

# Event-Triggered Distribution System State Estimation: Sparse Kalman Filtering with Reinforced Coupling

Alireza Akrami, *Student Member, IEEE* and Hamed Mohsenian-Rad, *Fellow, IEEE*

**Abstract**—A novel distribution system state estimation (DSSE) method is proposed for power distribution networks with low-observability, where the measurements come from only a few distribution-level phasor measurement units (D-PMUs). The proposed DSSE method is *event-triggered*, which means the state variables are updated based on the information that is extracted from the events in the power distribution system. In this regard, the estimations of the state variables during the previous events are used as *a priori* information to predict the state variables at the current event. Accordingly, a novel data-driven method based on elastic net regression analysis is proposed to learn the event-triggered state transition matrix. The DSSE problem is formulated as a *generalized group Lasso* problem, which is augmented based on the knowledge on the sparsity patterns of the state variables that are extracted from the analysis of the events. Here, in the absence of direct power measurements, we enhance our ability in sparse recovery by developing a new reinforced physics-based coupling method among the state variables, in which we add a novel set of linear *differential* power flow equations to the DSSE problem formulation in forms of virtual measurements. Finally, two different approaches are proposed to solve the formulated sparse event-triggered DSSE problem. The first approach is exact but computationally expensive, as it requires conducting a batch alternating direction method of multipliers (ADMM) analysis. The second approach is approximate, but it is much faster as it works based on a novel modified Kalman filter/smoothen in the presence of  $\ell_1$ -norm.

**Index Terms**—Event-triggered DSSE, low-observability, sparsity, physics-based coupling, virtual measurements, Kalman filter, elastic net regression, distribution synchrophasors, D-PMU.

## I. INTRODUCTION

### A. Motivations

Distribution system state estimation (DSSE) is a core module in monitoring and operation of power distribution systems. The purpose of DSSE is to estimate the power system state variables from the available power system measurements. The measurements may include slow reporting data from SCADA sensor devices and smart meters, and fast reporting and high resolution data from distribution-level phasor measurement units (D-PMUs) [1]. In the development of a DSSE method, there are key challenges that need to be addressed:

1) *Low-Observability*: Unlike the power transmission systems are that usually well-instrumented, power distribution systems are usually *not* fully-observable, i.e., they often lack sufficient measurements to the extent that we cannot uniquely estimate the state variables due to facing underdetermined equations [2]. Due to the expanded size of the feeders and the limitations of the communication infrastructure, the grid operator *cannot* install sensors at every location in the power

distribution system to achieve full-observability. Therefore, a practical DSSE method must deal with the inevitable low-observability circumstances in power distribution system.

2) *Dynamic and Time-Varying System*: A typical sudden change in loads or in the output power of renewable generators can cause a sudden deviation in the state variables in a power distribution system. Thus, *static* DSSE methods may not capture and track these frequent changes; because of their inherent assumption on steady-state operation of the power distribution systems [3]. This is particularly the case in DSSE methods that rely on *pseudo-measurements*, alongside other measurements to maintain full-observability [4]. Meanwhile, it is crucial to establish an accurate state transition model.

3) *Computation Burden*: While the high reporting rate of D-PMUs can potentially give the utility a unique opportunity to enhance situational awareness, processing the heavy streams of D-PMU data remains a challenge by itself. Interestingly, in practice, a major portion of the D-PMU data does *not* carry much useful information [5]–[8]. Therefore, a DSSE mechanism that focuses on extracting the most informative aspects of the measurements under the inherent low-observability in a power distribution system can inevitable also help with reducing the computational and communication burden.

### B. Technical Contributions

In this paper, our goal is to develop a DSSE method which can address the above aforementioned challenges. The main contributions of this paper can be summarized as follows:

- 1) The proposed DSSE method is *event-triggered* and it uses the voltage and current phasor measurements from only a few D-PMUs. Accordingly, the state variables are defined in differential mode under low-observability conditions, where all the changes in the state variables, i.e., state transitions, are due to the occurrences of physical events.
- 2) The introduced event-triggered DSSE problem is formulated as a *generalized group Lasso* problem, which is augmented by sparsity patterns, not only in voltage and current phasors, but also in power injections. While we do *not* include direct power measurements in our model, we reinforce the DSSE problem formulation by a novel set of linear differential power flow equations, to be used as virtual measurements to enhance the physics-based coupling that exists among the differential state variables.
- 3) A data-driven method is developed to learn the event-triggered state transition matrix among the differential state variables. It works as a discriminative elastic net regression method, and it can capture the sparsity in such matrix due to the radial topology and the spatial-temporal correlations of the state variables.

The authors are with the Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA. The corresponding author is H. Mohsenian-Rad, e-mail: hamed@ece.ucr.edu.

TABLE I  
SUMMARY OVERVIEW AND COMPARISON WITH THE EXISTING DSSE METHODS

Method	Event-Triggered	Observability	Setting	State Transition Matrix	Physics-Based Sparsity Features
[9]	No	Full-Observability <sup>†</sup>	Dynamic	Time-Variant	Zero Injection Nodes
[10][11]	No	Full-Observability <sup>†</sup>	Dynamic	Identity	No
[12]	No	Full-Observability <sup>†</sup>	Dynamic	Identity	Temporal Correlation
[13][14][15][16]	No	Full-Observability <sup>†</sup>	Static	N/A	No
[17][18]	No	Low-Observability	Static	N/A	No
[19][20]	Yes	Full-Observability <sup>†</sup>	Dynamic	Identity	No
[21]	Yes	Low-Observability	Static	N/A	Sparsity in Voltage and Current
Proposed	Yes	Low-Observability	Dynamic	Time-Variant	Sparsity in Differential Power Flow Equations

<sup>†</sup>Full-observability is achieved either through sufficient measurements or by using pseudo-measurements.

- 4) Two different approaches are proposed to solve the formulated sparse event-triggered DSSE problem. The first approach is exact and works based on the batch alternating direction method of multipliers (ADMM). The second method is approximate to further lower computational complexity. It works in three steps: Kalman filtering, incorporating sparsity, and backward smoothing.

### C. Literature Review

Table I shows an overview summary of the related literature, in comparison with the proposed method. So far, the most common approach to address low-observability has been to increase the observability by using pseudo-measurements, whether in static DSSE [10], [15], [16], or in dynamic DSSE [9]–[11]. Pseudo-measurements are typically generated from historical or proxy data and by using methods such as artificial neural networks [13] or Gaussian mixture model [14].

A more recent approach to tackle low-observability is to use sparse recovery tools from signal processing to solve the DSSE problem *without* making it fully-observable [17], [18], [21]–[23]. However, all the existing methods in this line of work so far seek to solve the DSSE problem in a *static* setting, i.e., they do not use any dynamic model, such as a state transition matrix, in the formulation of the problem. Therefore, they cannot capture the dynamic nature of the modern power distribution systems.

Importantly, a method that is dedicated to address the DSSE problem, not only under *low-observability* conditions but also in a *dynamic* setting, is still missing. In [10], a dynamic DSSE method is proposed using Iterated Kalman Filter (IKF). In [11], a past-aware DSSE method is proposed which uses the estimation results from the previous time slots to correct the estimation for the current time slot through the use of Ensemble Kalman filter (EnKF). In [12], a forecasting-aided DSSE method is proposed by using a robust EnKF to increase redundancy in measurements. However, the dynamic DSSE methods in [10]–[12] do *not* address low-observability and they do *not* utilize the inherent sparsity in the DSSE problem.

Another aspect that is relevant to dynamic DSSE methods is the *state transition model*. A common assumption in the existing limited literature on dynamic DSSE is that, the state transition matrix is an identity matrix; e.g., see [10]–[12]. This quasi-steady state assumption is valid for a *time-triggered* dynamic system with *smooth* changes in the state variables.

However, this simplified model is *not* valid for an *event-triggered* system with *sudden* and *sharp* changes. Of course, given the low-observability of the power distribution feeders, addressing this shortcoming is particularly challenging.

Although there are few studies that use event-triggered models in the broader field of power system state estimation, they have focused on wide area measurements in power *transmission* systems. In [24], a master-slave non-linear filtering structure is developed for the state estimation module for power transmission networks in order to lower the computation burden of state estimation task. In [25], an event-triggered state estimator is developed based on particle filter design to relieve the computation burden due to widespread use of distributed generators in power systems. In [26], an event-based cubature Kalman (EBCKF) filter is proposed, which guarantees that only the observations with new information are transmitted to the Kalman filter for running state estimation. In [27], an event-triggered state estimation is developed for monitoring the electromechanical dynamic states (dynamic modes) of the generators, where the measurements come from PMUs.

Recently, a DSSE method is proposed in [19], which is based on event-triggered data transmission. It uses a special closed-loop event triggering condition to utilize the limited data communication bandwidth more efficiently during state estimation. In [20], a forecasting-aided event-triggered DSSE method is proposed which uses a component-based event-triggered mechanism, and an unscented Kalman filter (UKF), to perform DSSE by using different types of measurements.

However, the above limited literature on event-triggered state estimation in [19], [20], [24]–[27] does *not* address the low-observability issue, which is the primary challenge in this paper. They also do *not* involve the learning mechanisms for the state space model which is another key focus in this paper.

## II. SYSTEM MODEL AND ASSUMPTIONS

Consider a three-phase power distribution network that is represented by graph  $\mathcal{G} := (\mathcal{N}, \mathcal{L})$ , where  $\mathcal{N}$  denotes the set of three-phase nodes and  $\mathcal{L}$  denotes the set of three-phase distribution lines. For notational simplicity, we skip using a specific index for the phases; however, all the phasors in our formulations are assumed to be in three-phase. We assume that the understudy network has a typical radial topology.

### A. State Variables

Traditionally, it is very common in the power systems literature to model the power system dynamics using *time-triggered* state space models, in which the state variables in the system may change on fixed time intervals. However, in principle, one can model a dynamic system also by using *event-triggered* state space models, in which the state variables in the system may change due to the occurrences of events [28]. The advantage of considering the event-triggered paradigm is that we can run and solve the DSSE problem only when a major event is detected in the system, rather than doing so every single time that a measurement is obtained.

Here, an ‘‘event’’ is defined broadly. For example, an event can be a load switching, or an equipment switching, such as a capacitor bank switching, a sudden change in the generation level of a renewable generator. In all these cases, the event may cause changes in the state variables in the system; which in turn triggers a new state estimation.

