UNIVERSITY OF CALIFORNIA
RIVERSIDE

Data-Driven Analysis and Dynamic Modeling of Inverter-Based Resources Under Grid Disturbances Enabled by Automated Event Region Identification

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Palberz Khaledian

June 2022

Dissertation Committee:

Dr. Hamed Mohsenian-Rad, Chairperson
Dr. Matthew J. Barth
Dr. Hyoseung Kim

The Dissertation of Palberz Khaledian is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to thank my supervisor Dr. Hamed Mohsenian-Rad, whose expertise was invaluable in formulating the research questions and methodology. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would like to thank Dr. Matthew J. Barth and Dr. Hyoseung Kim for their support and being committee members for my dissertation evaluation and defense. Also, I would like to thank my project collaborators and coauthors for their effective participation in my research.

Finally, my sincere gratitude to my family members and my friends for their love and support.

The content of this dissertation, in part, is a reprint of the materials that have appeared in the following publications:

1. **P. Khaledian**, A. Aligholian and H. Mohsenian-Rad, "Event-Based Analysis of Solar Power Distribution Feeder Using Micro-PMU Measurements," 2021 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2021, pp. 1-5 [1].

2. **P. Khaledian** and H. Mohsenian-Rad,"Automated Event Region Identification and Its Data-Driven Applications in Behind-the-Meter Solar Farms Based on Micro-PMU Measurements," in IEEE Transactions on Smart Grid, vol. 13, no. 3, pp. 2094-2106, May 2022 [2].

3. **P. Khaledian**, A. Shahsavari and H. Mohsenian-Rad, "Data-Driven Dynamic Response Modeling of Solar Farms under Grid Events," to be submitted to in IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (Smart Grid Comm), Singapore, June 2022.

4. **P. Khaledian**, A. Shahsavari and H. Mohsenian-Rad,"Dynamic Modeling and Model Reduction of Inverter-Based Resources under Grid Disturbances: A Data-Driven Approach with Limited Data," to be submitted to in IEEE Transaction of Sustainable Energy 2022.

This thesis is dedicated to all who has supported me throughout my education journey.

ABSTRACT OF THE DISSERTATION

Data-Driven Analysis and Dynamic Modeling of Inverter-Based Resources Under Grid Disturbances Enabled by Automated Event Region Identification

by

Palberz Khaledian

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, June 2022
Dr. Hamed Mohsenian-Rad, Chairperson

The increasing penetration of solar power generation continues to pose new challenges to the power grid; with respect to planning studies, power quality, system control, and operation. In an effort to address some of these challenges, this doctoral thesis focuses of two aspects: 1) improving the self aware capabilities of smart inverters in solar farms, and 2) analysis of solar farm's behavior and modeling its dynamic response to the grid disturbances. Research is done in three main stages to accomplish these objectives.

First, we capture power system events at solar farms, benefiting from the diagnostic application of distribution phasor measurement units (micro-PMUs) installed in the feeder head of a behind-the-meter solar farm. We determine the source region for each event, and analyze different types of events. According to the event's region, we either examine the impact of solar production level and other significant parameters to make statistical conclusions or we characterize the response of the solar farm. Our results reveal the smart inverter behavior by revealing the dynamics to the control system of the solar distribution feeder.

Second, we locate the region of the events with automatic approaches and build a foundation for *event-based situational awareness* and its *data-driven applications*. This provides knowledge of the system for not only the grid operators but also for the smart inverters to help with their self awareness to understand the impact of their operation on the power grid. Several examples of the applications of our methodology is presented.

Third, during and after grid disturbances, we model the *dynamic response* of the solar farm to the events that occur in the power grid. Accordingly, we predict the impact of the grid disturbance on the real and reactive power injections of smart inverters.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Due to the increase in *behind-the-meter* solar farms deployment in recent years [3] [4], *situational awareness in solar farms* through *distribution-level synchrophasors* measurements is of significant importance [2].

1.1.1 Behind-the-Meter Solar Farms

The term "behind-the-meter solar farms" refers to solar farm energy resources that are located behind the utility's revenue meter; thus, they are not operated by the utility. As shown in Fig. 1.1 in the highlighted area. There is an increasing trend of "behind-the-meter solar farms" integration to the utility grid in California and elsewhere. For example, three large behind-the-meter solar farms are currently operating in Riverside, CA, ranging from

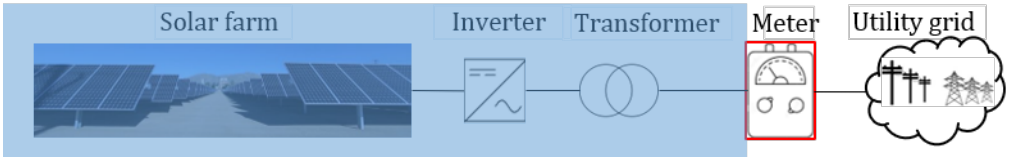

Figure 1.1: A behind-the-meter solar farm highlighted in the blue color.

3.2 MW to 7.3 MW [4]. Due to the more restricted environmental requirements, we even expect more deployment of "behind-the-meter solar farms" in future.

1.1.2 Situational Awareness in Solar Farms

Situational awareness in behind-the-meter solar farms is attaining sufficient knowledge about the behavior of the soar farm in different operational conditions and grid situations. This is an essential and challenging task in power systems [5]. If it is done right, the resultant awareness can be highly beneficial to both the utility and the operators of the behind-the-meter energy resources [6]. Of course gaining this knowledge requires to first have proper monitoring of these inverter-based distributed and renewable energy resources, which is possible through high resolution measurements that will be discussed in the next subsection. The situational awareness and its applications is strongly interlinked to the availability of measurements and the observability of the power distribution system [7]. Operators should be provided with the data that they need to comprehend the state of the system and expect the system's future behavior [8]. It is critical to understand solar farms influence on electrical system in order to recognize critical interaction and junction point of the system [9]. Similarly the power system has a direct impact on the solar farm real and reactive power injection to the system.

1.1.3 Distribution-Level Synchrophasors at Solar Farms

Phasor measurement units (PMUs) that are used for distribution systems are distribution-level synchrophasors. These meters are also called D-PMUs and micro-PMUs (μPMUs). The high reporting rates of micro-PMUs are sufficient to capture the high dynamics of power distribution systems, which are due to the intermittent resources integration to the system [10]. These high resolution measurements created a broad range of new applications in power distribution systems. For example, synchronized measurements of voltage and current phasors at a high resolution are now possible with micro-PMUs. One common measuring rate of micro-PMUs is 120 phasor readings per second [11, 12]. This level of reporting rate requires much higher reporting rate; micro-PMUs have a sampling rate of 512 samples per cycle [13]. Obtaining both magnitude and phase angle in high resolution by micro-PMUs, our ability to analyze and scrutinize events and disturbances in the distribution system is sharpened more than ever [14]. One example of micro-PMUs measurements is shown in Fig. 1.2. In sub-figure (a) the measurements include normal fluctuations of the solar farm production due to the variation in the irradiation caused by the cloud passing. In sub-figure (b), the event in Part(a) is magnified. Such events with fast dynamic are very common in distribution system and can only be captured by the micro-PMUs. Subsequently, they can be analyzed and the states of the system can be inferred, the health of the equipment can be recognized, and the disturbances can be detected and classified [15, 16]. As we discussed in the previous subsection, micro-PMUs significantly improved the ability of achieving situational awareness in power distribution systems [17–21].

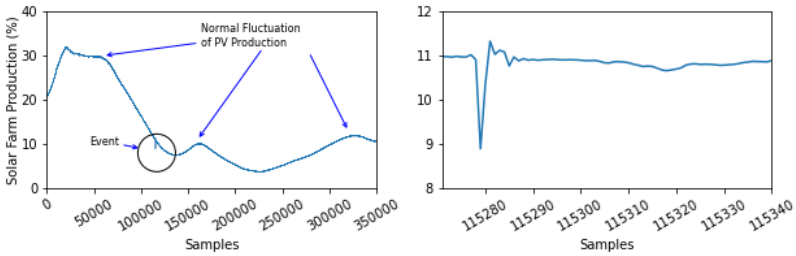

Figure 1.2: The voltage measurements during a cloudy day that are captured by a micro-PMU on a behind-the-meter solar farm: (a) behind-the-meter solar farm production with irradiation fluctuation and a single event; (b) the magnified high resolution version of the event in Part (a).

1.2 Related Literature

1.2.1 Methods for Event Region Identification

The massive investment in power distribution systems and the increased integration of renewable energy resources with their associated equipment's have influenced the power quality, grid reliability, stability, and safety issues [22]. The source of power disturbances is mostly electromagnetic transients, that are caused by lightning strokes, switching actions, self-clearing faults, and switching of end-user equipment's [23]. Hence a power system that incorporates the renewable energy sources into the utility networks is involved in challenges concerning the power disturbances [24]. Therefore, there should be intelligent approaches to detect the power quality disturbances and identify their sources in order to determine the foundations of these disturbances before any mitigation actions are engaged [25]. There are several approaches for characterizing detected events, such as techniques in [26], [27] [28], [29]. There are also some studies that address event region location identification by micro-PMUs measurements such as [30], [31]. Fault location identification based on high-impedance is studied in [32]. In [33], a stochastic high impedance fault monitoring and location scheme

using synchronized micro-PMUs data is proposed for distribution network protection. A considerable number of research is done on event source location identification; however, most of them are not related to micro-PMUs. The main methodologies are impedance-based methods, which work based on calculating the event impedance between the location of the event and the sensor location, which are mostly suitable to locate permanent faults [34], [35]. Wide-area monitoring is another group of applied methods that work by using data from several sensors across the distribution system. These methods, such as [36], [37], are mostly concerned with fault events. The wide-area methods are also employed to recognize the operation status for distributed resources, e.g., in [38] and [39], or identify the event region for major power quality events, such as in [40], and to identify islanding [41].

Various methods have considered data-driven frameworks with accurate micro-PMU measurements, such as the works in [42] and [43] that utilized an unsupervised machine learning approach for event detection and clustering. In [44] and [45], signal processing is used for detection and localization of multiple events. These studies, [42] [43] [44] [45], focus on an exhaustive training process with several micro-PMUs measurement to achieve situational awareness.

Another option in identifying the location of events based on micro-PMU measurements is the visual inspection. Even though, this option has not been discussed as an actual methodology in a formal setting, some papers occasionally utilize visual inspection method. In [12], the measurements from two micro-PMUs are used for cross-comparison to identify the local events for each micro-PMU. Further, visual inspection is used in [1, 29, 46]

to investigate the simultaneous impact of specific events on various components of a power distribution network, such as solar power inverters.

1.2.2 Methods for Dynamic Response Modeling

There are several approaches for dynamic response modeling. Two common techniques are equivalent physical model and precise physical model to represent the system dynamic characteristics [47] [48]. However, due to the high-order non-linearity of the PV stations including many internal states [49], the modeling cost is significant. Precise models makes the mathematical computation extremely complex. Thus, the equivalent models are introduced to reduce the complexity of the system and the cost of simulation, But, the equivalent might not reflect the dynamic characteristics accurately if the model ignores too many internal states. To correct the difference between the precise and equivalent models the data-driven model can be utilized. This is because the data of the dynamic process reflects the system dynamic characteristics. Nowadays, literature mainly focus on the following models: DC-DC converter [50] [51] [52], DC-AC inverter [53], static model [54], daily average power generation of PV station [55], PV array models [56] [57].

Next widely researched area is the equivalence modeling of the system dynamic process that contains PV stations. In [58], a single PV station was made to be the equivalent of a large-scale distributed PV stations. But due to the difference of the control parameters that was not considered in this model, the accuracy was not satisfactory. A simplified version of a boost converter, a two-staged PV station model, was proposed in [54], the precision was sufficient but the model ignored its grid-connected dynamic characteristics. In [59] and [60], the distributed PV stations were clustered based on the dynamic affinity

propagation. They introduced long short-term memory neural network to improve the model accuracy. Literature such as [61] [62] proposed similar dynamic model to a wind farm dynamic modeling. They built a multi equivalent model through clustering PV station by K-means algorithm. Many studies have been conducted int this area, equivalent model combined with the data-driven model, but pure data-driven modeling in solar farms is still in its infancy. Recently, the Western Electricity Coordinating Council (WECC) presented DER_A model specific to distribution-level solar farms, a positive sequence model to represent aggregated performance of DERs [63]. The WECC model parameters detailed definitions is presented in [64] and its mathematical state-space representations are presented in [65]. In [66], the DER_A model is used as distributed generator model in composite load modeling and parameter identification by deep reinforcement learning. In [67], characterization of DER_A model is done considering voltage ride through and dynamic voltage support.

1.3 Summary of Contributions

The contributions of doctoral research are as follows:

1. Using micro-PMU measurements, an event-based analysis is conducted at a solar distribution feeder that classifies the events according to their origin. This determines whether the events are locally-induced by the solar farm itself; or they are grid-induced. Several insightful observations are made; 1) most locally-induced events happen during the low solar production periods, 2) these events demonstrate more significant changes in power factor, 3) the response of the solar distribution feeder to grid-induced events are examined through comparing them to the response of a neighboring feeder to the exact same events, and 4) the event dynamics are characterized based on the control system mechanisms of the solar distribution feeder.

2. By formulating the event region identification for the behind-the-meter solar farms and employing it on real-world data, the major shortcomings of the conventional impedance-based method are identified (limited applicability, poor performance, and high sensitivity) and to address these shortcomings, a comprehensive study is conducted on a wide range of customized data-driven methods to solve the automated event region identification problem. Both statistical and machine learning methods are developed and examined on a multitude of extracted features. As a result the fundamental strengths and weaknesses in each class methods in solving the event region identification problem is identified.

3. From the learned lessons through the comprehensive analysis explained in the previous point, a new region identification method is proposed to make the best use of these various data-driven methods to significantly improve the applicability and performance of the automated event region identification in behind-the-meter solar farms. The results of the employment of the proposed model on real-world micro-PMU measurements demonstrates a significant performance improvement.

4. The outcome of the automated event region identification is utilized to build the foundation for event-based situational awareness and data-driven applications in behind-the-meter solar farms to unmask the practical value of our analysis. Moreover, for each identified region as grid-induced and locally-induced events, specific practical applications are proposed.

5. For the grid disturbances (grid-induced events), a data-driven event-based technique is developed to model IBRs' dynamic response; which is of interest for modeling dynamic response of the previously installed IBRs to provide insight of these systems to the utility; also it is critical in planning and feasibility study of solar farm -related projects.

6. The proposed methods makes use of the measurement devices that are previously installed for other purposes, e.g., micro-PMU data, instead of new sensors installation for inverters. Thus, the modeling of the dynamic response is done with limited input data of the plant and minimal extra asset.

7. A comprehensive active-learner library of models is build that constantly evolves through events that cannot be accurately estimated by the available models; a new

model will be build and add to the library for any new and unknown event. By this comprehensive active-learner library, accurate estimation of the dynamic response output in the post-event period is achieved. To achieve the best dynamic modeling, the best model is selected based on two different similarity measures, dynamic time warping and Pearson correlation. This makes sure of singling out the most accurate model in the library.

8. The active learning library will grow really fast. So, a dimension reduction technique is used to always keep the library in a reasonable size by classifying the models and choosing the best model in each class to represent the whole class.

9. The proposed dynamic modeling has several extra degrees of freedom compare to most of the data-driven approaches: 1) the lengths events can be different without impacting the accuracy or the process timing, 2) the estimation window is based on the occurrence time of the event so it is automatic and dynamic, 3) the parameter classification are also self-regulating according to the distribution of the parameters values, and 4) the estimation of the real and reactive powers are independent which creates a flexibility in the model selection process that leads to better estimation for each signal separately.

1.4 Thesis Outline

This dissertation is structured as follows:

Chapter 2 provides an in-detail analysis of the events related to the a real-world solar farm. These events are either resulted withing the solar farm, so their impact of the grid parameters are studied, or the events occur somewhere in the grid, so the response of the solar farm to these kind of events are studied. Also, the dynamic steps of the solar farm responses are revealed and related to the smart inverters control strategies.

In Chapter 3, several automatic approaches are developed to locate the region of the events which builds a foundation for *event-based situational awareness*. The performance of the proposed approaches are compared in details. The *data-driven applications* of the automated region identification are explained and several examples of the applications of our methodology is presented.

Chapter 4 introduced dynamic response modeling of the solar farms using single model, multi models, their combination, and the reduced models. The dynamic outputs of the models, real and reactive powers, are estimated and compared against the actual dynamic responses. The performance of each method is evaluated and the evolution of the models are presented.

In Chapter 5, the summary of the works in this doctoral research and the potential future works are presented.

Chapter 2

Event-Based Analysis of Solar Power Distribution Feeder Using Micro-PMU Measurements

2.1 Introduction

As the penetration of solar power generation continues to grow, system operators confront new challenges. Some of these challenges are introduced either by the power system *events* which have impact on solar farms; or by locally generated *events* that cause power quality issues due to the sharp drop and spike in solar power production [68].

To recognize these events and analyze their signatures and impacts, micro-PMU measurements can be of great value; given their high reporting rate of 120 phasor measure-

ments per second; and their synchronization capability. Availability of such data has enabled high-resolution event analysis by using data mining and machine learning techniques [43, 69].

2.1.1 Approach and Scope of Analysis

The study in this chapter is about a real-world solar distribution feeder that is integrated into a distribution substation. This solar distribution feeder is a large behind-the-meter solar farm, which is monitored by a micro-PMU at the distribution substation. We conduct an *event-based analysis* of the micro-PMU measurements and report and discuss our discoveries.

The events are first detected by an unsupervised machine learning method. Next, we conduct the following analysis:

1. For each captured event, we seek to first answer the following fundamental question: is the event caused by the solar farm, i.e., is it locally-induced? or is it initiated in the grid, which consequently caused a response by the solar farm, i.e., is it grid-induced? Our answer to this question is based on an impedance-based method that is applied to the differential phasor representation of each event, coupled with a signature inspection method.

2. Regarding the locally-induced events, we seek to understand their engineering implications. We observe that the majority of the locally-induced events happen during the low production periods of the solar farm. Furthermore, the events during the low production periods demonstrate more significant change in power factor.

3. Regarding the grid-induced events, we characterize the response of the solar distribution feeder to such events. We also make comparisons with the response of an *auxiliary neighboring feeder* to the same events.

4. We scrutinize multiple specific events that are particularly informative; such as by revealing the control system dynamics of the solar distribution feeder. The behavior of the solar farm is explained by the smart inverter control levels via dissecting two use cases.

The results in this study are insightful to utilities and solar power industry. They also provide new insight on the application of micro-PMU measurements in the study of behind-the-meter solar distribution feeders.

2.1.2 Literature Review

Power quality events that are associated with PV inverters in power distribution systems have been previously studied, such as in [22,70–72]. Some studies, such as in [70], use real-world measurements, while some others, such as in [22], use computer simulations. More importantly, *all* these prior studies have focused on typical *load-serving feeders*, with varying PV penetration levels. In fact, to the best of our knowledge, this study is the first data-driven event-based study of solar distribution feeders by using micro-PMU measurements.

In terms of the relevant data-driven methodologies, machine learning techniques have been already used in [69, 73–75] in order to detect power quality events in power distribution feeders. However, the previous studies have been concerned with load-serving distribution feeders. Therefore, while we used the same deep learning architecture as in [69]

for event detection in this study; we had to *train* the model with *different data* from a real-world solar distribution feeder.

Several studies have discussed identifying the source location of events in power distribution systems, e.g., in [76–78]. Here, our concern is only on whether the source of the event is the solar farm itself; or the source of the event is the grid. In case of the latter, the solar farm still responds to the event.

