UC Riverside
UC Riverside Previously Published Works

Title

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

Permalink

https://escholarship.org/uc/item/898197c7

Authors

Khaledian, Parviz
Mohsenian-Rad, Hamed

Publication Date

2022-01-28

Peer reviewed

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

Parviz Khaledian, *Student Member, IEEE,* and Hamed Mohsenian-Rad, *Fellow, IEEE*

Abstract—This paper is motivated by the fact that behind-the-meter solar farms are being increasingly deployed in California and elsewhere in recent years. The objective is to use real-world micro-PMU measurements at a 4.3 MW behind-the-meter solar Photovoltaic (PV) farm to build a foundation for *event-based situational awareness* and its *data-driven application*. Two essential tasks are conducted. First, through developing an automated event region identification mechanism, we identify whether an event at a behind-the-meter solar farm is "locally-induced", i.e., it is *caused* by the solar farm, thus potentially indicating *internal* issues in the solar farm, or "grid-induced", i.e., it is caused by something else on the grid, thus revealing how the solar farm *responded* to *external* disturbances. We show that this is a highly challenging task in practice: the conventional impedance-based method is ineffective, the statistical method and the machine learning method each has its weaknesses. Accordingly, a novel mixed-integrated method is proposed and tested that can achieve very high performance metrics. The proposed mixed-integrated method also closes the gap between the accuracies in identifying grid-induced events versus locally-induced events. Second, the outcome of automated event region identification is used to unmask the constructive use of the proposed analysis. Practical use cases are proposed to take advantage of the situational awareness that we gain from analyzing both types of events to provide critical reporting, unmask trends and relationships, adjust control parameters, or take remedial actions when needed.

Index Terms—Micro-PMU data, behind-the-meter solar farm, experimental results, event region identification, machine learning, event-based situational awareness, data-driven application.
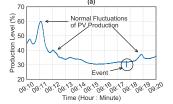
I. INTRODUCTION

PROPER monitoring of behind-the-meter inverter-based distributed and renewable energy resources is an essential and challenging task in power systems [1]. 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 [2].

In this paper, 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 [3]. 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.

A. 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 [4].
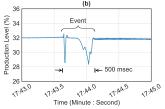
Fig. 1. Comparing normal fluctuation in solar power generation versus the type of events that are of concern in this study: (a) solar production over a period of 10 minutes; (b) the magnified version of the event in Part (a).

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. 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. 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.

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.

B. 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 paper. 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. 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 region on
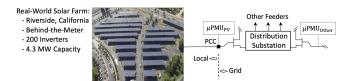
Fig. 2. The real-world test-bed in this study: μPMU$_{\text{PV}}$ is installed at the point-of-common coupling (PCC) of a 4.3 MW behind-the-meter solar farm in Riverside, CA; μPMU$_{\text{Other}}$ is installed at the other side of the substation on another feeder to serve as a reference for certain analysis.

the *right* side of μPMU$_{\text{PV}}$. The main technical contributions in this paper 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.

C. 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 [5], load modeling [6], cybersecurity [7], state estimation [8], and stability analysis [9]. A few studies have also focused on analyzing events in solar generation units and distributed energy resources. In [10], events in micro-PMU measurements are examined to detect irregularity in the operation of PV resources. In [11], 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 [4], [12]. However, as we will see later in this paper, 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 [13]–[15]. 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 [13], a combined impedance-based and data-driven method is proposed to locate faults in a power plant. In [14], 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 [4], 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 [16]–[18], 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 [19] or by analyzing the waveform measurements as in [20]. 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 paper.

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 [21], Volt-Var control [22], and monitoring equipment operation and state of health [23]. 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 paper for our proposed automated event location identification method.

II. LIMITATIONS OF THE CONVENTIONAL IMPEDANCE-BASED SOLUTION

This section briefly discusses the conventional *impedance-based* method [18], [19] to solve the event region identification problem. We apply this method to micro-PMU data from the
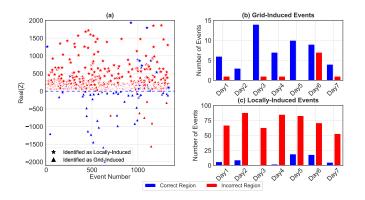
Fig. 3. The performance of the impedance-based method on *sustained* events: (a) scatter plot for Real$\{Z_{\text{Event}}\}$ for all sustained events during one week; the colors show the correct vs. incorrect identification; the markers show the grid vs. local identification; (b) the daily summary of the results for grid-induced events; (c) the daily summary of the results for locally-induced events.