Our study in this paper focuses only on the events that happen at the buses of the system. That is, our proposed DSSE does not support those events that may happen on the branches of the system, i.e., faults or topology changes. Most line faults are usually analyzed separately [29]; and topology identification too is often a separate task [30]. Given the fact that our focus is on networks with low-observability, it is indeed preferred to keep those tasks separately; leaving the focus of our sparse and dynamic DSSE algorithm on state estimation during normal operation.

At each event  $k$ , let  $\mathbf{V}_k$  denote the vector of voltage phasors at every node in set  $\mathcal{N}$ . Also, let  $\mathbf{I}_k$  denote the vector of current phasors at every line in set  $\mathcal{L}$ . In this study, we represent all phasors in the *rectangular* coordinates. Thus, each phasor has two associated quantities, i.e., the real part and the imaginary part. Similarly, let  $\mathbf{S}_k$  denote the vector that contains the injected apparent power to all the nodes in  $\mathcal{N}$  at event  $k$ .

Suppose  $\mathbf{x}_k$  denotes the vector that contains  $\mathbf{V}_k$ ,  $\mathbf{I}_k$ , and  $\mathbf{S}_k$ . Unlike most other formulations in the DSSE literature, we define the state variables in *differential* mode. We define  $\Delta\mathbf{x}_k$  as the vector of differential state variables as follows:

$$\Delta\mathbf{x}_k := \mathbf{x}_k - \mathbf{x}_{k-1}. \quad (1)$$

Here,  $\Delta\mathbf{x}_k$  denotes the vector of the changes in all the system state variables that are caused due to the occurrence of event  $k$ , in comparison with the status of the system at event  $k-1$ . The vector of state variables in our problem formulation contains all differential voltage phasors, all differential current phasors, and all differential apparent power injections:

$$\Delta\mathbf{x}_k := [ (\Delta\mathbf{V}_k)^\top (\Delta\mathbf{I}_k)^\top (\Delta\mathbf{S}_k)^\top ]^\top. \quad (2)$$

### B. Available Measurements and Network Observability

Since the state variables in this paper are defined in *differential* mode, we need to know their initial values, such that we can add them to the estimated differential phasors to obtain the regular state variables after each event. Hence, we assume that at  $k=0$ , the initial values of the state variables are known. Subsequently, during the events  $k=1, \dots, K$ , we assume that

*low-observability* condition holds, where the measurements are limited to the voltage and current synchrophasors at only a few D-PMUs on the power distribution feeder. We only rely on the measurements of D-PMUs, which have *high reporting rates*, because other types of measurements, such as from legacy meters or pseudo-measurements, do not capture the sudden changes that occur in the state variables due to the events [9].

At each event  $k$ , let  $\mathbf{y}_k$  denote the vector of voltage phasor measurements at the nodes in set  $\mathcal{N}_m \subseteq \mathcal{N}$  and the current phasor measurements at the lines in set  $\mathcal{L}_m \subseteq \mathcal{L}$ . Similar to (1), we can define the vector of differential phasor measurements in the rectangular coordinates as:

$$\Delta\mathbf{y}_k := \mathbf{y}_k - \mathbf{y}_{k-1}. \quad (3)$$

Therefore, measurement vector  $\Delta\mathbf{y}_k$  can be defined as:

$$\Delta\mathbf{y}_k := [ (\Delta\mathbf{V}_k^m)^\top (\Delta\mathbf{I}_k^m)^\top ]^\top, \quad (4)$$

where superscript  $m$  indicates the measurements.

*Remark 1:* As we can see in (4), we do *not* collect any power measurements in the proposed formulation of the DSSE problem. Yet we want to reinforce our DSSE problem formulation with physics-based couplings that exist between the differential state variables in (2), because of the advantages of doing so; which we will unmask throughout this paper. Therefore, despite not having power measurements in (4), we have intentionally included the differential power phasors in the vector of state variables in (2), such that their values are estimated as a part of our proposed DSSE problem.

Based on the available measurements, we introduce measurement matrix  $\mathbf{H}_k$  to relate the measurements to the state variables. In its basic form, matrix  $\mathbf{H}_k$  has three block rows, denoted by  $\mathbf{H}_k^1$ ,  $\mathbf{H}_k^2$ , and  $\mathbf{H}_k^3$ , as we will explain next. We will also discuss a fourth block, denoted by  $\mathbf{H}_k^4$ , in Section II-E.

In the first set of equations, the differential voltage phasor measurements are mapped to the differential voltage phasors in the vector of state variables via an identity mapping:

$$\mathbf{H}_k^1 := [ \mathbf{U} \ \vdots \ \mathbf{0} \ \vdots \ \mathbf{0} ], \quad (5)$$

where  $\mathbf{U}$  is the adequate identity block matrix.

In the second set of equations, the available differential current phasor measurements are mapped to the differential voltage phasors in the vector of state variables by applying the Kirchhoff’s current law (KCL), as follows:

$$\Delta\mathbf{I}_k^m = \mathbf{Y}\Delta\mathbf{V}_k, \quad (6)$$

where  $\mathbf{Y}$  is the admittance matrix. Accordingly, we have:

$$\mathbf{H}_k^2 := [ \mathbf{Y} \ \vdots \ \mathbf{0} \ \vdots \ \mathbf{0} ]. \quad (7)$$

The equations in (6) only include a small subset of line segments, i.e., those in  $\mathcal{L}_m$ . For the rest of the line segments, i.e., those in  $\mathcal{L} \setminus \mathcal{L}_m$ , we do not have any direct measurements. Thus, we instead add the KCL equations associated with the line segments that are not equipped with D-PMUs as auxiliary equations to the DSSE problem formulation as follows:

$$\mathbf{0} = \mathbf{Y}\Delta\mathbf{V}_k - \Delta\mathbf{I}_k. \quad (8)$$

Accordingly, we have:

$$\mathbf{H}_k^3 := [ \mathbf{Y} \ \vdots \ -\mathbf{U} \ \vdots \ \mathbf{0} ]. \quad (9)$$

### C. Event-Triggered State-Space Model

In this Section, we formulate the DSSE problem in a dynamic setting. Similar to the literature, such as in [4], [9], [11], we assume that the state-space equation is in *linear* form. As discussed in [4], a linear state-space model is commonly used in dynamic DSSE, also known as forecasting-aided state estimation (FASE). Linear state-space models can achieve acceptable accuracy, provided that the power distribution system operates under normal conditions, i.e., abnormal events such as faults or topology changes, that suddenly cause a major deviation in the operating points of the power distribution system, do *not* happen. For those abnormal conditions, which are beyond the scope of this work, one may need to use the more complex non-linear state-space models.

From this, and since our focus in this paper is on differential mode; we can express the state space model as:

$$\Delta \mathbf{x}_k = \mathbf{A}_k \Delta \mathbf{x}_{k-1} + \mathbf{q}_k, \quad (10)$$

where  $\mathbf{A}_k$  is the state transition matrix at event  $k$  and  $\mathbf{q}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$  is the zero-mean process noise with covariance  $\mathbf{Q}_k$ .

Since our focus is on *event-triggered state estimation*, the DSSE problem is solved *only if* a major event occurs in the network. The occurrences of such major events are detected by the existing *event-detection* methods in the literature. In fact, there is a rich literature not only for event-detection but also for event location identification which work under the *same* low-observability conditions that we consider in this paper. Examples of such methods are in [8], [31] for event detection and in [32]–[34] for event location identification. Specifically, the methods in [8] and [32] require the presence of one D-PMU at the substation and one D-PMU at the end of each lateral; which is the same requirement for the availability of the D-PMUs that we have considered in this paper.

Thus, for the rest of this paper, we assume that the occurrences of the events and their exact or at least approximate locations are known. We assume that *only one event can happen at a time*. The challenge is to formulate and solve the DSSE problem that is *triggered* by an event.

*Remark 2:* The study of events in this paper is based on the findings in [8], [31], [32], [35] about event-detection in real-world D-PMU data. They have shown that, in practice, it is almost always the case that only one major event may happen at a time on a typical distribution feeder. Accordingly, unless it is stated otherwise, we assume in this paper that only one event happens at a time. However, for the special case where one major event and multiple smaller events happen at the same time, the proposed method can still perform reasonably well, as we will see in a case study in Section V-F.

### D. Sparsity in Voltage and Current Phasors

Consider an event that occurs on the feeder. Consider the *path* between the substation and the node where the event occurs. Let us denote such path by a tree  $\mathcal{T} := (\mathcal{V}, \mathcal{E})$ . Let us denote the rest of the network by  $\mathcal{T}' := (\mathcal{V}', \mathcal{E}')$ . Note that:

$$\mathcal{T} \cup \mathcal{T}' = \mathcal{G} \quad \text{and} \quad \mathcal{T} \cap \mathcal{T}' = \emptyset. \quad (11)$$

As it is shown in [21], once a major event happens at the distribution feeder, the vector of differential voltage phasors and the vector of differential current phasors become *sparse*. Only the differential voltage phasors at the nodes in  $\mathcal{V}$ , and only the differential current phasors at the line segments in  $\mathcal{E}$  would be *non-zero*. The differential current phasors for all the line segments outside tree  $\mathcal{T}$ , i.e., those in  $\mathcal{E}'$  would be *zero*. As for the nodes in  $\mathcal{V}'$ , *group sparsity* holds among the differential voltage phasors, as there would be groups of nodes whose differential voltage phasors would be either *all* (approximately) zero or *all* non-zero. That is, in each group, if the differential voltage phasor is zero for one node, then it would be zero for the rest of nodes that are on the same group. For more details please refer to [21].

### E. Sparsity in Power Injections

Let us assume that all the load and generation components at *each node* on the power distribution feeder are collectively modeled as a constant power component. This is a common model, e.g., see [36], [37]. An important implication of this assumption is that, once an event occurs, the *differential power injection* is zero at all of the nodes across the power distribution feeder, *except* at the node where the event occurs. Since the event detection method gives us the location of the event, c.f. [32], we accordingly know at which nodes the differential power injection phasors are zero. By incorporating this information with the DSSE problem through the power flow equations, we can derive an additional set of equations to enhance the observability of the power distribution system.

Throughout this paper, we call the above additional set of differential zero-injection equations as *virtual measurements*, as we do not have any direct measurement for the power. While these additional equations do not necessarily make the underdetermined system of equations in the DSSE problem full-rank, they *reinforce* the DSSE problem formulation with *physics-based coupling* among the state variables, which helps us with recovering the state variables at each event.

The above idea can be materialized by incorporating the power flow equations into our analysis. Since we defined the state variables in differential mode, we need to modify the standard power flow equations accordingly. After that, we need to linearize the resulting differential power flow equations.

Interestingly, the linearization of the power flow equations in differential mode is justified mathematically beyond the typical linearization of the standard power flow equations. The details are given in the Appendix. The resulting linearized differential power flow equations are obtained in an abstract form as:

$$\mathbf{0} = \mathbf{J}_k \Delta \mathbf{V}_k - \Delta \mathbf{S}_k, \quad (12)$$

where  $\mathbf{J}_k$  is the Jacobian matrix as explained in the Appendix.

From (12), we can now obtain the last block row of the measurement matrix which we denote by  $\mathbf{H}_k^4$ , as follows:

$$\mathbf{H}_k^4 := \left[ \mathbf{J}_k \mid \mathbf{0} \mid -\mathbf{U} \right]. \quad (13)$$

The sparsity patterns that we have extracted in this paper with respect to the differential voltage phasors, differential current phasors, differential power injection phasors, and differential power flow equations are all with respect to the physics

of the power distribution system, i.e., the radial topology of the feeder, the Ohm's law, and the Kirchhoff's laws. That is the reason why we refer to these sparsity features and associated relationships as "physics-based" throughout this paper.

*Remark 3:* We do *not* use any power measurement or any power pseudo-measurement, yet we *do* take advantage of the power flow equations in differential mode by including the fourth block row in the measurement matrix. Thus, we catch the physics-based couplings that exist among the differential voltage phasors, differential current phasors, and the differential power injections in an event-triggered DSSE problem.

#### F. Event-Triggered DSSE Formulation

For notational simplicity, we introduce  $\Delta \mathbf{z}_k$  as a new vector that includes all the measurements in differential mode. Vector  $\Delta \mathbf{z}_k$  is an extension of vector  $\Delta \mathbf{y}_k$  in (4), where we also add the additional rows of *zeros* corresponding to the auxiliary equations in (8) and the virtual measurements in (12):

$$\Delta \mathbf{z}_k = [\Delta \mathbf{y}_k \quad \mathbf{0}]. \quad (14)$$

Accordingly, we can obtain the following ultimate relationship between the new vector of differential measurements and the vector of differential state variables:

$$\Delta \mathbf{z}_k = \mathbf{H}_k \Delta \mathbf{x}_k + \mathbf{r}_k, \quad (15)$$

where matrix  $\mathbf{H}_k$  in (15) is constructed by putting together matrices  $\mathbf{H}_k^1$ ,  $\mathbf{H}_k^2$ ,  $\mathbf{H}_k^3$ , and  $\mathbf{H}_k^4$ . The added term  $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$  is the measurement noise that is assumed to be zero-mean Gaussian with covariance matrix  $\mathbf{R}_k$ . The measurement noises at different D-PMUs are assumed to be mutually independent.

Together, the relationships in (10) and (15) provide the core formulation for the event-triggered DSSE problem. Given the fact that the state estimation problem is *dynamic*, the state estimation process is done in the following two stages.

**First Stage:** In this stage, at the beginning of each sequence, the state variables at the current sequence are predicted based on their values in the previous sequence and by using the state-space model in (10). The outcome of event detection is directly taken into consideration in (10). An important aspect in the analysis in this first stage is to have an *accurate state transition matrix* as part of the event-triggered state space model in (10). In this paper, we use Elastic Net Regression to learn the state transition matrix. We will discuss this subject in Section III.

**Second Stage:** In this stage, the measurement model in (15) is used to correct the prediction of the state variables. Typically, a Kalman filter is used, such as the Iterated Kalman filter [10] or the ensemble Kalman filter [11], to conduct the aforementioned correction task. In this paper, we use a novel sparse Kalman filter/smoothen, given that our focus is on sparse recovery in order to address the low-observability in the power distribution system. We will discuss this subject in Section IV.

### III. OFFLINE LEARNING OF TRANSITION MATRIX VIA ELASTIC NET REGRESSION

As we mentioned in Section II-F, the accuracy in predicting the current state variables based on the event-triggered state

space model in (10) depends on the accuracy of the state transition matrix  $\mathbf{A}_k$ . However, obtaining matrix  $\mathbf{A}_k$  is a challenging task. In fact, it is common in the literature to assume that the state transition matrix in the DSSE problem is always an *identity* matrix [10]. While there have been some recent efforts, such as in [38], to estimate the time-varying state transition matrix, no prior study has aimed to obtain such matrix under the *low-observability* conditions. Also, no prior method is designed for *event-triggered* DSSE formulations.

#### A. Event-Triggered State Transition Matrix

Suppose event  $k$  is detected and its location is identified as bus  $s_k$ . Suppose the location of the *previous* event, i.e., event  $k-1$ , is bus  $s_{k-1}$ . Our goal is to obtain:

$$\mathbf{A}_k = \mathbf{B}(s_{k-1}, s_k) \quad (16)$$

for the transition of the system in differential mode from the moment when an event occurs at bus  $s_{k-1}$  to the moment when a subsequent event occurs at bus  $s_k$ . In order to explain the role of matrix  $\mathbf{B}$ , let us first consider an example. Suppose  $K = 15$ , and the location of the events are identified as buses 5, 8, 15, 1, 7, 10, 4, 8, 15, 6, 4, 8, 2, 9, 12, respectively. Notice that, at event  $k = 3$ , we have  $s_{k-1} = 8$  and  $s_k = 15$ . Similarly, at event  $k = 9$ , we have  $s_{k-1} = 8$  and  $s_k = 15$ . However, at event  $k = 13$ , we have  $s_{k-1} = 8$  and  $s_k = 2$ . Therefore, from the relationship in (16), we set

$$\mathbf{A}_3 = \mathbf{A}_9 = \mathbf{B}(8, 15), \quad \text{and} \quad \mathbf{A}_{13} = \mathbf{B}(8, 2). \quad (17)$$

Similarly, we can obtain matrix  $\mathbf{A}_k$  for *any* event  $k$  based on matrix  $\mathbf{B}$  that we learn using historical data. Here, matrix  $\mathbf{B}(n, m)$  itself is defined as the event-triggered state-transition matrix that captures the changes in the state variables in the power system for any case where the previous event occurs at bus  $n$  and the current event occurs at bus  $m$ .

In total, we have  $(|\mathcal{N}| - 1) \times (|\mathcal{N}| - 1)$  possible transitions from one event location to another event location. Therefore, there are exactly  $(|\mathcal{N}| - 1) \times (|\mathcal{N}| - 1)$  matrices  $\mathbf{B}(n, m)$ .

For any two buses  $n$  and  $m$  in set  $\mathcal{N} \setminus \{1\}$ , we can obtain matrix  $\mathbf{B}(n, m)$  by examining the set of all historical events of index  $\kappa$  for which we have  $s_{\kappa-1} = n$  and  $s_\kappa = m$ . We can learn the event-triggered state transition matrix  $\mathbf{B}(n, m)$  by solving the following optimization problem:

$$\mathbf{B}(n, m) = \arg \min_{\mathbf{B}} \sum_{\kappa \in \mathcal{K}} \|\Delta \mathbf{x}_\kappa - \mathbf{B} \Delta \mathbf{x}_{\kappa-1}\|_2^2 \quad (18)$$

where

$$\mathcal{K} = \{\kappa \mid s_{\kappa-1} = n, s_\kappa = m\}. \quad (19)$$

The above training is done *offline* by analyzing the events in the historical data. Once we obtain  $\mathbf{B}(n, m)$  for all the  $(|\mathcal{N}| - 1) \times (|\mathcal{N}| - 1)$  possible event-triggered transitions, we can obtain  $\mathbf{A}_k$  for any event  $k$  by using (16). Note that, the terms "learning" and "training" that are mentioned above are associated with the process of calculating the entries of the event-triggered state transition matrix. We refer to this process as learning; because it is data-driven; it works by applying the elastic net regression method to the historical data.

### B. Elastic Net Regression

While the formulation in (18) is conceptually valid, we may face numerical issues in solving this optimization problem. Such numerical issues are due to the inherent sparsity in matrix  $\mathbf{B}$ , which is due to the same reasons that we mentioned in Section II-D. In fact, since each node in a radial network topology is connected to very few nodes, each row of matrix  $\mathbf{B}$  is highly sparse. Therefore, we propose to formulate the optimization problem in (18) as a *discriminative elastic net regression* [39]. The new formulation is obtained as:

$$\mathbf{B}(n, m) = \arg \min_{\mathbf{B}} \sum_{\kappa \in \mathcal{K}} \|\Delta \mathbf{x}_{\kappa} - \mathbf{B} \Delta \mathbf{x}_{\kappa-1}\|_2^2 + \lambda_2 \|\mathbf{B}\|_F^2 + \lambda_1 \|\mathbf{B}\|_{1,1}, \quad (20)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm and  $\|\cdot\|_{1,1}$  denotes the  $l_{1,1}$  matrix norm. Parameters  $\lambda_1$  and  $\lambda_2$  are weight factors.

The reason for using the above elastic net regression in this study is the fact that, each row of matrix  $\mathbf{B}$  is sparse due to the radial topology of the distribution feeder, yet the non-zero entries are *correlated* because of the spatial-temporal correlation of the state variables in a power distribution feeder [17]. Elastic net regression has the advantage to make a trade-off between these two properties by adjusting the penalty parameters  $\lambda_1$  and  $\lambda_2$ . Similar to any other regression model, the values of these weight parameters are determined based on the values of data in practice. The larger  $\lambda_1$  is, the more sparse and less correlated the entries of matrix  $\mathbf{B}$  would be. Also, larger values for  $\lambda_2$  would lead to higher correlation between the entries of matrix  $\mathbf{B}$ , which lowers the sparsity.

The solution for the elastic net regression is a matrix, whose element in row  $i$  and column  $j$  is obtained as:

$$b[ij] = \frac{\{|b_{LS}[ij]| - \lambda_1/2\}^+}{1 + \lambda_2} \text{sign}\{b_{LS}[ij]\} \quad (21)$$

where  $\{\cdot\}^+ = \max\{\cdot, 0\}$  and  $b_{LS}[ij]$  is the element in row  $i$  and column  $j$  of the following matrix:

$$\sum_{\kappa \in \mathcal{K}} \Delta \mathbf{x}_{\kappa-1} (\Delta \mathbf{x}_{\kappa})^T. \quad (22)$$

*Remark 4:* In a feeder with a large number of nodes, it may not be possible to capture all the  $(|\mathcal{N}| - 1) \times (|\mathcal{N}| - 1)$  transitions in the system. One solution is to group multiple nodes into one zone, such that instead of finding the transition from the nodal location of one event to the nodal location of the next event, we seek to find the transition from the zonal location of one event to the zonal location of the next event.