2.2 Background and Methodology

The test site in this study is a solar distribution feeder that is dedicated to integrate about 200 PV inverters in a 4 MW solar farm into a distribution substation. The solar farm is behind-the-meter; however, a micro-PMU is available at the feeder-head at the distribution substation that provides us the voltage and current phasor measurements of this solar distribution feeder. This feeder does not have any load and all its solar power production is injected into the distribution substation.

There is another micro-PMU that monitors a nearby feeder that contains a mix of major PV generation and major load. This *auxiliary neighboring feeder* is sometimes a *net load* and sometimes a *net generator* during the period of our analysis. Our focus in this study is *not* on this auxiliary neighboring feeder. However, in one part of our study, we use the synchronized micro-PMU measurements at this auxiliary feeder to better understand the behavior of the solar distribution feeder.

Next, we briefly overview the three key methodologies that we plan to use for our various analysis in this study.

2.2.1 Event Detection

The first step in our analysis is to detect the events from the micro-PMU measurements. This is a challenging task because events are inherently *infrequent*, *unscheduled*, and *unknown*. Hence, there is no prior knowledge about their types and time of occurrence. Accordingly, in our study, we used an unsupervised deep learning model from our previous work in [69], which implements Generative Adversarial Network (GAN) models. There are two main components in the event detection model; namely the *generator* and and the *discriminator*; which play a min-max game over the following function:

$$\begin{aligned} V(G,D) =& \mathbb{E}_x \sim p_{data}(x)[log(D(x))] + \\ & \mathbb{E}_x \sim p_z(z)[log(1 - D(G(z)))], \end{aligned} \tag{2.1}$$

where V is the objective function, G is the generator, D is the discriminator, $p_{data}(x)$ is teh distribution of the real samples, and $p_z(z)$ is the noise probability distribution function.

For the GAN model, the optimal value of the min-max game over $V(G,D)$ in (2.1) must satisfy the following two conditions:

- **C1:** For any fixed G, the optimal discriminator D^* is:

$$D^*_G(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}. \tag{2.2}$$

- **C2:** There exists a global solution such that:

$$\min(\max_D(V(G,D))) \Longleftrightarrow p_g(x) = p_{data}(x). \tag{2.3}$$

where $p_g(x)$ is the distribution of the generated sample by the generator and D^*_G is the optimal value for the output of the discriminator.

The model in [69] was trained for load monitoring; therefore, we used the micro-PMU measurements from the solar distribution feeder to train new GAN models for the purpose of event detection at the solar distribution farm. See [69] for more detail.

The dataset under study consists of measurements from both PMUs for a period of ten days. A total of 229 events are detected at the solar distribution feeder; out of which 88 events are of interest because they happened during solar production. A total of 215 events are detected at the auxiliary neighboring feeder, which 87 of them happened during solar production of the solar distribution feeder.

2.2.2 Event Region Identification

In reference to the location of the micro-PMU, an event occurs either in the *upstream* or in the *downstream* of the micro-PMU. The former is a *locally-induced* event; while the latter is a *grid-induced* event. To determine the source region of the event, we apply the following two different methods:

Impedance-based Method

For each event, we can calculate the *equivalent impedance*, denoted by Z, that is seen in the *differential mode* in the upstream of the micro-PMU:

$$Z = \frac{\Delta V}{\Delta I} = \frac{V^{\text{post}} - V^{\text{pre}}}{I^{\text{post}} - I^{\text{pre}}} \tag{2.4}$$

where I^{pre} and V^{pre} are the current and voltage phasors that are seen in the steady-state condition right *before* the event starts; and I^{post} and V^{post} are the current and voltage phasors that are seen during the steady state condition right *after* the event settles down.

Both ΔV and ΔI are *differential phasors* [78, 79]. Our focus is on the resistive component of the Z. In particular, the event is considered to be *locally-induced* if

$$\text{Real}\{Z\} > 0; \tag{2.5}$$

otherwise, the event is considered to be *grid-induced* [78]. We will use the impedance-based method in Section 2.3.

Comparison with Auxiliary Measurements

This method takes advantage of the synchronized micro-PMU measurements from the auxiliary neighboring feeder. By comparing the signatures of an event on both feeders, i.e., the solar distribution feeder and the auxiliary neighboring feeder, we can determine that the event is grid-induced if it creates similar signatures on the voltage measurements on both feeders. If the event is seen only on one feeder and there is no major signature on the other feeder, then it is a locally-induced event [12]. We will use the combination of the two methods in Section 2.4.

2.2.3 Event Dynamic

Some of the events that are captured in this study can reveal the dynamic behavior of the solar farm's control system. The control systems of the inverters on a solar distribution feeder are highly convoluted. To dissect the event dynamics, we use the following four general control components [80]:

- Current regulation loop is the fastest loop that controls the injected currant by each inverter to the grid.

- Voltage regulation loop is slower than the current regulation loop. It provides the setpoint for the current regulation loop; upon changes in inverter terminal voltage.
- Maximum Power Point Tracking (MPPT) optimizes the utilization of input power for maximum power output. For an individual inverter, this is the slowest controller.
- Plant-level controller maintains the scheduled voltage and power factor (PF) of the system by coordinating the set points of individual inverter voltage or reactive power. Plant-level control speed is coordinated with the controls of individual inverters and is normally slower.

During an event, the *major disturbances* are mostly controlled by faster controls; while *minor disturbances* are mainly controlled by the plant-level controller response. Based on the above control components, two use cases are scrutinized in Section 2.5.

2.3 Analysis of Locally-Induced Events

In this section, we study locally-induced events, make statistical conclusions, and discuss representative example events.

2.3.1 Event Correlation with PV Production Level

Our analysis of the captured locally-induced events reveals a relationship between event occurrence and the solar production level. In particular, we have observed that the majority of the locally-induced events occur during low production period.

It is worth clarifying that what we refer to as event in the micro-PMU data is very different from the regular fluctuations in the solar production level that are due to the

changes in solar irradiance. This point is explained in Fig. 1. Here, we show two example events, denoted by Event 1 in Fig. 1(a) and Event 2 in Fig. 1(b). These two events are much *smaller in magnitude* and much *shorter in duration* compared to the typical fluctuation in solar production level. These kinds of events are captured only by the installed micro-PMU. They cannot be captured by regular meters. The solar production level is 9.2% during Event 1 and 14.8% during Event 2.

Fig. 2.2 shows the scatter plot for all the captured events that are identified as locally-induced. Here the focus is on the events with Real$\{Z\} > 0$, see the methodology in Section 2.2.2. There is an inverse correlation between the production level and the number of events, which can be expressed as an exponential decay function, as presented in (2.6) and Fig. 2.2(a).

$$y = a.x^b + c, \tag{2.6}$$

where x and y are the production level and the real part of the impedance Z that we defined in (2.4). Parameters a, b, and c are obtained through curve fitting as 850, -1, and 50, respectively.

As shown in Fig. 2.2(b), about 70% of the locally-induced events happened when the solar production was at %30 or less. That means, either more control actions took place at the solar farm during low production periods; or the control actions are more impactful during such periods and therefore their signatures are more visible. In either case, these

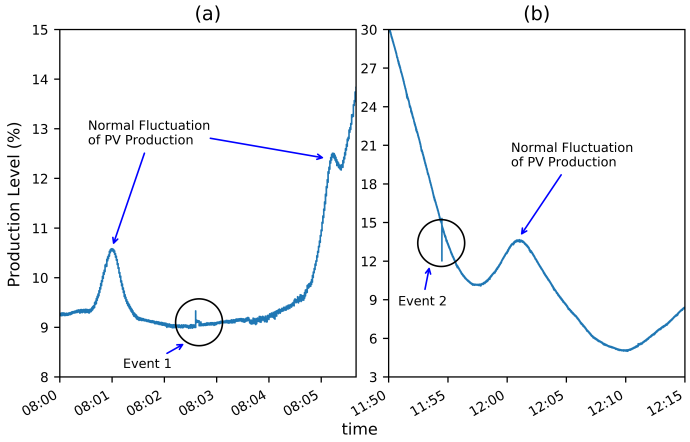

Figure 2.1: Two sample locally-induced events: (a) Event 1 shortly after production starts; (b) Event 2 occurs around noon on a cloudy day.

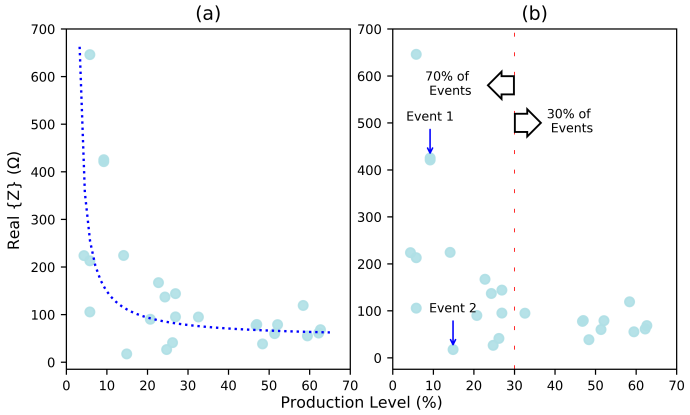

Figure 2.2: Locally-induced events on the solar distribution feeder: (a) exponential decay relation between production level and number of events; (b) inverse correlation between the PV production level and the number of events.

results highlight the importance of monitoring the operation of the solar distribution feeder during low production periods.

A closer look of Events 1 and 2 is provided in Fig. 2.3. For Event 1 in Fig. 2.3(a), we can obtain $\Delta V = 238.6 + 1789.5j$ and $\Delta I = -1.835 - 1.934j$. From (2.4), we have: Real$\{Z\} = 425.37$. Therefore, from (2.5), we can conclude that Event 1 is a locally-induced event. For Event 2 in Fig. 2.3(b), we can obtain $\Delta V = 2.8 - 29.5j$ and $\Delta I = -0.363 - 1.561j$.

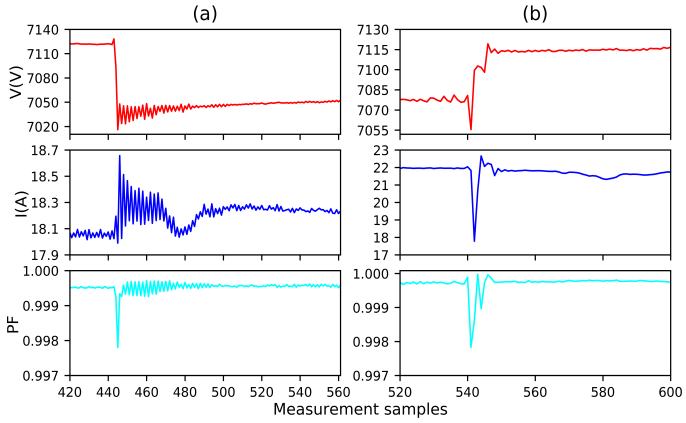

Figure 2.3: High resolution event signatures: (a) Event 1; and (b) Event 2.

From (2.4), we have: Real$\{Z\} = 17.57$. Thus, from (2.5), we can conclude that Event 2 is also a locally-induced event.

2.3.2 Changes in Power Factor

We further observed that, not only the majority of the locally-induced events occurred during low production periods, but also the events that occurred during low production periods demonstrated more significant changes in power factor (PF). In Fig. 2.4(a), we can see the change in the *phase angle difference* between voltage and current phasors that are caused by the locally-induced events versus the production level.

Note that, the cosine of the quantity in the y-axis provides the *change* in power factor. Here, θ_V and θ_I denote the phase angle measurements in voltage and in current, respectively. The changes that caused by the events in the phase angle differences *declines exponentially* with the increase in production level. See the formulation in (2.7) and Fig. 2.4(a).

$$y = d.x^e \tag{2.7}$$

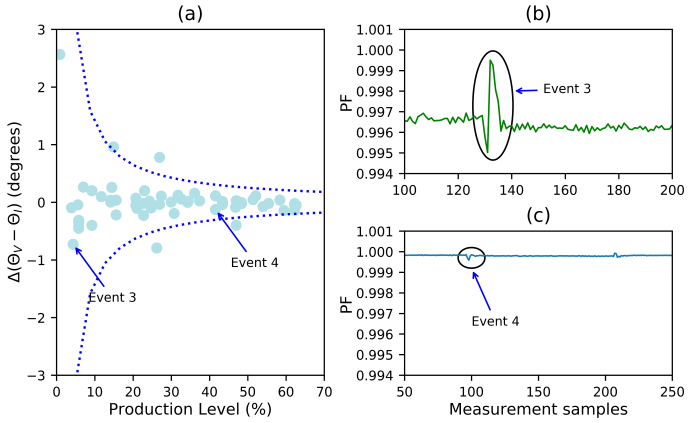

Figure 2.4: (a) Relationship between disturbance in phase-angle differences and production level; (b) small change in PF in Event 3 during low production; and (c) almost no change in PF in Event 4 during high production.

where parameters d and e are obtained through curve fitting. Parameter d is 4.5 and -4.5 for the exponential analysis and its inverse, respectively. Parameter e is obtained as -1.

Next, we look at two events which took place at different production levels and led to different changes in phase angle differences, and accordingly in PF. In Fig. 2.4(b), Event 3 happened at a low production level and caused a *major* change in PF, however in Fig. 2.4(c), Event 4 happened at a higher production level and resulted in an insignificant PF agitation.

The critical low production periods are the loss of voltage support at sunset, particularly at large PV penetration, during *startup* and *shutdown* of the inverters. Events during these periods are impacting the power system more, by affecting the phase angle and consequently the power factor of the system.

2.4 Analysis of the Response to Grid-Induced Events

Since we have access to the *synchronized* micro-PMU measurements at the auxiliary neighboring feeder, we can also apply the *signature inspection* method, described in Section 2.2.2, to see if a grid-induced event also creates a signature on the voltage phasor measurements at the auxiliary neighboring feeder.

Among all 88 events associated with the solar distribution feeder, only 8 events are recognized as grid induced events based on the Real$\{Z\} < 0$ criteria. Interestingly, *all* these eight events create signature also on the auxiliary neighboring feeder. Therefore, they are confirmed with *both* methods to be grid-induced events. Moreover, these observations verify the reliability of the impedance-based method.

The response of the two feeders to an example grid-induced event is shown in Fig. 2.5. Let us refer to this event as Event 5. Notice how the voltage suddenly increases on both set of measurements. The two responses have major differences. In particular, the event causes only a *very minor transient* change in the power factor in Fig. 2.5(a). However, the same event causes a *major steady* change in the power factor in Fig. 2.5(b). The change in the magnitude of current are also in the opposite directions in Figs. 2.5(a) and (b). It should be noted that the auxiliary neighboring feeder is a net load during this event. The behavior of the solar distribution feeder can be potentially attributed to the MPPT behavior of the PV inverters. As for the auxiliary neighboring feeder, it appears to act as a constant-impedance load; while its power factor is also affected.

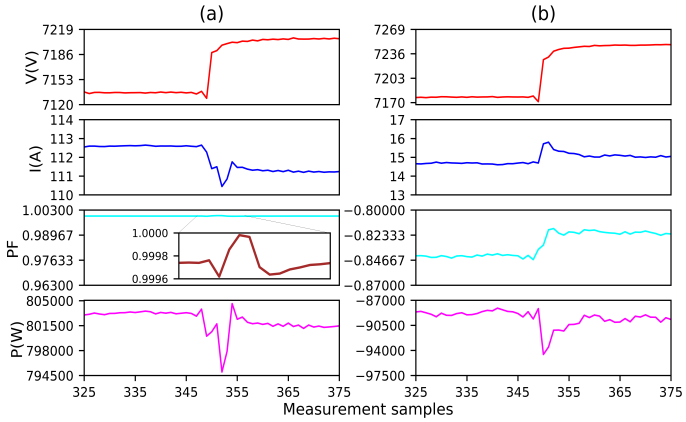

Figure 2.5: Response of the two feeders to grid-induced event: (a) response of the solar distribution feeder; (b) response of the auxiliary neighboring feeder.

2.5 Analysis of Event Dynamics with PV Inverters Operation

As discussed in Section 2.2.3, the control system in a solar distribution feeder has multiple control loops. Therefore, the dynamic response of the solar distribution feeder to a grid-induced event can constitute various stages.

In this section, we examine the dynamic response of the solar distribution feeder to two distinct grid-induced events; as shown in Fig. 2.6(a) and Fig. 2.6(b). We refer to these events as Event 6 and Event 7, respectively. The response of the solar distribution feeder to each event is broken down into *five stages* that are marked from 1 through 5 on each graph.

In Stage 1, the event occurs. Event 6 is a sudden rise voltage; see Fig. 2.6(a). is a sudden drop in voltage; see Fig. 2.6(b); thus, triggering a response by the PV control system due to detecting the changes at the inverter voltage terminals. In Stage 2, the immediate reaction of the system is to keep the output power stable; this causes a prompt current drop in Fig. 2.6(a) to decrease the dc-bus voltage; and a prompt current rise in Fig. 2.6(b) to

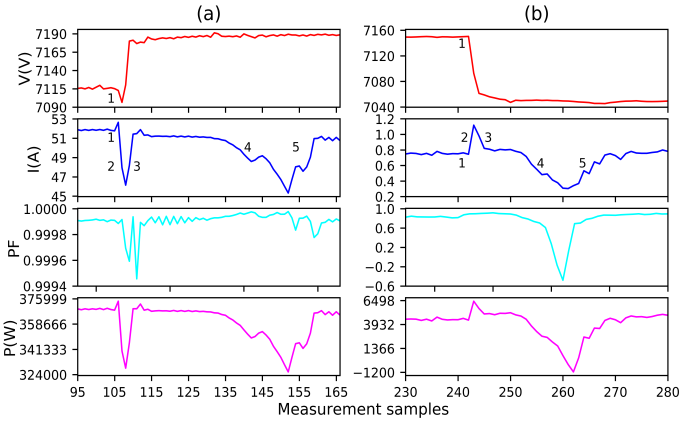

Figure 2.6: Comparing and characterizing the dynamic response of the solar distribution feeder to two different grid-induced events: (a) Event 6, which is a step-up change in voltage; and (b) Event 7; which is a step-down in voltage.

increase the dc-bus voltage. In Stage 3, the new dc-bus voltage level modifies the MPPT output; which is the input to the voltage regulation loop. Subsequently, the reference is fine-tuned for the current regulation loop. As a result, current increases in Event 6 and decreases in Event 7. Thus, the dc-bus voltage goes back to its pre-disturbance value. In Stage 4, after passing the initial transient conditions, the plant level control regulates the process and applies the ramp rate limitation. This results in a momentary decrease of the current on both events. Finally, in Stage 5, by a moderate ramp rate, the plant level controller brings back the current to the regulated set point.

Chapter 3

Automated Event Region Identification and its Data-Driven Applications in Behind-the-Meter Solar Farms Based on Micro-PMU Measurements

3.1 Introduction

Proper monitoring of behind-the-meter inverter-based distributed and renewable energy resources is an essential and challenging task in power systems [5]. If it is done

right, the results can be highly beneficial to both the utility and the operators of the behind-the-meter energy resources [6].

In this study, our focus is on monitoring behind-the-meter solar farms, which are being increasingly deployed in California and elsewhere in recent years. For example, three large behind-the-meter solar farms are currently operating in Riverside, CA, ranging from 3.2 MW to 7.3 MW [4]. As suggested by the term "behind-the-meter", the energy resources in such solar farms are located behind the utility's revenue meter; thus, they are not operated by the utility.

3.1.1 Motivation

We are interested in monitoring and scrutinizing the *events* in such systems, which are captured using distribution-level phasor measurement units (PMUs), a.k.a., micro-PMUs [12].