4.3 MW behind-the-meter PV farm in Section I to demonstrate the severe limitations of such impedance-based method.

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

$$Z_{\text{Event}} = \frac{\Delta V}{\Delta I} = \frac{V^{\text{post}} - V^{\text{pre}}}{I^{\text{post}} - I^{\text{pre}}}, \tag{1}$$

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 (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

$$\text{Real}\{Z_{\text{Event}}\} > 0; \tag{2}$$

otherwise, the event is deemed to be *grid-induced* [19].

B. Shortcomings in Impedance-Based Method

1) 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.

2) 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. The impedance-based method leads to *incorrect* region identification for a large portion of events.

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

Fig. 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.

The performance summary of the impedance-based method for the results in Fig. 3 is given in Table I. 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 [24]:

$$F_1\text{-Score(Grid)} = 2 \times \frac{0.83 \times 0.09}{0.83 + 0.09} = 0.16, \tag{3a}$$

$$F_1\text{-Score(Local)} = 2 \times \frac{0.1 \times 0.84}{0.1 + 0.84} = 0.18. \tag{3b}$$

3) 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. 4. This example is a grid-induced event, as shown in Fig. 4(a). From Section II-A, such grid-induced event is supposed to have 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 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.

III. AUTOMATIC REGION IDENTIFICATION: STATISTICAL AND MACHINE LEARNING METHODS

A. 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 I 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.

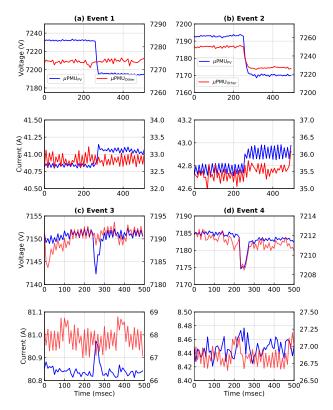Four examples are shown in Fig. 5. Event 1 (sustained) and Event 3 (transient) created clear signatures at $\mu\text{PMU}_{\text{PV}}$. But

Fig. 5. Visual inspection: (a) a local sustained event captured by μPMU$_{PV}$ with *no* signature at μPMU$_{Other}$; (b) a grid-induced sustained event captured by μPMU$_{PV}$ with *similar* signature at μPMU$_{Other}$; (c) a local transient event captured by μPMU$_{PV}$ with *no* signature at μPMU$_{Other}$; (d) a grid-induced transient event captured by μPMU$_{PV}$ with *similar* signature at μPMU$_{Other}$.

they did *not* create noticeable signatures at μPMU$_{Other}$. Thus, they must be locally-induced. Event 2 (sustained) and Event 4 (transient) created clear signatures *not only* at μPMU$_{PV}$ *but also* at μPMU$_{Other}$. Thus, they must be grid-induced. We observe that, in general, locally-induced events only create a signature on μPMU$_{PV}$; and they do *not* create any noticeable signature on μPMU$_{Other}$. 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$_{Other}$. Nevertheless, the signature at μPMU$_{Other}$ would be significantly less severe than the signature at μPMU$_{PV}$. 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 [4], [17], [18]. 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 III, we seek to replace human visual inspection with automated data-driven methods.

B. Features to Assess Signature Similarity

Let vector $\mathbf{X}_{PV}$ denote the time series of the measurements that are obtained by μPMU$_{PV}$ during an event. For example, for the case of an event that creates a signature in voltage, vector $\mathbf{X}_{PV}$ may include the time series of the voltage measure-

ments that are obtained by μPMU$_{PV}$ 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 [25]. Furthermore, let vector $\mathbf{X}_{Other}$ denote the time series of the measurements that are *simultaneously* obtained by μPMU$_{Other}$ during the *same* time window as of the event that is captured by μPMU$_{PV}$ in vector $\mathbf{X}_{PV}$.

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

$$\mathbf{X} = \mathbf{X}_{PV} \quad \text{and} \quad \mathbf{Y} = \mathbf{X}_{Other}. \tag{4}$$

1) The *root mean square (RMS) similarity* is defined as [26]:

$$rtSim(\mathbf{X}, \mathbf{Y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[1 - \frac{|x_i - y_i|}{|x_i| + |y_i|} \right]^2}, \tag{5}$$

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 [26]:

$$cos(\beta_{\mathbf{X},\mathbf{Y}}) = \frac{\mathbf{X}^T \mathbf{Y}}{\|\mathbf{X}\| \|\mathbf{Y}\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}. \tag{6}$$

