Review of Emerging Concepts in Distribution System State Estimation: Opportunities and Challenges

AJAY PRATAP YADAV¹, (Member, IEEE), JAMES NUTARO², (Senior Member, IEEE), BYUNGKWON PARK², (Member, IEEE), JIN DONG¹, (Member, IEEE), BOMING LIU¹, (Member, IEEE), SRIKANTH B. YOGINATH¹, (Member, IEEE), HE YIN³, (Senior Member, IEEE), JIAOJIAO DONG³, (Senior Member, IEEE), YUQING DONG³, (Member, IEEE), YILU LIU¹, (Fellow, IEEE), TEJA KURUGANTI¹, (Senior Member, IEEE), AND YAOSUO XUE¹, (Senior Member, IEEE)

¹Oak Ridge National Lab, Oak Ridge, TN 37830, USA
²Department of Electrical Engineering, Soongsil University, Seoul 06978, South Korea
³Department of Electrical Engineering and Computer Science, The University of Tennessee at Knoxville, Knoxville, TN 37996, USA

Corresponding author: Ajay Pratap Yadav (yadavap@ornl.gov)

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

ABSTRACT Distribution System State Estimation (DSSE) infers the states (e.g., voltage phasors) using measurements from the electric distribution grid. As the distribution network becomes more active due to the increasing penetration of Distributed Energy Resources (DER), a deeper investigation into conventional DSSE approaches is needed. This review paper focuses on state-of-the-art techniques and challenges for DSSE in the context of emerging concepts, such as new sensor technologies, data-driven approaches, and high DER penetration.

INDEX TERMS Distribution system state estimation, data-driven approach, high renewable penetration, new sensor technology, pseudo measurements, sensor placement.

NOMENCLATURE

Abbreviations

AMI Advanced metering infrastructure.
DER Distributed energy resources.
DR Demand response.
DSSE Distribution system state estimation.
ERCOT Electric Reliability Council of Texas.
GA Genetic algorithm.
GMM Gaussian mixture model.
NN Neural networks.
PMU Phasor measurement unit.
SCADAb Supervisory control and data acquisition.
SDP Semidefinite programming.
VPP Virtual power plant.

Variables

WLS Weighted least square.
$\lambda$ Lagrange multipliers.
$e(x)$ Equality constraints.
$e$ Measurement error.
$h(x)$ Measurement function.
$H_k$ Jacobian of the objective function in (1).
i Current injection.
p Active power injection.
q Reactive power injection.
v Node voltages.
W Weight matrix.
x Distribution system states.
Y Admittance matrix.
z Measurement vector.
$\sigma$ Standard deviation of the measurement error.
$B$ Imaginary component of admittance matrix.

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Bellan.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/
I. INTRODUCTION

Distribution system state estimation (DSSE) infers state information about an electric distribution network from measurement data. In other words, distribution state estimation maps measurements, from e.g., phasor measurement units (PMUs) or supervisory control and data acquisition (SCADA), to the system state (e.g., voltage phasors).

Seminal work on DSSE proposed in the early nineties was mostly derived from transmission system state estimation [1], [2]. Despite some similarities, the algorithms developed for state estimation in the transmission system are not directly applicable to DSSE. Compared to the transmission system, the distribution system poses unique challenges such as unobservability (i.e., the lack of sufficient measurements), low x/r ratio (i.e., ignoring line resistance will make DSSE inaccurate), and unbalanced operation (i.e., the requirement of multiphase formulations) [3]. A detailed review of standard techniques and relevant challenges associated with DSSE can be found in [3] and [4].

The recent modernization of the electrical grid has brought about significant changes in the distribution network, evidenced by the integration of Distributed Energy Resources (DERs), the incorporation of new sensor technologies, enhanced communication systems, and advanced data analytics [5], [6]. These developments have transformed load behavior from passive to active, enabling greater consumer participation through initiatives such as demand response programs. The growing adoption of electric vehicles further empowers households to support the grid during peak demand hours.

Integration of new technologies also increases the grid’s vulnerability to cyber threats. As a result, it is crucial to ensure that Distribution System State Estimation (DSSE) remains resilient against various cyber attacks, such as false data injection or denial of service. Developing robust and secure DSSE frameworks will be essential for maintaining the stability and reliability of the modernized grid, safeguarding both critical infrastructure and consumer interests [7], [8].

In this paper, we provide a summary of the current literature on DSSE, with a particular emphasis on research from the last few years. We examine the challenges in DSSE and review recent trends, specifically the integration of renewable energy sources such as solar photovoltaic (PV) systems. The paper is organized as follows: Section II offers a brief overview of the DSSE problem formulation, algorithms, and challenges. Section III discusses works related to sensors, followed by data-driven techniques developed for DSSE in Section IV. Section V presents a discussion on challenges and opportunities for future research. Finally, Section VI concludes the paper.

II. DSSE PROBLEM

This section reviews the standard DSSE formulation, followed by a discussion of other DSSE formulations. Detailed discussions on pseudo-measurement generation methods are provided, along with a summary of some recent works on DSSE in a tabulated list.