These events are very different from the typical fluctuations in the production level of solar generators that are due to the intermittency in solar irradiance. For example, consider the power generation level at a behind-the-meter solar farm in Fig. 3.1(a). Except for one instance, all the fluctuations in this figure *are* due to the intermittency in solar generation. However, the singleton drop that is encircled is *not* related to solar generation intermittency. It is instead an *event* that is more relevant to the operation of the inverters at the solar generator. The magnified view of this event is shown in Fig. 3.1(b). This event takes only a few hundred milliseconds. The reporting rate of the micro-PMU measurements in this figure is 120 readings per second.

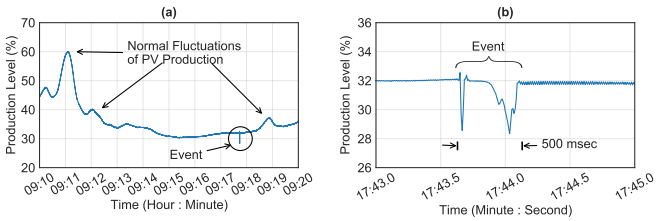

Figure 3.1: Comparing normal fluctuation in solar power generation versus the type of events that are of concern in this study: (a) solar production over a period of 10 minutes; (b) the magnified version of the event in Part (a).

Events at a behind-the-meter solar farm can be divided into two types, the events that are *caused* by the solar farm, i.e. *locally-induced* events, and the events that are caused by something else somewhere on the grid, i.e. *grid-induced* events. The first type shows the *internal* issues that may occur in the behind-the-meter solar farm. The second type shows how the solar farm *responded* to *external* disturbances. Depending on the type of an event, i.e., whether it is locally-induced or grid-induced, remedial actions might be needed.

3.1.2 Technical Contributions

It is crucial to *distinguish* the above two different types of events correctly. It is important also to develop practical use cases to take advantage of the situational awareness that we can gain from analyzing both types of events.

Addressing the above open problems is the focus of this study. Our study is based on an extensive analysis of real-world micro-PMU measurements at a 4.3 MW behind-the-meter solar farm in Riverside, CA; see Fig. 3.2. Two micro-PMUs are used in this study: $\mu\text{PMU}_{\text{PV}}$ and $\mu\text{PMU}_{\text{Other}}$; both are installed outside the premises of the behind-the-meter solar farm to be monitored by the utility. A locally-induced event has a root cause in the region on the *left* side of $\mu\text{PMU}_{\text{PV}}$; while a grid-induced event has a root cause in the

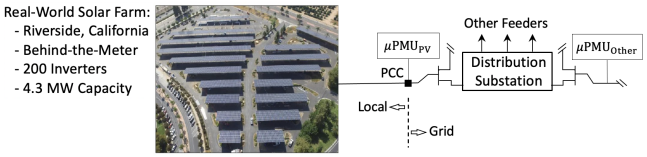

Figure 3.2: The real-world test-bed in this study: μPMU$_{\text{PV}}$ is installed at the point-of-common coupling (PCC) of a 4.3 MW behind-the-meter solar farm in Riverside, CA; μPMU$_{\text{Other}}$ is installed at the other side of the substation on another feeder to serve as a reference for certain analysis.

region on the *right* side of μPMU$_{\text{PV}}$. The main technical contributions in this study can be summarized as follows:

- The *event region identification* problem is introduced and formulated in the emerging practical context of behind-the-meter solar farms. By analyzing real-world data, we show that the conventional impedance-based method does *not* work well; as it has three major shortcomings: limited applicability, poor performance, and high sensitivity.

- To address the above shortcomings, and inspired by visual inspection and domain expert knowledge, a comprehensive analysis is conducted on a wide range of data-driven methods that are customized to solve the automated event region identification problem. Both statistical and machine learning methods (supervised and unsupervised) are examined on a multitude of extracted features. The most capable methods are identified. Importantly, this comprehensive analysis also identifies the fundamental *strengths* and *weaknesses* in each class of these methods in solving the event region identification problem.

- Built upon the *lessons learned* from the comprehensive analysis in the previous bullet point, a new method is proposed to make the best use of the *complementary characteristics* of these various data-driven methods to significantly enhance the

applicability and performance of the automated event region identification in behind-the-meter solar farms. After applying the proposed mixed-integrated algorithm to real-world micro-PMU measurements, the performance metrics are significantly improved.

- To unmask the practical value of our analysis, the outcome of the automated event region identification is utilized to build the foundation for *event-based situational awareness* and *data-driven applications* in behind-the-meter solar farms. Specific applications are proposed for the identified grid-induced and locally-induced events.

3.1.3 Literature Review

Event-based analysis of micro-PMU measurements has received increasing attention in recent years. The majority of the work in this area has focused on monitoring the utility equipment [81], load modeling [82], cybersecurity [83], state estimation [84], and stability analysis [85]. A few studies have also focused on analyzing events in solar generation units and distributed energy resources. In [86], events in micro-PMU measurements are examined to detect irregularity in the operation of PV resources. In [80], recommendations are made to utilize event data to conduct disturbance-based model verification for inverter-based resources.

Some studies are concerned with identifying the location (region) of the events that are observed in micro-PMU data. Many of the methods that are developed in this area can be broadly categorized as impedance-based methods; e.g., see [12, 32]. However, as we will see later in this chapter, impedance-based methods are not suitable to identify the region of events in behind-the-meter solar farms. As an alternative to impedance-based

methods, data-driven approaches have also received increasing attention in recent years, e.g., see [87–89]. Most of these methods focus on localizing different types of *faults*, with no or little concern about other power system disturbances. Many of these methods are also not related to power distribution systems. In [87], a combined impedance-based and data-driven method is proposed to locate faults in a power plant. In [88], an event location identification method is discussed based on k-means clustering in transmission systems. We will examine the above and other data-driven methods in the context of the event location identification problem and we will explore their weaknesses and strengths.

Visual inspection is another option in identifying the location of events based on micro-PMU measurements. However, this option has *not* been discussed in a formal setting or as an actual methodology. Nevertheless, few papers occasionally bring up visual inspection on specific examples. In [12], visual inspection is used to cross-compare the measurements from micro-PMUs on two different load feeders to identify the local events for each feeder. In [1, 29, 46], visual inspection is used to investigate the simultaneous impact of specific events, such as lightning strikes, on various components of a power distribution network, such as solar power inverters. For example, they visually verified the results from impedance-based event region identification methods.

We shall emphasize that, under the hypothetical scenario that the utility *does* have access to *inside the premises* of the behind-the-meter solar farm in order to install additional sensors, we could solve the event region identification problem by using some existing methods in the literature, either by analyzing phasor measurements as in [78] or by analyzing the waveform measurements as in [90]. However, such hypothetical scenario often

does not take place in practice; because of the nature of these behind-the-meter solar farms. In fact, not having any such access is one of the main challenges in the problem that we seek to address in this study.

Finally, some studies assume that the location of the event is *already known*; therefore, they rather focus on the applications of analyzing events with known locations. Some of the applications in this regard include power system stability [91], Volt-Var control [92], and monitoring equipment operation and state of health [93]. While these studies are not specific to behind-the-meter solar farms, they do inspire us in some of the applications that we will use in this study for our proposed automated event location identification method.

3.2 Limitations of the Conventional Impedance-Based Solution

This section briefly discusses the conventional *impedance-based* method [1, 78] to solve the event region identification problem. We apply this method to micro-PMU data from the 4.3 MW behind-the-meter PV farm in Section 3.1 to demonstrate the severe limitations of such impedance-based method.

3.2.1 Conventional Impedance-Based Method

In the impedance-based method, we examine the *event impedance* that is seen by $\mu\text{PMU}_{\text{PV}}$ to decide the region of the cause of the event. For each event, we calculate the *equivalent impedance*, denoted by Z, that is seen by $\mu\text{PMU}_{\text{PV}}$ in the *differential mode* in the

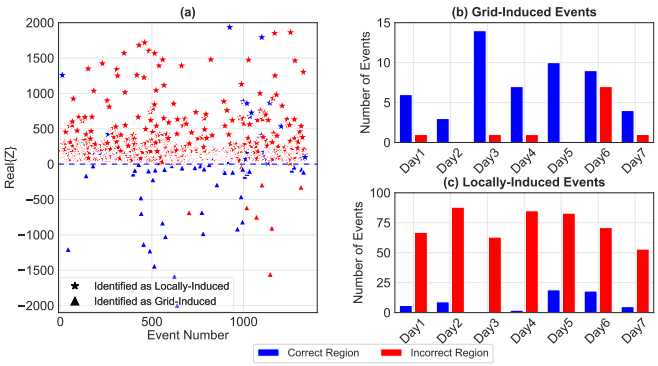

Figure 3.3: The performance of the impedance-based method on *sustained* events: (a) scatter plot for Real$\{Z_{\text{Event}}\}$ for all sustained events during one week; the colors show the correct vs. incorrect identification; the markers show the grid vs. local identification; (b) the daily summary of the results for grid-induced events; (c) the daily summary of the results for locally-induced events.

upstream of $\mu\text{PMU}_{\text{PV}}$:

$$Z_{\text{Event}} = \frac{\Delta V}{\Delta I} = \frac{V^{\text{post}} - V^{\text{pre}}}{I^{\text{post}} - I^{\text{pre}}}, \tag{3.1}$$

where I^{pre} and V^{pre} are the current phasors and the voltage phasors that are seen by $\mu\text{PMU}_{\text{PV}}$ in the steady-state condition right *before* the event starts; and I^{post} and V^{post} are the current phasors and the voltage phasors that are seen by by $\mu\text{PMU}_{\text{PV}}$ at the steady-state condition right *after* the event settles down.

Once the event impedance is obtained as in (3.1), one can use its resistive component to identify the source of the event. In particular, the event is deemed to be *locally-induced* if

$$\text{Real}\{Z_{\text{Event}}\} > 0; \tag{3.2}$$

otherwise, the event is deemed to be *grid-induced* [78].

Table 3.1: Performance Summary of the Impedance-Based Method in Region Identification of Sustained Events

		Grid	Local	Precision %	Recall %	F_1 − Score %
Reality	Grid	53	510	83	9	16
	Local	11	59	10	84	18

3.2.2 Shortcomings in Impedance-Based Method

Limited Applicability

By construction, the impedance-based method only works on *sustained* events, i.e., the events that create steady-state impact in the system. This method does *not* work on *transient* events, i.e., the events that are momentary; where the system returns to its pre-event steady-state conditions. The reason is that we cannot define Z_{Event} for a transient event; because ΔV and ΔI are both almost zero for a transient event. This fundamental shortcoming is problematic in achieving situational awareness; specially if the transient event is locally-induced and it is caused by an abnormality in the operation of the behind-the-meter PV farm.

Poor Performance

Even for the sustained events, where the impedance-based method *is* applicable, its performance is sometimes poor in practice. For example, consider the real-world results in Fig. 3.3. The impedance-based method leads to *incorrect* region identification for a large portion of events.

The performance summary of the impedance-based method for the results in Fig. 3.3 is given in Table 3.1. Precision is the ratio of the true positive to the sum of true positive and false positive. Recall is the ratio of the true positive to the sum of true positive and false negative. Accordingly, we can calculate the F_1-Scores for the impedance-based method as [94]:

$$F_1\text{-Score(Grid)} = 2 \times \frac{0.83 \times 0.09}{0.83 + 0.09} = 0.16, \tag{3.3a}$$

$$F_1\text{-Score(Local)} = 2 \times \frac{0.1 \times 0.84}{0.1 + 0.84} = 0.18. : (a) \tag{3.3b}$$

High Sensitivity

One reason for the poor performance of the impedance-based method is the difficulty in deciding *when* is the right moment to be considered as *after*. This is particularly an issue when we work with real-world measurements. If we select a moment *too early*, then the event may not have settled down yet. If we select a moment *too late*, then we may capture not only the impact of the event in question, but also the other changing factors in the system. This can affect calculating Z_{Event}; and the result of event region identification.

An example is shown in Fig. 3.4. This example is a grid-induced event, as shown in Fig. 3.4(a). From Section 3.2.1, such grid-induced event is supposed to have $\text{Real}\{Z_{\text{Event}}\} < 0$. That is the case during the time instances that are marked in green. However, for most time instances, we have $\text{Real}\{Z_{\text{Event}}\} > 0$; as shown in the red areas. The outcome of the impedance-based method would be wrong in all those red areas.

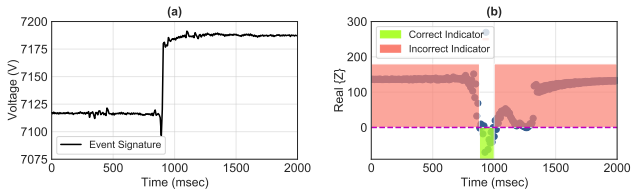

Figure 3.4: High-sensitivity of the impedance-based method: (a) the event signature is in voltage measurements for a sustained grid-induced event; (b) the value of Real$\{Z_{\text{Event}}\}$ is calculated according to various choices of the *after* time instance. The sign of Real$\{Z_{\text{Event}}\}$ fluctuates between indicating locally-induced and indicating grid-induced events.

3.3 Automatic Region Identification: Statistical and Machine Learning Methods

3.3.1 Solution Based on Human Visual Inspection

One can resolve the shortcomings of the impedance-based method by *visually inspecting* the event signature that is captured by $\mu\text{PMU}_{\text{PV}}$; and then *comparing* it with the signature of the same event that is captured by $\mu\text{PMU}_{\text{Other}}$. Recall from Section 3.1 that $\mu\text{PMU}_{\text{Other}}$ is another micro-PMU that is installed at a nearby feeder. It can serve as a *point* of reference.

Four examples are shown in Fig. 3.5. Event 1 (sustained) and Event 3 (transient) created clear signatures at $\mu\text{PMU}_{\text{PV}}$. But they did *not* create noticeable signatures at $\mu\text{PMU}_{\text{Other}}$. Thus, they must be locally-induced. Event 2 (sustained) and Event 4 (transient) created clear signatures *not only* at $\mu\text{PMU}_{\text{PV}}$ *but also* at $\mu\text{PMU}_{\text{Other}}$. Thus, they must be grid-induced. We observe that, in general, locally-induced events only create a signature on $\mu\text{PMU}_{\text{PV}}$; and they do *not* create any noticeable signature on $\mu\text{PMU}_{\text{Other}}$. Of course, if the locally-induced event is very severe, such as in the case of a *major fault* in the solar farm, then it could be possible that such event creates a noticeable signature also on

μPMU$_{\text{Other}}$. Nevertheless, the signature at μPMU$_{\text{Other}}$ would be significantly less severe than the signature at μPMU$_{\text{PV}}$. We shall add that we did not observe any such severe locally-induced event in our data set.

Other examples of visual inspection are discussed in the literature, e.g., in [1,12,46]. However, the problem with this approach is the need for constant and real-time human supervision of the micro-PMU measurements, which is cost-prohibitive. Besides, many transient events last for only a few seconds, making it practically impossible for human eyes to reason and react accordingly. Human error is another issue.

For the rest of Section 3.3, we seek to replace human visual inspection with automated data-driven methods.

3.3.2 Features to Assess Signature Similarity

Let vector $\mathbf{X}_{\text{PV}}$ denote the time series of the measurements that are obtained by μPMU$_{\text{PV}}$ during an event. For example, for the case of an event that creates a signature in voltage, vector $\mathbf{X}_{\text{PV}}$ may include the time series of the voltage measurements that are obtained by μPMU$_{\text{PV}}$ during a time window that starts a few milliseconds *before* and ends a few milliseconds *after* the event occurs. The event can be detected and captured by using the existing methods in the literature, such as by using the methods in [95]. Furthermore, let vector $\mathbf{X}_{\text{Other}}$ denote the time series of the measurements that are *simultaneously* obtained by μPMU$_{\text{Other}}$ during the *same* time window as of the event that is captured by μPMU$_{\text{PV}}$ in vector $\mathbf{X}_{\text{PV}}$.

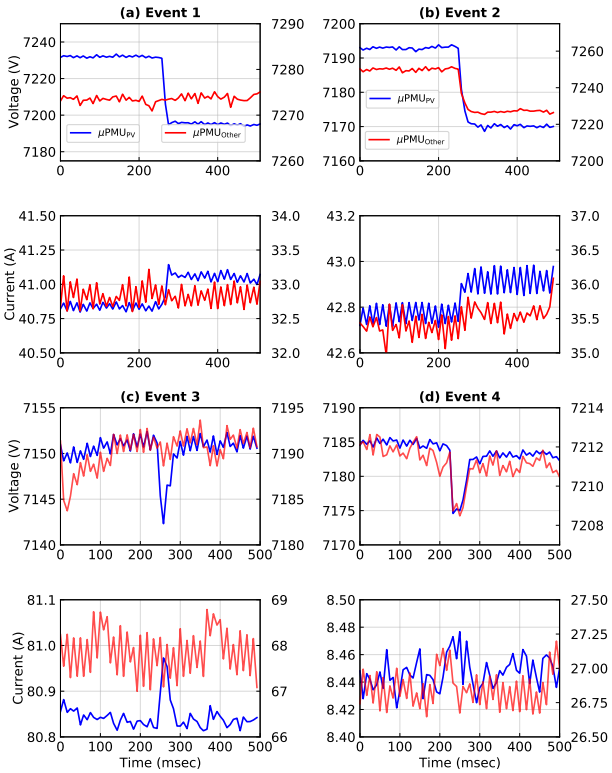

Figure 3.5: Visual inspection: (a) a local sustained event captured by μPMU$_{\text{PV}}$ with *no* signature at μPMU$_{\text{Other}}$; (b) a grid-induced sustained event captured by μPMU$_{\text{PV}}$ with *similar* signature at μPMU$_{\text{Other}}$; (c) a local transient event captured by μPMU$_{\text{PV}}$ with *no* signature at μPMU$_{\text{Other}}$; (d) a grid-induced transient event captured by μPMU$_{\text{PV}}$ with *similar* signature at μPMU$_{\text{Other}}$.

In this section, we seek to examine the *similarity* between the two event signatures that are simultaneously captured by μPMU$_{\text{PV}}$ and μPMU$_{\text{Other}}$. To achieve this objective, we need to use proper features that can quantify the similarity of the event signatures in the time-series in $\mathbf{X}_{\text{PV}}$ and $\mathbf{X}_{\text{Other}}$. For notational simplicity, for the rest of this section, we assume that

$$\mathbf{X} = \mathbf{X}_{\text{PV}} \quad \text{and} \quad \mathbf{Y} = \mathbf{X}_{\text{Other}}. \tag{3.4}$$

1) The *root mean square (RMS) similarity* is defined as [96]:

$$rtSim(\mathbf{X}, \mathbf{Y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[1 - \frac{|x_i - y_i|}{|x_i| + |y_i|}\right]^2}, \tag{3.5}$$

where x_i is row i $\mathbf{X}$, y_i is row i in $\mathbf{Y}$, and n is the length of the time window during which $\mathbf{X}$ and $\mathbf{Y}$ are obtained. Higher values for *rtSim* means higher similarities between $\mathbf{X}$ and $\mathbf{Y}$.