3) The similarity features in (6) and (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 [26]:

$$\text{LCSS}(\mathbf{X}, \mathbf{Y}) = \frac{n + m + 2 \, \Phi(\mathbf{X}, \mathbf{Y})}{n + m}, \tag{7}$$

where $\Phi(\mathbf{X}, \mathbf{Y})$ is the relaxed LCSS recurrence function between $\mathbf{X}$, $\mathbf{Y}$; as defined in [26]. 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 [25]:

$$D(\mathbf{X}, \mathbf{Y}) = D_{n,m}, \tag{8}$$

where we recursively apply the following:

$$D_{i,j} = (x_i - y_j)^2 + min\{D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1}\}. \tag{9}$$

As in (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* [26]:

$$p_{\mathbf{X},\mathbf{Y}} = \frac{\sum_{i=1}^{n} (x_i - \overline{\mathbf{X}})(y_i - \overline{\mathbf{Y}})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{\mathbf{Y}})^2 (y_i - \overline{\mathbf{Y}})^2}}, \tag{10}$$

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 [27]:

$$s_{\mathbf{X},\mathbf{Y}} = p_{r_{\mathbf{X}}, r_{\mathbf{Y}}}, \qquad (11)$$

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 $\mathbf{Y}$ [28]:

$$\tau_{\mathbf{X},\mathbf{Y}} = \frac{2}{n^2 - n} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbb{I}\left((x_i - y_i)(x_j - y_j) \right), \quad (12)$$

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. 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 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. 6, each box indicates where the majority of the given features appear for each class of the events. The two
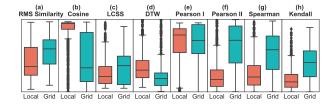

Fig. 6. The distribution of various similarity features for locally-induced events and for grid-induced events: (a)-(h) the similarity features in (5)-(12).

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.

C. 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 III-B. 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 (5), (6), or (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:

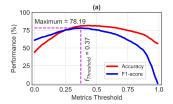$$f_u(\mathbf{X}, \mathbf{Y}) \geq f_{u,\text{Threshold}}; \qquad (13)$$

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 (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:

$$\max_{f_{u,\text{Threshold}}} \quad P\left\{ \text{Grid-Induced} \mid f_u(\mathbf{X}, \mathbf{Y}) \geq f_{u,\text{Threshold}} \right\}$$
$$\text{s.t.} \quad f_{u,\text{Threshold}}^{\min} \leq f_{u,\text{Threshold}} \leq f_{u,\text{Threshold}}^{\max}, \qquad (14)$$

where $f_{u,\text{Threshold}}$ is the optimization variable. The objective in (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 (13) holds. The maximization is done between $f_{u,\text{Threshold}}^{\min}$ and $f_{u,\text{Threshold}}^{\max}$; 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:

$$P\left\{ \text{Locally-Induced} \mid f_u(\mathbf{X}, \mathbf{Y}) < f_{u,\text{Thrfeshold}} \right\}. \qquad (15)$$
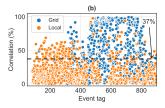
Fig. 7. An example for training and then testing the statistical method: (a) the optimal choice of the threshold parameter based on the *training* data and when two different objective functions are used; (b) the results of using the obtained optimal threshold parameter in Part (a) in the condition in (13) to conduct automated event region identification for the *test* data set.

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

An example is shown in Fig. 7. Here, the similarity feature is the Pearson correlation between voltage at μPMU_{PV} and voltage at μPMU_{Other}. Therefore, in this example, we have:

$$f_{u,\text{Threshold}}^{\min} = 0 \quad \text{and} \quad f_{u,\text{Threshold}}^{\max} = 1. \tag{16}$$

Two curves are shown in Fig. 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 (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 II is the same data set that was used in Section II-A to obtain Table I. 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 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. 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 (13).

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

TABLE II
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

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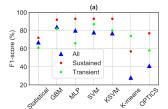
D. Machine Learning Methods

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 III-B.

Importantly, machine learning methods can resolve some of the weaknesses of the statistical methods that we previously identified at the end of Section III-C. 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 [29].

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 [30].

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. [31].

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 [32]. In this study, we use the radial basis function (rbf) kernel; because it shows the best results among other kernels.
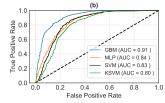
Fig. 8. Performance comparison of different data-driven methods: (a) based on F_1-Score; (b) based on ROC curves with their respected AUC.

TABLE III
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

IV. 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.

A. Strengths and Weaknesses of the Statistical Method

While the results in Table II in Section III-C 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.

1) 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 (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. 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 (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. 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.