*Remark 5:* In addition to the locations of the previous event and the current event, the state transition matrix  $\mathbf{A}_k$  may depend on other factors, such as the type of the previous event and the type of the current event. Such additional factors can be taken into consideration by revising (16) as follows:

$$\mathbf{A}_k = \mathbf{B}(s_{k-1}, s_k, \gamma_{k-1}, \gamma_k), \quad (23)$$

where  $\gamma_{k-1}$  denotes the type of the previous event and  $\gamma_k$  denotes the type of current event. Considering these additional factors may lead to some improvements. However, there are *at least three issues* to be considered with respect to any such

revised model. First, the type of the event is likely not known for most events. This is particularly a concern given the low-observability of the network which is the main challenge in this paper. Second, the formulation in (23) requires learning the state transition matrix in a four dimensional space (instead of two); because matrix  $\mathbf{B}$  becomes a four dimensional matrix. Apart from the computational complexity, one may not have sufficient measurements to properly learn the matrix under such large dimensionality. Third, over-fitting can become an issue given that each element of matrix  $\mathbf{B}$  would become too specific for each transition scenario. Therefore, for the rest of this paper, we consider the original formulation in (16).

## IV. EVENT-TRIGGERED DSSE PROBLEM FORMULATION AS SPARSE SIGNAL RECOVERY AND ITS SOLUTION

After learning the transition matrix, we can now solve the event-triggered DSSE problem. In this section, we provide the problem formulation and also two different solution methods.

### A. Problem Formulation

At each event  $k$ , our goal is to estimate the most recent differential state variables, i.e.,  $\Delta \mathbf{x}_k$ , as well as to refine and update the estimation of the previous differential state variables, i.e.,  $\Delta \mathbf{x}_1, \dots, \Delta \mathbf{x}_{k-1}$ . In this regard, suppose we stack up the vectors of differential state variables  $\Delta \mathbf{x}_1, \dots, \Delta \mathbf{x}_k$  and denote the resulting vector by  $\Delta \mathbf{x}_{1:k}$ . Similarly, suppose we stack up the vectors of differential measurements  $\Delta \mathbf{z}_1, \dots, \Delta \mathbf{z}_k$  and denote the resulting vector by  $\Delta \mathbf{z}_{1:k}$ .

From (10) and (15), and given the fact that the vector of the differential state variables is a sparse vector, we can estimate  $\Delta \mathbf{x}_{1:k}$  by solving the event-triggered DSSE problem which is formulated as the following generalized group Lasso problem:

$$\begin{aligned} \underset{\Delta \mathbf{x}_{1:k}}{\text{minimize}} \quad & \frac{1}{2} \sum_{\kappa=1}^k \|\mathbf{R}_{\kappa}^{-1/2} (\Delta \mathbf{z}_{\kappa} - \mathbf{H}_{\kappa} \Delta \mathbf{x}_{\kappa})\|_2^2 \\ & + \frac{1}{2} \sum_{\kappa=2}^k \|\mathbf{Q}_{\kappa}^{-1/2} (\Delta \mathbf{x}_{\kappa} - \mathbf{A}_{\kappa} \Delta \mathbf{x}_{\kappa-1})\|_2^2 \\ & + \frac{1}{2} \|\mathbf{G}_1^{-1/2} (\Delta \mathbf{x}_1 - \mathbf{m}_1)\|_2^2 \\ & + \lambda \sum_{\kappa=1}^k \sum_{p=1}^P w_{\kappa,p} \|\Delta \mathbf{x}_{\kappa,p}\|_1. \end{aligned} \quad (24)$$

Here,  $\mathbf{Q}_{\kappa}$  and  $\mathbf{R}_{\kappa}$  are the covariance matrices that are used in (10) and (15), respectively;  $\mathbf{m}_1$  and  $\mathbf{G}_1$  are the mean and covariance of the initial differential state variable, which are assumed to be known; and  $\lambda$  is the sparsity regularization parameter. Subscript  $p$  is the index for each partition of the differential state variables with respect to the concept of group sparsity that we discussed in Section II-D; and  $w_{\kappa,p}$  is the weight associated with partition  $p$  at event  $\kappa$ . More details on selecting the partitions and their weights can be found in [21].

For notation simplicity, we can further rewrite the optimization problem in (24) in the following more compact form:

$$\begin{aligned} \underset{\Delta \mathbf{x}_{1:k}}{\text{minimize}} \quad & \frac{1}{2} \|\mathbf{R}^{-1/2} (\Delta \mathbf{z}_{1:k} - \mathbf{H} \Delta \mathbf{x}_{1:k})\|_2^2 \\ & + \frac{1}{2} \|\mathbf{Q}^{-1/2} (\Phi \Delta \mathbf{x}_{1:k} - \mathbf{m})\|_2^2 \\ & + \lambda \|\mathbf{W} \Delta \mathbf{x}_{1:k}\|_1 \end{aligned} \quad (25)$$

where matrix  $\mathbf{R}$  is the block-diagonal representation of matrices  $\mathbf{R}_1, \dots, \mathbf{R}_k$ , matrix  $\mathbf{H}$  is the block-diagonal representation of matrices  $\mathbf{H}_1, \dots, \mathbf{H}_k$ , matrix  $\mathbf{Q}$  is the block-diagonal representation of matrices  $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_k$ , matrix  $\Phi$  is the block-diagonal representation of matrices  $\mathbf{A}_2, \dots, \mathbf{A}_k$ , matrix  $\mathbf{W}$  is the adequate block-diagonal representation of the weights in (24), and vector  $\mathbf{m}$  is obtained by stacking up vector  $\mathbf{m}_1$  following by  $k-1$  vectors denoted by  $\mathbf{0}$ , which are zero vectors of the same size as  $\mathbf{m}_1$ . Next, we discuss two different approaches to solve the optimization problem in (24).

### B. Approach 1: Batch ADMM Solution

A common approach to solve a generalized group Lasso optimization problem is to use the ADMM algorithm [40].

We can obtain the augmented Lagrangian function for the problem in (25) under constraint  $\mathbf{W} \Delta \mathbf{x}_{1:k} - \boldsymbol{\mu} = \mathbf{0}$  as:

$$\begin{aligned} \mathcal{L}_\rho(\Delta \mathbf{x}_{1:k}, \boldsymbol{\mu}, \mathbf{u}) = & \frac{1}{2} \|\mathbf{R}^{-1/2} (\Delta \mathbf{z}_{1:k} - \mathbf{H} \Delta \mathbf{x}_{1:k})\|_2^2 \\ & + \frac{1}{2} \|\mathbf{Q}^{-1/2} (\Phi \Delta \mathbf{x}_{1:k} - \mathbf{m})\|_2^2 \\ & + \lambda \|\boldsymbol{\mu}\|_1 + \langle \mathbf{u}, \mathbf{W} \Delta \mathbf{x}_{1:k} - \boldsymbol{\mu} \rangle \\ & + \frac{\rho}{2} \|\mathbf{W} \Delta \mathbf{x}_{1:k} - \boldsymbol{\mu}\|_2^2 \end{aligned} \quad (26)$$

where  $\mathbf{u}$  is the vector of dual variables and  $\rho > 0$  is a regularization parameter. Here,  $\lambda > 0$  is the sparsity regularization parameter which is used to adjust the trade-off between the measurement data fidelity and the penalty in the sparsity. A large  $\lambda$  would push the partitions of the state variables towards zero; while a very small  $\lambda$  would relax the sparsity penalty term on the partitions. Therefore, per the discussion made in Chapter 3 of reference [40], the value of  $\lambda$  can be selected based on cross validation. We will assess the sensitivity of our proposed method to the choice of the sparsity regularization parameter  $\lambda$  in Section V-G.

For each set of the variables, ADMM alternatively solves the following minimization problems and obtains the optimal values for the  $(\tau+1)$ -th iteration of alternation as:

$$\begin{aligned} \Delta \mathbf{x}_{1:k}^{(\tau+1)} = \arg \min_{\Delta \mathbf{x}_{1:k}} \quad & \frac{1}{2} \|\mathbf{R}^{-1/2} (\Delta \mathbf{z}_{1:k} - \mathbf{H} \Delta \mathbf{x}_{1:k})\|_2^2 \\ & + \frac{1}{2} \|\mathbf{Q}^{-1/2} (\Phi \Delta \mathbf{x}_{1:k} - \mathbf{m})\|_2^2 \\ & + \langle \mathbf{u}^\tau, \mathbf{W} \Delta \mathbf{x}_{1:k} - \boldsymbol{\mu}^\tau \rangle \\ & + \frac{\rho}{2} \|\mathbf{W} \Delta \mathbf{x}_{1:k} - \boldsymbol{\mu}^\tau\|_2^2, \end{aligned} \quad (27)$$

$$\begin{aligned} \boldsymbol{\mu}^{(\tau+1)} = \arg \min_{\boldsymbol{\mu}} \quad & \lambda \|\boldsymbol{\mu}\|_1 + \langle \mathbf{u}^\tau, \mathbf{W} \Delta \mathbf{x}_{1:k}^{(\tau+1)} - \boldsymbol{\mu} \rangle \\ & + \frac{\rho}{2} \|\mathbf{W} \Delta \mathbf{x}_{1:k}^{(\tau+1)} - \boldsymbol{\mu}\|_2^2, \end{aligned} \quad (28)$$

$$\mathbf{u}^{(\tau+1)} = \mathbf{u}^\tau + \rho \left( \mathbf{W} \Delta \mathbf{x}_{1:k}^{(\tau+1)} - \boldsymbol{\mu}^{(\tau+1)} \right). \quad (29)$$

By taking the derivative of the least square problem in (27), the optimal solution of  $\Delta \mathbf{x}_{1:k}$  in iteration  $(\tau+1)$  of the ADMM algorithm takes the following closed form:

$$\begin{aligned} \Delta \mathbf{x}_{1:k}^{(\tau+1)} = & [\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} + \Phi^\top \mathbf{Q}^{-1} \Phi + \rho \mathbf{W}^\top \mathbf{W}]^{-1} \\ & [\mathbf{H}^\top \mathbf{R}^{-1} \Delta \mathbf{z}_{1:k} + \Phi^\top \mathbf{Q}^{-1} \mathbf{m} + \rho \mathbf{W}^\top \boldsymbol{\mu}^k + \mathbf{W}^\top \mathbf{u}^k]. \end{aligned} \quad (30)$$

We can also solve the problem in (28) as:

$$\boldsymbol{\mu}^{(\tau+1)} = \mathcal{S}_{\lambda/\rho} \left( \mathbf{W} \Delta \mathbf{x}_{1:k}^{(\tau+1)} + \frac{\mathbf{u}^\tau}{\rho} \right), \quad (31)$$

where  $\mathcal{S}(\cdot)$  is the soft-thresholding operator [40].

### C. Approach 2: Kalman Filter and Smoother

Although the ADMM algorithm is exact and effective, it can be computationally expensive; because it requires calculating products and inverses for large matrices in (30). Hence, in this section, we propose an alternative approach to solve (25).

Suppose we *remove* the last line in (25), i.e., the line with the  $\ell_1$ -norm. What would be left in (25) would take the form of the optimization problem which can be solved by a conventional Kalman filter/smoother [41]. This is of interest due to the computational efficiency in Kalman filtering.

Based on the above observation, we propose to first make changes that we mentioned in the previous paragraph in (25), and obtain the corresponding solution for  $\Delta \mathbf{x}_k$  by using Kalman filter. Next, to take into account the impact of sparsity, we project the obtained solution for  $\Delta \mathbf{x}_k$  into a sparse domain. Finally, we use a conventional Kalman smoother to refine the previous estimation for  $\Delta \mathbf{x}_{1:k-1}$  that is already obtained at event  $k-1$  based on the projection of solution for  $\Delta \mathbf{x}_k$ .

These steps will provide us with an approximate solution for  $\Delta \mathbf{x}_{1:k}$ , i.e., the optimal solution for (25). As we will discuss in Section V, the difference between the approximate solution in Approach 2 and the exact solution in Approach 1 is negligible.

1) *Kalman Filtering*: At each event  $k > 1$ , we can use the state space model in (10) to conduct a preliminary one-step ahead prediction of the state variables as follows:

$$\Delta \mathbf{x}_{k|k-1} = \mathbf{A}_k \Delta \mathbf{x}_{k-1}. \quad (32)$$

The covariance matrix for  $\Delta \mathbf{x}_{k|k-1}$  is obtained as:

$$\mathbf{G}_{k|k-1} = \mathbf{A}_k \mathbf{G}_{k-1} \mathbf{A}_k^\top + \mathbf{Q}_k. \quad (33)$$

From (15), we can similarly make a one-step ahead prediction of the differential measurement as follows:

$$\Delta \mathbf{z}_{k|k-1} = \mathbf{H}_k \Delta \mathbf{x}_{k|k-1}. \quad (34)$$

We can obtain the measurement residual as  $\Delta \mathbf{z}_k - \Delta \mathbf{z}_{k|k-1}$ . Therefore, we can correct the preliminary one-step ahead prediction in (32) by applying the Kalman filter as follows:

$$\Delta \mathbf{x}_k = \Delta \mathbf{x}_{k|k-1} + \mathbf{K}_k (\Delta \mathbf{z}_k - \Delta \mathbf{z}_{k|k-1}), \quad (35)$$

where  $\mathbf{K}_k$  is the gain of the Kalman filter [42]:

$$\mathbf{K}_k = \mathbf{G}_{k|k-1} \mathbf{H}_k^\top \mathbf{S}_k^{-1}, \quad (36)$$

and  $\mathbf{S}_k$  is the covariance of  $\Delta \mathbf{z}_{k|k-1}$  as

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{G}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k. \quad (37)$$

We can obtain the covariance matrix of  $\Delta \mathbf{x}_{k|k}$  as:

$$\mathbf{G}_{k|k} = \mathbf{G}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top. \quad (38)$$

Note that, the process to obtain  $\Delta \mathbf{x}_{k|k}$  does *not* take into consideration the inherent sparsity in the solution.

2) *Incorporating Sparsity*: In the next step, we project the obtained estimation from Kalman filter, i.e.,  $\Delta \mathbf{x}_{k|k}$ , to the sparse domain with respect to its corresponding weight vector  $\mathbf{W}_k$  by solving the following optimization problem:

$$\underset{\Delta \mathbf{x}_k}{\text{minimize}} \frac{1}{2} \|\Delta \mathbf{x}_k - \Delta \mathbf{x}_{k|k}\|_2^2 + \lambda \|\mathbf{W}_k^\top \Delta \mathbf{x}_k\|_1. \quad (39)$$

Here, we obtain  $\Delta \mathbf{x}_k$  such that it is as close as possible to  $\Delta \mathbf{x}_{k|k}$  while it meets the sparsity constraints.

Since problem (39) does not have a closed-form solution, we cannot obtain the covariance of its solution in closed-form. Thus, we assume that the covariance of  $\Delta \mathbf{x}_k$  is  $\mathbf{G}_k = \mathbf{G}_{k|k}$ .

3) *Backward Smoothing*: After obtaining a computationally efficient sparse solution for  $\Delta \mathbf{x}_k$ , we *successively* apply the Kalman smoother to all the previous state estimation results in order to update  $\Delta \mathbf{x}_{1:k-1}$ . First, we replace  $\Delta \mathbf{x}_{k-1}$  with

$$\Delta \mathbf{x}_{k-1} + \mathbf{F}_{k-1} (\Delta \mathbf{x}_k - \Delta \mathbf{x}_{k|k-1}), \quad (40)$$

where  $\mathbf{F}_{k-1}$  is the gain of the Kalman smoother [42]:

$$\mathbf{F}_{k-1} = \mathbf{G}_{k-1} \mathbf{A}_k^\top \mathbf{G}_{k|k-1}^{-1}. \quad (41)$$

Finally, we update  $\mathbf{G}_{k-1}$  by replacing it with

$$\mathbf{G}_{k-1} + \mathbf{F}_{k-1} (\mathbf{G}_k - \mathbf{G}_{k|k-1}) \mathbf{F}_{k-1}^\top. \quad (42)$$

We then continue with applying the above backward smoother similarly to events  $k-2, k-3, \dots, 1$  to update vector  $\Delta \mathbf{x}_{1:k-1}$ .

Before we end this section, we add a brief note about the issue of convergence with regards to the above two approaches. Approach 1 is iterative; because it is a batch ADMM method. The convergence of Approach 1 is guaranteed as long as the objective function in the optimization problem is a sum of several convex functions with linear constraints [40]; which is the case in this paper. As for Approach 2, it was proposed to reduce the computation time by avoiding any iteration. That is, unlike Approach 1, Approach 2 is *not* iterative. Therefore, the issue of convergence is *not* applicable to Approach 2.

#### D. Real-Life Applications of the Proposed Methods

DSSE is a key module in the monitoring systems in Automated Distribution Management Systems (ADMS) [7]. Despite the advancements made in the sensor technologies, such as transition from analog meters to smart meters at the customer end or the emergence of D-PMUs, in practice, the power distribution systems still suffer from low-observability due to the various reasons that were discussed in Section I. Our DSSE method addresses low-observability as the central focus; thus advancing the field of power system monitoring by addressing a challenging issue with direct practical relevance.

Another real-life concept that is of importance in this paper is the notion of “event”, which is any abrupt change in the elements of the power system. Events happen very frequently in practice. In fact, one of the main reasons that monitoring of

power distribution systems is becoming increasingly important is due to the need to monitor occurrence of such changes across the power distribution system. This requires going beyond the traditional assumptions with regards to having static and non-changing power distribution systems. This fact, combined with the advancements in sensor technology, which have enabled the grid operator to capture and study the events, has been the main reason for the current growing interest in the study of events at the power distribution networks. In the same essence, this work has put the focus on the study of events as the main factor which alters the state variables of the power distribution system during its operation.

Finally, the analysis in this paper is aligned with the recent advancements in the field of smart grid sensors and the expectations for the availability of such sensors in practice. Most notably, while the use of D-PMUs is increasingly adopted in practice, it is important to note that such advanced sensors cannot be deployed in large numbers in a single feeder due to their high cost. Therefore, while we take advantage of the availability of D-PMUs in this paper, we assume that only very few of such sensors are available. This puts our design in accordance with the trends in practice for the availability of D-PMUs. Meanwhile, the event-triggered nature of our proposed approach can also help with properly reducing the computation burden that the use of D-PMUs poses to the system.

## V. CASE STUDIES

Unless stated otherwise, the case studies in this section are done based on the IEEE 33-Bus power distribution test system [43], where we keep the default load profiles. MATPOWER in MATLAB R2018B is used as the simulation environment to also generate the events. The magnitude of the events are set to be at most up to 50% of the associated load in the pre-event condition. Unless stated otherwise, we follow Remark 4, and we define five zones in the network, as in Fig. 1. Accordingly, there are  $5 \times 5 = 25$  possible choices for matrix  $\mathbf{B}$  to model transitions from  $s_{k-1}$  to  $s_k$ . For each transition, we train an offline model to learn the corresponding transition matrix.

Only five D-PMUs are assumed to be available, at buses 9, 18, 22, 25, and 33. Each D-PMU reports nodal voltage and line current phasors, once every 100 milliseconds. If the event detection method, as in [32], detects an event, then the proposed event-triggered DSSE is conducted. For the D-PMUs, the standard deviation in measurement error is  $\sigma_V = 0.1\%$  for voltage and  $\sigma_I = 1\%$  for current. The noise covariance matrices are obtained offline during the training of the event-triggered state transition matrices. We defined 100 scenarios. Each scenario includes a sequence of 1000 randomly generated events that are in form of sudden load changes at random buses, and with random magnitudes up to 50% of the original load amount at the bus where the event occurs. Therefore, in each scenario, the DSSE problem is solved 1000 times due to the occurrence of 1000 events.

#### A. Performance Comparison

Performance comparison is done against the DSSE methods in [10], [21], [44]. The method in [10] is a dynamic DSSE

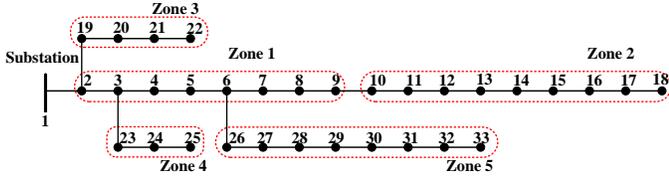


Fig. 1. The default definition of the zones for the examples in this paper.

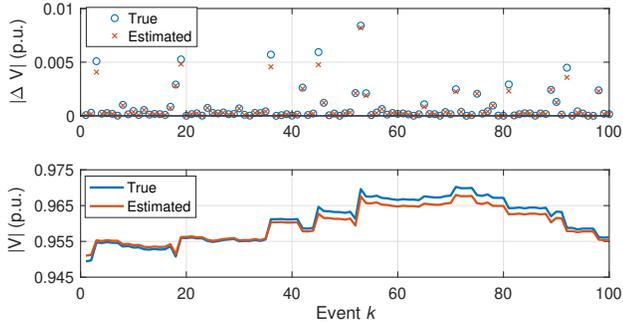


Fig. 2. The estimated and actual values for the magnitude of voltage phasors at bus 29 for 100 events: (a) differential mode (b) regular mode.

based on conventional iterated Kalman filtering. It uses non-linear power flow equations. The method in [21] is a static sparse DSSE method. The method in [44] is a static DSSE based on conventional weighted least square analysis. Because the DSSE methods in [10] and [44] require full-observability, the initial measurements by legacy meters are used as pseudo-measurements for the rest of the events for these two methods.

We use *mean absolute percentage error* (MAPE) to evaluate the accuracy of estimation for differential state variables. Also, we use *root mean square error* (RMSE) to evaluate the estimation of the state variables over the sequence of events.

The results are summarized in Table II. As we can see, the estimation accuracy is better for the proposed method than the other three methods, whether in terms of MAPE for the differential mode or in terms of RMSE for the regular mode. It is evident from Table II that our ability to integrate sparse recovery, dynamic state estimation, and virtual measurements has improved the DSSE performance across all variables.

An example for the outcome of the proposed method is shown in Fig. 2, where the results are at bus 29, for one random scenario, and across 100 consecutive events. The magnitude of the true and the estimated voltage phasors in differential mode are shown in Fig. 2(a). Notice the sparsity in the results, i.e., the fact that the differential voltage is non-zero at only a small subset of the events. Furthermore, the proposed method can identify the sparsity almost perfectly; and overall estimate the differential voltage phasors with a high accuracy. The magnitude of the true and the estimated voltage phasors in regular mode are shown in Fig. 2(b). We can see that the proposed method can reasonably follow the state variables.

### B. Performance Comparison: IEEE 123-bus system

To show the scalability of the proposed method, we also test it on the IEEE 123-bus test system [45], which has single phase, two phase, and unbalanced three phase loads with Wye

TABLE II  
COMPARING DIFFERENT DSSE METHODS IN ESTIMATING SYNCHROPHASORS AND DIFFERENTIAL SYNCHROPHASORS

Method	Dynamic	Sparsity	MAPE $\Delta x$	RMSE $\dagger x$
Proposed Method	✓	✓	3.82%	7.25
[10]	✓	×	21.72%	32.67
[21]	×	✓	4.84%	8.58
[44]	×	×	19.93%	27.78

$\dagger$  The RMSE values come with a  $10^{-3}$  coefficient.

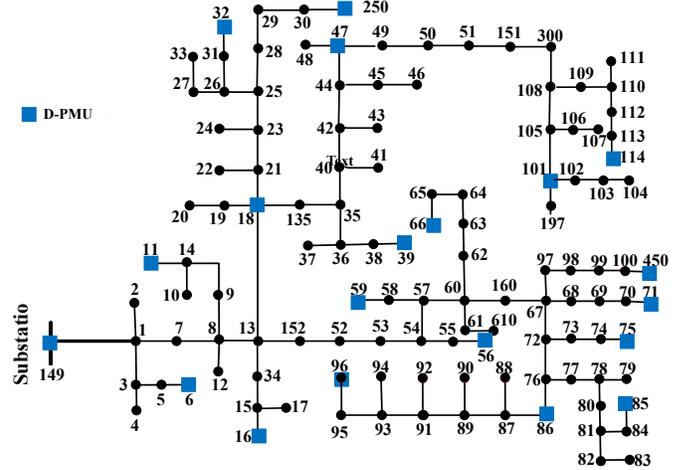


Fig. 3. The one line diagram of the IEEE 123-bus test system with 20 D-PMUs (at 16% of the buses). The network is three-phase and unbalanced.

connection. We assume there are 20 D-PMUs available. One at substation (node 149) and the rest at nodes 6, 11, 16, 18, 32, 39, 47, 56, 59, 66, 71, 75, 85, 86, 96, 101, 114, 250, and 450; as shown in Fig. 3. The results are shown in Table III. As we can see, the performance of the proposed method on this large unbalanced network is very good, and it outperforms the other methods in the literature. Importantly, the physics-based sparsity features that we have extracted in this paper are valid for both balanced networks and unbalanced networks.

### C. Comparing Approach 1 and Approach 2

Next, we compare the performance of the two methods that we proposed in Section IV-B (Approach 1) and Section IV-C (Approach 2). Table IV shows the results based on three different metrics. As we can see, both methods converge in 100% of the simulated random scenarios. Approach 1 performs better in terms of the state estimation accuracy, but Approach 2 performs much better in terms of the computation time. The computation time in Table IV corresponds to the total time for estimating the differential state variables for 1000 events.

### D. Effect of Learning the Transition Matrix

In this section, we investigate the importance of learning the event-triggered state transition matrix. We compare two cases in Table V. First, an *identity* state transition matrix, as in [11]. In this case, we always assume that  $\mathbf{A}_k = \mathbf{I}$ . Second, we learn the state transition matrix, as in this paper. In this case, we set  $\mathbf{A}_k$  as in (16). As we can see, learning the state transition matrix significantly improves the state estimation

TABLE III  
PERFORMANCE COMPARISON FOR IEEE 123-BUS TEST SYSTEM

Method	MAPE $\Delta \mathbf{x}$	RMSE <sup>†</sup> $\mathbf{x}$
Proposed Method	7.87%	10.12
[10]	27.36%	36.05
[21]	9.14%	11.64
[44]	25.39%	32.89

<sup>†</sup> The RMSE values come with a  $10^{-3}$  coefficient.

TABLE IV  
COMPARING THE CONVERGENCE RATE, COMPUTATION TIME, AND ACCURACY OF THE DIFFERENT DSSE METHODS

Method	Convergence Rate	Computation Time	MAPE
Approach 1	100%	5.03 Seconds	3.80%
Approach 2	N/A	0.75 Seconds	4.12%
[10]	83.57%	24.56 Seconds	21.72%

performance, which is understood from the improvement in MAPE of  $\Delta \mathbf{x}_k$ . Here, we also check the ability of the state transition matrix  $\mathbf{A}_k$  in accurately predicting the differential state variables at the next event as in (10). From (32), such prediction is denoted by  $\mathbf{x}_{k|k-1}$ . As we can see, the MAPE of  $\mathbf{x}_{k|k-1}$  can drastically improve, meaning that the accuracy of the state space model in (10) can drastically improve by using the proposed method to learn the state transition matrix  $\mathbf{A}_k$ .

We also studied how the results may change if we change the number of zones in learning the state transition matrix in Fig. 4. We check the ability of the state transition matrix  $\mathbf{A}_k$  in accurately predicting the differential state variables at the next event in (10); thus we again use the MAPE of  $\mathbf{x}_{k|k-1}$  as the metric. We can see that as we increase the number of zones, we can enhance the accuracy of prediction, and ultimately the accuracy of estimation. However, it would come with an increased computation time to learn the state transition matrix. Also, increasing the number of zones might lead to overfitting and higher sensitivity to the outcome of the event location identification algorithm, both of which should be avoided.

#### E. Effect of Virtual Power Measurements

A major feature of the proposed method is the use of virtual differential power measurements in matrix  $\mathbf{H}_k^4$ . In order to directly evaluate the importance of this innovative feature, next, we compare the state estimation *with* and *without* using virtual power measurements. The results are shown in Table VI. Here, we also show the results for estimating the line power flow and nodal injection power. As we can see, the use of virtual power can highly improve the accuracy in estimating both of these power quantities that are *not* measured directly. It also helps improve the overall accuracy of the DSSE results, as seen here in terms of RMSE  $\mathbf{V}$  and RMSE  $\mathbf{I}$ .

#### F. Multiple Simultaneous Events

As a follow up to Remark 2, in this section, we examine the rare case wherein *multiple* additional (less major) events occur simultaneously with the main major event. The main event is a sudden change in the net power of one node and between 10% to 50% of its pre-event value; while the minor events are less than 10% changes in the net power at some other nodes.

TABLE V  
THE IMPORTANCE OF LEARNING THE STATE TRANSITION MATRIX USING DISCRIMINATIVE ELASTIC NET REGRESSION

Method	MAPE of $\Delta \mathbf{x}_k$	MAPE of $\Delta \mathbf{x}_{k k-1}$
Using Identity Matrix	5.36%	18.36%
Learning the Matrix	3.98%	6.77%

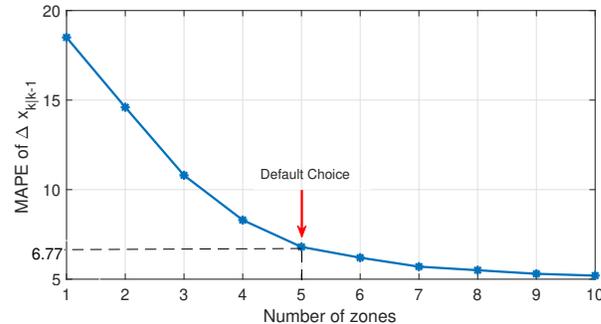


Fig. 4. The impact of increasing the number of zones on the power distribution system on the performance of the proposed event-triggered state transition matrix in predicting the differential state variables at the next event.

The results are shown in Fig. 5. As the number of additional minor events increases, the performance of the proposed method gradually degrades. However, the performance degradation is much higher for the other methods in the literature in comparison with the proposed method. Note that, there is a trade-off between adding more sensors to the network to make network fully-observable versus relying on a much lower number of sensors to obtain less exact but still accurate estimation of the state variables in a more practical setting.

#### G. Sensitivity Analysis: Sparsity Regularization Parameter

In this section, we conduct a sensitivity analysis for the performance evaluation of the proposed method with respect to the choice of the sparsity regularization parameter  $\lambda$  in equation (24). The results are shown in Fig. 6 for different values of  $\lambda$ . As we can see, if  $\lambda$  is *not too high*, i.e.,  $\lambda \leq 10^{-2}$ , and it is *not too low*, i.e.,  $\lambda \geq 10^{-6}$ , then the results are reasonable. In this study, we have set  $\lambda = 10^{-4}$ .

#### H. Sensitivity Analysis: Magnitude of the Event Event

The magnitude of the event may slightly affect the accuracy of the proposed event-triggered DSSE method. This is shown in Fig. 7. Here, we plot the average RMSE versus the size of the change in the load that causes the event. The change in the load is presented in percentage. Notice the very small scale of the numbers on the y-axis. The changes in the size of the event causes the RMSE to vary only by about 0.005; which is very small. Therefore, we can conclude that the size of the event is not a major factor in affecting the accuracy of the DSSE algorithm. The event-triggered DSSE algorithm works well for different sizes of events.

## VI. CONCLUSIONS AND FUTURE WORK

To tackle low-observability in power distribution systems, a novel event-triggered DSSE method is proposed, where the

TABLE VI  
EFFECT OF USING VIRTUAL POWER MEASUREMENTS

Method	RMSE $S^\dagger$ (Node)	RMSE $S^\dagger$ (Line)	RMSE $V^\dagger$	RMSE $I^\dagger$
With Virtual Power Measurements	6.12	6.46	7.39	6.44
Without Virtual Power Measurements	8.25	7.86	7.88	7.12

<sup>†</sup> The RMSE values come with a  $10^{-3}$  coefficient.

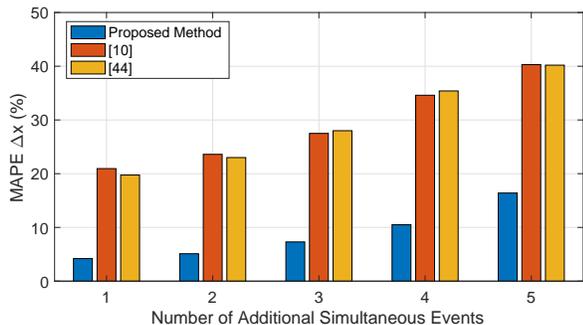


Fig. 5. Performance evaluation of DSSE when multiple events occur simultaneously; one major event with a few additional relatively minor events.

measurements come from a handful of D-PMUs. The proposed method simultaneously addresses three main challenges that exists in the DSSE problem. First, the event-triggered nature of our design can help to avoid imposing unnecessary burden on the operation center. Second, to properly capture the dynamic nature of the modern distribution systems under the low-observability conditions, the problem is formulated over the differential state variables as a generalized group Lasso optimization, which leverages the sparsity features that exist in the system under the event-triggered paradigm. To further improve our ability in conducting a sparse recovery, the DSSE problem formulation is reinforced by a novel set of linear differential power flow equations, and in forms of virtual measurements. Third, to improve the accuracy of the state-space model, the event-triggered state transition matrix is learned in an off-line fashion through discriminative elastic net regression. Extensive performance evaluations confirmed the effectiveness of the proposed methodologies. Representing the state variables in differential mode is advantageous in this analysis. In particular, it provides us with the opportunity to extract physics-based sparsity patterns in the DSSE problem. This accordingly allows us to formulate and solve the DSSE problem under the low-observability conditions. Moreover, the use of differential phasors results in the linearization of the power flow equations and including them in the DSSE problem, without the need for power measurements.

The study in this paper can be extended in various directions. First, one may consider a non-linear state space model for the event-triggered DSSE problem, which could be more accurate than the linear model, which is commonly used in the literature and also in this paper. Second, the way that the event-triggered state transition matrix is estimated can be improved. While our approach to estimate such matrix instead of using the identity matrix is a major step forward compared to the literature, we only considered the location of the event

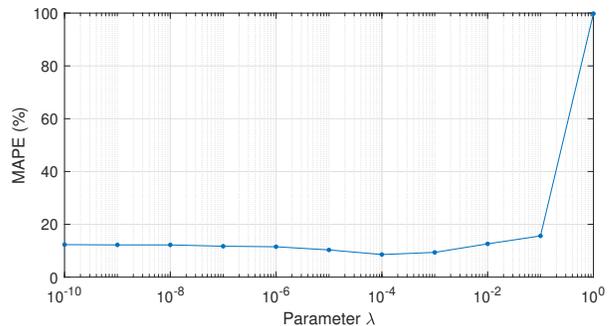


Fig. 6. Sensitivity with respect to the sparsity regularization parameter.

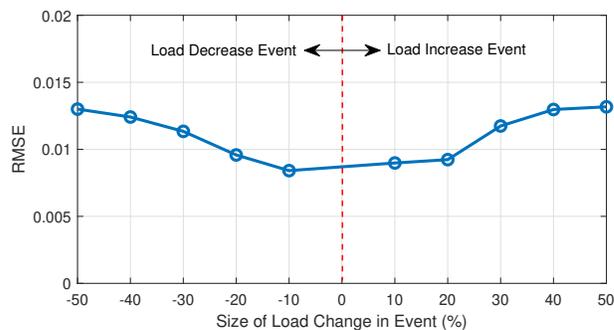


Fig. 7. The impact of the magnitude of the event on the DSSE performance.

as the key factor in the estimation of the state transition matrix. However, other factors such as the type or other characteristics of the event may also be considered. Of course, one would have to address the various challenges and concerns that we previously raised in *Remark 5* on this issue. Third, the proposed event-triggered DSSE can be improved such that it could support other types of events which may happen on the branches of the system such as faults or topology changes.

#### APPENDIX: LINEARIZED DIFFERENTIAL POWER FLOW EQUATIONS

In this Appendix, we explain the linearized differential power flow equations in (12). Let us start with writing the active power injection at bus  $i$  in differential mode at the occurrence of event  $k$  in an event-triggered formulation:

$$P_{k,i} = P_{k-1,i} + \Delta P_{k,i}. \quad (43)$$

Here, we express  $P_{k,i}$  at event  $k$  in terms of its difference  $\Delta P_{k,i}$  compared to  $P_{k-1,i}$  at event  $k-1$ . We can similarly express the reactive power injection as follows:

$$Q_{k,i} = Q_{k-1,i} + \Delta Q_{k,i}. \quad (44)$$

From (43) and (44), we can obtain:

$$S_{k,i} = (P_{k-1,i} + \Delta P_{k,i}) + j(Q_{k-1,i} + \Delta Q_{k,i}), \quad (45)$$

where  $S_i$  denotes the complex injected power at bus  $i$ . Let  $v_{k,i}$  denote the voltage phasor at bus  $i$  at the occurrence of event  $k$ . In differential mode, we have:

$$v_{k,i} = v_{k-1,i} + \Delta v_{k,i}. \quad (46)$$

From the definition of complex power, we have:

$$S_{k,i} = v_{k,i} \sum_{j \in \mathcal{N}} (R_{ij} - jX_{ij}) v_{k,j}^*, \quad (47)$$

where  $\{\cdot\}^*$  denotes the complex conjugate operator. Notations  $R_{ij}$  and  $X_{ij}$  denote the resistance and the admittance of the line segment between node  $i$  and node  $j$ , respectively. By replacing (46) in (47), we have:

$$S_{k,i} = \left[ \Re\{v_{k-1,i} + \Delta v_{k,i}\} + j\Im\{v_{k-1,i} + \Delta v_{k,i}\} \right] \times \left[ \sum_{j \in \mathcal{N}} (R_{ij} - jX_{ij}) (\Re\{v_{k-1,j} + \Delta v_{k,j}\} - j\Im\{v_{k-1,j} + \Delta v_{k,j}\}) \right] \quad (48)$$

where  $\Re\{\cdot\}$  denotes the real part and  $\Im\{\cdot\}$  denotes the imaginary part. From (45) and (48), and after reordering the terms, we can obtain:

$$\begin{aligned} \Delta P_{k,i} \approx \sum_{j \in \mathcal{N}} \left[ \Re\{\Delta v_{k,i}\} (R_{ij} \Re\{v_{k-1,j}\} - X_{ij} \Im\{v_{k-1,j}\}) \right. \\ + \Im\{\Delta v_{k,i}\} (R_{ij} \Im\{v_{k-1,j}\} + X_{ij} \Re\{v_{k-1,j}\}) \\ + \Re\{\Delta v_{k,j}\} (R_{ij} \Re\{v_{k-1,i}\} + X_{ij} \Im\{v_{k-1,i}\}) \\ \left. + \Im\{\Delta v_{k,j}\} (R_{ij} \Im\{v_{k-1,i}\} - X_{ij} \Re\{v_{k-1,i}\}) \right], \quad (49) \end{aligned}$$

$$\begin{aligned} \Delta Q_{k,i} \approx - \sum_{j \in \mathcal{N}} \left[ \Re(\Delta v_{k,i}) (R_{ij} \Im\{v_{k-1,j}\} + X_{ij} \Re\{v_{k-1,j}\}) \right. \\ + \Im\{\Delta v_{k,i}\} (-R_{ij} \Re\{v_{k-1,j}\} + X_{ij} \Im\{v_{k-1,j}\}) \\ + \Re\{\Delta v_{k,j}\} (-R_{ij} \Im\{v_{k-1,i}\} + X_{ij} \Re\{v_{k-1,i}\}) \\ \left. + \Im\{\Delta v_{k,j}\} (R_{ij} \Re\{v_{k-1,i}\} + X_{ij} \Im\{v_{k-1,i}\}) \right]. \quad (50) \end{aligned}$$

In (49) and (50), the differential power injections and differential voltage phasors for the current event  $k$  are unknowns, while the values of voltage phasors for the previous event, i.e.,  $k-1$  are known. Accordingly, we can rewrite them in an abstract form as shown in (12).

The approximations in (49) and (50) are due to the fact that we discarded the terms that are the *products of two differential voltage phasors*. This approximation is reasonable; since such products are very small, in case of typical loading events that we have considered in this paper. Note that under severe events such as short-circuit faults which may cause extreme changes in the nodal voltage phasors, such assumption may not be true anymore, which is of course out of scope of this paper.