2) We can also use the cosine of the *angle* between the two vectors $\mathbf{X}$ and $\mathbf{Y}$ as another similarity feature [96]:

$$cos(\beta_{\mathbf{X},\mathbf{Y}}) = \frac{\mathbf{X}^T\,\mathbf{Y}}{\|\mathbf{X}\|\,\|\mathbf{Y}\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}. \tag{3.6}$$

3) The similarity features in (3.6) and (3.5) require the same lengths for $\mathbf{X}$ and $\mathbf{Y}$. However, due to missing data or other reasons, $\mathbf{X}$ and $\mathbf{Y}$ may have different lengths. Suppose n and m denote the lengths of $\mathbf{X}$ and $\mathbf{Y}$, respectively. The Longest Common Sub-Sequence (LCSS) distance is defined as [96]:

$$\text{LCSS}(\mathbf{X}, \mathbf{Y}) = \frac{n + m + 2\,\Phi(\mathbf{X}, \mathbf{Y})}{n + m}, \tag{3.7}$$

where $\Phi(\mathbf{X}, \mathbf{Y})$ is the relaxed LCSS recurrence function between $\mathbf{X}$,$\mathbf{Y}$; as defined in [96]. Higher values for *LCSS* indicate *less* similarities between the time-series in $\mathbf{X}$ and $\mathbf{Y}$.

4) The Dynamic Time Warping (DTW) feature is an elastic similarity measure that optimally aligns (or *warps*) the time series in $\mathbf{X}$ and $\mathbf{Y}$ in the temporal domain such that the accumulated cost of alignment is minimal. This accumulated cost can be obtained by dynamic programming [95]:

$$D(\mathbf{X}, \mathbf{Y}) = D_{n,m}, \tag{3.8}$$

where we recursively apply the following:

$$D_{i,j} = (x_i - y_j)^2 + min\{D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1}\}. \tag{3.9}$$

As in (3.7), parameters n and m are the lengths of vectors $\mathbf{X}$ and $\mathbf{Y}$, $D_{i,j}$ is the similarity between the entry i of $\mathbf{X}$ and the entry j of $\mathbf{Y}$. The initial condition is $D_{1,1} = (x_1 - y_1)^2$.

5) Another similarity feature is the *Pearson correlation* [96]:

$$p_{\mathbf{X},\mathbf{Y}} = \frac{\sum_{i=1}^{n}(x_i - \overline{\mathbf{X}})(y_i - \overline{\mathbf{Y}})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{\mathbf{Y}})^2(y_i - \overline{\mathbf{Y}})^2}}, \tag{3.10}$$

where $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ denote the *mean* over the entries of vectors $\mathbf{X}$ and $\mathbf{Y}$, respectively. A key property of Pearson correlation is that it is *invariant* under separate changes in location of the entries and the scales of the two time series. This property is critical for our purpose due to the differences in the variation and amplitude of the measurements from $\mu\text{PMU}_{\text{PV}}$ and $\mu\text{PMU}_{\text{Other}}$. The Pearson correlation is a number between -1 to 1; but here we use its absolute value.

6) The Pearson correlation between two ranked vectors is defined as their *Spearman correlation* coefficient [97]:

$$s_{\mathbf{X},\mathbf{Y}} = p_{r_{\mathbf{X}},r_{\mathbf{Y}}}, \tag{3.11}$$

where $r_{\mathbf{X}}$ and $r_{\mathbf{Y}}$ are the ranked versions of vectors $\mathbf{X}$ and $\mathbf{Y}$, respectively. Here, the entries of each vector are ranked, either both in a descending order or both in an ascending order.

7) The Kendall rank correlation measures the strength of the similarity between the entries of two vectors $\mathbf{X}$ and $\mathbf{Y}$ [98]:

$$\tau_{\mathbf{X},\mathbf{Y}} = \frac{2}{n^2 - n}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{I}\left((x_i - y_i)(x_j - y_j)\right), \tag{3.12}$$

where $\mathbb{I}(\cdot)$ is an indicator function; If $(x_i - y_i)(x_j - y_j) \geq 0$, then $\mathbb{I}(\cdot) = 1$; and if $(x_i - y_i)(x_j - y_j) < 0$, then $\mathbb{I}(\cdot) = -1$.

The performances of the above similarity features are summarized in Fig. 3.6. Here, we apply each similarity feature to several *labeled* event signatures that are already visually inspected and are accordingly identified as locally-induced events or grid-induced events. Using this *box-plot* representation, we assess and compare how each similarity feature can *differentiate* between the locally-induced events and grid-induced events. Note that the Pearson correlation is obtained in two different ways. Pearson I is the correlation coefficient between the event's signature on voltage measured by $\mu\text{PMU}_{\text{PV}}$ and the event's signature on voltage measured by $\mu\text{PMU}_{\text{Other}}$. Pearson II is the correlation coefficient between the event's signature on voltage measured by $\mu\text{PMU}_{\text{PV}}$ and the event's signature on current measured by $\mu\text{PMU}_{\text{PV}}$.

We can see that Pearson II and the cosine similarity show the best performance due to their *minimal overlap* between the two box-plots. They clearly differentiate between the locally-induced events and grid-induced events. For Pearson correlation, the good performance could be due to the fact that Pearson correlation is *invariant* under separate changes in the locations and the scales of the entries in the two vectors. This results in higher correlation between the voltage vectors at $\mu\text{PMU}_{\text{PV}}$ and $\mu\text{PMU}_{\text{Other}}$ for grid-induced events; because of their relatively more similar signatures; while it leads to lower correlation for locally-induced events; because they have relatively more dissimilar signatures. As for the cosine, the similarity is measured irrespective of the of magnitude of the two vectors. The similarity is evaluated rather based on the orientation of the two vectors. On the contrary, some other features, such as the DTW similarity, show major overlap; which means they

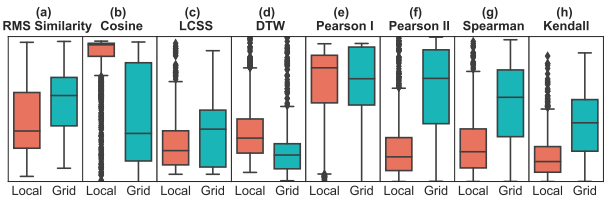

Figure 3.6: The distribution of various similarity features for locally-induced events and for grid-induced events: (a)-(h) the similarity features in (3.5)-(3.12).

cannot perform well in differentiating between the locally-induced events and grid-induced events.

To clarify the above discussion on the implications of the results in Fig. 3.6, each box indicates where the majority of the given features appear for each class of the events. The two classes are best separable by a given feature in each sub-figure if the two boxes have minimal overlap, i.e., the given feature takes a range of values for the first class that are different from the range of values for the second class.

It should be noted that, the type of measurements that are placed in vectors $\mathbf{X}$ and $\mathbf{Y}$ depends on the characteristics of the event that is captured by $\mu\text{PMU}_{\text{PV}}$. One option is to use the measurements with the most dominant signature for the purpose of similarity analysis. For example, if an event has its most dominant signature in the voltage measurements, then we can construct $\mathbf{X}$ and $\mathbf{Y}$ based on the voltage measurements.

3.3.3 Statistical Method

One option to automatically solve the event region identification problem is to conduct a statistical analysis based on the similarity features that we discussed in Section 3.3.2. Consider the event signatures in the time-series of the raw measurements in $\mathbf{X} = \mathbf{X}_{\text{PV}}$ and $\mathbf{Y} = \mathbf{X}_{\text{Other}}$. Suppose $f_u(\mathbf{X}, \mathbf{Y})$ is a similarity feature, such as in (3.5), (3.6), or (3.12),

for $u = 1, \ldots, U$, where U is the number of similarity features. We identify an event as a *grid-induced* event if the following condition holds:

$$f_u(\mathbf{X}, \mathbf{Y}) \geq f_{u,\text{Threshold}}; \tag{3.13}$$

otherwise, the event is identified as a *locally-induced* event. Here, $f_{u,\text{Threshold}}$ is a threshold parameter to indicate the *minimum similarity that is required* between the event signature that is captured at $\mu\text{PMU}_{\text{PV}}$ and the event signature that is simultaneously captured at $\mu\text{PMU}_{\text{Other}}$, such that the event can be identified as grid-induced. Of course, the threshold should be defined separately for *each* similarity feature; because different similarity features may require different thresholds.

The main challenge in using the condition in (3.13) is to properly select the threshold parameter $f_{u,\text{Threshold}}$ for each similarity feature $f_u(\mathbf{X}, \mathbf{Y})$. This can be done by solving an optimization problem, such as the following:

$$\begin{aligned} \underset{f_{u,\text{Threshold}}}{\textbf{max}} \quad & P\left\{\text{Grid-Induced} \mid f_u(\mathbf{X}, \mathbf{Y}) \geq f_{u,\text{Threshold}}\right\} \\ \textbf{s.t.} \quad & f^{\min}_{u,\text{Threshold}} \leq f_{u,\text{Threshold}} \leq f^{\max}_{u,\text{Threshold}}, \end{aligned} \tag{3.14}$$

where $f_{u,\text{Threshold}}$ is the optimization variable. The objective in (3.14) is to maximize the *accuracy* of the event region identification solution, i.e., the probability that the event is indeed a grid-induced event; subject to the condition that the inequality in (3.13) holds. The maximization is done between $f^{\min}_{u,\text{Threshold}}$ and $f^{\max}_{u,\text{Threshold}}$; which are the lowest and the highest acceptable values for $f_{u,\text{Threshold}}$, respectively. Of course, the objective function could be equivalently expressed as:

$$P\left\{\text{Locally-Induced} \mid f_u(\mathbf{X}, \mathbf{Y}) < f_{u,\text{Thrfeshold}}\right\}. \tag{3.15}$$

Other objective functions can also be considered, such as maximizing the F_1-Score or maximizing the recall.

An example is shown in Fig. 3.7. Here, the similarity feature is the Pearson correlation between voltage at $\mu\text{PMU}_{\text{PV}}$ and voltage at $\mu\text{PMU}_{\text{Other}}$. Therefore, in this example, we have:

$$f_{u,\text{Threshold}}^{\min} = 0 \quad \text{and} \quad f_{u,\text{Threshold}}^{\max} = 1. \tag{3.16}$$

Two curves are shown in Fig. 3.7(a). They show the values of the accuracy and F_1-Score, both as a function of $f_{u,\text{Threshold}}$. Importantly, these two curves are obtained based on the *training* data set, i.e., a portion of the captured and labeled events that are used to solve the optimization problem in (3.14). The *peak* in each curve is the optimal choice for $f_{u,\text{Threshold}}$; as far as the specific objective function associated with the curve is considered. In this example, the optimal choice for $f_{u,\text{Threshold}}$ corresponding to the maximization of the accuracy is 0.43, and the F_1-Score is 0.37. They result in 81.72% accuracy and 78.19% F_1-Score, respectively.

It should be noted that, the data set that is used in order to obtain Table 3.2 is the same data set that was used in Section 3.2.1 to obtain Table 3.1. Of course, here we needed to dedicate a portion of the data set for training and the rest of the data set for testing; because unlike the impedance-based method, the statistical method requires training. Out of the one week of data, the data in five days are selected for training and the data in the other two days are selected for testing. This is done carefully, such that we can observe the most challenging cases in order to best identify the weaknesses and the strengths of the proposed methods. The test data set includes the event data from one weekday and one

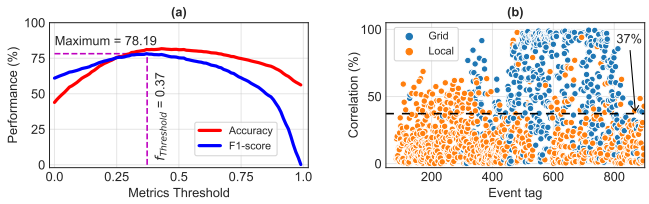

Figure 3.7: An example for training and then testing the statistical method: (a) the optimal choice of the threshold parameter based on the *training* data and when two different objective functions are used; (b) the results of using the obtained optimal threshold parameter in Part (a) in the condition in (3.13) to conduct automated event region identification for the *test* data set.

weekend. Accordingly, the training data set includes the event data from both weekdays and a weekend. To assure consistency in the analysis, we will continue to use the same training data set and the same test data set in Sections III-D, IV-A, IV-B, and IV-C.

The application of the obtained optimal threshold parameter is shown in Fig. 3.7(b). Here, we use $f_{u,\text{Threshold}} = 0.37$ to decide which events are grid-induced and which events are locally-induced, i.e., by using the condition in (3.13).

The performance metrics of the statistical method are shown in Table 3.2. Note that, there are fewer sustained events in Table 3.2 than in Table 3.1; because as we mentioned earlier, the statistical method requires a *training* data set. While, the results in Table 3.1 include *all* the sustained events, the results for the sustained events in Table 3.2 include only the sustained events that are part of the *test* data set.

It is clear that the statistical method overcomes the three fundamental limitations of the impedance-based method.

3.3.4 Machine Learning Methods

Another option to automatically solve the event region identification problem is to use machine learning based on the same similarity features that we discussed in Section 3.3.2.

Table 3.2: The Performance Summary of the Statistical Method

			Grid	Local	Precision %	Recall %	$F_1 - Score$ %
Reality	Sustained	Grid	97	72	96	57	72
		Local	5	14	16	74	27
	Transient	Grid	248	188	65	57	61
		Local	132	733	80	85	82

Importantly, machine learning methods can resolve some of the weaknesses of the statistical methods that we previously identified at the end of Section 3.3.3. While this is promising, the machine learning methods, too, have their own weaknesses.

In this section, we examine six different machine learning methods. The first four methods are based on *supervised* learning. They require prior labeling of several events. This is done by conducting visual inspection of the event signatures at μPMU_{PV} and μPMU_{Other} to label each event as either grid-induced or locally-induced. The last two methods are based on *unsupervised* learning and do not require prior labels. They rather cluster the events into two groups, to separate grid-induced events from locally-induced events.

1) Gradient Boosting Model (GBM): This is a supervised machine learning technique that builds a single estimator from a collection of weak learners, i.e., decision trees. The learning objective is to minimize a loss function based on the similarity features such that the model can correctly decide whether an event is grid-induced or locally-induced [99].

2) Multi-Layer Perception (MLP): This supervised learning method is based on an artificial neural network (ANN) with similarity features as neurons in the input layer, two hidden layers of five and three neurons, and event regions as the two neurons of the output layer, representing the locally-induced and grid-induced events. These neurons are interconnected via a respective weighted sum of the similarity features and a bias to form an affine function. The weight vector and the affine function are updated using back-propagation until the target results of event region identification are achieved [100].

3) Support-Vector machines (SVMs): This supervised machine learning method separates the events into two classes, grid-induced events and locally-induced events. This is done by obtaining proper separating hyperplanes that are calculated based on the similarity features for the two classes of events. The SVM runs an optimization problem to maximize the distance between the separation hyperplanes. [101].

4) Kernel SVM (KSVM): This supervised learning method is an extension of the standard SVM, in which we use kernels, i.e., non-linear boundaries for separation. This is done by mapping the input features into high-dimensional feature spaces [102]. In this study, we use the radial basis function (rbf) kernel; because it shows the best results among other kernels.

5) K-Means Clustering: This unsupervised learning method separates the events into two clusters. Clustering is done based on the similarity features in Section 3.3.2. The objective is to put the events for which the similarities are high between the measurements at $\mu\text{PMU}_{\text{PV}}$ and $\mu\text{PMU}_{\text{Other}}$ in one cluster; and the events for which the similarities are low between the measurements at $\mu\text{PMU}_{\text{PV}}$ and $\mu\text{PMU}_{\text{Other}}$ in another cluster. First, we

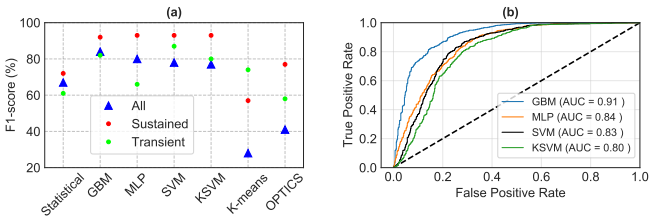

Figure 3.8: Performance comparison of different data-driven methods: (a) based on F_1-Score; (b) based on ROC curves with their respected AUC.

randomly initialize the centroids of the two clusters. Next, we recurrently assign each event to its closest centroid and update the centroid for each cluster until we reach a point that the positions of the two centroids do not change [103].

6) Ordering Points to Identify the Clustering Structure (OPTICS): This unsupervised learning method is an algorithm for finding density-based clusters. Two clusters represent grid-induced and locally-induced events. The clustering is done based on the similarity features. The points in the training data set are linearly ordered such that spatially closest events (as far as their similarity features are concerned) become neighbors in the ordering. Additionally, a special distance is stored for each event that represents the density that must be accepted for a cluster to enhance clustering accuracy [104].

The summary comparison of the performance of the above various machine learning methods is given in Fig. 3.8(a). We observe that the supervised learning methods perform better than the unsupervised learning methods. Among the supervised learning methods, GBM has a slightly better performance and more consistent results. This is better illustrated in Fig. 3.8(b). This figure shows the Receiver Operating Characteristic (ROC) curves for the four supervised learning methods. For each curve, the Area Under Curve (AUC) is also shown in the legend [105]. AUC provides an aggregate measure of performance across all

Table 3.3: The Performance Summary of the GBM Method, which has the Best Performance among Machine Learning Methods

			Grid	Local	Precision %	Recall %	$F_1 - Score$ %
Reality	Sustained	Grid	165	4	94	98	96
		Local	10	9	69	47	56
	Transient	Grid	250	186	86	57	69
		Local	42	823	82	95	88

possible classification thresholds. The highest AUC for GBM means that the probability of correctly identifying the region of a random event is the highest for GBM.

Given that the GBM method demonstrated the best relative performance among the machine learning methods, we select GBM as the representative machine learning method to tackle the problem of event region identification.

The performance of the GBM method is summarized in Table 3.3. We can see that the machine learning method too can highly improve the performance compared to the impedance-based method. It appears to also improve the performance compared to the statistical method; although there is a caveat here that we will explain in the next section.

3.4 Improving Performance with a New Mixed-Integrated Method

In this section, we make the case that the statistical and the machine learning methods have *complementary* strengths and weaknesses. Therefore, we propose to identify and accordingly utilize their strengths and weaknesses, such that we can achieve a new *mixed* method that takes advantage of the strengths of both classes of the data-driven methods.

3.4.1 Strengths and Weaknesses of the Statistical Method

While the results in Table 3.2 in Section 3.3.3 provide the overall summary of the performance of the statistical method, one can further scrutinize the cases where the statistical method *was* successful as well as the cases where the statistical method was *not* successful, in order to identify the strengths and weaknesses of the statistical method, as we explain next.

Strengths

First, the statistical method is particularly strong in identifying the transient grid-induced events that are captured by their voltage signature as the dominant signature and the similarity among voltage measurements is high, i.e., the condition in (3.13) holds. In such cases, the statistical method indicates that the transient event is grid-induced; which is often correct under these circumstances. An example is shown in Fig. 3.9(a), where the statistical method works better than the machine learning method. That is, the statistical method identifies the event correctly but the machine learning method does *not* identify

the event correctly. The statistical method shows good results also when the transient locally-induced events are captured by their current signature as the dominant signature and the similarity among voltage measurements is low, i.e., condition (3.13) does *not* hold. In such cases, the statistical method indicates that the transient event is locally-induced, which is often correct under these circumstances. An example is shown in Fig. 3.9(b), where the statistical method works better than the machine learning method. That is, the statistical method identifies the event correctly but the machine learning method does *not* identify the event correctly.

While the above conclusions are data-driven, one can also comment on the likely rational for these observations. The statistical method, which works by optimizing the value of the similarity features that separate the two classes of events, tends to properly capture the sufficient condition for the event to be grid-induced, when the dominant signature is in voltage and the voltage similarity is high. This method also tends to properly capture the sufficient condition for the event to be locally-induced, when the dominant signature is in current and the voltage similarity is low.

Another overall advantage of the statistical method is that it is easy to implement and it is computationally efficient. Once the threshold is obtained, the event region identification becomes as simple as checking the inequality in (3.13).

Weaknesses

First, while the statistical method outperforms the impedance-based method, its ability to correctly identify sustained events is not as good as the machine learning method. It particularly often fails to correctly identify the region of locally-induced sustained events.

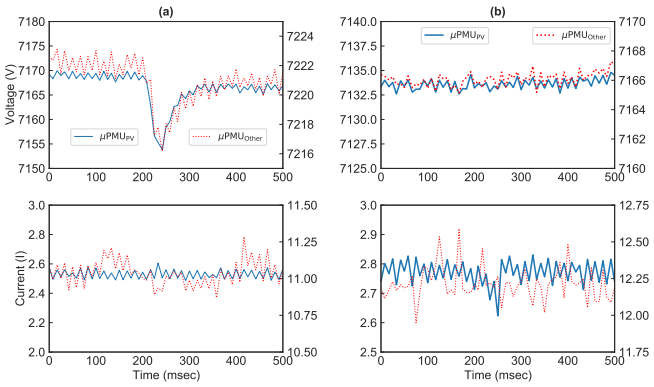

Figure 3.9: Two examples for transient events that *are* correctly identified by the statistical method but are *not* correctly identified by the machine learning method: (a) a grid-induced event; (b) a locally-induced event.

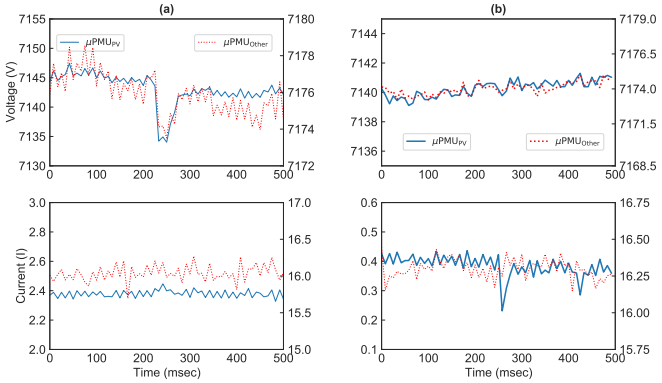

Figure 3.10: Two examples for transient events that *are* correctly identified by the machine learning method but are *not* correctly identified by the statistical method: (a) a grid-induced event; (b) a locally-induced event.

Second, the statistical method often cannot identify the correct event region for the locally-induced transient events that are captured by the signature in current while the similarity among voltage measurements is high; and also when a grid-induced transient event is captured by its voltage signature while the similarity among voltage measurements is low. Third, the performance of the statistical method is often poor for the grid-induced events that have low voltage similarities.

3.4.2 Strengths and Weaknesses of the Machine Learning Method

Recall from Section 3.3.4 that the GBM method performed better than the rest of the machine learning methods that we examined. Therefore, we took the GBM method as the representative machine learning method. While the results in Table 3.3 provide the overall summary of the performance of this machine learning method, one can further scrutinize the cases where this machine learning method *was* successful and the cases where it was *not* successful in order to identify the strengths and weaknesses of this method, as follows. First, this method, too, can overcome all the three fundamental limitations of impedance-based method. Second, it performs better than the statistical method when it comes to identifying the *sustained* events. Third, the machine learning method is particularly strong in identifying the transient grid-induced events that are captured by their voltage signature as the dominant signature while the similarity among voltage measurements is low. An example is shown in Fig. 3.10(a), where the machine learning method works better than the statistical method. This transient event is a grid-induced event. It is identified correctly by the machine learning method but it is *not* identified correctly by the statistical method.

The machine learning method shows good results also when locally-induced transient events are captured by their current signature as the dominant signature while the similarity among the voltage measurements is high. An example is shown in Fig. 3.10(b), where the machine learning method works better than the statistical method. This event is a locally-induced event. It is identified correctly by the machine learning method but it is *not* identified correctly by the statistical method.

While the above conclusions are data-driven, one can also comment on the likely rational for these observations. In this regard, we note that, the machine learning method

learns the trend of the events and examines the overall relationships between all the event features, rather than using a single dominant similarity feature as the only factor. For the events that are captured in voltage but have low similarity, the trends and the relationships between the features are still noticeable, hence the machine learning can identify the right region. The events that are captured in current are often locally-induced, but there can still be high similarity in voltage; because there may not be major agitation created on the voltage as the result of a locally-induced event. This could be missed by the statistical method, because it uses only one feature to make the classification. On the contrary, the combined impact of the similarity features in the machine learning method tends to provide the correct result in such cases.

Weaknesses

First, for those transient grid-induced events where the most dominant signature is in voltage measurements and the similarity between the voltage at $\mu\text{PMU}_{\text{PV}}$ and the voltage at $\mu\text{PMU}_{\text{Other}}$ is high, the machine learning method does *not* perform as good as the statistical method. Second, for those transient locally-induced events where the most dominant signature is in current measurements and the similarity between the voltage at $\mu\text{PMU}_{\text{PV}}$ and the voltage at $\mu\text{PMU}_{\text{Other}}$ is low, the machine learning method is again *not* as good as the statistical method.

3.4.3 Mixed-Integrated Algorithm

Based on the analysis in Sections 3.4.1 and 3.4.2, we are now ready to propose a new algorithm that can take advantage of the identified strengths in both statistical and machine learning methods. The algorithm is given in Algorithm 1.

Algorithm 1 : Mixed-integrated Event Region Identification

Input: Captured event at $\mu\text{PMU}_{\text{PV}}$ and $\mu\text{PMU}_{\text{Other}}$.
Output: Identified region of the event.

if the event is transient **then**
    if the dominant signature is in voltage **then**
        if the similarity check in (3.13) holds **then**
            Use statistical method
        else
            Use machine learning method
        end if
    end if
    if the dominant signature is in current **then**
        if the similarity check in (3.13) does not hold **then**
            Use statistical method
        else
            Use machine learning method
        end if
    end if
else
    Use machine learning method
end if

The performance summary of Algorithm 1 is given in Table 3.4. This table is comparable with Tables II and III. We can see that the shortcomings of the statistical method and the machine learning method are now resolved by their complementary strengths. As a result, the proposed method is very accurate in correctly identifying the region of the event; both for sustained events and also for transient events.

Table 3.4: The Performance Summary of the Mixed-integrated Method

			Grid	Local	Precision %	Recall %	F_1 – Score %
Reality	Sustained	Grid	165	4	94	98	96
		Local	10	9	69	47	56
	Transient	Grid	403	33	91	92	92
		Local	41	824	96	95	96

The results in Tables 3.2, 3.3, and 3.4 can be summarized in terms of their overall performance with respect to grid-induced events and their overall performance with respect to locally-induced events. While the F_1-Score in identifying the grid-induced events is 67% for the statistical method and 79% for the machine learning method, it is much higher at 93% for the mixed-integrated method. Similarly, while the F_1-Score in identifying the locally-induced events is 80% for the statistical method and 88% for the machine learning method, it is considerably higher at 95% for the mixed-integrated method.

3.4.4 Case Study: New Data Set

To further examine the performance of the proposed mixed-integrated method, in this section, we apply Algorithm 1 to a *completely new data set*, i.e., a data set that is different from the data set that we previously used in Sections 3.2, 3.3, 3.4.1, 3.4.2, and 3.4.3 to develop Algorithm 1. The new data set is from the *same* behind-the-meter solar farm. However, it is for a different period of time, i.e., *one week later*. The new data set is for the period of one week and it includes a total of 3874 events; all of which are new events.

Importantly, we do *not* update the training of the proposed method in this new case study. In other words, the entire new data set in this section is used only as a *test*

data set. Furthermore, while the test data set in Sections 3.2, 3.3, 3.4.1, 3.4.2, and 3.4.3 was smaller than the training data set, which is common in data-driven analysis, the test data set in the case study in this section is much larger, even larger than the training data set in Sections 3.2, 3.3, 3.4.1, 3.4.2, and 3.4.3. Therefore, conducting the automated event region identification task based on this new data set is challenging, which makes this data set suitable for our performance evaluation in this section.

The results for the aforementioned new data set are summarized in Fig. 3.11. Here we compare the performance of all the three data-driven methods that we discussed in this study, i.e., the statistical method that was designed in Section 3.3.3, the machine learning method that was designed in Section 3.3.4, and the mixed-integrated method that was proposed in Section 3.4.3. We use the F_1-Score in percentage as the performance metric, and we calculate it separately for the grid-induced events and for the locally-induced events.

We can make three observations. First, the mixed-integrated method shows the best performance in identifying the correct event region, both for grid-induced events and for locally-induced events. In particular, while the F_1-Score in identifying the grid-induced events is 72% for the statistical method and 73% for the machine learning method, the F_1-Score for the mixed-integrated method is 83%; which is significantly higher. Similarly, while the F_1-Score in identifying the locally-induced events is 82% for the statistical method and 82% for the machine learning method, the F_1-Score for the mixed-integrated method is 89%; which is again significantly higher.

Second, the F_1-Scores for all the three methods are slightly lower compared to their corresponding F_1-Scores that we saw at the end of Section IV-C. This is because

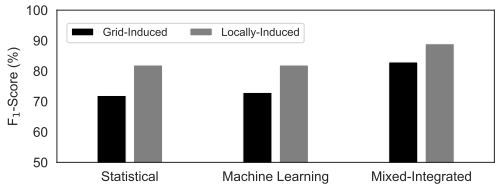

Figure 3.11: Comparing the F_1-Score for all three data-driven methods that are discussed in this study, for grid-induced events and locally-induced events.

the case study in this section uses a long and entirely new data set without conducting any new training. Nevertheless, the F_1-Scores here are still high for the mixed-integrated method. More importantly, the exact same patterns in terms of the advantages of the mixed-integrated method are again observed here, despite using a completely new data set. This can confirm the robustness of the proposed mixed-integrated method.

Third, the mixed-integrated method closes the gap between the results in identifying the grid-induced events versus in identifying the locally-induced events. While there is 14% and 12% gap between the F_1-Scores in identifying the grid-induced events versus locally-induced events for the statistical method and for the machine learning method, respectively, such gap is only 7% for the mixed-integrated method.

3.5 Event-Actuated Applications

Automated event region identification builds the foundation for event-based situational awareness and event-actuated operation in the understudy behind-the-meter solar farm. In this section, we discuss multiple representative applications that use the results from automated event region identification.

3.5.1 Applications of Analyzing Grid-Induced Events

Adaptive Volt-Var Control

One application of identifying grid-induced events is in fine-tuning Volt-Var control (VVC) at PV inverters. This is particularly important for large behind-the-meter solar farm. PV inverters can automatically *absorb* or *inject* reactive power to regulate voltage in power distribution systems. The common approach in inverter-based VVC is to use *piece-wise linear control* curves [92, 106].

Consider the *adaptive* VVC method that is proposed in [107] and shown in Fig. 3.12. Parameters $q_{\max}$ and $q_{\min}$ denote the inverter's maximum reactive power limits, and μ is the reference voltage set point. Reactive power is injected to (absorbed from) the power grid in the capacitive (inductive) zone. Based upon the system conditions during *external disturbances*, the adaptive VVC method in [107] either shifts the VVC curve to left or right; or rotates the VVC curve clock-wise or counter-clockwise. The former is referred to as *error adjustment*, and it is done by changing parameter q, as shown in Fig. 3.12(a). The latter is referred to as *slope adjustment* and it is done by changing parameter m, as shown in Fig. 3.12(b).

A key step in both of the above adaptive VVC methods is to first *identify* the external events. Therefore, we can conduct adaptive VVC by examining the sustained grid-induced events with dominant signature in voltage that are identified by the proposed automated event region identification method in this study. For any such event, we obtain the *steady state error* (SSE) and the *voltage flicker* (VF) as follows [107]:

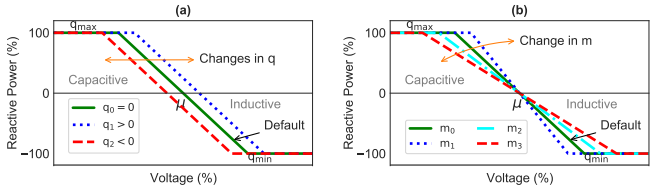

Figure 3.12: The adaptive VVC according to the sustained grid-induced voltage events: (a) VVC error adjustment; (b) VVC slope adjustment.

$$\text{SSE} = \sum_{t=1}^{T} (V_t - \mu)/T, \tag{3.17a}$$

$$\text{VF} = \sum_{t=1}^{T} \frac{(V_t - V_{t-1})/V_t}{T} \times 100, \tag{3.17b}$$

where V_t is the measured voltage and T is the period over which we measure the impact of the event. If $|\text{SSE}| < \epsilon$, then we do not change q. Otherwise, we do change q by the amount of $-\kappa\,\text{SSE}$. Parameters ϵ and κ are determined by either the utility or the solar farm operator [107]. The results for changing VVC parameter q in the case of 20 examples of sustained grid-induced voltage events are shown in Fig. 3.13(a).

As for the slope adjustment method, it uses two thresholds on the value of VF, per the IEEE 141 standard [108]. If $|\text{VF}| < \zeta$, then we may choose to make no change in m. If $|\text{VF}| \geq \zeta$ but $|\text{VF}| < \xi$, then a relatively small change is made in m. If $|\text{VF}| > \xi$, then a relatively large change is made in m. The results for changing VVC parameter m in 20 examples of sustained grid-induced voltage events are shown in Fig. 3.13(b).

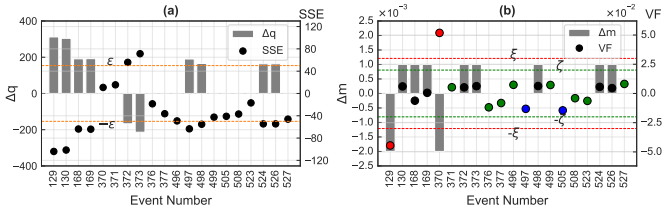

Figure 3.13: Adaptive VVC based on 20 examples of the sustained grid-induced voltage events: (a) the changes in q based on the calculation of SSE at each event; (b) the changes in m based on the calculation of VF at each event.

Dynamic Response

Another application of identifying grid-induced events is in analyzing the dynamic response of the PV inverters in the solar farm to external disturbances. Two examples are shown in Fig. 3.14(a)-(b) and Fig. 3.14(c)-(d).

The grid-induced event in the first example is a *step-change* in voltage, as shown in Fig. 3.14(a). The dynamic response of the behind-the-meter solar farm to this grid-induced disturbance is shown in Fig. 3.14(b). It includes multiple stages, marked from 1 to 5, which are due to the operation of different control loops of the solar farm [1]. Stage 1 is when the event occurrence, which changes the inverter voltage and triggers a response by the solar farm's control system. Stage 2 is the immediate reaction of the system to keep the output power stable; this causes a prompt drop in current to decrease the DC-bus voltage. In Stage 3, the new DC-bus voltage level modifies the output of the Maximum Power Point Tracker (MPPT); and it fine-tunes the reference for the current regulation loop. This causes an increase in the current and sets the DC-bus voltage back to its pre-disturbance value. In Stage 4, after passing the initial transient conditions, the plant level control applies the ramp rate limitation. This results in a momentary decrease in the current. In Stage 5, at a

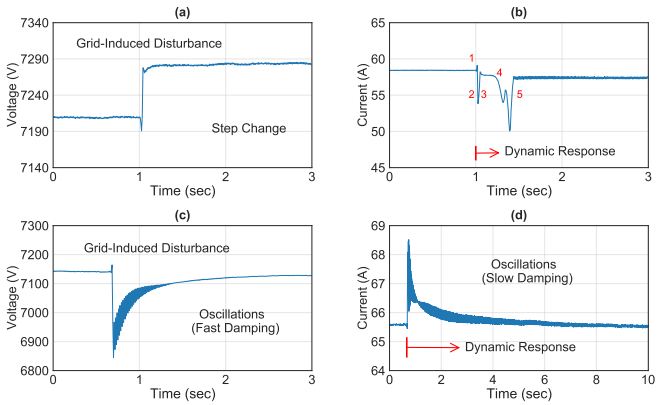

Figure 3.14: Dynamic response of the solar farm to two grid-induced events: (a) a sustained step change in voltage; (b) the dynamic response of the solar farm (and its various stages) to such sustained step change; (c) a temporary voltage sag with subsequent momentary oscillations that stabilize quickly; (d) the dynamic response of the solar farm to such temporary voltage sag; the oscillations in the dynamic response last longer and stabilize rather slowly.

moderate ramp rate, the plant-level controller brings the current back to the regulated set point.

The grid-induced disturbance in the second example is a major but temporary *voltage sag*, followed by some momentary damping oscillations, as shown in Fig. 3.14(c). Notice that the voltage is stabilized quickly, within two seconds after the event occurred. However, this disturbance creates a major dynamic response in the solar farm that lasts much longer to stabilize. It took over 10 seconds before the current at the solar farm gradually reaches stability, i.e., its oscillations damp down, until it finally reaches a stage at which the level of fluctuations is comparable to the pre-disturbance conditions, see Fig. 3.14(d).

Similar analysis can be done for all major grid-induced events to capture the dynamic response of the solar farm to various disturbances. The results can be used in dynamic modeling of the inverters [109], and to evaluate the compliance of the inverters with the inter-connection rules [110].

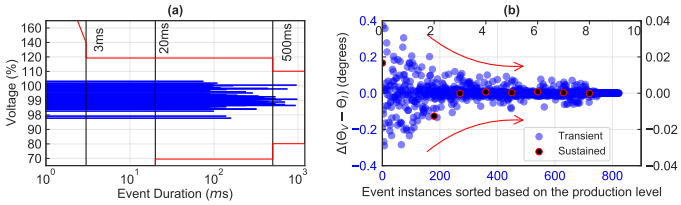

Figure 3.15: Applications of analyzing locally-induced events: (a) voltage magnitude and duration for the locally-induced transient events enveloped in the ITIC curve. Mitigation actions must be done if the event exceed the green area; (b) identifying the relationship between the occurrence and significance of the locally-induced events and the production-level of the solar farm.

3.5.2 Applications of Analyzing Locally-Induced Events

Compliance with Equipment Requirements

One of the applications of identifying locally-induced events is to examine the compliance of the behind-the-meter solar farm with the Information Technology Industry Council (ITIC) performance curve [93]. This is done by examining the *amplitude* and *duration* of the major locally-induced transient events. Of course, a key step here is to first *identify* such local events.

The amplitude and duration of the locally-induced transient events during one example day are shown in Fig. 3.15(a). All events fall in the ITIC tolerance envelope that is shown in green. This curve is the limitation of the safe operation zone, beyond which, the corrective actions are mandatory.

Analysis of Trends and Relationships

The analysis of the locally-induced events may reveal various trends regarding the internal operation of the behind-the-meter solar farm. For example, it may reveal a relationship between the occurrence of the locally-induced events and the *production level* of the solar farm, as shown in Fig. 3.15(b). We can see that most of the locally-induced events,

especially the transient locally-induced events, occurred during the *low production* periods. Also, the events that occurred during low production periods demonstrated more significant changes in power factor, as shown here in terms of the change in the phase angle difference between the voltage phasors and the current phasors at each event, see Fig. 3.15(b). Here, θ_V and θ_I denote the phase angle measurements in voltage and in current, respectively. Understanding these and other trends can help with achieving situational awareness and identifying potential issues in the operation of the under-study behind-the-meter solar farm [111].

Chapter 4

Distribution Solar Farm Dynamic Response Modeling Under Grid Events

4.1 Introduction

Environmental issues are rapidly evolving and the grid is changing rapidly than ever before [112]; ; the time between interconnection request and the service operation has shortens more than ever before and plants, specifically solar PV farms and Battery Energy Storage System (BESS), are coming online so fast. Solar farms integration to the distribution system are increasingly due to the trend in the green energy requirements. As the penetration of solar power generation continues to grow, new challenges emerge and new modeling techniques are required [113].

