to investigate and monitor power failures. However, with the emergence of DERs, this clear demarcation between producers and consumers has blurred. During periods of power surplus at the consumer end, they transition into power producers, supplying electricity to the power grid, thereby adding more complexity to the network grid.

The proliferation of DERs has underscored the need for robust and reliable power network monitoring systems capable of continuously overseeing the electrical

---

<sup>\*</sup> The authors would like to extend their sincere appreciation to Neuville Grid Data for their generous provision of the data, which greatly contributed to the successful completion of this research project.

grid for any eventuality. Such systems must also furnish grid operators with high-quality data to facilitate comprehensive root-cause analyses for both minor and major events affecting grid stability.

The challenge of constant network monitoring was partially resolved through the introduction of Micro Phasor Measurement Units ( $\mu$ PMUs) [2]. These units offer exceedingly high-resolution data, with a standard  $\mu$ PMU typically linked to either single- or three-phase distribution circuits for measuring voltage and current phasors' magnitudes and phase angles, all accurately timestamped with GPS precision. Operating at a rate of 100 to 120 readings per second (depending on the network's frequency),  $\mu$ PMU devices not only yield data at a remarkably granular level but are also capable of accumulating highly precise electrical phasor measurements, ideal for monitoring network event behaviour and supporting various other applications [3].

However, getting swamped with these data can be challenging for the grid operators. As processing and analysing enormous data requires powerful system for processing along with efficient analytical models for driving meaningful insights. It is often the case that statistical and analytical models which work well with small datasets do not perform as expected when the datasets are large [4], [5]. Also, the  $\mu$ PMU data lacks event information [6], limiting the use of supervised machine learning models. Additionally, the diversity of distribution network events makes the categorisation of different frequency events even more challenging [7]. Researchers are trying to label data by understanding different signatures and with the help of domain experts [6]. Also, there is a lack of accurate models for detecting events in the electrical grid network [8].

## 2 Literature Review

Many studies have concentrated on exploring  $\mu$ PMU data for various electrical event detection and diagnosis purposes. A few of them have focused on statistical methods, while others have used machine learning; however, the majority of studies have used a data-driven approach in conjunction with machine learning. In, [9] a data-driven approach utilizing median absolute deviation for labeling the different types of events has been used. Then classification machine learning algorithms such as Support Vector Machine (SVM), k-Nearest Neighbours (kNN), and Decision Tree had been used to classify the events based on event signature. In [6], the author used data driven power quality rules to first identify the regions of interest and then utilized unsupervised machine learning algorithm, namely multiple K- Means Clustering with voting strategy, to cluster the events. They have also attempted to categorize different events based on event signatures into six event categories. Similarly, in [10], researchers explored  $\mu$ PMU data for identifying Subsynchronous Oscillations (SSO) using data-driven methods along with multiple Support Vector Machines (SVMs). Additionally, [11] used a mixed integrated approach combining statistical methods with machine learning for identifying grid-induced and locally induced events for current and voltage measurements using solar  $\mu$ PMU data. Also, in [12] the author employed unsupervised

machine learning based on a Generative Adversarial Network with minimal human intervention to detect events in current, voltage, and power, using deep neural networks.

Continuing from the previous literature, this paper delves deeper into fifty days of  $\mu$ PMU data collected at a solar farm in England to explore patterns and similarities in frequency events. Extensive data analysis techniques have been employed which has aided in understanding the data, identifying various patterns in frequency events, and attempting to categorize events based on different event signatures. An event here refers to any occurrence causing the frequency of the power grid network to deviate from the nominal frequency of 50 Hz, as any fluctuation exceeding  $\pm 1\%$  can potentially result in significant damage to generators [13]. Subsequently, the study extends its focus to the Isolation Forest Machine Learning method [14] for anomaly detection, known for its effectiveness in unsupervised anomaly detection. The model is trained on 30 days of  $\mu$ PMU data, comprising 8.64 million data points per day, and tested on 20 days of unseen data. Finally, the captured anomalous events detected by the model are verified against power quality events provided through the Grid Data Unit (GDU), a device which captures Power Quality Data, thus ensuring the reliability and accuracy of the anomaly detection process.

### 3 Proposed Methodology

This research comprises of two major steps: (1) the first step is understanding and processing large amounts of real-time  $\mu$ PMU data from a solar farm located in England and understanding the relationship between different electrical parameters (voltage, current, and frequency); (2) the second step is analysing and detecting abnormal frequency events from the  $\mu$ PMU data.