To investigate how discarding the products of differential voltage phasors may impact the differential power injections, we plot the linearization error in Fig. 8 for the IEEE 33-bus test system that we considered in Section V. We calculate the linearization error at each bus as the difference between the power injection that is calculated by using the original nonlinear power flow equations in (47) and the power injection that is calculated by using the proposed linear power flow equations in (49) and (50). The analysis is done under the scenario where the “event” is very major, in which the largest load in the system, which is at bus 24, is suddenly disconnected. We considered two cases: the case where the network loading is the same as the default/standard loading in the IEEE 33-bus

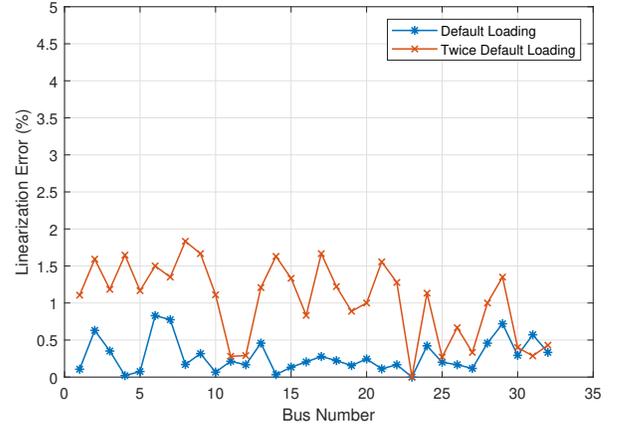


Fig. 8. Linearization error in power flow equations in differential mode for the test system with the default/standard loading of the IEEE 33-bus test system as well as the case with twice the default/standard loading.

test system; and the case where the network loading is *twice* the default/standard loading. As we can see, the linearization error in the differential power flow equations is very small; less than 0.8% under the default/standard loading and less than 2% under the double the default/standard loading.

## REFERENCES

- [1] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, “A survey on state estimation techniques and challenges in smart distribution systems,” *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2312–2322, Mar. 2019.
- [2] S. Bhela, V. Kekatos, and S. Veeramachaneni, “Enhancing observability in distribution grids using smart meter data,” *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5953–5961, Nov. 2018.
- [3] M. Hassanzadeh, C. Y. Evrenosoglu, and L. Mili, “A short-term nodal voltage phasor forecasting method using temporal and spatial correlation,” *IEEE Trans. Power Syst.*, vol. 31, no. 5, p. 3881–3890, Sep. 2016.
- [4] J. Zhao *et al.*, “Power system dynamic state estimation: Motivations, definitions, methodologies, and future work,” *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3188–3198, Jul. 2019.
- [5] A. Ahmed, K. Sajan, A. Srivastava, and Y. Wu, “Anomaly detection, localization and classification using drifting synchrophasor data streams,” *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3570–3580, Jul. 2021.
- [6] C. Hannon, D. Deka, D. Jin, M. Vuffray, and A. Y. Likhov, “Real-time anomaly detection and classification in streaming pmu data,” in *2021 IEEE Madrid PowerTech*, 2021, pp. 1–6.
- [7] H. Mohsenian-Rad, *Smart Grid Sensors: Principles and Applications*. Cambridge University Press, UK, Apr. 2022.
- [8] 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. Smart Grid*, vol. 10, no. 6, pp. 6167–6177, Nov. 2019.
- [9] J. Song, E. Dall’Anese, A. Simonetto, and H. Zhu, “Dynamic distribution state estimation using synchrophasor data,” *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 821–831, Jan. 2020.
- [10] S. Sarri, M. Paolone, R. Cherkaoui, A. Borghetti, F. Napolitano, and C. A. Nucci, “State estimation of active distribution networks: Comparison between WLS and iterated kalman-filter algorithm integrating PMUs,” in *IEEE PES ISGT Europe*, 2012.
- [11] C. Carquex, C. Rosenberg, and K. Bhattacharya, “State estimation in power distribution systems based on ensemble kalman filtering,” *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6600–6610, Nov. 2018.
- [12] M. Huang, Z. Wei, J. Zhao, R. A. Jabr, M. Pau, and G. Sun, “Robust ensemble kalman filter for medium-voltage distribution system state estimation,” *IEEE Trans. Instrum. Meas.*, vol. 69, pp. 4114–4124, 2020.
- [13] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, “Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling,” *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1888–1896, Nov. 2012.

- [14] R. Singh, B. C. Pal, and R. A. Jabr, "Distribution system state estimation through gaussian mixture model of the load as pseudo-measurement," *IET Gen., Transm. Distrib.*, vol. 4, no. 1, pp. 50–59, Jan. 2009.
- [15] K. A. Clements, "The impact of pseudo-measurements on state estimator accuracy," in *Proc. IEEE PES General Gen. Meeting*, Detroit, MI, 2011.
- [16] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2919–2928, Nov. 2015.
- [17] S. M. S. Alam, B. Natarajan, and A. Pahwa, "Distribution grid state estimation from compressed measurements," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1631–1642, Jul. 2014.
- [18] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang, "Matrix completion for low-observability voltage estimation," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2520–2530, May 2020.
- [19] C. Cheng and X. Bai, "Robust forecasting-aided state estimation in power distribution systems with event-triggered transmission and reduced mixed measurements," *IEEE Trans. Power Syst.*, vol. 38, no. 5, pp. 4343–4354, Sep. 2021.
- [20] X. Bai, X. Zheng, L. Ge, F. Qin, and Y. Li, "Event-triggered forecasting-aided state estimation for active distribution system with distributed generations," *Frontiers in Energy Research*, vol. 9, Jul. 2021.
- [21] A. Akrami, M. S. Asif, and H. Mohsenian-Rad, "Sparse tracking state estimation for low-observable power distribution systems using D-PMUs," *IEEE Trans. Power Syst.*, Accepted for publication, Jul. 2021.
- [22] F. Ahmadi Gorjaji and H. Mohsenian-Rad, "Physics-aware sparse harmonic state estimation in power distribution systems," in *2022 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Apr. 2022, pp. 1–5.
- [23] F. Ahmadi Gorjaji and H. Mohsenian-Rad, "A physics-aware miqp approach to harmonic state estimation in low-observable power distribution systems using harmonic phasor measurement units," *IEEE Trans. Smart Grid*, vol. 14, no. 3, pp. 2111–2124, Sep. 2022.
- [24] S. Li *et al.*, "Event-trigger heterogeneous nonlinear filter for wide-area measurement systems in power grid," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2752–2764, May 2019.
- [25] X. Liu *et al.*, "Event-trigger particle filter for smart grids with limited communication bandwidth infrastructure," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6918–6928, Nov. 2018.
- [26] S. Li *et al.*, "Event-based cubature kalman filter for smart grid subject to communication constraint," in *Proc. of the IFAC World Congress*, vol. 50, no. 1, Jul. 2017, pp. 49–54.
- [27] X. Bai, F. Qin, L. Ge, L. Zeng, and X. Zheng, "Dynamic state estimation for synchronous generator with communication constraints: An improved regularized particle filter approach," *IEEE Trans. Sustainable Computing*, vol. 9, To be published, Nov. 2022.
- [28] H. Kopetz, "Event-triggered versus time-triggered real-time systems," in *Operating Systems of the 90s and Beyond*. Springer, 1991, pp. 86–101.
- [29] I. Srivastava, S. Bhat, B. S. Vardhan, and N. D. Bokde, "Fault detection, isolation and service restoration in modern power distribution systems: A review," *Energies*, vol. 15, no. 19, Oct. 2022.
- [30] G. Cavarro and R. Arghandeh, "Power distribution network topology detection with time-series signature verification method," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 3500–3509, Dec. 2017.
- [31] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5152–5162, Sep. 2018.
- [32] 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. Power Syst.*, vol. 33, no. 6, pp. 6343–6354, Nov. 2018.
- [33] J. C. Mayo-Maldonado *et al.*, "Data-driven framework to model identification, event detection and topology change location using D-PMUs," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, Mar. 2020.
- [34] Q. Cui and Y. Weng, "Enhance high impedance fault detection and location accuracy via  $\mu$ PMUs," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 797–809, Jan. 2020.
- [35] A. Aligholian, A. Shahsavari, E. M. Stewart, E. Cortez, and H. Mohsenian-Rad, "Unsupervised event detection, clustering, and use case exposition in micro-PMU measurements," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3624–3636, Jul. 2021.
- [36] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Smart inverter grid probing for learning loads: Part I — identifiability analysis," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 3527–3536, Sep. 2019.
- [37] A. Shahsavari, M. Farajollahi, and H. Mohsenian-Rad, "Individual load model parameter estimation in distribution systems using load switching events," *IEEE Trans. Power Syst.*, vol. 34, pp. 4652–4664, Nov. 2019.
- [38] X. Kong, X. Zhang, X. Zhang, C. Wang, H. Chiang, and P. Li, "Adaptive dynamic state estimation of distribution network based on interacting multiple model," *IEEE Trans. Sustain. Energy (accepted)*, 2021.
- [39] Z. Zhang, Z. Lai, Y. Xu, L. Shao, J. Wu, and G. Xie, "Discriminative elastic-net regularized linear regression," *IEEE Trans. Image Processing*, vol. 26, no. 3, pp. 1466–1481, Mar. 2021.
- [40] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, Jul. 2011.
- [41] B. Bell, "The iterated kalman smoother as a gauss-newton method," *SIAM J. Optim.*, vol. 4, no. 3, pp. 626–636, Aug. 1994.
- [42] S. Sarkka, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
- [43] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Trans. Power Del.*, vol. 4, no. 2, pp. 1401–1407, Apr. 1994.
- [44] D. A. Houghton and G. T. Heydt, "A linear state estimation formulation for smart distribution systems," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1187–1195, May 2013.
- [45] <https://site.ieee.org/pes-testfeeders/resources>.



**Alireza Akrami** (S'18) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2014, the M.Sc. degree in electrical engineering from University of Tehran, Tehran, Iran, in 2018, and the Ph.D. degree in electrical engineering from University of California, Riverside, CA, USA, in 2022. His research interests include power system state estimation, smart grid initiatives, power system optimization, and applications of machine learning in power systems. He was a recipient of MPCE journal best paper award in the year 2019 and was selected among the 2019 Best Reviewers of the IEEE TRANSACTIONS ON SMART GRID.



**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 and Energy Society General Meeting, the Best Paper Award from the IEEE Conference on Smart Grid Communications, and a Technical Achievement Award from the IEEE Communications Society. Recently, his work on distribution system state estimation was selected as a Best Paper Award Finalist at the IEEE PES ISGT Conference 2023. He has been the PI or co-PI on 15 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.