A. CONVENTIONAL FORMULATION

In a distribution network, current injection i and node voltages v are mapped as i = Yv, where Y is the admittance matrix. Active and reactive power injection can be formulated as p = Re(vi*) and q = Im(vi*), respectively. Vectors i and v contain complex values, and the symbol * denotes the complex conjugate operation.

DSSE is solved by minimizing

$$\hat{x} = \arg\min_x (z - h(x))^TW(z - h(x)), \quad (1)$$

where $x \in \mathbb{R}^n$ is the system states, $z \in \mathbb{R}^m$ denotes measurements and $h(\cdot)$ is the measurement function that maps states to measurements. The choice of system state for a distribution network is an important consideration as it affects the analysis and control of the network. Node voltages are a popular choice for system state as they provide a direct measurement of the electrical properties of the network. However, other quantities, such as line flows can also be used as system state variables depending on the specific application. Section II-B presents detailed formulations for x,
TABLE 1. Recent works on DSSE.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Methodology/Focus</th>
<th>Contributions</th>
<th>Research Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>[9]</td>
<td>Uncertainty measurement theory; Robustness against large measurement error</td>
<td>Simultaneous solution for DSSE and bad data detection is performed. Maximum normal measurement rate-based bad data detection is utilized.</td>
<td>Validation on more IEEE test systems is needed including a more realistic test case such as IEEE 123 instead of the modified 13-bus system. All testing scenarios consider heavy noise at a specific node.</td>
</tr>
<tr>
<td>[10]</td>
<td>Gradient-based multi-area algorithm; Parallel &amp; distribution computation of large DSSE problems</td>
<td>Gradient-based WLS algorithm is designed for multi-area case with two objectives: 1) robust against large measurement errors 2) Tree/subtree topological structure of distribution system is used for computational benefits 3) Real-time state estimation under changing load conditions. Tested on IEEE 37-node system and a 11000-node feeder (combined IEEE 8500 and EPRI C97).</td>
<td>Linearization used during modeling is valid only for balanced and lossless systems—bad assumptions for a distribution system.</td>
</tr>
<tr>
<td>[11]</td>
<td>WLS for low voltage distribution grid (LVDG)</td>
<td>Given significant voltage drop in neutral conductor in the LVDGs, the WLS algorithm is modified by including neutral voltage as state variables. Measurement functions and the Jacobian matrix are adjusted accordingly.</td>
<td>Proposed approach needs to be validated on other test cases.</td>
</tr>
<tr>
<td>[12]</td>
<td>WLS based on symmetrical components: Compensation currents are utilized for decoupling of the sequence networks</td>
<td>WLS problem is decomposed into three separate problems for each phase. Incorporates different types of measurements and topologies. Computationally efficient for large systems.</td>
<td>Mutual couplings between sequences is neglected.</td>
</tr>
<tr>
<td>[13]</td>
<td>DSSE formulation using virtual reference bus.</td>
<td>DSSE problem is formulated assuming that the root bus has unbalanced voltages and a balanced virtual bus is used as reference bus. Standard assumptions of balanced feeder head or using angle of only one phase is challenged.</td>
<td>Proposed approach should be further validated using real measurements.</td>
</tr>
<tr>
<td>[14]</td>
<td>Focus is to avoid running multiple Monte Carlo simulations</td>
<td>Standard WLS is reformulated using Taylor series expansion. The resulting form offers computational benefits.</td>
<td>Small voltage drops and normal voltage limits are assumed for distribution system.</td>
</tr>
<tr>
<td>[15]</td>
<td>DSSE with combined WLS and Levenberg-Marquardt algorithm</td>
<td>LM algorithm is incorporated within the WLS problem. DSSE is formulated using power measurements.</td>
<td>Slow sampling rate of smart meters.</td>
</tr>
<tr>
<td>[16]</td>
<td>Semidefinite programming-based DSSE</td>
<td>DSSE is reformulated as a rank-constraint SDP problem. Rank reduction and convex iteration approaches have been used to obtain low rank solution.</td>
<td>Accuracy of the solution suffers due to omission of the rank one condition.</td>
</tr>
<tr>
<td>[17]</td>
<td>Constrained low-rank matrix completion</td>
<td>DSSE is formulated as matrix completion problem under low-observability conditions.</td>
<td>AC power flow equations are linearized. Under exact formulations, the employed SDP may not give good solutions.</td>
</tr>
<tr>
<td>[18]</td>
<td>DSSE considering multiple system uncertainties.</td>
<td>Interval state estimation is solved, obtaining bounds on states in the presence of uncertainties due to measurement noise, imprecise line parameters, and uncertain DG outputs.</td>
<td>Applications of the proposed work could be extended.</td>
</tr>
<tr>
<td>[19]</td>
<td>DSSE for three-phase four-conductor distribution network with grounded wye-connected loads</td>
<td>Scalability of DSSE is addressed by considering state variables as voltages of non-neutral and non-zero injection phases. Smart meter measurements are used to estimate loads (pseudo measurement).</td>
<td>The method is only evaluated on a modified IEEE 123-bus distribution system and needs to be tested on other distribution systems. Real-time data from smart meters are needed (usually not available).</td>
</tr>
<tr>
<td>[20]</td>
<td>Fault location algorithm, PMU-based state estimation</td>
<td>In the SE algorithm is revised considering faulted scenarios. Furthermore, to speed up the calculations, the distribution system is partitioned into multiple zones—SE applied in each zone using PMU data.</td>
<td>Results for large test cases are needed.</td>
</tr>
</tbody>
</table>