5) K-Means Clustering: This unsupervised learning method separates the events into two clusters. Clustering is done based on the similarity features in Section III-B. The objective is to put the events for which the similarities are high between the measurements at μPMU_{PV} and μPMU_{Other} in one cluster; and the events for which the similarities are low between the measurements at μPMU_{PV} and μPMU_{Other} in another cluster. First, we 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 [33].

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 [34].

The summary comparison of the performance of the above various machine learning methods is given in Fig. 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. 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 [35]. AUC provides an aggregate measure of performance across all 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 III. 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.

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 (13).

2) 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. 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.

B. Strengths and Weaknesses of the Machine Learning Method

Recall from Section III-D 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 III 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. 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. 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,
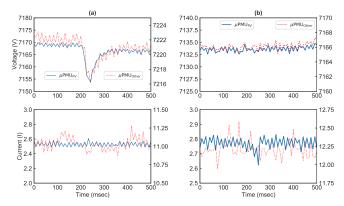

Fig. 9. Two examples for transient events that *are* correctly identified by the statistical method but are *not* correctly identified by the machine learning method: (a) a grid-induced event; (b) a locally-induced event.
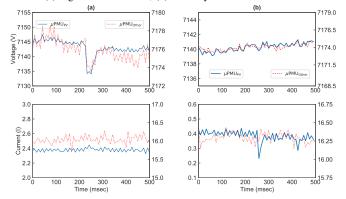

Fig. 10. Two examples for transient events that *are* correctly identified by the machine learning method but are *not* correctly identified by the statistical method: (a) a grid-induced event; (b) a locally-induced event.

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.

1) 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.

C. Mixed-Integrated Algorithm

Based on the analysis in Sections IV-A and IV-B, 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.

The performance summary of Algorithm 1 is given in Table IV. This table is comparable with Tables II and III. We can see that the shortcomings of the statistical method

---

Algorithm 1 : Mixed-integrated Event Region Identification

Input: Captured event at μPMU_{PV} and μPMU_{Other}.
Output: Identified region of the event.

1: **if** the event is transient **then**
2:     **if** the dominant signature is in voltage **then**
3:         **if** the similarity check in (13) holds **then**
4:             Use statistical method
5:         **else**
6:             Use machine learning method
7:         **end if**
8:     **end if**
9:     **if** the dominant signature is in current **then**
10:         **if** the similarity check in (13) does not hold **then**
11:             Use statistical method
12:         **else**
13:             Use machine learning method
14:         **end if**
15:     **end if**
16: **else**
17:     Use machine learning method
18: **end if**

TABLE IV
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

Fig. 11. Comparing the F_1-Score for all three data-driven methods that are discussed in this paper, for grid-induced events and locally-induced events.

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.

The results in Tables II, III, and IV 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.

D. 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 II, III, IV-A, IV-B, and IV-C 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.
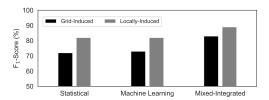
Importantly, we do *not* update the training of the proposed method in this new case study. In other words, the entire new data set in this section is used only as a *test* data set. Furthermore, while the test data set in Sections II, III, IV-A, IV-B, and IV-C was smaller than the training data set, 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 II, III, IV-A, IV-B, and IV-C. 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. 11. Here we compare the performance of all the three data-driven methods that we discussed in this paper, i.e., the statistical method that was designed in Section III-C, the machine learning method that was designed in Section III-D, and the mixed-integrated method that was proposed in Section IV-C. 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 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.

V. 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.

A. Applications of Analyzing Grid-Induced Events

1) 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 [22], [36].

Consider the *adaptive* VVC method that is proposed in [37] and shown in Fig. 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 [37] 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. 12(a). The latter is referred to as *slope adjustment* and it is done by changing parameter m, as shown in Fig. 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 paper. For any such event, we obtain the *steady state error* (SSE) and the *voltage flicker* (VF) as follows [37]:

$$\text{SSE} = \sum_{t=1}^{T}(V_t - \mu)/T, \tag{17a}$$

$$\text{VF} = \sum_{t=1}^{T}\frac{(V_t - V_{t-1})/V_t}{T} \times 100, \tag{17b}$$

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 [37]. The results for changing VVC parameter q in the case of 20 examples of sustained grid-induced voltage events are shown in Fig. 13(a).