One major challenge is to determine how the solar farm will behave in case of disturbances on the grid and how this response is going to impact the grid as a result. We need to clarify that grid disturbances are those disturbances that are due to events occurring in the grid beyond the micro-PMU on the solar farm. These events are called *grid-induced*

events and are studied in great details and are identified with multiple methodologies in our previous works [1] and [2]. When solar farm injects power to the grid as the response to a *grid-induced* event, the resultant agitations will impact the power system [114]. Having knowledge of the dynamic of this agitation is significantly important in proper planing and designing of the system. Thus, accurate dynamic modelling is a need for both utilities and ISO. Also, the compliance with the utility checklist requirements, e.g. rule 21, UL1741, IEEE1547. Utilities such as SDG&E and California ISO have requirements to understand the dynamics of solar farm interconnection to the power grid. These recent rules is the motivation to understand the dynamics of interconnected large inverter-based resources (IBRs). However, currently there are several IBRs that are not following these rules because they were installed in prior to the rule establishment. There are two solutions to address this challenge; first installing individual measurement devices in each input/output of the inverters; second is the utilization of the data-driven methods with limited inputs that we suggest. Moreover, in case of any new modification in the system, such as adding a new section to the solar and add or remove a bulk load, a model is required to predict the model, so we have understanding of the feasibility of the intended modification and its impact on the system.

The main question that we want to answer is that how to model the dynamic response of IBRs using data-driven techniques with limited measurements data? Later we propose an event-based methodology by developing an algorithm that analyzes the dynamic response in presence of wide area disturbances (grid events). The main output of the work

is a tool (based on an algorithm) that provides the characteristics of the post-event dynamic in solar farms.

4.1.1 Approach and Scope of Analysis

This chapter is about dynamic response of a real-world solar farm that is integrated into a distribution network as shown in Fig. 4.1. This solar distribution feeder is a large behind-the-meter solar farm, which is monitored by a micro-PMU at the distribution substation. We construct dynamic models of this solar farm under different grid events reported by the micro-PMU measurements and then estimate the dynamic response of the solar farm to these events and evaluate the performance. The raw data has high frequency noise on top of the actual dynamic and trend of the signals, as shown in Fig. 4.2. These noise will confuse the modeling algorithms and introduce uncertainties to the parameter estimations of the models. Therefore we will use some pre-processing techniques, such as low-pass filtering and scaling to not only do noise cancellation but also bring all the agitations of the input and output signals to the range of 0 to 1, which is called normalized signals.

As shown in Fig. 4.3 and equations (4.1), in the training stage, the inputs to the system are full size of the voltage and frequency of the events and the outputs of the system are full size of the real and reactive powers of the solar farm during of the events. Note that

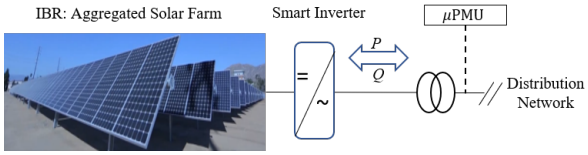

Figure 4.1: The distribution solar farm under study.

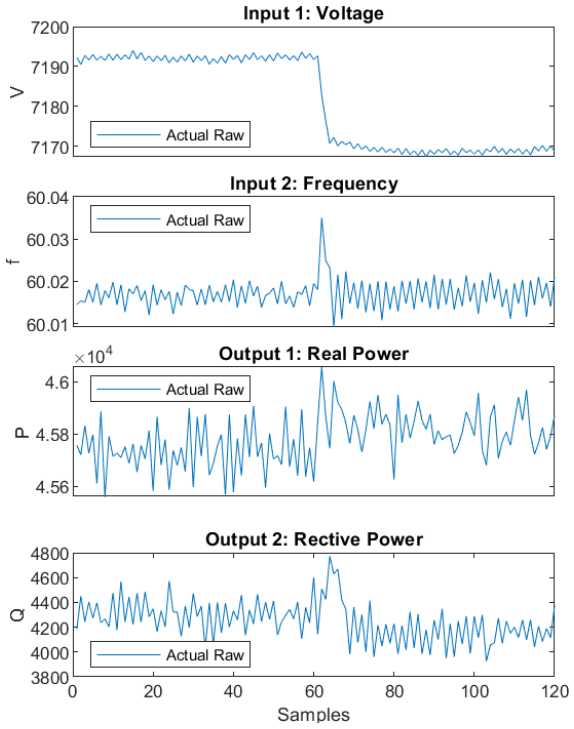

Figure 4.2: An example event with two inputs and two dynamic outputs in the test stage. The raw data with noise and actual range is presented.

the signals are first filtered to remove the high frequency noises and then the agitations are scaled to be between 0 and 1. The training events are denoted by $\mathbf{R}$ and the testing events are denoted by $\mathbf{E}$.

In the testing stage, as described in (4.2), the inputs to the system are full size of the voltage, full size of the frequency, and the pre-event period, that is specified by K, of the real and reactive power. The current input is also used as an auxiliary input, which is slightly improving the output estimation results, specially the real power estimation. The outputs of the system are the post-event period of the real and reactive powers of the solar farm.

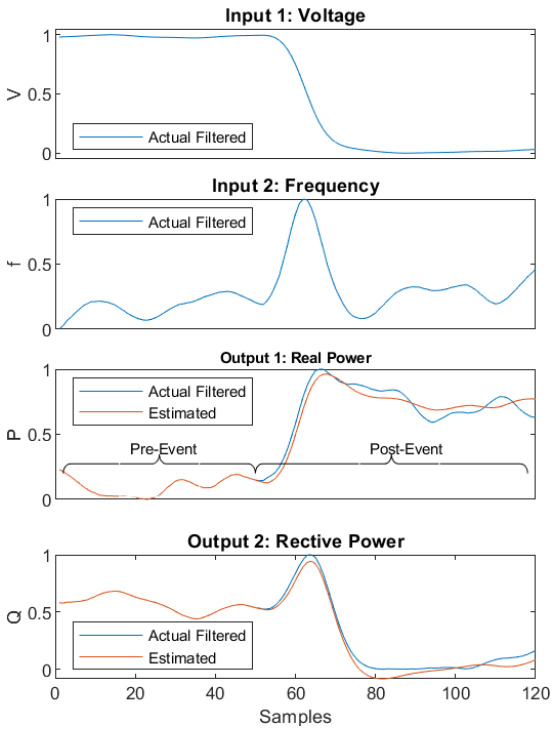

Figure 4.3: An example event with two inputs and two dynamic outputs in the test stage: the pre-event is the initialization and the post-event is the dynamic response period that is estimated.

$$
\begin{array}{c}
\text{Training} \\
\\
\underline{\text{Inputs}} \qquad \underline{\text{Outputs}} \\
\\
\text{event } R_1: \quad \begin{matrix} [v_{11}, \cdots, v_{1L_1}] \\ [f_{11}, \cdots, f_{1L_1}] \end{matrix} \rightarrow \begin{matrix} [P_{11}, \cdots, P_{1L_1}] \\ [Q_{11}, \cdots, Q_{1L_1}] \end{matrix} \\
\\
\vdots \\
\\
\text{event } R_N: \quad \begin{matrix} [v_{N1}, \cdots, v_{NL_N}] \\ [f_{N1}, \cdots, f_{NL_N}] \end{matrix} \rightarrow \begin{matrix} [P_{N1}, \cdots, P_{NL_N}] \\ [Q_{N1}, \cdots, Q_{NL_N}] \end{matrix}
\end{array}
\tag{4.1}
$$

where L_i is the length of the event i and v_{ij}, f_{ij}, P_{ij}, and Q_{ij} are the voltage, frequency, real power, and the reactive power of the event i at timestamp j, respectively.

Testing

$$
\begin{array}{lll}
 & \underline{\text{Inputs}} \quad \underline{\text{Initializing}} & \underline{\text{Outputs}} \\
\text{event } E_1: & \begin{array}{l}[v_{11},\cdots,v_{1L_1}]\,[P_{11},\cdots,P_{1K_1}] \\ [f_{11},\cdots,f_{1L_1}][Q_{11},\cdots,Q_{1K_1}]\end{array} \rightarrow & \begin{array}{l}[P_{1K_1+1},\cdots,P_{1L_1}] \\ [Q_{1K_1+1},\cdots,Q_{1L_1}]\end{array} \\
 & \vdots & \\
\text{event } E_M: & \begin{array}{l}[v_{M1},\cdots,v_{ML_M}]\,[P_{M1},\cdots,P_{MK_M}] \\ [f_{M1},\cdots,f_{ML_M}][Q_{M1},\cdots,Q_{MK_M}]\end{array} \rightarrow & \begin{array}{l}[P_{MK_M+1},\cdots,P_{ML_M}] \\ [Q_{MK_M+1},\cdots,Q_{ML_M}]\end{array}
\end{array}
\tag{4.2}
$$

The events are first detected by any available detection method, then using the techniques in our paper in [2], we identify the sustained events that occur somewhere in the grid and divide them into the train and test events' datasets. Both train and test datasets include two input signals (voltage and frequency) and two output signals (real and reactive powers) as described in (1) and (2). Then, we conduct the following:

1. For the training events, we build single model and multiple models. The single model treats all the train events as a single time series and then accordingly constructs a model. In the second case, we treat each event as a single time series and fit a corresponding model model. This results in a library of model that has one model per each training event.

2. Also, for the train dataset, we build linear models to estimate the output of the models. These models are added to the library of models for more flexibility in case of very harsh dynamic response.

3. An augmented library is developed that consists of both multiple and single models that are constructed using linear fittings.

4. For each event in the test dataset, we first find the most similar event in the train dataset using the two different similarity measures, Dynamic Time Wrapping (DTW) and Pearson Correlation. Then, the related model to the most similar event is used to estimate the dynamic response output of the model to the test event.

5. For each step of the model development, we provide the results and show the evolution of our models and the improvement of estimation of the dynamic response of the models to the events.

The list of contributions in this study are provided as following:

- This study provides a data-driven event-based technique to model IBRs' dynamic response; which is of interest to 1) modeling dynamic response of the previously installed IBRs to provide insight of these systems to the utility; 2) planning and feasibility study of solar farm projects. Furthermore, the data use is from a real-world solar farm which makes the results applicable to other practical studies in the field.
- The proposed method uses the measurement devices that are installed for other purposes, e.g., micro-PMU data, instead of installing asset sensors for inverters.
- The modeling of the dynamic response is done with limited input data of the plant.
- A comprehensive library of linear models are build that is also involved in an active learning process. For events that the models in the library cannot estimate a reasonable outputs, a new model will be build and add to the library.
- Accurate estimation of the dynamic response output in the post-event period. This is due to the availability of comprehensive single and multi build models.

- The selection of the best models are based on two different similarity measures, DTW and Pearson correlation, that makes sure of find the most comparable event and model in the library.

- The active learning library will grow really fast. So, a dimension reduction technique is use to always keep the library in a reasonable size by classifying the models and choosing the best model in each class to represent the whole class.

- The approach in this study has three more level of flexibility compare to most of the data-driven approaches; the lengths events can be different, the estimation window is dynamic and is automatically chosen based on the occurrence time of the event, and the parameter classification are also self-regulating according to the distribution of the parameters values.

- The estimation of the real and reactive powers are independent. This creates a flexibility in the model selection process that leads to better estimation for each signal separately.

4.1.2 Literature Review

Planing and feasibility study, and dynamic modeling

Single modeling

Library of models

4.2 Single Dynamic Modeling

In this section, as described in (4.3). all the events in the training dataset are considered as one continuous time series that leads to building a single comprehensive model.

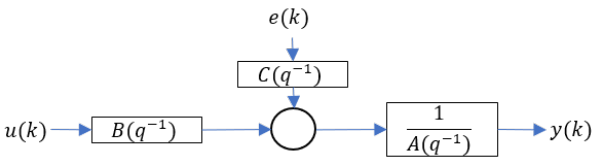

Figure 4.4: The ARMAX model structure.

This single model is then used to estimate the dynamic output of the solar farm in response to each and all events in the test deadset.

$$
\begin{array}{c}
\textit{Training} \\
\underline{\textit{Inputs}} \quad\quad \underline{\textit{Outputs}} \\
\textit{event } 1-N: \quad \begin{array}{ll} [v_{11},\cdots,v_{NL}] & [P_{11},\cdots,P_{NL}] \\ [f_{11},\cdots,f_{NL}] & [Q_{11},\cdots,Q_{NL}] \end{array} \rightarrow
\end{array}
\tag{4.3}
$$

To construct this single model, we employ two different ways based on the event time series and linear methods. These models are explained in the following two subsections.

In this study, we focus on the linear models for the parameter estimation due to challenges related to overfitting in the nonlinear models and their significantly bigger number of parameters.

In the linear model, the number of parameters that are estimated for the model are very limited and the fitted model is a rough estimation of the real and reactive powers. Here, the error of the fitting model is not negligible. In this study, we employ Autoregressive Moving Average eXogenous (ARMAX). The structure of the ARMAX model is shwon in Fig. 4.4

The parameters of ARMAX models are estimated using the prediction-error method and the polynomial orders specified in [na nb nc nk] in (4.4). The model properties include

estimation covariances (parameter uncertainties) and goodness of fit between the estimated and the measured data.

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k), \tag{4.4a}$$

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}, \tag{4.4b}$$

$$B(q^{-1}) = b_1 q^{-1} + \cdots + b_{nb} q^{-nb}, \tag{4.4c}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}, \tag{4.4d}$$

$$\begin{aligned} y(k) + a_1 y(k-1) + \cdots + a_{na} y(k-na) = b_1 u(k-1) + \cdots \\ + b_{nb} u(k-nb) + e(k) + c_1 e(k-1) + \cdots + c_{nc} e(k-nc). \end{aligned} \tag{4.4e}$$

For an ARMAX model with Ny outputs and Nu inputs: na is the order of the polynomial A(q), specified as an Ny-by-Ny matrix of non-negative integers; nb is the order of the polynomial B(q) + 1, specified as an Ny-by-Nu matrix of non-negative integers; nc is the order of the polynomial C(q), specified as a column vector of non-negative integers of length Ny; nk is the input-output delay, specified as an Ny-by-Nu matrix of non-negative integers.

Using this linear model to estimate the real and reactive powers of 50 test events, the accuracy of he model is shown in Fig. 4.5

4.3 Multiple Dynamic Modeling

In this section, we build one model per event and construct a library of events equal to the number of the events in the training dataset. In the testing stage, first, we

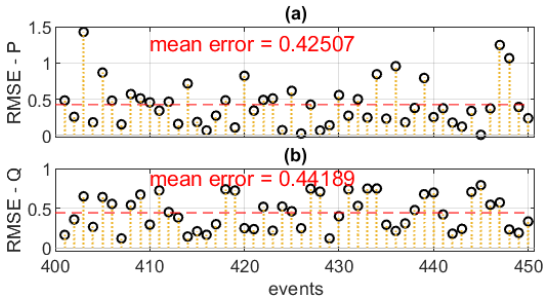

Figure 4.5: The estimation errors of the single model for the test events: (a) the real power estimation errors; (b) the reactive power estimation errors.

find the similarity of the events to the events in the training set, then, the model of the chosen training event is used to estimate the real and reactive power output of the solar farm resulting from the test event.

The two similarity measures, that are used cooperatively, are listed below:

1) First the Dynamic Time Warping (DTW) feature which is an elastic similarity measure that optimally warps the time series in the test event (we denote it by $\mathbf{E}$) and the train events (we denote each of them by $\mathbf{R}$) in the temporal domain in a way that the accumulated error of alignments is minimized. This accumulated error (cost) is obtained by a dynamic programming [95]:

$$D(\mathbf{E}, \mathbf{R}) = D_{m,n}, \tag{4.5}$$

where we recursively apply the following:

$$D_{j_E,j_R} = (E_{j_E} - R_{j_R})^2 + min\{D_{j_E,j_R-1}, D_{j_E-1,j_R}, D_{j_E-1,j_R-1}\}. \tag{4.6}$$

Parameters m and n are the lengths of vectors $\mathbf{E}$ and $\mathbf{R}$, D_{j_E,j_R} is the similarity between the entry j_E of $\mathbf{E}$ and the entry j_R of $\mathbf{R}$. The initial condition is $D_{1,1} = (E_1 - R_1)^2$.

Algorithm 2 :Best Model Selection for P and Q, Individually.

Input: Test event $\mathbf{E}$ from the test dataset and events $\mathbf{R}$s from the train dataset.

Output: $\mathbf{M_P^*}$ and $\mathbf{M_Q^*}$, the best model for estimating P and Q, respectively.

1: **for** each event $\mathbf{R}$ in train dataset **do**
2: Find $D(\mathbf{E}, \mathbf{R})$ from (4.5).
3: Find $p_{\mathbf{E},\mathbf{R}}$ from (4.7).
4: Store D and p.
5: **end for**
6: Find D_{max} and p_{min} and their related models $\mathbf{M}_{D_{max}}$, and $\mathbf{M}_{p_{min}}$.
7: **if** $\mathbf{M}_{D_{max}} \equiv \mathbf{M}_{p_{min}}$ **then**
8: $\mathbf{M_P^*}$ and $\mathbf{M_Q^*} = \mathbf{M}_{D_{max}}$.
9: **else**
10: Estimate P and Q using $\mathbf{M}_{D_{max}}$, and $\mathbf{M}_{p_{min}}$.
11: $\mathbf{M_P^*}$ = model with the closes range of P to 0 and 1.
12: $\mathbf{M_Q^*}$ = model with the closes range of Q to 0 and 1.
13: **end if**

2) The other similarity feature is the *Pearson correlation* [96]:

$$p_{\mathbf{E},\mathbf{R}} = \frac{\sum_{i=1}^{m}(E_i - \overline{\mathbf{E}})(R_i - \overline{\mathbf{R}})}{\sqrt{\sum_{i=1}^{m}(E_i - \overline{\mathbf{R}})^2(R_i - \overline{\mathbf{R}})^2}}, \tag{4.7}$$

where $\overline{\mathbf{E}}$ and $\overline{\mathbf{R}}$ denote the *mean* over the entries of vectors $\mathbf{E}$ and $\mathbf{R}$, respectively. An important feature of Pearson correlation is that it is *invariant* under separate changes in location of the entries and the scales of the two time series. This feature is critical for this study due to the differences in the variation and amplitude of the measurements. Using Algorithm 2, these similarity measures are used to find an event among the events in the train dataset that is most similar to the event under study in the test dataset.

The independent estimation of the real and reactive powers creates a flexibility in the model selection process that leads to better estimation for each both signals. Once, that

Algorithm 3 : Multi Model Parameter Identification

Input: Voltage, frequency, and current of during the captured event.

Output: Real and reactive powers injection the dynamic response of the solar farm.

```
Construct one model per event
Form a library of the multi models
for each event E in test dataset do
    Select the best model using Algorithm 2.
    Estimate the dynamic response using this model
    Evaluate the performance
    if performance metrics less than required then
        Build a new model and add to the library
    end if
end for
```

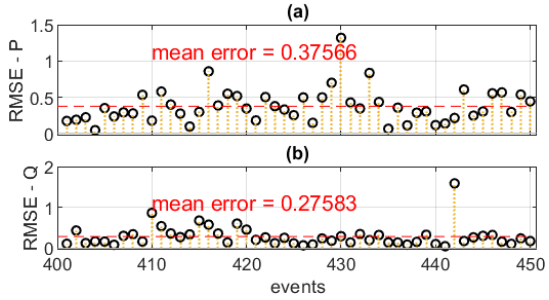

Figure 4.6: The estimation errors of the multi models for the test events: (a) the real power estimation errors; (b) the reactive power estimation errors.

the process of selecting the best model based on the similarity of the events is done, the multi model algorithm can be formed. The process of the multi model construction and testing is shown in Flowchart 3.

We apply Algorithm 3 and observe the estimation errors for the test dataset. As shown in Fig. 4.6, the errors of estimations is much smaller than the single model.

The improvement for the estimation of the real power is 11.62 % and the improvement on the reactive power estimation is 37.58 %. In the Single-Model case, the errors of

the real power estimation are more focused around the mean except for a few cases that are calculated to be around upper limit of 1. While, the distribution of the reactive power estimation error is scattered almost evenly across the range from 0 to 1. Therefore, when using the Multi-Model, due to the bigger reduction in the bigger errors of the reactive power estimation, the overall improvement is much larger. Moreover, the overall better estimation of the reactive power is due to the strong channel from the voltage profile of the network to the reactive power exchange from the IBRs; the *Volt-Var* control strategy of the smart inverters are directly related to this interconnection of voltage of the grid and the reactive power of the solar farm. In other words, the voltage magnitude is more influential factor in the reactive power.

4.4 Library of Single and Multiple Models

The single model performance is not significantly high, however for all type of events, its estimation results are in acceptable range. The multi models performance results are high for most of the events but for some types of events they diverge and the estimation result becomes unreliable. Thus, we combine them to benefit from the reliability of the single model and the accuracy of the multi models. During our study on the single and multi models, we realized that in some cases the single model provided better estimation of the dynamic response of the solar farm. For example, the single model estimation of real power for event 430, in Fig. 4.5(a), is much better than the multi model estimation of real power for event 430, in Fig. 4.6(a). Similarly, the single model estimation of reactive power for event 442, in Fig. 4.5(b), is much better than the multi model estimation of real power for event 442, in Fig. 4.6(b).

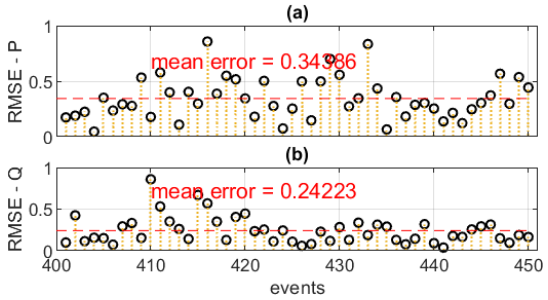

Figure 4.7: The estimation errors of the multi-single models for the test events: (a) the real power estimation errors; (b) the reactive power estimation errors.

Thus, in this section we use the collection of the multi models and the single model as our new library of models. Then, we seek the best fitted model from this new library. Here the error of the estimation is even smaller than the multi model for both real and reactive power dynamic estimations, as shown in Fig. 4.7.

4.5 Models Library Reduction

As the number of training events increase, the number of related models increase and consequently complication, computation time and cost increases. Therefore, to handle the challenge of library size expansion we reduce the size of the library by clustering the models and select the best model in each cluster. Then, the selected models form a new library and will be used as reference for dynamic response estimation of the test events. The process of library reduction is shown in Algorithm 4.

By library reduction and the fact that we can cluster the models in small groups, not only we reduce the computation cost, but also we gain very important insight and perspective of the solar farm; we understand that even though each event's data leads to

Algorithm 4 : Models Library Dimensional Reduction

Input: Constructed models $\mathbf{M_{Original}}$ and target number of models, $\mathbf{T}$.

Output: Target representative models, $\mathbf{M_{Reduced}}$.

1: **for** each parameter in $\mathbf{M_{Original}}$ **do**
2: Cluster the models into $\mathbf{T}$ groups using K-Means.
3: Find distances between cluster centers.
4: **end for**
5: Find the best parameter that creates the maximum clusters distance, as in Fig. 4.8.
6: Cluster the models using the best parameter.
7: Create a new model library, $\mathbf{M_{Reduced}}$.
8: **for** each cluster **do**
9: Find the model with the closest parameter to the
10: cluster's median.
11: Append the selected model to $\mathbf{M_{Reduced}}$.
12: **end for**

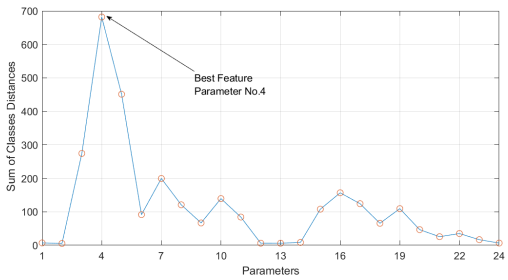

Figure 4.8: The best parameter selection as the classifier feature.

a separate model, deep down, the models have very similar parameters and also the solar farm has specific performance and dynamic response to these events.

The best parameter is used as the feature that classifies models into $\mathbf{T}$ groups. The resultant clusters are shown in Fig. 4.10.

Using the reduced models, the results are as shown in Fig. 4.10. We can see that the estimation error is slightly increased compare to the library with all models. However, the results are still much better than the multi and the single model.

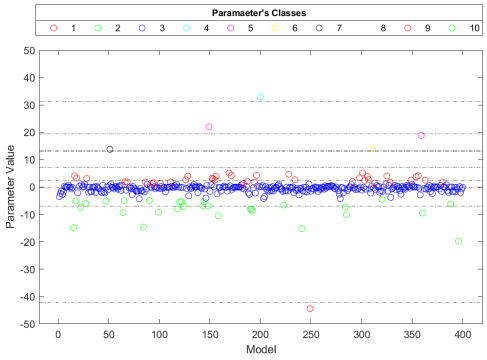

Figure 4.9: The resultant classes of employing the best selected parameter.

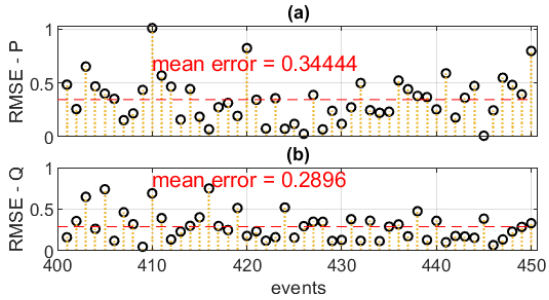

Figure 4.10: The estimation errors of the multi-single models with the reduced library for the test events: (a) the real power estimation errors; (b) the reactive power estimation errors.

4.6 *DER_A*: A Grey Box Model-Based Approach

To model aggregated solar farm using model-based techniques, mostly the Western Electricity Coordinating Council (WECC) *generic* model is recommended. However, the generic model is intended for utility scale PV plants ($P_{rated} > 10MW$) which are connected to transmission systems ($V_{rated} > 60KV$) [115]. WECC suggests a simplified version of the generic model, called *DER_A*, to be used for distribution and commercial level PV plants [116].

The *DER_A* model stands for the Distributed Energy Resource Model Version A and is a model used for modeling the positive-sequence dynamic behavior of aggregated distributed solar PV plants. This model is derived from the large-scale generic PV model.

The large-scale PV model contains 121 parameters and 16 states, thus is too complex for representation of aggregated solar PV plants. Moreover, this model was developed to depict the behavior of a single large renewable plant and thus may not be easily adaptable to incorporate the aggregation of distributed generators. Therefore, the DER_A model was developed to provide a straightforward approach to modeling the aggregation of distributed solar PV plants, and also a reduction in parameters and states was established without diminishing the core functionalities to adequately represent the dynamic behavior.

Identifying the unknown parameters comes down to finding optimal parameters by reducing the estimation residual in (4.8) [117]:

$$min_{\Theta}\ell(Y,\Theta,V) = min_{\Theta}\frac{1}{2}(\|Y - f(\Theta,V)\|_2^2), \tag{4.8}$$

where Y denotes active and reactive power measurement vector, θ represents the vector of parameters to be identified, V denotes voltage measurement vector, ℓ represents calculating the estimation residual, and $\|.\|$ is the ℓ 2-norm [65]. The process of residual minimization and DER_A model tuning and testing is shown in Fig. 4.11.

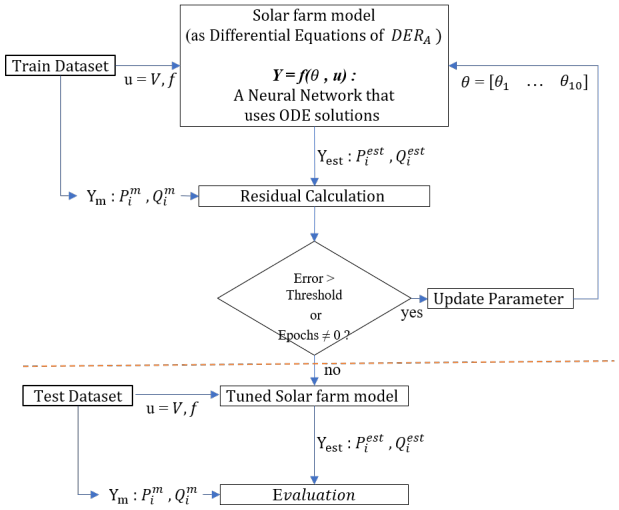

Figure 4.11: The Process of DER_A model tuning.

Chapter 5

Conclusions and Future Works

5.1 Summary of the Conclusions

In this dissertation, we conducted a comprehensive analysis of events and disturbances at a real-world solar distribution feeder and propose novel and automated identification methods of the events regions using real-world micro-PMU measurements. Having the events regions automatically identified, we also modeled the dynamic response of the inverter-based resources under grid disturbances, i.e., the events in the grid region.

In chapter 2, an event-based analysis is conducted based on real-world micro-PMU measurements at a solar distribution feeder. After detecting all the events by training an unsupervised deep learning model, the events are classified based on their origin; either locally-induced by the solar farm itself; or they are grid-induced. It was observed that 70% of the events happen when the solar production is at %30 or less. Furthermore, the events during the low solar production periods demonstrate more significant changes in power factor. In low production, the range of the phase angle change is $\pm 3^{\circ}$, while in high production, it is mostly around 0°. Among the 88 events that were detected at the solar distribution feeder, only 8 events were recognized as grid-induced. We also examined the response

of the solar distribution feeder to all these grid-induced events; and compared it to the response of a neighboring feeder to the exact same events. Finally, we characterized the event dynamics into 5 steps based on the control system mechanisms of the solar distribution feeder. The analysis in this chapter provides awareness about the operation of solar distribution feeders. This study is also beneficial in monitoring the health of equipment in the solar farm. Therefore, it is of economic value to the utility due to providing with insight on the state of the health of the inaccessible behind-the-meter solar farm equipment. Importantly, this type of study is also essential to address the cascading effect of ongoing solar energy increment on the stability and operation of the distribution systems.

In chapter 3, we automated the process of event region identification in behind-the-meter solar farms by establishing a novel mixed-integrated data-driven approach that combines the strengths of both statistical and machine learning methods. This method overcame the deficiencies of the conventional impedance-based method and the human visual inspection. Very high performance is demonstrated by using real-world micro-PMU measurements from a behind-the-meter solar farm in Riverside, CA. Multiple practical applications are discussed for both locally-induced events and grid-induced events, contributing to the behind-the-meter solar farm's situational awareness, control, and operation. For events that are identified as grid-induced, we performed adaptive Volt-Var control as well as dynamic response analysis. For events that are identified as locally-induced, we performed compliance analysis for equipment requirements and the analysis of trends and relationships. These applications help with achieving situational awareness and efficient operation of the behind-the-meter solar farms.

In Chapter 4, the dynamic behavior of a real-world solar farm, as an IBR, in response to grid disturbances is modeled using data-driven techniques. The modeling of the dynamic response is done with limited input data of the plant which replicates the level of IBR's information that utilities access. A comprehensive library of linear models are build that is also involved in an active learning process so the library size grows really fast. To avoid the library size expansion, a reduction technique is used based on K-Means clustering and representative model selection. Two different similarity measures, DTW and Pearson correlation are jointly used for model selection to ensure of finding the most comparable event and model in the library. The modeling in our approach is flexible to the lengths events, the estimation window, and the parameter classification. Further, the estimation of the real and reactive powers are independent to provide better estimation for each signal separately. The estimation of the dynamic response output in the post-event period is improved by 20% for the dynamic estimation of the real power and 40% the dynamic estimation of the reactive power. This is due to the availability of comprehensive single and multi build models.

5.2 Future Works

In the future, the results and findings in Chapter 2 can be used in signature mapping for the purpose of diagnostics and prognostics application in higher-penetration solar distribution feeders. The analysis in Chapter 3 can be extended in several directions. First, one can further explore the other applications of the proposed automated event region identification problem. For example, the outcome of the event region identification method can be used for dynamic modeling of the solar farms, as we have done it in Chapter 4. This will be done based on examining the response of the solar farm to the grid-induced events

and it would involve event-based parameter estimation and model training. Of course, this task requires to first identify and separate such events, which can be done by using the method that was proposed in Chapter 3. Second, the automated region identification process can be used in an online mode on the micro-PMU data for the purpose of event classification. For example, one can try to further identify the sub-classes of the locally-induced events to help even more when it comes to taking remedial actions. Third, if proper data will be available, one can extend the proposed method to identify the region of the short sub-cycle events or even some harmonic issues that are only visible in waveform measurements and are not visible to phasor measurements. This would require expanding the analysis from the phasor domain to the waveform domain. The IBRs' dynamic response modeling, that is done in Chapter 4, is of interest for dynamic response modeling of the previously installed IBRs to obtain insight of these systems to the utility. Also, it is essential for planning and feasibility study of solar farm projects. These results are practical for effective IBRs model development by the interconnection customers and facilitates dynamic behaviour estimation for the transmission planners to ensure the inverter-based generators meet various interconnection requirements.

Bibliography

[1] P. Khaledian, A. Aligholian, and H. Mohsenian-Rad, "Event-based analysis of solar power distribution feeder using micro-PMU measurements," in *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2021, pp. 1–5.

[2] P. Khaledian and H. Mohsenian-Rad, "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*, 2022.

[3] D. Chen and D. Irwin, "Sundance: Black-box behind-the-meter solar disaggregation," in *Proceedings of the eighth international conference on future energy systems*, 2017, pp. 45–55.

[4] A. Martinez-Morales, T. Nielsen, and H. Mohsenian-Rad, "Integration of a DER management system in riverside," *Deliverable Report 9.1.1, UC Riversde, U.S. DoE Award no. DE-EE0008001*, Mar. 2021.

[5] D. E. Olivares, A. Mehrizi-Sani, A. H. Etemadi, C. A. Cañizares, R. Iravani, M. Kazerani, A. H. Hajimiragha, O. Gomis-Bellmunt, M. Saeedifard, R. Palma-Behnke *et al.*, "Trends in microgrid control," *IEEE Trans. on Smart Grid*, vol. 5, no. 4, pp. 1905–1919, Jul. 2014.

[6] I. S. Bayram and T. S. Ustun, "A survey on behind the meter energy management systems in smart grid," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 1208–1232, May 2017.

[7] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[8] P. Cuffe, "Optimization and visualization tools for situational awareness in highly renewable power systems," in *2020 6th IEEE International Energy Conference (ENERGYCon)*. IEEE, 2020, pp. 930–933.

[9] Q. Wang, S. Bu, Z. He, and Z. Y. Dong, "Toward the prediction level of situation awareness for electric power systems using cnn-lstm network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6951–6961, 2020.

[10] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796–4805, 2019.

[11] L.-A. Lee and V. Centeno, "Comparison of μPMU and PMU," in *2018 Clemson University Power Systems Conference (PSC)*, 2018, pp. 1–6.

[12] H. Mohsenian-Rad, E. Stewart, and E. Cortez, "Distribution synchrophasors: Pairing big data with analytics to create actionable information," *IEEE Power and Energy Magazine*, vol. 16, no. 3, pp. 26–34, May 2018.

[13] Micro-synchrophasors for distribution grids: instrumentation lessons learned. [Online]. Available: https://www.naspi.org/

[14] IEEE, "IEEE standard for synchrophasor measurements for power systems," *IEEE Std C*, vol. 37, pp. 1–61, 2011.

[15] Y. Zhou, R. Arghandeh, I. Konstantakopoulos, S. Abdullah, A. von Meier, and C. J. Spanos, "Abnormal event detection with high resolution micro-PMU data," in *2016 Power Systems Computation Conference (PSCC)*. IEEE, 2016, pp. 1–7.

[16] A. Shahsavari, M. Farajollahi, E. M. Stewart, E. Cortez, and H. Mohsenian-Rad, "Situational awareness in distribution grid using micro-PMU data: A machine learning approach," *IEEE Trans. on Smart Grid*, vol. 10, no. 6, pp. 6167–6177, Feb 2019.

[17] M. Saini and R. Kapoor, "Classification of power quality events–a review," *Int. J. of Electrical Power & Energy Sys.*, vol. 43, pp. 11–19, Dec. 2012.

[18] O. Samuelsson, M. Hemmingsson, A. H. Nielsen, K. O. H. Pedersen, and J. Rasmussen, "Monitoring of power system events at transmission and distribution level," *IEEE Trans. on Power Systems*, vol. 21, no. 2, pp. 1007–1008, May 2006.

[19] E. Stewart, M. Stadler, C. Roberts, J. Reilly, D. Arnold, and J.-Y. Joo, "Data-driven approach for monitoring, protection, and control of distribution system assets using micro-PMU technology," *CIRED-Open Access Proceedings Journal*, pp. 1011–1014, Oct. 2017.

[20] M. Farajollahi, A. Shahsavari, and H. Mohsenian-Rad, "Location identification of high impedance faults using synchronized harmonic phasors," in *Proc. of the IEEE PES ISGT*, Washington, DC, Apr. 2017.

[21] A. Shahsavari, M. Farajollahi, E. Stewart, C. Roberts, and H. Mohsenian-Rad, "A data-driven analysis of lightning-initiated contingencies at a distribution grid with a PV farm using micro-PMU data," in *Proc. of the IEEE PES NAPS*, Morgantown, WV, Nov. 2017.

[22] A. G. Shaik and O. P. Mahela, "Power quality assessment and event detection in hybrid power system," *Electric power systems research*, vol. 161, pp. 26–44, 2018.

[23] H. Erişti and Y. Demir, "A new algorithm for automatic classification of power quality events based on wavelet transform and svm," *Expert systems with applications*, vol. 37, no. 6, pp. 4094–4102, 2010.

[24] S. Khokhar, A. A. B. M. Zin, A. S. B. Mokhtar, and M. Pesaran, "A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 1650–1663, 2015.

[25] M. Valtierra-Rodriguez, R. de Jesus Romero-Troncoso, R. A. Osornio-Rios, and A. Garcia-Perez, "Detection and classification of single and combined power quality disturbances using neural networks," *IEEE transactions on industrial electronics*, vol. 61, no. 5, pp. 2473–2482, 2013.

[26] A. Shahsavari, M. Farajollahi, E. Stewart, A. von Meier, L. Alvarez, E. Cortez, and H. Mohsenian-Rad, "A data-driven analysis of capacitor bank operation at a distribution feeder using micro-PMU data," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2017, pp. 1–5.

[27] A. Shahsavari, A. Sadeghi-Mobarakeh, E. M. Stewart, E. Cortez, L. Alvarez, F. Megala, and H. Mohsenian-Rad, "Distribution grid reliability versus regulation market efficiency: An analysis based on micro-PMU data," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2916–2925, 2017.

[28] D. B. Arnold, C. Roberts, O. Ardakanian, and E. M. Stewart, "Synchrophasor data analytics in distribution grids," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2017, pp. 1–5.

[29] A. Shahsavari, M. Farajollahi, E. Stewart, C. Roberts, and H. Mohsenian-Rad, "A data-driven analysis of lightning-initiated contingencies at a distribution grid with a PV farm using micro-PMU data," in *Proc. IEEE NAPS*, 2017, pp. 1–6.

[30] M. Farajollahi, A. Shahsavari, and H. Mohsenian-Rad, "Location identification of distribution network events using synchrophasor data," in *2017 North American Power Symposium (NAPS)*. IEEE, 2017, pp. 1–6.

[31] ——, "Tracking state estimation in distribution networks using distribution-level synchrophasor data," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2018, pp. 1–5.

[32] ——, "Location identification of high impedance faults using synchronized harmonic phasors," in *Proc. IEEE PES ISGT*, 2017, pp. 1–5.

[33] Q. Cui and Y. Weng, "Enhance high impedance fault detection and location accuracy via μ-PMUs," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 797–809, 2019.

[34] R. Krishnathevar and E. E. Ngu, "Generalized impedance-based fault location for distribution systems," *IEEE transactions on power delivery*, vol. 27, no. 1, pp. 449–451, 2011.

[35] J. Ren, S. Venkata, and E. Sortomme, "An accurate synchrophasor based fault location method for emerging distribution systems," *IEEE Transactions on Power Delivery*, vol. 29, no. 1, pp. 297–298, 2013.

[36] M. Majidi, M. Etezadi-Amoli, and M. S. Fadali, "A novel method for single and simultaneous fault location in distribution networks," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3368–3376, 2014.

[37] M. Pignati, L. Zanni, P. Romano, R. Cherkaoui, and M. Paolone, "Fault detection and faulted line identification in active distribution networks using synchrophasors-based real-time state estimation," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 381–392, 2016.

[38] K.-R. Shih and S.-J. Huang, "Application of a robust algorithm for dynamic state estimation of a power system," *IEEE Transactions on Power Systems*, vol. 17, no. 1, pp. 141–147, 2002.

[39] Y. Seyedi, H. Karimi, and S. Grijalva, "Distributed generation monitoring for hierarchical control applications in smart microgrids," *IEEE Trans. on Power Syst.*, vol. 32, no. 3, pp. 2305–2314, 2016.

[40] D.-J. Won, I.-Y. Chung, J.-M. Kim, S.-I. Moon, J.-C. Seo, and J.-W. Choe, "A new algorithm to locate power-quality event source with improved realization of distributed monitoring scheme," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1641–1647, 2006.

[41] A. Eshraghi and R. Ghorbani, "Islanding detection and over voltage mitigation using controllable loads," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 125–135, 2016.

[42] A. Aligholian, A. Shahsavari, E. Stewart, E. Cortez, and H. Mohsenian-Rad, "Unsupervised event detection, clustering, and use case exposition in micro-PMU measurements," *IEEE Trans. on Smart Grid*, March 2021.

[43] A. Shahsavari, M. Farajollahi, E. M. Stewart, E. Cortez, and H. Mohsenian-Rad, "Situational awareness in distribution grid using micro-PMU data: A machine learning approach," *IEEE Trans. on Smart Grid*, vol. 10, no. 6, pp. 6167–6177, November, 2019.

[44] R. Yadav, A. K. Pradhan, and I. Kamwa, "Real-time multiple event detection and classification in power system using signal energy transformations," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1521–1531, 2018.

[45] S. S. Negi, N. Kishor, K. Uhlen, and R. Negi, "Event detection and its signal characterization in PMU data stream," *IEEE Trans. on Industrial Informatics*, vol. 13, no. 6, pp. 3108–3118, 2017.

[46] A. Shahsavari, M. Farajollahi, E. Stewart, C. Roberts, F. Megala, L. Alvarez, E. Cortez, and H. Mohsenian-Rad, "Autopsy on active distribution networks: A data-driven fault analysis using micro-PMU data," in *Proc. IEEE NAPS*, 2017, pp. 1–7.

[47] E. I. Batzelis, "Simple PV performance equations theoretically well founded on the single-diode model," *IEEE Journal of Photovoltaics*, vol. 7, no. 5, pp. 1400–1409, 2017.

[48] M. Abido and M. S. Khalid, "Seven-parameter PV model estimation using differential evolution," *Electrical Engineering*, vol. 100, no. 2, pp. 971–981, 2018.

[49] H. Cai, J. Xiang, and W. Wei, "Modelling, analysis and control design of a two-stage photovoltaic generation system," *IET Renewable Power Generation*, vol. 10, no. 8, pp. 1195–1203, 2016.

[50] G. R. C. Mouli, J. H. Schijffelen, P. Bauer, and M. Zeman, "Design and comparison of a 10-kw interleaved boost converter for PV application using si and sic devices," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 5, no. 2, pp. 610–623, 2016.

[51] S. Siouane, S. Jovanović, and P. Poure, "Open-switch fault-tolerant operation of a two-stage buck/buck–boost converter with redundant synchronous switch for PV systems," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3938–3947, 2018.

[52] J. W. Zapata, S. Kouro, G. Carrasco, H. Renaudineau, and T. A. Meynard, "Analysis of partial power dc–dc converters for two-stage photovoltaic systems," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 7, no. 1, pp. 591–603, 2018.

[53] J.-S. Kim, J.-M. Kwon, and B.-H. Kwon, "High-efficiency two-stage three-level grid-connected photovoltaic inverter," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2368–2377, 2017.

[54] M. C. Di Piazza, M. Luna, G. Petrone, and G. Spagnuolo, "Translation of the single-diode PV model parameters identified by using explicit formulas," *IEEE Journal of Photovoltaics*, vol. 7, no. 4, pp. 1009–1016, 2017.

[55] M. Q. Raza, M. Nadarajah, and C. Ekanayake, "On recent advances in PV output power forecast," *Solar Energy*, vol. 136, pp. 125–144, 2016.

[56] R. Hariharan, M. Chakkarapani, G. S. Ilango, and C. Nagamani, "A method to detect photovoltaic array faults and partial shading in PV systems," *IEEE Journal of Photovoltaics*, vol. 6, no. 5, pp. 1278–1285, 2016.

[57] C. Shao, W. Wang, L. Hang, A. Tong, and S. Wang, "A novel PV array connection strategy with PV-buck module to improve system efficiency," in *2018 International Power Electronics Conference (IPEC-Niigata 2018-ECCE Asia)*. IEEE, 2018, pp. 2866–2870.

[58] D. Remon, A. M. Cantarellas, and P. Rodriguez, "Equivalent model of large-scale synchronous photovoltaic power plants," *IEEE Transactions on Industry Applications*, vol. 52, no. 6, pp. 5029–5040, 2016.

[59] P. Li, W. Gu, H. Long, G. Cao, Z. Cao, B. Xu, and J. Pan, "High-precision dynamic modeling of two-staged photovoltaic power station clusters," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4393–4407, 2019.

[60] P. Li, W. Gu, L. Wang, B. Xu, M. Wu, and W. Shen, "Dynamic equivalent modeling of two-staged photovoltaic power station clusters based on dynamic affinity propagation clustering algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 95, pp. 463–475, 2018.

[61] J. Zou, C. Peng, H. Xu, and Y. Yan, "A fuzzy clustering algorithm-based dynamic equivalent modeling method for wind farm with dfig," *IEEE transactions on energy conversion*, vol. 30, no. 4, pp. 1329–1337, 2015.

[62] W. Li, P. Chao, X. Liang, J. Ma, D. Xu, and X. Jin, "A practical equivalent method for dfig wind farms," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 610–620, 2017.

[63] D. Ramasubramanian, I. Alvarez-Fernandez, P. Mitra, A. Gaikwad, and J. Boemer, "The new aggregated distributed energy resources (DER_A) model for transmission planning studies: 2019 update," *Electric Power Research Institure (EPRI): Washington, DC, USA*, 2019.

[64] W. E. C. Council, "WECC dynamic composite load model (CMPLDW) specifications," WECC, Tech. Rep, Tech. Rep., 2015.

[65] Z. Ma, Z. Wang, Y. Wang, R. Diao, and D. Shi, "Mathematical representation of wecc composite load model," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 5, pp. 1015–1023, 2020.

[66] F. Bu, Z. Ma, Y. Yuan, and Z. Wang, "Wecc composite load model parameter identification using evolutionary deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5407–5417, 2020.

[67] I. Alvarez-Fernandez, D. Ramasubramanian, A. Gaikwad, and J. Boe-Mer, "Parameterization of aggregated distributed energy resources (DER_A) model for transmission planning studies," *Cigre Sci. Eng.*, vol. 15, pp. 158–168, 2019.

[68] K. Jain, A. Bhargava, and B. Mewara, "Power quality enhancement using unified power flow controller in standalone grid connected solar PV system," in *Proc. of the IEEE ICIRCA*, Coimbatore, India, July, 2018.

[69] A. Aligholian, A. Shahsavari, E. Cortez, E. Stewart, and H. Mohsenian-Rad, "Event detection in micro-PMU data: A generative adversarial network scoring method," in *Proc. of the IEEE PES General Meeting*, Montreal, QB, Canada, July 2020.

[70] S. Seme, N. Lukač, B. Štumberger, and M. Hadžiselimović, "Power quality experimental analysis of grid-connected photovoltaic systems in urban distribution networks," *Energy*, vol. 139, pp. 1261–1266, November, 2017.

[71] R. Tonkoski, D. Turcotte, and T. H. El-Fouly, "Impact of high PV penetration on voltage profiles in residential neighborhoods," *IEEE Trans. on Sustainable Energy*, vol. 3, no. 3, pp. 518–527, May, 2012.

[72] O. P. Mahela and A. G. Shaik, "Detection of power quality events associated with grid integration of 100kw solar PV plant," in *Proc. of the IEEE ICEEE*, Chennai, India, July, 2015.

[73] D. Wang, X. Wang, Y. Zhang, and L. Jin, "Detection of power grid disturbances and cyber-attacks based on machine learning," *Journal of Information Security and Applications*, vol. 46, pp. 42–52, June, 2019.

[74] B. Biswal, M. Biswal, S. Mishra, and R. Jalaja, "Automatic classification of power quality events using balanced neural tree," *IEEE Trans. on Industrial Electronics*, vol. 61, no. 1, pp. 521–530, February, 2014.

[75] M. Valtierra-Rodriguez, R. de Jesus Romero-Troncoso, R. A. Osornio-Rios, and A. Garcia-Perez, "Detection and classification of single and combined power quality disturbances using neural networks," *IEEE Trans. on Industrial Electronics*, vol. 61, pp. 2473–2482, May, 2014.

[76] M. Biswal, S. M. Brahma, and H. Cao, "Supervisory protection and automated event diagnosis using PMU data," *IEEE Trans. on power delivery*, vol. 31, no. 4, pp. 1855–1863, January, 2016.

[77] G. Sanitha and R. Shereef, "Micro-PMU data based location identification and classification of events," in *Proc. IEEE PESGRE India*, 2020, pp. 1–6.

[78] M. Farajollahi, A. Shahsavari, E. M. Stewart, and H. Mohsenian-Rad, "Locating the source of events in power distribution systems using micro-PMU data," *IEEE Trans. on Power Syst.*, vol. 33, no. 6, pp. 6343–6354, Nov. 2018.

[79] A. Akrami, S. Asif, and H. Mohsenian-Rad, "Sparse distribution system state estimation: An approximate solution against low observability," in *Proc. of the IEEE PES ISGT*, Washington, DC, May 2020.

[80] NERC Reliability Guideline, "BPS-Connected Inverter-Based Resource Performance," Sep. 2018.

[81] L. Zhang, H. Chen, Q. Wang, N. Nayak, Y. Gong, and A. Bose, "A novel on-line substation instrument transformer health monitoring system using synchrophasor data," *IEEE Trans. on Power Deli.*, vol. 34, no. 4, pp. 1451–1459, Aug. 2019.

[82] Y. Ge, A. J. Flueck, D.-K. Kim, J.-B. Ahn, J.-D. Lee, and D.-Y. Kwon, "An event-oriented method for online load modeling based on synchrophasor data," *IEEE Trans. on Smart Grid*, vol. 6, no. 4, pp. 2060–2068, Jul. 2015.

[83] M. Kamal, M. Farajollahi, H. Nazaripouya, and H. Mohsenian-Rad, "Cyberattacks against event-based analysis in micro-PMUs: Attack models and counter measures," *IEEE Trans. on Smart Grid*, vol. 12, no. 2, pp. 1577–1588, Mar. 2020.

[84] A. Akrami, S. Asif, and H. Mohsenian-Rad, "Sparse tracking state estimation for low-observable power distribution systems using d-PMUs," *IEEE Trans. on Power Syst.*, Jul. 2021.

[85] J. Gonzalez, P. N. Papadopoulos, J. V. Milanović, G. Peskir, and J. Moriarty, "Risk-constrained minimization of combined event detection and decision time for online transient stability assessment," *IEEE Trans. on Smart Grid*, vol. 12, no. 5, pp. 4564–4572, Jun. 2021.

[86] Y. Seyedi, H. Karimi, and S. Grijalva, "Irregularity detection in output power of distributed energy resources using PMU data analytics in smart grids," *IEEE Trans. on Industrial Informatics*, vol. 15, no. 4, pp. 2222–2232, Apr. 2019.

[87] P. Gopakumar, M. J. B. Reddy, and D. K. Mohanta, "Fault detection and localization methodology for self-healing in smart power grids incorporating phasor measurement units," *Elec. Power Comp. and Syst.*, vol. 43, no. 6, pp. 695–710, 2015.

[88] D.-I. Kim, A. White, and Y.-J. Shin, "PMU-based event localization technique for wide-area power system," *IEEE Trans. on Power Syst.*, vol. 33, no. 6, pp. 5875–5883, Nov. 2018.

[89] H. Ren, Z. J. Hou, B. Vyakaranam, H. Wang, and P. Etingov, "Power system event classification and localization using a convolutional neural network," *Frontiers in Energy Research*, vol. 8, p. 327, 2020.

[90] M. Izadi and H. Mohsenian-Rad, "Synchronous waveform measurements to locate transient events and incipient faults in power distribution networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4295–4307, 2021.

[91] V. Kumar, A. Pandey, and S. Sinha, "Grid integration and power quality issues of wind and solar energy system: A review," in *Proc. IEEE ICETEESES*, 2016, pp. 71–80.

[92] M. Kraiczy, B. York, M. Bello, D. Montenegro, S. Akagi, and M. Braun, "Coordinating smart inverters with advanced distribution voltage control strategies," in *Proc. IEEE PES General Meeting*, 2018.

[93] R. Seguin, J. Woyak, D. Costyk, J. Hambrick, and B. Mather, "High-penetration PV integration handbook for distribution engineers," National Renewable Energy Lab, Golden, CO, Tech. Rep., 2016.

[94] L. A. Jeni, J. F. Cohn, and F. De La Torre, "Facing imbalanced data–recommendations for the use of performance metrics," in *Proc. IEEE ACII*, 2013, pp. 245–251.

[95] J. Serra and J. L. Arcos, "An empirical evaluation of similarity measures for time series classification," *Knowledge-Based Syst.*, vol. 67, pp. 305–314, Sep. 2014.

[96] C. Cassisi, P. Montalto, M. Aliotta, A. Cannata, and A. Pulvirenti, "Similarity measures and dimensionality reduction techniques for time series data mining," *Advances in data mining knowledge discovery and applications, InTech, Rijeka, Croatia*, pp. 71–96, 2012.

[97] S.-D. Bolboaca and L. Jäntschi, "Pearson versus Spearman, Kendall's tau correlation analysis on structure-activity relationships of biologic active compounds," *Leonardo J. of Sciences*, vol. 5, no. 9, pp. 179–200, Jul. 2006.

[98] A. Geller, "Calculating Kendall's Tau with multiple measurements," Ph.D. dissertation, Texas A&M University, 2018.

[99] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of Gradient Boosting algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, Aug. 2020.

[100] H. Ramchoun, M. A. J. Idrissi, Y. Ghanou, and M. Ettaouil, "Multilayer Perceptron: Architecture optimization and training." *Int. J. Interact. Multim. Artif. Intell.*, vol. 4, no. 1, pp. 26–30, Jan. 2016.

[101] G. Doran and S. Ray, "A theoretical and empirical analysis of support vector machine methods for multiple-instance classification," *Machine learning*, vol. 97, no. 1-2, pp. 79–102, Oct. 2014.

[102] C. Cowan, P. Grant, and S. Chen, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. on neural networks*, vol. 2, no. 2, pp. 302–309, Mar. 1991.

[103] K. A. Nazeer and M. Sebastian, "Improving the accuracy and efficiency of the k-means clustering algorithm," in *Proc. of Citeseer WCE*, 2009.

[104] K. Agrawal, S. Garg, S. Sharma, and P. Patel, "Development and validation of OPTICS based spatio-temporal clustering technique," *Information Sciences*, vol. 369, pp. 388–401, Nov. 2016.

[105] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[106] J. Smith, W. Sunderman, R. Dugan, and B. Seal, "Smart inverter Volt/Var control functions for high penetration of PV on distribution systems," in *Proc. IEEE PES Power Syst. Conference and Exposition*, 2011.

[107] A. Singhal, V. Ajjarapu, J. Fuller, and J. Hansen, "Real-time local Volt/Var control under external disturbances with high PV penetration," *IEEE Trans. on Smart Grid*, vol. 10, no. 4, pp. 3849–3859, Jul. 2018.

[108] IEEE, "IEEE Recommended Practice for Electric Power Distribution for Industrial Plants," *IEEE Std. 141-1993*, pp. 1 –768, Apr. 1994.

[109] K. Yamashita, H. Renner, S. M. Villanueva, G. Lammert, P. Aristidou, J. C. Martins, L. Zhu, L. D. P. Ospina, and T. Van Cutsem, "Industrial recommendation of modeling of inverter-based generators for power system dynamic studies with focus on photovoltaic," *IEEE PETS-J*, vol. 5, no. 1, pp. 1–10, Feb. 2018.

[110] J. T. Johnson, "Draft electric rule 21 test protocols for advanced inverter functions." Sandia National Lab, Albuquerque, NM, Tech. Rep., 2014.

[111] H. Mohsenian-Rad, *Smart Grid Sensors: Principles and Applications*. Cambridge University Press, Cambridge, UK, 2021.

[112] K. Nwaigwe, P. Mutabilwa, and E. Dintwa, "An overview of solar power (PV systems) integration into electricity grids," *Materials Science for Energy Technologies*, vol. 2, no. 3, pp. 629–633, 2019.

[113] M. S. Alam, F. S. Al-Ismail, A. Salem, and M. A. Abido, "High-level penetration of renewable energy sources into grid utility: Challenges and solutions," *IEEE Access*, vol. 8, pp. 190 277–190 299, 2020.

[114] L. Callegaro, G. Konstantinou, C. A. Rojas, N. F. Avila, and J. E. Fletcher, "Testing evidence and analysis of rooftop PV inverters response to grid disturbances," *IEEE Journal of Photovoltaics*, vol. 10, no. 6, pp. 1882–1891, 2020.

[115] W. E. C. Council, "Central station photovoltaic power plant model validation guideline," WECC, Jun. 17 2015. [Online]. Available: https://www.wecc.org/Pages/results.aspx?k=modeling

[116] WECC, "The new aggregated distributed energy resources (DER_A) model for transmission planning studies," WECC, Dec. 9, 2019. [Online]. Available: https://www.wecc.org/Reliability/Forms/Default%20View.aspx

[117] K. Zhang, H. Zhu, and S. Guo, "Dependency analysis and improved parameter estimation for dynamic composite load modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3287–3297, 2016.