where \( k \) denotes iteration counts, \( H_k \) is the residual Jacobian (w.r.t. \( x \)), and \( G_k = H_k^TH_k \). State update during optimization can be obtained as \( x_{k+1} = x_k + \Delta x_k \).

Note that the problem (1) suffers from non-convexity due to nonlinear measurement functions. Hence, Newton-based methods face issues such as convergence, finding proper initialization, and matrix ill-conditioning. Other tools such as convex optimization have been used to address these limitations [23]. Quite often, the conditioning of an optimization problem can be improved by providing additional information to the optimizer. In the distribution system, there are many nodes/buses that have zero power injection. Providing this information as equality constraints could be helpful and is known as virtual measurements. The equality constraints can be incorporated into the objective
function via Lagrange multipliers,
\[
L = (z - h(x))^T \mathbf{W}(z - h(x)) + \lambda^T c(x)
\]
\[
[x, \lambda] = \text{arg min} \ L
\]
where \( L \) is the Lagrangian, \( c(x) \) denotes the equality constraints representing virtual measurements, and \( \lambda \) is the Lagrange multiplier. The Karush-Kuhn-Tucker conditions can be then obtained by finding \( \frac{\partial L}{\partial x} = 0 \) and \( \frac{\partial L}{\partial \lambda} = 0 \). Herein, the Gauss-Newton formulation becomes [24], [25]

\[
\begin{bmatrix}
\Delta x_k \\
\lambda_k
\end{bmatrix} = 
\begin{bmatrix}
\mathbf{H}^T \mathbf{W} \mathbf{C}^T & \mathbf{C} \\
\mathbf{C}^T & 0
\end{bmatrix}^{-1}
\begin{bmatrix}
\mathbf{H}^T \mathbf{W}(z - h(x)) \\
-c(x)
\end{bmatrix},
\]

(5)

where \( C = \frac{\partial C}{\partial x} \). For the standard WLS in (1), the objective function can be sensitive to bad data. Therefore, other formulations on residuals have also been employed such as least median squared or normalized residuals functions [26]. Similarly, robust estimation techniques which reduce the impact of bad data or outliers consider an objective function of form \( F(z - h(x)) \), where \( F \) is a robust loss function, e.g., Huber’s functions [27].

Similar to (2) and (5) that use the Gauss-Newton approach, other Newton-based techniques have also been extended for distribution system state estimation (shown in (1)). For example, the quasi-Newton BFGS method was used for state estimation to simplify the Hessian matrix computation [28]. Similarly, optimization tools such as genetic algorithm [29] and convex optimization [30] have also been extended to DSSE. Table 1 lists some recent works on DSSE literature along with plausible research gap for future works.

B. STRUCTURE AND ESTIMATION METHODS

Given the WLS formulation in (1), different formulations lead to different measurement functions which map \( z \) to the system states \( x \). Consequently, the choice of states determines the function \( h(.) \). For the distribution system, popular choices for states include bus voltages and branch currents. Moreover, these variables can be represented as phasors or in rectangular forms, each leading to a distinct formulation [2], [30].

1) NODE VOLTAGES AS STATES

In the following, we show a sample formulation for this case. The state variables can be defined as,

\[
x = [e_1^a, f_1^b, e_1^b, f_1^a, \ldots, e_n^a, f_n^b, e_n^b, f_n^a, \ldots, e_1^c, f_1^c, e_1^c, f_1^c, \ldots, e_n^c, f_n^c].
\]

(6)

The three-phase measurement equation and Jacobian matrix of the power injection measurement is shown below. \( e/f \) are the voltage phasor’s real/imaginary component; \( G/B \) are the real/imaginary component of the admittance matrix; \( i/k \) are the considered nodes; and \( \phi_i \) is the set of nodes connected to \( i \). \( B_1 \) is the set of phases connected to phase \( t \). Superscript \( d \) denote the phases \( a, b, c \).