As for the slope adjustment method, it uses two thresholds on the value of VF, per the IEEE 141 standard [38]. 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. 13(b).
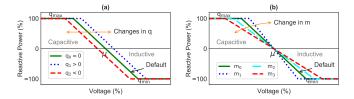

Fig. 12. The adaptive VVC according to the sustained grid-induced voltage events: (a) VVC error adjustment; (b) VVC slope adjustment.
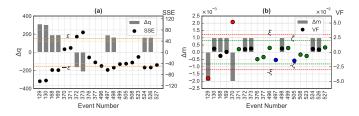

Fig. 13. Adaptive VVC based on 20 examples of the sustained grid-induced voltage events: (a) the changes in q based on the calculation of SSE at each event; (b) the changes in m based on the calculation of VF at each event.

2) 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. 14(a)-(b) and Fig. 14(c)-(d).

The grid-induced event in the first example is a *step-change* in voltage, as shown in Fig. 14(a). The dynamic response of the behind-the-meter solar farm to this grid-induced disturbance is shown in Fig. 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 [18]. 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 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. 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. 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 [39], and to evaluate the compliance of the inverters with the inter-connection rules [40].
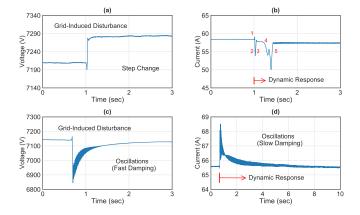
Fig. 14. Dynamic response of the solar farm to two grid-induced events: (a) a sustained step change in voltage; (b) the dynamic response of the solar farm (and its various stages) to such sustained step change; (c) a temporary voltage sag with subsequent momentary oscillations that stabilize quickly; (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.
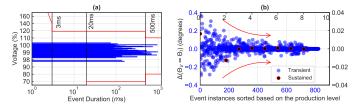

Fig. 15. Applications of analyzing locally-induced events: (a) voltage magnitude and duration for the locally-induced transient events enveloped in the ITIC curve. Mitigation actions must be done if the event exceed the green area; (b) identifying the relationship between the occurrence and significance of the locally-induced events and the production-level of the solar farm.

B. Applications of Analyzing Locally-Induced Events

1) 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 [23]. 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. 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.

2) 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. 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. 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 [41].

VI. CONCLUSIONS AND FUTURE WORK

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.

The analysis in this paper 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. 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 this paper. 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.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] M. Farajollahi, A. Shahsavari, and H. Mohsenian-Rad, "Location identification of high impedance faults using synchronized harmonic phasors," in *Proc. IEEE PES ISGT*, pp. 1–5, 2017.

[13] 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.

[14] 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.

[15] 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.

[16] 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*, pp. 1–6, 2017.

[17] 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*, pp. 1–7, 2017.

[18] P. Khaledian, A. Aligholian, and H. Mohsenian-Rad, "Event-based analysis of solar power distribution feeder using micro-PMU measurements," in *Proc. IEEE ISGT*, pp. 1–5, Washington DC, February, 2021.

[19] 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.

[20] 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.

[21] 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*, pp. 71–80, 2016.

[22] 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.

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

[24] 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*, pp. 245–251, 2013.

[25] 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.

[26] 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.

[27] 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.

[28] A. Geller, *Calculating Kendall's Tau with multiple measurements*. PhD thesis, 2018.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

[34] 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.

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

[36] 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.

[37] 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.

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

[39] 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.

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

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

Parviz Khaledian (S'17) received the B.Sc. degree in electrical engineering from University of Isfahan, Isfahan, Iran, in 2009, and the M.Sc. degree in electrical engineering from University of Idaho, Moscow, ID, USA, in 2018. He is working towards a Ph.D. degree in electrical engineering at the University of California,Riverside, CA, USA. His research interests include renewable energies, power system monitoring and situational awareness, and data analysis.

Hamed Mohsenian-Rad (M'09-SM'14-F'20) received the Ph.D. degree in electrical and computer engineering from the University of British Columbia, Vancouver, BC, Canada, in 2008. He is currently a Professor of electrical engineering and a Bourns Family Faculty Fellow at the University of California, Riverside, CA, USA. His research is on monitoring, data analysis, and optimization of power systems and smart grids. He is the author of the textbook *Smart Grid Sensors: Principles and Applications* by Cambridge University Press - 2022.
He was the recipient of the National Science Foundation (NSF) CAREER Award, the Best Paper Award from the IEEE Power & Energy Society General Meeting, and the Best Paper Award from the IEEE Conference on Smart Grid Communications. He has been the PI on ten million dollars research grants in the area of smart grid. He has served as Editor for the IEEE TRANSACTIONS ON POWER SYSTEMS, IEEE TRANSACTIONS ON SMART GRID and the IEEE POWER ENGINEERING LETTERS.