#### 3.1 Data Gathering

The  $\mu$ PMU has been gathered by Neuville Grid Data (NGD) - Small-Medium Enterprise (SME) working extensively in the Power Sector to improve the Electrical Network Capabilities- using their patented hardware device: Grid Data Unit (GDU) [15]. The GDU includes a  $\mu$ PMU, Power Quality Monitor (PQM), Global Positioning System (GPS) based time-synchronisation to sub-100 ns, data-buffering memory, edge computing, and secure bidirectional 3G/4G cellular data telemetry. NGD has commissioned GDUs on five utility-scale solar sites across the England (two 33 kV and three 11 kV sites), connected to the UK Power Network (UKPN) for condition monitoring of those solar sites [16].

#### 3.2 Frequency Events Identification

The UK's nominal grid frequency is 50Hz, thus National Grid is obliged to control the frequency range within  $\pm 1\%$  of 50Hz so it can fluctuate between 49.5Hz to 50.5Hz [17]. However, the operational limits are 49.8Hz to 50.2Hz

[18]. The collected frequency data from the GDU has been further analysed based on the above condition. When the frequency is beyond the operational range (49.8 Hz to 50.2 Hz) is considered as an operational event (OE) and when the frequency goes beyond (49.5 Hz to 50.5 Hz) is considered as a hazardous event (HE). The event duration and the event patterns during the event are also recorded for further investigation to determine the source of the event. Based on the grid frequency limits, this research focuses on identifying different patterns of both the frequency events based on operational range and serious hazardous range.

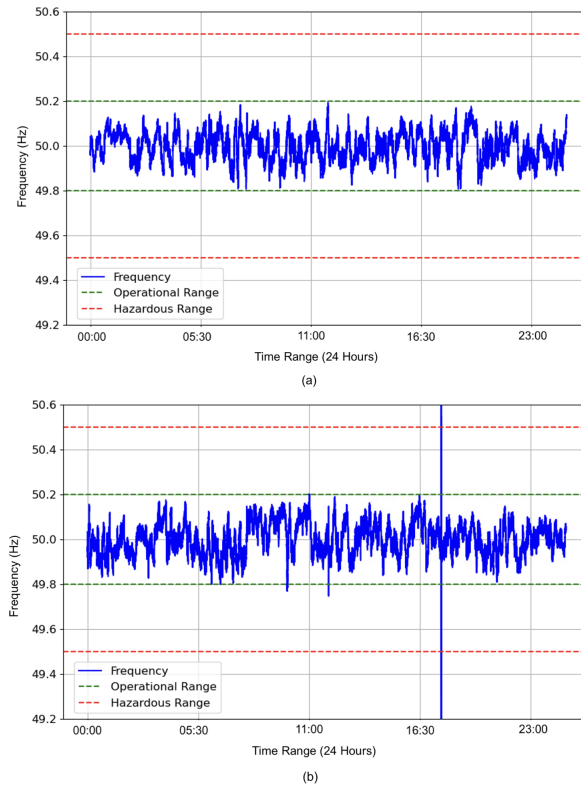


Fig. 1: Plot of Frequency for full day (a) for a normal day on April 29th, 2023, without any frequency events. (b) for a hazardous event day of May 05th, 2023.

Fig. 1 shows daily frequency graphs, where the red dotted lines represent the hazardous event range and the green dotted lines represent the operational event range. Fig. 1(a) shows the data of 29th April 2023, when there is no event has been captured. Whereas Fig. 1(b) shows the data of 5th May 2023, when a

hazardous event was noted at 17:44 hours, with the frequency fluctuating in the range from 48.03 Hz to 51.47 Hz

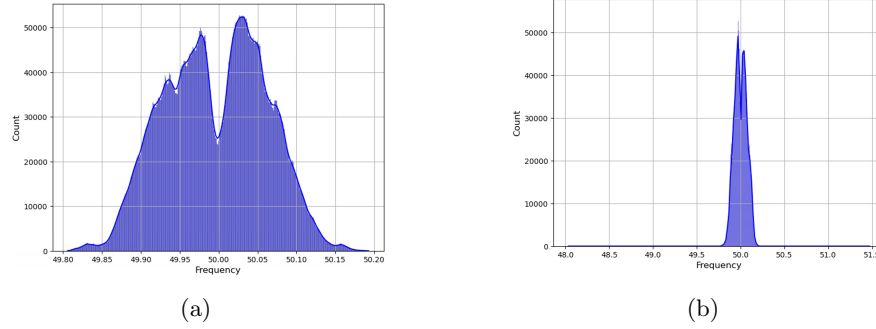


Fig. 2: Distribution plot of Frequency (a) for a normal day of April 29th, 2023 (without any frequency events) (b) for a hazardous event on May 05th, 2023