Measurement equation for power:

\[
P_i^d = f_i^d \sum_{k \in \phi_i} \sum_{l \in B_1} (G_{ikl}^d f_l^d + B_{ikl}^d e_l^d)
\]

(7)

Jacobian:

\[
\frac{\partial P_i^d}{\partial e_k^d} = \begin{cases}
    f_i^d g_{ik}^d + e_i^d b_{ik}^d & \text{if } i = k \text{ and } d = t \\
    f_i^d b_{ik}^d + e_i^d g_{ik}^d & \text{else}
\end{cases}
\]

(8)

\[
\frac{\partial P_i^d}{\partial f_k^d} = \begin{cases}
    f_i^d g_{ik}^d - e_i^d b_{ik}^d & \text{if } i = k \text{ and } d = t \\
    f_i^d g_{ik}^d + e_i^d b_{ik}^d & \text{else}
\end{cases}
\]

(9)

\[
\frac{\partial Q_i^d}{\partial e_k^d} = \begin{cases}
    f_i^d g_{ik}^d + e_i^d b_{ik}^d - \sum_{l \in B_1} (g_{ikl}^d b_{il}^d + b_{ikl}^d e_{il}^d) & \text{if } i = k \text{ and } d = t \\
    f_i^d g_{ik}^d - e_i^d b_{ik}^d & \text{else}
\end{cases}
\]

(10)

\[
\frac{\partial Q_i^d}{\partial f_k^d} = \begin{cases}
    -f_i^d b_{ik}^d - e_i^d g_{ik}^d + \sum_{l \in B_1} (b_{ikl}^d g_{il}^d - b_{ikl}^d e_{il}^d) & \text{if } i = k \text{ and } d = t \\
    -f_i^d b_{ik}^d - e_i^d g_{ik}^d & \text{else}
\end{cases}
\]

(11)

2) BRANCH CURRENTS AS STATES

Herein, the branch currents are chosen as the state variables. All the measurements are converted to the equivalent branch current dependent functions. In this case, a major advantage comes out that the Jacobian matrix is constant for all iteration unlike the previous case where the Jacobian matrices need to be calculated in each iteration [31].

C. PSEUDO-MEASUREMENT

Unlike transmission systems, distribution systems suffer from poor observability due to lack of reliable measurements. Usually, the measurement devices are only placed at important nodes such as a substation and a large part of the distribution network is not actively monitored. Note that having an observable system is a prerequisite in estimation theory. This problem is further exacerbated by the presence of ‘bad’ data in measurements which can be caused by faults or sensor malfunction. Traditionally, system observability can be assessed by studying the rank of the Jacobian matrix. For
example, in case of (4), the system is observable if the matrix \[
\begin{bmatrix}
H^T & WH & C^T
\end{bmatrix}
\] is full rank (i.e. invertible) [32]. Another approach to this observability issue would be using graph theory (spanning tree) to represent this Jacobian invertibility condition [33], [34]. It often becomes necessary to further augment the available measurements with synthetic (but realistic) data, called pseudo-measurement, obtained from historical data records. Usually, pseudo-measurements are forecasts of active and reactive power injection/consumption at different buses. Pseudo-measurements are an economical alternative to installing additional measurement devices.

Fig. 2 outlines a method to obtain pseudo-measurements for different buses based on probability theory. By studying the recorded load data from different consumers, load probability density histograms can be obtained for each bus. For example, consider the hourly load data (ERCOT year 2014 [35]) and its histogram for a bus shown in Fig. 2. This can further be used to identify the distribution functions that best represents system loads.

Traditionally, normal distributions are used to model load profiles [25]. However, in most cases, system loads cannot be described by a normal distribution function as evident from Fig. 2. A popular approach to tackle this problem is by using Gaussian mixture model (GMM) where the probability distribution of loads (at each bus) is denoted by weighted sum (convex combination) several Gaussian components [36], [37]. Note that the load profiles and the GMM components are calculated and stored prior to state estimation. Hereafter, the actual probability density function is modeled as a convex combination of multiple normal Gaussian distributions.

III. SENSORS

A. SENSOR TECHNOLOGIES

Some of the important sensor technologies for a distribution grid are listed below:

1) Phasor measurement units (PMU): PMUs are designed to provide phasor estimates for currents and voltages with a resolution of around 30-60 Hz. These measurements are accompanied by GPS-enabled time stamps. PMU comprises modules for data collection (CT/PT for current & voltage sensing, filters, and analog-digital converter), CPU block, GPS, and communication interface blocks [51]. This can further be augmented with additional features depending upon requirements. For example, a C37.118 compliant PMU for wide-area monitoring applications will also contain a Feature Extraction block [52].

2) Micro PMU: The traditional PMUs find most application in transmission networks. Given their high costs, they are rarely used in distribution networks. Nonetheless, \( \mu \)PMUs have 100 times the resolution of a conventional PMU and so are better suited for distribution networks [53].

3) Smart power meters: Conventional power meter comprises of current and voltage sensors, data type converters, and data processing capabilities. Smart meters augment these with features such as remote access, telemetry, data storage via remote servers. Such features can provide information on active/reactive power, power factor, THDs, etc, for real-time and historical values [54].

4) Current/Voltage sensors: These can be classified based on their operating principles, such as electromagnetic and optical sensors [52]. Electromagnetic sensors use electromagnetic principles such as the conventional current and voltage transformers with primary/secondary windings on a ferromagnetic core. Rogowski coil current sensors are used where conventional current and voltage transformers with primary/secondary windings on a ferromagnetic core. Rogowski coil current sensors are used where the cores are non-magnetic with output voltage being proportional to the derivative of the current in coil. Hall effect sensors use the named principle where the sensor produces the voltage in proportion to the strength of magnetic field. Giant magnetoresistance sensors utilize the property of certain materials which change their resistances under magnetic fields. Additionally, there are optical sensors that can be used for both current and voltage sensors. To sense currents, the magnetic field produced by the conductor is measured by observing

\[
\text{REFERENCES}
\]

\[
\text{FIGURE 2. Overview of a pseudo-measurement generation technique. Here, the actual probability density function is modeled as a convex combination of multiple normal Gaussian distributions.}
\]
TABLE 2. Sensors & DSSE.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Methodology/Focus</th>
<th>Contributions</th>
<th>Research Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>[40], [41]</td>
<td>Non-dominated sorting genetic algorithm II</td>
<td>Multi-objective optimization problem aiming to minimize various objectives such as the number of PMUs, state estimation uncertainty, sensitivity, and stability.</td>
<td>State estimation is done using only information from PMU and zero injection buses. Data from AMI and SCADA are ignored. High computation time (as high as 10 hours for 18 bus system and 112 hours for 141 buses [40]).</td>
</tr>
<tr>
<td>[42]</td>
<td>Integer linear programming</td>
<td>Minimizes capital cost of PMU installation and maximizes system observability. Assumes an active distribution network with DRs.</td>
<td>Only considers PMU measurements. A higher number of measuring devices are needed for similar case studies. Similarly, cost of communication or storage not considered.</td>
</tr>
<tr>
<td>[43]</td>
<td>Greedy algorithm, genetic algorithm, hybrid GA</td>
<td>Optimization tool for sensor and recloser placement.</td>
<td>Large networks should also be considered.</td>
</tr>
<tr>
<td>[44]</td>
<td>Convex optimization</td>
<td>Multiple metrics for optimal PMU placement are evaluated to obtain a set of bounds for the optimal solution.</td>
<td>Better search algorithms could be used to obtain solutions closer to low bounds. Network reconfiguration is ignored.</td>
</tr>
<tr>
<td>[22]</td>
<td>Mixed integer SDP</td>
<td>Scalability of the optimal meter placement problem is improved using the Barrier method.</td>
<td>Network reconfiguration is ignored.</td>
</tr>
<tr>
<td>[45]</td>
<td>Mixed integer SDP</td>
<td>Maximizes worst-case estimation accuracy – Robust placement of sensors.</td>
<td>Uncertainty in measurement malfunction is ignored. Although computationally efficient, relaxation only provides a suboptimal solution and a local optimizer is used.</td>
</tr>
<tr>
<td>[46]</td>
<td>Modified smart meter (MSM)</td>
<td>Existing smart meter hardware can be augmented with a precision time protocol to obtain synchronized measurements (similar to PMUs).</td>
<td>The measurement accuracy and resolution of these meters are low.</td>
</tr>
<tr>
<td>[47]</td>
<td>Integer linear programming, optimal sensor (µPMU) placement</td>
<td>Optimal placement of µPMUs considering the number of µPMUs. A cost model for µPMUs incorporating the number of channels is also included.</td>
<td>Only considers µPMU measurements and their installation costs.</td>
</tr>
<tr>
<td>[48]</td>
<td>Random forests (based on CART &amp; Bagging algorithms) for DSSE</td>
<td>Voltage estimation is performed for a low-voltage distribution grid using sensors from cable television (CATV) networks (5 min resolution). A high-bandwidth and secure communication infrastructure already exists for this network.</td>
<td>CATV sensors are decoupled from grid nodes.</td>
</tr>
<tr>
<td>[49]</td>
<td>Evaluate performance classes of PMUs (T0, T1, ..., T5) w.r.t DSSE accuracy</td>
<td>Estimates time synchronization accuracy for PMUs in distribution network for acceptable DSSE performance. Multiple indices such as confidence level, total vector error, RMSE, and absolute error of P_{loss} are provided for DSSE performance evaluation.</td>
<td>Only considers the impact of PMU accuracy on DSSE performance. Other factors, such as the number of PMU or their location, can also have a significant impact on DSSE performance. Assuming full observability during testing is unrealistic.</td>
</tr>
<tr>
<td>[50]</td>
<td>Meter placement, error analysis of DSSE</td>
<td>Studies the impact of low measurements (current and power) on the estimation of branch currents. Test case using an 11 kV 95-bus distribution network.</td>
<td>The DSSE needs to be in a specific formulation for the provided analysis (with branch currents as states). Similar study on DSSE with node voltages as states could be a good research direction.</td>
</tr>
</tbody>
</table>

the change in light angle (an optical fiber surrounds the conductor). Voltage sensors utilize the Pockels effect in which a crystal is used that rotates the polarized light proportional to the electric field.

B. SENSOR PLACEMENT

It is infeasible to install measurement devices at all the buses. In reality, the DSSE is performed using only a few real measurements along with pseudo-measurements. Herein, the problem of optimal placement of measuring devices at different nodes of the distribution system becomes important [22], [42].