Fig. 2 shows the distribution of frequency data for the same days referenced in Fig. 1. Furthermore, Fig. 2(a) shows the distribution plot of frequency data from April 29, 2023, during which no events were recorded. It is evident from the distribution plot that it follows a normal distribution pattern with no skewness on the either sides, adhering to the empirical rules of normal distribution. Conversely, Fig. 2(b) shows the distribution plot of frequency data from May 5, 2023, when a hazardous event and multiple operational events were recorded by the  $\mu$ PMU. From the distribution plot of 5th May, it is apparent that the data is more dispersed, exhibiting longer tails attributed to the presence of outlier(s) [19]. These outliers are isolated in high or low-frequency values which could be the result of sudden spikes in renewable energy generation, unexpected power generation failure, or equipment malfunctions. To comprehensively grasp the nature of these events, further investigation into the frequency fluctuations, including their patterns and durations, is necessary. This aspect is elaborated upon in section 4.2 of this paper.

### 3.3 Anomaly Detection Using IFM

An anomaly is a deviation from the standard behavior that is observed within a dataset. Anomalies are indicators of potential issues, or significant events which require additional investigation. In this context, our aim is to identify abnormalities/discrepancies within the grid frequency data that were collected by the GDU, potentially indicating faults. Various methods had been used for anomaly detection, including statistical approaches, machine learning algorithms, and domain-specific techniques as in [16], [17], and [18].

Isolation Forest (IF) is an unsupervised machine learning algorithm for anomaly detection introduced by Liu et al. in [20]. The Isolation Forest Model (IFM) is

based on the idea that anomalies are generally “few and different” [20], it uses this core idea to isolate anomalies through constructing ensemble of binary trees or Isolation Tree (iTree). Due to their tendency to stand out, anomalies are prone to be singled out nearer to the root of an iTree, while regular points are typically isolated towards the deeper parts of an iTree. Anomalies are detected in the IFM using anomaly scores which are based on the average path length in the ensemble of trees and the average path length for an unsuccessful search in a binary tree (as mentioned in the Eq. 1). For anomalous data points the anomaly scores are negatives and for regular data points it is on the positive sides.

The equation representing the basic principle of the Isolation Forest algorithm is as follows:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

Where,

- $s(x, n)$  represents the anomaly score for data point  $x$  in a dataset of size  $n$
- $h(x)$  denotes the path length of data point  $x$  in the tree
- $E(h(x))$  is the average path length in the ensemble of trees.
- $c(n)$  is the average path length for an unsuccessful search in a binary tree.

It is important to note here that IFM works only to detect the anomalies in the frequency data, it does not categorise data into different types of anomalies such as OE(s) or HE(s).

## 4 Result Analysis and Discussion

This proposed method has been investigated using  $\mu$ PMU data collected from one utility-scale solar site located in the Norfolk, England connected with 33kV UK power network [16].

### 4.1 Data Description

There are 8.64 millions of frequency datapoints collected in every 10ms per day from the above mentioned solar site, providing two sets of phasor-based measurements per AC cycle (100 samples per second) data reporting rate. Thus the resolution of data is 100 readings per second. To cover as many different frequency events as possible, fifty days of  $\mu$ PMU data is selected for the analysis starting from 1st April 2023 until 20th May 2023. In total we analysed 4.2 billions of frequency data points spread across 50 days.

The complete data is first analysed for detecting OE and HE based on their fluctuations from the nominal frequency of 50 Hz on the basis of Operational Range (OR) and Hazardous Range (HR). The timestamp of the OE and HE, duration of the events and the total number of OE and HE are recorded for the further investigation. Individual OE and HE were also analysed for their fluctuation patterns. To obtain a deep understanding of the frequency data it is

analysed for its distribution and other statistical measures such as skewness and kurtosis. The recorded frequency data and events provided by the  $\mu$ PMU are also examined through graphs to identify the presence of events and understand patterns.

Once all the events have been analysed, the IFM has been chosen as an unsupervised machine learning method to detect anomalies that were recorded as an event and verified with the power quality data.

## 4.2 Frequency Pattern Analysis

For frequency pattern analysis, frequency data for all fifty days has been explored to identify any hazardous events (HEs) or operational events (OEs). The days on which any HE or OE is found are further examined to understand distinct patterns during the time of the event. Additionally, the frequency pattern during a normal day is compared with the corresponding pattern on an event day to further understand the type of fluctuations during an event. The detailed analysis is as follows.

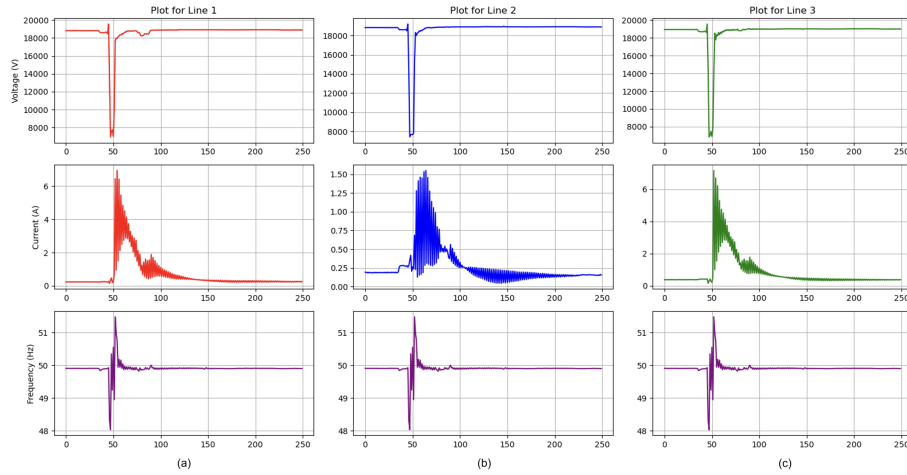


Fig. 3: Plot of Voltage and Current Magnitude during the Hazardous Frequency Event on May 05 at 15.52 for all three lines.

As previously shown in Fig. 1, the frequency pattern for a normal day differs from that of an event day. For further analysis of the patterns in frequency data, the dates April 29, 2023, and May 5, 2023, have been considered in this study, representing a normal day and an event day, respectively. Fig. 1(a) and Fig. 1(b) show the frequency as recorded by  $\mu$ PMU for a normal day (April 29th, 2023) and a day with a frequency event (May 5th, 2023), respectively. Multiple

OEs were observed on May 5th, where the frequency touched the green line representing the operational range, while on April 29th, the frequency remained within the green lines. Additionally, a sudden spike in frequency was observed at 17:44 hours, where the frequency plot crossed the hazardous range, represented by the red line in Fig. 1(b), on May 5th.

To further understand the fluctuation pattern during a day with a HE, frequency data for all days with HE have been explored. Among all the six days when HE(s) were observed, May 5th and May 15th have been selected in this study to illustrate the different patterns during the event. Fig. 3 shows the hazardous frequency event occurring on May 5th, 2023, and its effects on the associated electrical parameters; 3-phase (line-1, line-2 and line-3) voltage and currents, as these are also pivotal parameters to monitor electrical grid stability. Fig. 3(a) shows the line-1 voltage and current patterns along with the frequency patterns. There are significant pattern changes in the voltage and currents during the frequency event. Similarly, in Fig. 3(b), the current and voltage magnitude of the line-2 show a sudden voltage dip and current oscillation. Fig. 3(c) shows similar voltage dip and current oscillations for the line-3. However, of all three lines, line-2 current magnitude shows more sustained oscillations during the HE. Moreover, for all three lines, voltage magnitude shows nearly similar patterns of voltage dip in terms of level of dip and time duration of the event.

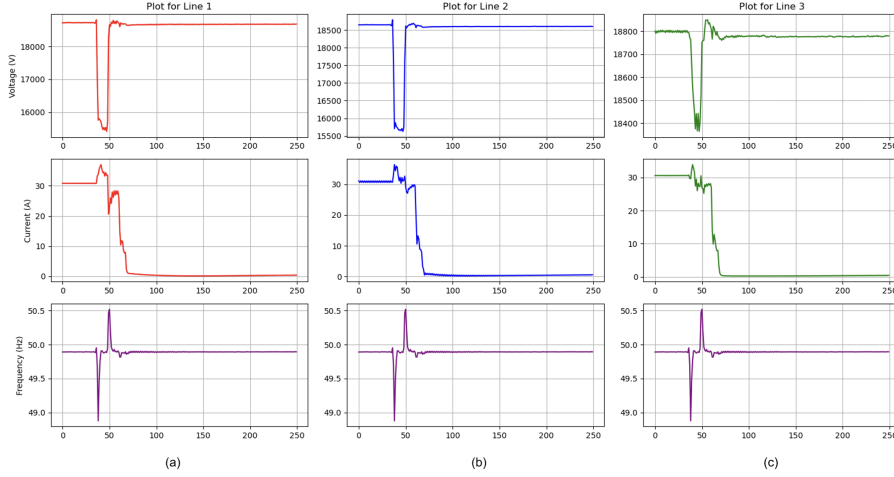


Fig. 4: Plot of Voltage and Current Magnitude during the Hazardous Frequency Event on May 15 at 03:51 for all three lines.

Further, Fig. 4 shows the time window when the hazardous frequency event occurred on 15th May 2023 along with its effects on the associated 3-line's current and voltage magnitudes. Fig. 4(a) shows the line-1 current and voltage

patterns during the event window along with the frequency. A sudden dip is seen in the voltage magnitude during the event while the current drops to zero. Similarly, Fig. 4(b) and Fig. 4(c) show similar patterns in the current and voltage magnitude for line-2 and line-3, respectively, with a sudden dip in the voltage magnitude while the current magnitude drops to zero.

Furthermore, upon investigating Hazardous Events for similarities in their patterns, distinct behaviors were observed, suggesting potentially different sources or origins of events. This phenomenon is evident in Fig. 3 and 4. While Fig. 3 shows oscillations in current magnitude alongside a sudden dip in voltage magnitude during HE, Fig. 4 shows minimal to no oscillations in current magnitude, with it dropping to zero during the event. Additionally, when comparing voltage magnitudes for May 5th and May 15th during HE, a notable difference is observed: the voltage dip on May 5th is substantial, around 10,000 Volts, compared to approximately 2000 Volts on May 15th. This observation further corroborates our initial hypothesis.

From overall observation of 50 days of data it is found that frequency fluctuates within the operational range more often, while hazardous events are less. In fifty days of data that has been analysed 48 days of frequency was recorded in the operational range while 6 days was recorded in the serious hazardous range by the  $\mu$ PMU. The duration during which the overall network frequency remained within the operational range is longer compared to its duration within the hazardous range, ranging from 70 milliseconds to as large as a *few minutes* for the operational range, and from 10 milliseconds to 80 milliseconds for the hazardous range. Furthermore, the highest number of frequency readings in the operational range were recorded on 6th April 2023 with as many as 41436 readings. While the highest number of readings in the hazardous range were recorded on 5th May with 8 readings.

It is to be noted that all these events that have been explored here in this study have been verified through PQM data.

### 4.3 IFM result analysis

This study also employs the Isolation Forest [14] Machine Learning method for anomaly detection which is one of the most popular and used unsupervised methods for anomaly detection.

The IFM is implemented to detect these events automatically, as the data does not have any prior information about the events. The IFM used 50 days of  $\mu$ PMU data for training and testing the model. While training the IFM model, the contamination value was chosen to be 0.02 which was experimentally found to be the most accurate for anomaly detection. Contamination refers to the proportion of anomalies or outliers present in a dataset relative to the total number of data points. As contamination refers to the percentage of anomalies within the data, it can range from zero to 100 percent. In this study, it was experimentally observed that varying the contamination score within the range of 0.02 to 0.00049 does not significantly alter the results. However, increasing the contamination value beyond 0.02 adversely affects the model performance, as

it begins to classify normal data points as anomalies. Therefore, contamination value has been kept at 0.02 for the final model and the results are shown in Fig. 5. It shows the IFM results for two different days. Fig. 5 (a) shows the IFM results for 5th May 2023 and Fig. 5(b) for 15th May 2023. Both days have HEs which are successfully captured by the IFM and are clearly seen in Fig. 5(a) and (b). The anomalous data points are shown in dark blue colour, and the anomaly score for anomalous data points for both days was -0.05 indicated by the colour bar, which signifies the existence of outliers within the data. The model has been trained on multiple days of data comprising millions of data points. The trained model has been tested on new days of unseen data and the captured anomalous events by the model were verified against the power quality events provided through the GDU.

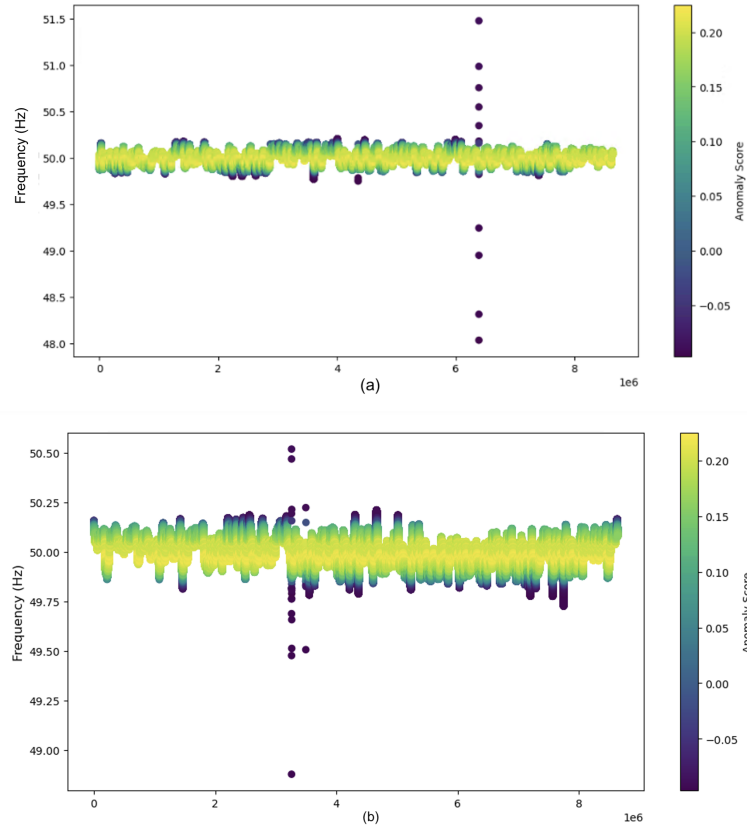


Fig. 5: Plot of IFM-based anomaly detection for two hazardous frequency events.

The proposed methodologies in this paper have showcased their capability in effectively extracting valuable insights concerning diverse frequency events, shedding light on patterns observed during both operational and hazardous frequency events. Furthermore, the Isolation Forest Model (IFM) has exhibited remarkable robustness in accurately identifying anomalous events, underscoring the effectiveness of machine learning approaches in augmenting the monitoring and analysis of electrical grid systems. In a comparative analysis with power quality data, the IFM not only successfully detected all frequency events as recorded in the power quality data but also identified additional anomalous data points. This comprehensive coverage suggests the versatility and reliability of the IFM in discerning subtle deviations and irregularities within the grid, thus offering a powerful tool for enhancing grid monitoring and management practices.

## 5 Conclusion And Future Work

This study for auto-detection of frequency events has demonstrated the efficacy of utilizing  $\mu$ PMU-generated data and employing data exploration techniques. Through this analysis, insights into the occurrence and characteristics of both Operational Events and Hazardous Events have been obtained, bolstering system monitoring capabilities in the long term. Furthermore, the integration of outlier and Hazardous Events awareness facilitates a deeper investigation into event occurrences and durations, thereby enhancing system monitoring capabilities. Additionally, the study has revealed the successful application of the Isolation Forest Model (IFM) in detecting anomalies within the frequency data. This underscores the potential of employing anomaly detection techniques for enhancing system reliability and performance.

Future research endeavours will focus on the application of classification techniques following auto-data labelling based on operational and hazardous events. This will necessitate working with larger datasets to ensure an adequate number of data points for each event class, thereby enabling the development of effective classification models. Through this research work, further advancements in identifying the root cause of anomaly and system monitoring capabilities can be achieved, ultimately contributing to the improvement of power grid operations and reliability.

## References

1. UK Government: Net zero: The uk government's emissions reduction plan (2020), <https://assets.publishing.service.gov.uk/media/6569cb331104cf000dfa7352/net-zero-government-emissions-roadmap.pdf>
2. Von Meier, A., Culler, D., McEachern, A., Arghandeh, R.: Micro-synchrophasors for distribution systems. In: ISGT 2014. pp. 1–5. IEEE (2014)
3. Von Meier, A., Stewart, E., McEachern, A., Andersen, M., Mehrmanesh, L.: Precision micro-synchrophasors for distribution systems: A summary of applications. IEEE Transactions on Smart Grid **8**(6), 2926–2936 (2017)

4. Bhadani, A.K., Jothimani, D.: Big data: challenges, opportunities, and realities. Effective big data management and opportunities for implementation pp. 1–24 (2016)
5. Zhou, Y., Arghandeh, R., Konstantakopoulos, I., Abdullah, S., von Meier, A., Spanos, C.J.: Abnormal event detection with high resolution micro-pmu data. In: 2016 Power Systems Computation Conference (PSCC). pp. 1–7. IEEE (2016)
6. Lan, T., Lin, Y., Wang, J., Leao, B., Fradkin, D.: Unsupervised power system event detection and classification using unlabeled pmu data. In: 2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe). pp. 01–05. IEEE (2021)
7. Zhou, Y., Arghandeh, R., Konstantakopoulos, I., Abdullah, S., von Meier, A., Spanos, C.J.: Abnormal event detection with high resolution micro-pmu data. In: 2016 Power Systems Computation Conference (PSCC). pp. 1–7 (2016). <https://doi.org/10.1109/PSCC.2016.7540980>
8. Rodrigues, N.M., Janeiro, F.M., Ramos, P.M.: Deep learning for power quality event detection and classification based on measured grid data. IEEE Transactions on Instrumentation and Measurement **72**, 1–11 (2023). <https://doi.org/10.1109/TIM.2023.3293555>
9. Shahsavari, A., Farajollahi, M., Stewart, E.M., Cortez, E., Mohsenian-Rad, H.: Situational awareness in distribution grid using micro-pmu data: A machine learning approach. IEEE Transactions on Smart Grid **10**(6), 6167–6177 (2019)
10. Liu, H., Qi, Y., Zhao, J., Bi, T.: Data-driven subsynchronous oscillation identification using field synchrophasor measurements. IEEE transactions on power delivery **37**(1), 165–175 (2021)
11. Khaledian, P., Mohsenian-Rad, H.: Automated event region identification and its data-driven applications in behind-the-meter solar farms based on micro-pmu measurements. IEEE Transactions on Smart Grid **13**(3), 2094–2106 (2022)
12. Aligholian, A., Shahsavari, A., Cortez, E., Stewart, E., Mohsenian-Rad, H.: Event detection in micro-pmu data: A generative adversarial network scoring method. In: 2020 IEEE Power & Energy Society General Meeting (PESGM). pp. 1–5. IEEE (2020)
13. Cheng, Y., Azizpanah-Abarghoee, R., Azizi, S., Ding, L., Terzija, V.: Smart frequency control in low inertia energy systems based on frequency response techniques: A review. Applied Energy **279**, 115798 (2020)
14. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 eighth IEEE international conference on data mining. pp. 413–422. IEEE (2008)
15. SIMMONS, C.V.: Methods and apparatus for the sensing, collecting, transmission, storage, and dissemination of high-resolution power grid electrical measurement data (Jul 1 2021), uS Patent App. 17/057,611
16. Dey, M., Rana, S.P., Simmons, C.V., Dudley, S.: Solar farm voltage anomaly detection using high-resolution  $\mu$ pmu data-driven unsupervised machine learning. Applied Energy **303**, 117656 (2021)
17. Dey, M., Rana, S., Wylie, J., Simmons, C.V., Dudley, S.: Detecting power grid frequency events from  $\mu$ pmu voltage phasor data using machine learning. In: 11th International Conference on Renewable Power Generation-Meeting net zero carbon (RPG 2022). vol. 2022, pp. 125–129. IET (2022)
18. Dey, M., Wickramarachchi, D., Rana, S.P., Simmons, C.V., Dudley, S.: Power grid frequency forecasting from  $\mu$ pmu data using hybrid vector-output lstm network. In: 2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE). pp. 1–5. IEEE (2023)

19. Moore, D.S., McCabe, G.P.: Introduction to the practice of statistics. WH Freeman/Times Books/Henry Holt & Co (1989)
20. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation-based anomaly detection **6**(1) (2012). <https://doi.org/10.1145/2133360.2133363>, <https://doi.org/10.1145/2133360.2133363>