There are two main approaches in this area. Firstly, having a fixed number of measurement devices while performing optimal placements for the minimum estimation error [55], [56]. Secondly, finding both the optimal number of measurement devices and their location in the network [57], [58], [59]. This problem is often formulated as an optimization problem with different objective functions. For example, for the DSSE problem in (1) and (2), the inverse of the gain matrix \( G \) denotes the error covariance matrix which can be used here to indicate the quality of estimation accuracy. Similarly, some works in the literature have used meter cost or network observability as an objective function [43]. Table 2 shows some recent works in this field from the DSSE perspective. Table 4 in [3] provides a summary on the body of works in this field. Reference [60] provides a detailed review on the application of different heuristic algorithms for meter placement problems.

IV. DATA DRIVEN APPROACHES

Data-driven approaches have been extensively applied in every facet of DSSE tools in recent years. Such methods use recorded data to find solutions without resorting to modeling its dynamics. For example, a neural network can be fed with a set of inputs and corresponding actual outputs in order to learn their relationship with no knowledge about the system dynamics (black-box model). Nonetheless, these techniques can be very helpful for real-time applications as most computations are done during offline training. Data-driven approaches have been successfully used for many problems such as pseudo-measurement generation [61], state estimation [62], [63] and sensor placement. Researchers have also explored probabilistic approaches to address these issues, e.g., Bayesian techniques or Game theory [25], [64].
The general structure for machine learning-based techniques is shown in Fig. 3. Usually, this consists of training data sets, a function approximator like neural networks, and a training algorithm. Training data is a set of inputs and outputs created from recorded data from the systems. The function approximator is selected by the user as per requirements. Fig. 3 shows a feed forward neural network that is trained offline and later (after successfully trained and validated) used to perform estimation. In the following, we discuss some data-driven approaches applied for DSSE and pseudo-measurement generation.

Efficacy of the trained model depends on the quality of training data. For a distribution network, where data is coming from multiple sources such as AMI or SCADA with different sampling times, making the best use of such diverse data can be challenging. Reference [73] formulates a hybrid DSSE formulation that utilizes both AMI and SCADA data. A set of feedforward NNs are trained for different topologies (of a specific distribution network) that use SCADA measurements (sampled every 15 minutes). The AMI data, which arrives at an hourly rate, enables the use of weighted least absolute value to perform state estimation. Reference [64] proposes the use of smart meter data to generate injection distributions, which are then used to train a deep neural network to perform state estimation (for unobservable distribution systems). This hybrid approach aims to leverage the complementary nature of both AMI and SCADA data, leading to more accurate and reliable state estimation results. Furthermore, incorporating advanced machine learning techniques, such as deep neural networks, can help improve the robustness and adaptability of the state estimation process, particularly in the presence of various network topologies and data uncertainties. Table 3 presents some recent works in literature that have used data-driven methods to solve different DSSE-related problems such as estimation, pseudo-measurements generation, initialization of Newton-based methods, and fault identification.

V. CHALLENGES AND OPPORTUNITIES

Given the scope of topics discussed in this paper, there can be many avenues that present great opportunities in the future. Some of these are presented in the following sections. Note that a lot of issues and challenges for DSSE for different applications will be similar as the core functionality remains same.

A. DEMAND RESPONSE

Demand response (DR) enables loads to alter their consumption motivated by economics. Herein, there can be two scenarios, 1) The user receives incentives for allowing load control and load curtailment by the network operator, 2) The user schedules its DR-enabled loads as per the energy prices [74], [75], [76]. Each of these types of DR programs creates its own challenges and opportunities. Although the direct control of loads by the network operator is an effective tool against system instability, it can raise privacy concerns among users [74], [77]. In the second approach, the network operator formulates the energy prices to motivate the consumers to actively participate. Although this approach gives the end-user full control, the network operator will often struggle to fulfill the DR requests of all users, given that the priority will always be to follow network constraints. DSSE can support the demand response by providing the distribution grid’s operating condition based on available measurements from sensors, smart meters, or other measurement devices. Based on the state variable estimates from DSSE, portion of the grid experiencing voltage swings or high demand can be quickly identified. Utilities can use this information to perform appropriate demand response measures, such as using DERs or energy storage to generate additional power, or reduce energy consumption for certain areas. Similarly, consumers can use this information in real-time to monitor their energy usage and exploit the grid incentives of demand response.

Some of the challenges that DSSE can face when handling demand response loads include:

1) Measurement accuracy: The accuracy of DSSE, as explained in Section II, directly depends on the quality of the measurements provided, as evident from (1). Poor data quality or insufficient data can result in inaccurate estimations of the system state, rendering demand response ineffective or even detrimental. Therefore, it is crucial that DSSE algorithms are robust against measurement noise. To this end, there have
been limited studies on topics such as missing data estimation [78], [79], [80], [81] and outlier detection [82], [83] for demand response-enabled loads. Another major challenge is managing data collected in various formats and from different devices [85], [86], [87], e.g., data in the distribution grid could be coming from µPMUs and SCADA with different sampling rate and formats [88].

2) Privacy and Security: There are several security challenges associated with deploying DSSE algorithms. DSSE requires data from multiple sources, including sensitive information such as power consumption data. Aggregated data can reveal information about consumers or the grid, which can be exploited by attackers [89]. In addition to this, data injection attacks (i.e., adding fake data into the measurements) can result in incorrect state estimation, which can lead to incorrect demand response adjustments. Furthermore, the models used by DSSE for state estimation should also be protected against unauthorized access to prevent attackers from manipulating the estimation process [90], [91].

3) Scalability: DSSE needs to process data in real-time to support demand response actions. As the number of devices increases, there will be an additional burden of computation and communication. Hence, estimation techniques that incorporate data loss in communication become relevant [92], [93]. Similarly, estimation approaches that are inherently scalable should be investigated [94], [95], [96], [97].

4) Regulatory issues: Given the involvement of different entities, there can be various aspects of regulatory issues related to demand response and DSSE. DSSE algorithms require access to sensitive data, which are regulated by data protection laws such as the California Consumer Privacy Act (CCPA) [98], [99]. Regulatory bodies such as the North American Electric Reliability Corporation (NERC) have established cybersecurity standards for power utilities to ensure compliance [100], [101]. In addition, some utilities may lose revenue due to demand response and may thus discourage its implementation. Regulators must consider how to incentivize utilities to invest in demand response and DSSE technologies, while balancing the need for utilities to maintain their revenue streams.

B. RESILIENCE

DSSE provides the real-time view of distribution system from available measurements and data [102]. As a result, this capability is used for various applications, such as fault detection, restoration, and situational awareness [20], [103], [104], [105], [106]. Some of the less obvious applications for DSSE in this regard include:

1) Contingency analysis: The purpose of contingency analysis is to examine the grid response in the event of any contingency, such as equipment failure. Herein, the existing condition of the grid for contingency analysis can be obtained from the DSSE algorithm as discussed in [107].
2) **Black start**: Black start is the process of restoring a power system from total shutdown with no external power. DSSE can be used during this process to identify states from incomplete information and identify critical components that need to be energized first. Furthermore, during the black start process, DSSE can detect any anomalies. For example, DSSE can identify a generator whose output is other than the expected amount [102].

One aspect of system resiliency is the grid’s response to high-impact and low-probability events. The recovery problem should be flexible enough to include resources, such as DR programs, microgrids, and network reconfiguration, while ensuring that the formulated resilience problem consistently converges to optimal or near-optimal solutions [108], [109].

**C. HEAVY RENEWABLE ENERGY SOURCES PENETRATION**

The presence of renewable power in distribution grids poses several challenges to DSSE. The output power of renewable sources is difficult to predict, as it depends on environmental factors, and the voltage profiles of buses can be affected by a high number of distributed energy resources (DERs) [110]. To address these challenges, various approaches have been proposed. For instance, [110] proposes a DSSE for medium voltage networks using the Bayes rule to accommodate measurements and pseudo-measurements with any statistical model. The estimation observability largely depends on the pseudo-measurements, and thus [111] studies the correlation between state estimation and load profile data. In a similar vein, [112] presents a nonlinear weighted least squares approach that uses a scenario construction combining load and solar data to obtain net-load profiles as pseudo measurements. Herein, the sensitivity of state estimator accuracy to forecast uncertainties, sensor accuracy, and sensor coverage level is also studied. The impact of high levels of photovoltaic (PV) penetration and load on the accuracy of state estimation is studied in using load and PV stochastic models [113], [114]. The study shows that higher PV penetration has a negative impact on state estimation, both in magnitude and phase. In addition, [115] focuses on incorporating PV models, including power electronic devices, within the DSSE framework for active distribution systems, and [116] includes PV modeling in the DSSE formulation. Increased fluctuations in the power system due to renewable energy and demand response can also affect dynamic security assessment techniques. Instead of the conventional steady-state assumptions of the current operating points, [117] proposes a non-equilibrium initialization for dynamic security assessment.

**D. VIRTUAL POWER PLANT**

Virtual power plants (VPPs) are a collection of multiple DERs that are managed as a single entity [118], [119]. The VPP aggregates sources, such as PV, wind, and battery storage, through advanced control and communication technologies. Due to this heterogeneous aggregation, VPPs can help with better integration of renewable or intermittent energy sources and even perform ancillary services for the grid, such as frequency regulation and reactive power control. The aggregation of DERs by VPP is made possible solely due to the control and communication infrastructure. Hence, state estimation by DSSE algorithms play a massive role in VPP monitoring [120]. Different DERs have different characteristics, and accommodating them all can be challenging. Renewable sources, like solar and wind, are intermittent in nature, while different storage solutions have their own charging and discharging characteristics. Moreover, each source has its own capacity, response time, and ramp rates. Hence, the DSSE algorithms employed for VPP applications need to incorporate such diversity into their formulations [121].

**VI. CONCLUSION**

In this paper, we present a thorough discussion on distribution system state estimation along with its formulations and limitations. To provide a comprehensive document, we also discuss relevant related topics, such as pseudo-measurements and sensor placement. Significant references are also tabulated for multiple topics along with research gaps for further future work. Finally, we discuss some challenges and opportunities for DSSE from the point of view of some emerging topics such as demand response and virtual power plants.

**ACKNOWLEDGMENT**

The U.S. government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

**REFERENCES**


M. G. Damavandi, V. Krishnamurthy, and J. R. Martí, “Robust meter
placement for state estimation in active distribution systems,” IEEE

R. Singh, B. C. Pal, and R. B. Vinter, “Measurement placement in

P. A. Pegrat and S. Sulis, “Robustness-oriented meter placement for
distribution system state estimation in presence of network parameter
May 2013.

P. Janssen, T. Sezi, and J.-C. Maun, “Meter placement impact on
distribution system state estimation,” in Proc. IET Conf., vol. 1, 2013,
pp. L1.1–L1.4.

L. Lee and V. Centeno, “Comparison of
state estimation approaches,” in Proc. IEEE Conf., vol. 1, 2014,
pp. 1–6.

M. G. Damavandi, V. Krishnamurthy, and J. R. Martí, “Robust meter
placement for state estimation in active distribution systems,” IEEE

R. Singh, B. C. Pal, and R. B. Vinter, “Measurement placement in

P. A. Pegrat and S. Sulis, “Robustness-oriented meter placement for
distribution system state estimation in presence of network parameter
May 2013.


AJAY PRATAP YADAV (Member, IEEE) received the bachelor’s and master’s degree in electrical engineering from the Indian Institute of Technology Roorkee and the Indian Institute of Technology Kanpur, respectively, and the Ph.D. degree in electrical engineering from The University of Texas at Arlington. He is currently a Postdoctoral Research Associate at the Oak Ridge National Laboratory. His research interests include smart grids, power systems, and machine learning.

JAMES NUTARO (Senior Member, IEEE) received the Ph.D. degree in computer engineering from the University of Arizona. He is currently a Group Lead for the Computational Systems Engineering and Cybernetics Group, Oak Ridge National Laboratory. His research interests include event systems, systems modeling and simulation, and hybrid dynamic systems.

BYUNGKWON PARK (Member, IEEE) received the B.S. degree from Chonbuk National University, South Korea, in 2011, and the M.S. and Ph.D. degrees from the University of Wisconsin–Madison, Madison, WI, USA, in 2014 and 2018, respectively, all in electrical engineering. He is currently an Assistant Professor with the Department of Electrical Engineering, Soongsil University (SSU), Seoul, South Korea. Before joining SSU, he was a Research Staff with the Oak Ridge National Laboratory, Computational Sciences and Engineering Division. His research interests include modeling, simulation, control, and optimization of electrical energy systems.
JIN DONG (Member, IEEE) received the B.S. degree from the Harbin Institute of Technology, Harbin, China, in 2010, and the Ph.D. degree from The University of Tennessee at Knoxville, Knoxville, TN, USA, in 2016. He is currently a Research and Development Staff with the Grid Interactive Controls Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA. His research interests include transactive energy systems, learning-assist control, building-to-grid integration, and power system decarbonization.

YUQING DONG (Member, IEEE) received the B.S. and Ph.D. degrees in the electrical engineering from Sichuan University, China, in 2017 and 2022, respectively. She is currently a Postdoctoral Researcher at The University of Tennessee at Knoxville, Knoxville, TN, USA. Her research interests include high voltage direct current and power system stability.

BOMING LIU (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Pittsburgh, in 2020. He is currently a Research and Development Associate Staff with the Grid-Interactive Controls Research Group, Oak Ridge National Laboratory. His research interests include large-scale parallel and distributed computing, modeling and simulations, and machine learning.

JIAOJIAO DONG (Senior Member, IEEE) received the B.S. degree in information engineering and the M.S. and Ph.D. degrees in automation control from Xi’an Jiaotong University, China, in 2008, 2011, and 2016, respectively. She is currently a Postdoctoral Researcher at The University of Tennessee at Knoxville, Knoxville, TN, USA. Her research interests include power system planning and operation, renewable energy integration, and microgrids.

TEJA KURUGANTI (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from The University of Tennessee at Knoxville, Knoxville, TN, USA, in 2003 and 2012, respectively. He is currently a Distinguished Research and Development Staff Member and the Section Head for the Advanced Computing and Electromagnetics and Controls Section at ORNL. He joined ORNL, in 2015, from the U.S. National High Magnetic Field Laboratory, which he co-founded, as its Deputy Director. Prior to that, he was a Professor with Virginia Tech, Blacksburg, VA, USA.

He currently serves in the editorial boards of IEEE TRANSACTIONS ON POWER ELECTRONICS, IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN POWER ELECTRONICS, IEEE POWER ENGINEERING LETTERS, and IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY.