Robust Distribution System State Estimation with Hybrid SCADA and AMI Measurements

SanthoshKumar C, *Student Member, IEEE,* Ketan Rajawat, *Member, IEEE,*
Saikat Chakrabarti, *Senior Member, IEEE,* Bikash C. Pal, *Fellow, IEEE*

***Abstract*—With growing connection of Distributed Energy Resources (DERs), availability of Demand Side Response (DSR) technologies, deployment of Smart Meters (SMs), the distribution system needs advanced network automation for running the system efficiently. State estimation is the core driver of network automation. While the output from SMs will make the state estimation more accurate, advanced metering infrastructures (AMI) come with several challenges such as noisy, erroneous measurement including lost or missed measurements, exposure to cyber attack etc. This paper proposes a Projection Statistics (PS) based three-phase unbalanced distribution system state estimation (DSSE) which is robust against noisy distribution system measurements and bad-data attacks. This method considers measurement from hybrid sources such as SCADA, micro-Phasor Measurement Units (μPMUs), and SMs. Kalman smoother is used to fill the missing SM measurements and Expectation-Maximisation (EM) based forecasting is used to interpolate the hybrid measurements to a common timestamp and compensate for the delay in SM measurements. Extensive numerical comparisons are made on IEEE 13, 37 and 123 bus systems to test the robustness of the proposed DSSE against bad or noisy data and delayed SM measurements.**

***Index Terms*—Distribution system state estimation, expectation maximisation, hybrid measurements, projection statistics, robust estimator**

I. Introduction

STATE estimation is an indispensable part of operation and control of the modern power networks. With the growing developments in smart grid technologies, DSSE is now essential to reliably run distribution networks. Estimated states provide necessary information for control center and energy management applications. The inherent characteristic differences and the challenges of three-phase DSSE, when compared to Transmission System State Estimation (TSSE), such as high R/X ratio, unobservable system, complex measurement functions due to unbalanced phases, scalability are explained in [1].

In [2], Bose and Clements review the topology processing, numerical algorithms of state estimation, network observability and bad-data detection. The linearized model for balanced and unbalanced distribution system power flow is provided in [3]. The modeling of DSSE is provided in [4] and [5]. Various numerical algorithms of TSSE and DSSE with their convergence characteristics under various conditions are compared in [6], [7] . These methods provide numerically stable solution since, they prevent inverting Jacobian matrix in the normal equation. Also, the use of orthogonal factorization to avoid numerical ill-conditioning is provided in [8]. Reference [9] uses Hachtel's augmented matrix method to avoid ill-conditioning caused by bad-data processing and zero-injection measurements.

The bad-data detection considering leverage points is discussed by [4] and it also compares the effectiveness of Generalised Studentized Residuals (GSR) and Diagnostic Robust Generalized Potential (DRGP) methods in both TSSE and DSSE. However the robustness against missing measurements and hybrid measurement sources is not discussed in this paper. References [10] and [11] provide the Projection Statistics (PS) based Iterative Reweighted Least Squares (IR-WLS) estimator for TSSE that requires very little modification to the conventional state estimators. This method is robust against multiple outliers and leverage points, and can be started from flat-start. Reference [11] also describes the forecasting aided robust estimators in TSSE. The effect of False Data Injection (FDI) attacks on the power system state estimation and various attack strategies considering the resource constraints of the adversaries are discussed in [12] and [13].

Smart meters will soon be essential part of every consumer premise, and their measurements will be used in DSSE to improve the observability and accuracy of the system, provided communication infrastructure is in place for high frequency data transfer. SM measurements in low voltage (LV) distribution system are most prone to attacks due to diverse communication networks, attack-prone low-voltage systems, etc., Though the SMs are mostly available in LV distribution system, three-phase unbalanced DSSE on LV side is not practically feasible since the number of nodes are enormous, leading to more computation resource requirement, scalability issues, and Jacobian ill-conditioning [14]. On the other hand, the medium volage (MV) system may lack the required measurement devices and therefore the system may be unobservable and under-determined, unless SMs are present at every distribution transformer. To overcome the above problem, pseudo-measurements (historical data) [15], [16] were used in state estimation by assigning low weights to these measurements; but it inherently reduced the DSSE accuracy and reliability. References [17], [18] utilise the measurements obtained from SMs on LV side at the MV feeders, to make the MV distribution system observable, and obtain better accuracy of the estimates. The use of μPMUs in the distribution system

SanthoshKumar C, Ketan Rajawat and Saikat Chakrabarti are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, India, 208 016 (e-mail: csksanthosh93@gmail.com; ketan@iitk.ac.in; saikatc@iitk.ac.in).

B.C. Pal is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: b.pal@imperial.ac.uk).

is explained in [19]. Reference [20] provides an integrated state and load estimation method using k-means clustering; but it doesnot consider the asynchronous measurement arrival time of SM measurements.

One main problem in using the SM measurements is that they have asynchronous measurement arrival time. The delay in SM measurements is due to various operational and technical problems such as terrain, placement of SMs, repeaters and Data Concentrator Unit (DCU), lack of communication signal strength, limited bandwidth between SM and DCU, packet loss, malfunctioning of meters, and consumer participation. All the above factors contribute to communication delay between DCU and SMs, hence it is not possible to have a snapshot of the entire system. Reference [21] proposed a method in which the variance of error of the measured signal is modified to compensate for the non-synchronous AMI measurements. Since the accuracy of the DSSE depends on the quality of the data from the meters, measurement delay compensation for each measurement is provided based on the measurement delay and the load variation during that delay [22]. It also verifies that the load variation from one sampling point to the other follows a normal Gaussian distribution, using Shapiro-Wilk and Anderson-Darling tests. Though the methods proposed in [22] show significant improvement in estimates accuracy, the load variation is calculated based on historical data and the uncertainty in load variations between the actual and historical load is not taken into consideration. Also, the missing SM measurements and hybrid measurement sources are not considered in the previous work.

This paper proposes a robust DSSE considering hybrid measurement sources such as μPMUs, SCADA Remote Terminal Units (RTUs), and SM measurements. The main contribution of this paper are as follows.

1) The robustness of the three-phase DSSE is improved against noisy SM measurements and bad-data by attackers. Projection Statistics based iteratively reweighted DSSE is applied to three-phase unbalanced distribution system and its performance under bad-data and noisy measurements are observed.
2) Missing and non-synchronised SM measurements are compensated by Kalman smoothing and Expectation-Maximisation method is used to forecast and interpolate the SM data to a common timestamp. Kalman-smoothed and EM interpolated measurements are used in DSSE and the accuracy of DSSE are compared with existing delay compensation methods available in the literature.
3) Accuracy of the distribution system is improved by using the robust estimator incorporating (1) and (2) above, designed for MV DSSE considering hybrid measurement devices. The proposed estimator is tested extensively on IEEE 13, 37 and 123 node and its accuracy and robustness under various conditions are observed.

The rest of the paper is organized as follows: Section II explains the formulation and the need for factorisation methods in unbalanced three-phase DSSE. The various bad-data detection methods that are available in literature are discussed with the merits of Projection Statistics based IR-WLS in Section III. Section IV describes the need for hybrid measurement sources and delay compensation of SM measurements. Kalman smoothing for missing measurements and expectation-maximisation based load forecasting is used to interpolate the measurements to a common timestamp. Section V discusses the test system and simulation results. It also highlights the improvement in accuracy and robustness of the proposed method when compared to the existing methods. Section VI summarizes the results and concludes the paper.

II. Formulation of Three Phase Unbalanced DSSE

The unbalanced three-phase DSSE estimates the system states based on available measurements. The WLS state estimate is given by,

$$\Delta \mathbf{x} = [\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} [\mathbf{H}^T \mathbf{R}^{-1} (\mathbf{z} - \mathbf{f}(\mathbf{x}))] \tag{1}$$

where $\mathbf{x}$ represents the system states (voltage magnitude and phase angle per phase), $\mathbf{z}$ denotes the real-time measurements (real, reactive power and voltage magnitude), $\mathbf{R}$ is the error covariance matrix ($\mathbf{W} = \mathbf{R}^{-1}$), $\mathbf{f}(\mathbf{x})$ represents the measurement functions, and $\mathbf{H}$ denotes the Jacobian matrix, given by,

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & | & \mathbf{H}_2 \end{bmatrix} \tag{2}$$

$$\mathbf{H}_1 = \begin{bmatrix} \frac{\partial f_1^a}{\partial \delta_1^a} & \frac{\partial f_1^a}{\partial \delta_1^b} & \frac{\partial f_1^a}{\partial \delta_1^c} & \cdots & \frac{\partial f_1^a}{\partial \delta_n^a} & \frac{\partial f_1^a}{\partial \delta_n^b} & \frac{\partial f_1^a}{\partial \delta_n^c} \\ \frac{\partial f_1^b}{\partial \delta_1^a} & \frac{\partial f_1^b}{\partial \delta_1^b} & \frac{\partial f_1^b}{\partial \delta_1^c} & \cdots & \frac{\partial f_1^b}{\partial \delta_n^a} & \frac{\partial f_1^b}{\partial \delta_n^b} & \frac{\partial f_1^b}{\partial \delta_n^c} \\ \frac{\partial f_1^c}{\partial \delta_1^a} & \frac{\partial f_1^c}{\partial \delta_1^b} & \frac{\partial f_1^c}{\partial \delta_1^c} & \cdots & \frac{\partial f_1^c}{\partial \delta_n^a} & \frac{\partial f_1^c}{\partial \delta_n^b} & \frac{\partial f_1^c}{\partial \delta_n^c} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial f_{nz}^a}{\partial \delta_1^a} & \frac{\partial f_{nz}^a}{\partial \delta_1^b} & \frac{\partial f_{nz}^a}{\partial \delta_1^c} & \cdots & \frac{\partial f_{nz}^a}{\partial \delta_n^a} & \frac{\partial f_{nz}^a}{\partial \delta_n^b} & \frac{\partial f_{nz}^a}{\partial \delta_n^c} \\ \frac{\partial f_{nz}^b}{\partial \delta_1^a} & \frac{\partial f_{nz}^b}{\partial \delta_1^b} & \frac{\partial f_{nz}^b}{\partial \delta_1^c} & \cdots & \frac{\partial f_{nz}^b}{\partial \delta_n^a} & \frac{\partial f_{nz}^b}{\partial \delta_n^b} & \frac{\partial f_{nz}^b}{\partial \delta_n^c} \\ \frac{\partial f_{nz}^c}{\partial \delta_1^a} & \frac{\partial f_{nz}^c}{\partial \delta_1^b} & \frac{\partial f_{nz}^c}{\partial \delta_1^c} & \cdots & \frac{\partial f_{nz}^c}{\partial \delta_n^a} & \frac{\partial f_{nz}^c}{\partial \delta_n^b} & \frac{\partial f_{nz}^c}{\partial \delta_n^c} \end{bmatrix}$$

$$\mathbf{H}_2 = \begin{bmatrix} \frac{\partial f_1^a}{\partial V_1^a} & \frac{\partial f_1^a}{\partial V_1^b} & \frac{\partial f_1^a}{\partial V_1^c} & \cdots & \frac{\partial f_1^a}{\partial V_n^a} & \frac{\partial f_1^a}{\partial V_n^b} & \frac{\partial f_1^a}{\partial V_n^c} \\ \frac{\partial f_1^b}{\partial V_1^a} & \frac{\partial f_1^b}{\partial V_1^b} & \frac{\partial f_1^b}{\partial V_1^c} & \cdots & \frac{\partial f_1^b}{\partial V_n^a} & \frac{\partial f_1^b}{\partial V_n^b} & \frac{\partial f_1^b}{\partial V_n^c} \\ \frac{\partial f_1^c}{\partial V_1^a} & \frac{\partial f_1^c}{\partial V_1^b} & \frac{\partial f_1^c}{\partial V_1^c} & \cdots & \frac{\partial f_1^c}{\partial V_n^a} & \frac{\partial f_1^c}{\partial V_n^b} & \frac{\partial f_1^c}{\partial V_n^c} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial f_{nz}^a}{\partial V_1^a} & \frac{\partial f_{nz}^a}{\partial V_1^b} & \frac{\partial f_{nz}^a}{\partial V_1^c} & \cdots & \frac{\partial f_{nz}^a}{\partial V_n^a} & \frac{\partial f_{nz}^a}{\partial V_n^b} & \frac{\partial f_{nz}^a}{\partial V_n^c} \\ \frac{\partial f_{nz}^b}{\partial V_1^a} & \frac{\partial f_{nz}^b}{\partial V_1^b} & \frac{\partial f_{nz}^b}{\partial V_1^c} & \cdots & \frac{\partial f_{nz}^b}{\partial V_n^a} & \frac{\partial f_{nz}^b}{\partial V_n^b} & \frac{\partial f_{nz}^b}{\partial V_n^c} \\ \frac{\partial f_{nz}^c}{\partial V_1^a} & \frac{\partial f_{nz}^c}{\partial V_1^b} & \frac{\partial f_{nz}^c}{\partial V_1^c} & \cdots & \frac{\partial f_{nz}^c}{\partial V_n^a} & \frac{\partial f_{nz}^c}{\partial V_n^b} & \frac{\partial f_{nz}^c}{\partial V_n^c} \end{bmatrix}$$

where n denotes the number of states, nz denotes the number of measurement functions per phase ($ph \in a, b, c$), $\mathbf{f}_i^{ph}$ represents the measurement functions $\mathbf{P}_i^{ph}$, $\mathbf{Q}_i^{ph}$, $\mathbf{P}_{ij}^{ph}$, $\mathbf{Q}_{ij}^{ph}$, $|\mathbf{V}_i^{ph}|$ for each phase from different measurement devices such as SCADA, μPMUs, and SMs.

Using flat-start and Newton's method, the states are estimated iteratively till specified tolerance is obtained.

$$\mathbf{x}_{new} = \mathbf{x}_{old} + \Delta \mathbf{x} \tag{3}$$

The DSSE has poor convergence due to high condition number of the gain matrix, caused by highly unbalanced phases, high R/X ratio due to shorter line lengths, missing lines, zero injection, power flow in short line as measurement, and unbalanced loads [1], [6]. Therefore, matrix factorisation methods need to be employed to reduce the ill-conditioning in the inversion of the gain matrix, $\mathbf{G}(\mathbf{x}) = [\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]$ which is required to perform the iterations shown below.

$$\Delta\mathbf{x} = [\mathbf{G}(\mathbf{x})]^{-1}[\mathbf{H}(\mathbf{x})^T\mathbf{R}^{-1}\Delta\mathbf{z}] \tag{4}$$

Various factorisation methods such as orthogonal, Peter-Wilkinson, and hybrid (QR) method are used to reduce the condition number of the measurement Jacobian matrix [14].

Apart from the above methods, equality constrained Hachtel augmented matrix method, as mentioned in [9] and [14], is also used in DSSE, since it is effective against ill-conditioning due to virtual measurements. Since determining the optimal value of α (Hachtel weighing factor) requires additional iterations and varies from system to system, it is not considered in this paper.

III. Robust Unbalanced DSSE

Though WLS based state estimators are more common, the estimates are sensitive to measurement noise and bad-data. The measurements from the distribution system are prone to noise, due to various operational and technical problems and bad-data attacks, which significantly affect the estimates of the system, irrespective of the factorisation methods used [20]. Due to easily accessible measurement devices, unsecured wireless data, infected devices, and misconfigured firewalls, cyber attacks can affect the DSSE to a great extent [23]–[25]. False Data Injection (FDI) attacks can introduce erroneous measurements to the data, thereby reducing the accuracy of the estimates.

A. Bad-Data Detection and Correction Techniques

Numerous bad-data detection and correction methods are existing for TSSE. Very few works are available for DSSE [4]. To improve the reliability of the estimates, it is observed that the estimator is made robust at the cost of additional computation time and complexity. Following are few of the methods available in the literature [10] to detect bad-data and their response to iterative re-weighting.

For Gaussian distributions, Median Absolute Deviation (MAD) and Mahalanobis Distance (MD) are used to detect outliers. But they are not suited for muliple outliers and prone to masking and swamping effects [10].

Hat matrix $\mathbf{K} = \mathbf{H}[\mathbf{H}^T\ \mathbf{H}]^{-1}\mathbf{H}^T$ is another popular method to detect leverage points by making use of diagonal elements of the measurement Jacobian ($\mathbf{H}$). However, it is prone to masking and swamping effects, as mentioned in [4]. Also, for a highly unbalanced distribution system, calculation of $[\mathbf{H}^T\ \mathbf{H}]^{-1}$ may lead to further ill-conditioning.

The comparative study on robustness of Largest Normalized Residual (LNR), Generalised Studentized Residuals (GSR) and Diagnostic Robust Generalized Potential (DRGP) to leverage points in both TSSE and DSSE, is discussed in [4]. However it doesnot discuss the robustness to hybrid and missing measurements in three-phase unbalanced distribution state estimation.

B. Projection Statistics based Robust IR-WLS DSSE

Since all the above estimators have limitations in detecting multiple bad-data, some method which is robust to leverage points, multiple outliers is needed. Projection Statistics (PS) offers better robustness to multiple outliers, identifies good leverage points and reduces effects of multiple outliers at the cost of higher computation time. Though PS is highly robust to multiple outliers and widely used in TSSE [10] [11], it is not well-explored in three-phase unbalanced DSSE. In this paper, PS is employed to make DSSE robust against noisy measurements and bad-data from distribution system measurement sources.

Projection Statistics for the measurement i is given by [10],

$$\mathbf{PS}_i = \max_{\mathbf{h}_k} \frac{|\mathbf{h}_i^T\mathbf{h}_k|}{\boldsymbol{\beta}_{PS}} \quad k = 1, ..., nz \tag{5}$$

$$\boldsymbol{\beta}_{PS} = 1.1926 \times \underset{i}{\text{med}}\{\underset{i\neq j}{\text{med}}\{|\mathbf{h}_i^T\mathbf{h}_k + \mathbf{h}_j^T\mathbf{h}_k|\} \tag{6}$$

where $\underset{i}{\text{med}}$ represent the median of the ith data-set, $\mathbf{h}_k$ represents the set of nz row vectors of the weighted Jacobian matrix $\hat{\mathbf{H}}$ [$\hat{\mathbf{H}} = \mathbf{R}^{-1/2}\mathbf{H}$].

The value of $\mathbf{PS}_i$ calculated in (5), can be applied to IR-WLS algorithm and the new weights are calculated for each measurement as follows.

$$\mathbf{w}_i = \min\left\{1, \left(\frac{\chi^2_{k,0.975}}{\mathbf{PS}_i}\right)^2\right\} \tag{7}$$

where $\chi^2_{k,0.975}$ is the Chi-square estimate for the ith measurement, k is the degrees of freedom (number of non-zero entry in that row), and $\mathbf{PS}_i$ is the Projection Statistic value for the ith measurement. If $\mathbf{PS}_i > \chi^2_{k,0.975}$, then the ith measurement is considered to be leverage measurement.

The problem with $\mathbf{w}_i$ is that, it is down-weighted irrespective of its residual and the nature of measurement. Therefore, from Schweppe-Huber type Generalised M-Estimator (SHGM),

$$\mathbf{r}_{si} = \frac{\mathbf{r}_i}{\mathbf{w}_i\mathbf{R}_i^{-1/2}} \tag{8}$$

$$\psi(\mathbf{r}_{si}) = \begin{cases} \mathbf{r}_{si}, & |\mathbf{r}_{si}| \leq c \\ c \times \text{sign}(\mathbf{r}_{si}), & \text{elsewhere} \end{cases} \tag{9}$$

where $\mathbf{r}_i = \mathbf{z}_i - \mathbf{f}_i(\mathbf{x})$ is the measurement residual. The value of c varies between 1 and 3 (c = 1.5 in simulations).

$$\mathbf{q}(\mathbf{r}_{si}) = \frac{\psi(\mathbf{r}_{si})}{\mathbf{r}_{si}} \tag{10}$$

$$\mathbf{W}_{PS,i} = \text{diag}(\mathbf{q}(\mathbf{r}_{si})) \tag{11}$$

The new weights are calculated using the weighing factor $\mathbf{W}_{PS,i}$ as follows.

$$\mathbf{R}_{new,i} = \mathbf{R}_{old,i}\ /\mathbf{W}_{PS,i} \tag{12}$$

The new error covariance matrix, $\mathbf{R}_{new}$, is used in (4) to obtain the robust estimates against bad and noisy measurements.

IV. Hybrid Measurement Sources and Delay Compensation

A. Hybrid Measurement Sources in Distribution System

Traditional distribution system lacks sufficient measurement sources. To improve the observability of the system and the reliability of the estimates, pseudo-measurements are typically used [15]. But it leads to poor reliability of the estimates. With the modernisation of the distribution system, various measurement sources such as RTU-SCADA, μPMUs, and SMs are available. Among these measurement devices, SMs are the main source of measurements in LV distribution system, as most utilities started the installation of SMs, replacing the conventional meters. These measurements are also used in DSSE to improve the system observability and accuracy of the estimates. Though the SMs are synchronized with internet time, due to various technical and operational factors such as terrain, placement of SMs, repeaters and Data Concentrator Unit (DCU), lack of communication signal strength, limited bandwidth between SM and DCU, packet loss, malfunctioning of meters, packet drop, and consumer participation, the measurements from SMs are delayed. Therefore, the snapshot of the entire system is not available, which may affect the DSSE accuracy. This leads to the need for measurement delay compensation for SM measurements.

References [21] and [22] explored the use of SMs in DSSE and proposed a method for delay compensation based on tuning the variance of the measurement devices. It also studies the short-term load variation patterns and confirmed that the load variation follows a Gaussian normal distribution. Reference [22] considers that SMs are present on distribution transformers and the DSSE is performed only up to the distribution transformer level.

In practical distribution system, neither all the MV feeders are equipped with SCADA or μPMUs, nor all the LV feeders are equipped with SMs. Therefore, hybrid measurement devices are needed to make the system observable and reduce the need for pseudo-measurements. The proposed method uses all the three measurement sources viz., SCADA, μPMUs, and SMs in MV DSSE to improve the accuracy of the estimates. Since the DSSE is carried out for MV distribution system, for nodes without SCADA and μPMU measurements, the SM measurements are aggregated as follows.

$$\mathbf{P}_{i,MV} = \sum_{j=1}^{N} \mathbf{P}_{j,SM} + \mathbf{P}_{loss,LV} \tag{13}$$

$$\mathbf{Q}_{i,MV} = \sum_{j=1}^{N} \mathbf{Q}_{j,SM} \tag{14}$$

where N is the number of SMs downwards to a feeder, $\mathbf{P}_{i,MV}$ and $\mathbf{Q}_{i,MV}$ denotes the real and reactive power values at the MV distribution feeders, $\mathbf{P}_{j,SM}$ and $\mathbf{Q}_{j,SM}$ represents the real and reactive power measurements obtained by SMs present in each LV node, $P_{loss,LV}$ accounts for the line losses in LV distribution network considering that the LV distribution network is known. Alternatively, when the SMs are available at the LV distribution transformer, the corresponding measurements can be considered.

B. Delay compensation for smart meter measurements using EM method

A key problem in using the SM measurements is that there is a possibility of missing measurements due to network interruption, packet loss and the delay in measurement arrival time. Since the loads are varying in nature, the measurement provided by SMs after that delay may not be reliable. Therefore one needs to compensate for missing and delayed measurements from SMs. Reference [22] proposes a method in which the mean and variance of the error of the out-of-date (OD) measured signal is modified to compensate for the non-synchronous AMI measurements as below.

$$\boldsymbol{\mu}_{OD_i}(t) = \frac{t}{t_{max}} \times \Big(\boldsymbol{\mu}_{OD_i}(t_{max}) - \boldsymbol{\mu}_{OD_i}(0)\Big) + \boldsymbol{\mu}_{OD_i}(0) \tag{15}$$

$$\boldsymbol{\sigma}^2_{OD_i}(t) = \frac{t}{t_{max}} \times \Big(\boldsymbol{\sigma}^2_{OD_i}(t_{max}) - \boldsymbol{\sigma}^2_{OD_i}(0)\Big) + \boldsymbol{\sigma}^2_{OD_i}(0) \tag{16}$$

where t represent the time since last measurement update (in seconds), t_{max} denote the maximum interval between two consecutive samples (900s for SMs, 60s for RTUs), $\boldsymbol{\mu}_{OD_i}(0)$, $\boldsymbol{\mu}_{OD_i}(t)$, $\boldsymbol{\sigma}_{OD_i}(0)$, $\boldsymbol{\sigma}_{OD_i}(t)$ are the delay compensated mean and standard deviation of the first and last points of the measurement interval. However the above method does not consider missing measurements, and hybrid non-synchronized measurement sources.

Since the short-term load variations follows a normal distribution [22], in this paper, using the sequence of load measurements from the SMs, Kalman filtering based forward-backward recursion (Rauch-Tung-Streibal smoother) is used to fill the missing measurements and Expectation-Maximisation (EM) method is used to interpolate the hybrid measurements to a common timestamp [26].

The hidden state variables are computed using the sequence of measurements in Kalman smoothing, which are later used by EM method to calculate the updated parameters. Let $\mathbf{z}$ denote the observations or measurements, $\mathbf{A}$ denote the state transition matrix, $\mathbf{C}$ represent the generative or observation weight matrix, $\boldsymbol{\Gamma_s}$ and $\boldsymbol{\Gamma_o}$ represent the state noise and observation noise respectively, $\boldsymbol{\mu_t^0}$ and $\boldsymbol{\sigma_t^0}$ represent the initial mean and covariance of the state variable respectively, T denote the number of observations, ϵ represent the tolerance for change in log-likelihood, L denote the log-likelihood, and α, β, γ, γ_1, γ_2, $\mathbf{J}$ represent the intermediary model variables in EM learning. The state space model is given below [26].

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w} \qquad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_s) \tag{17a}$$

$$\mathbf{z}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v} \qquad \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_o) \tag{17b}$$

Algorithm 1 Kalman smoothing - Inference

Require: measurements $(\mathbf{z})$, and initial model parameters $(\mathbf{A}, \mathbf{C}, \mathbf{\Gamma}_s, \mathbf{\Gamma}_o, \boldsymbol{\mu}_1^0, \boldsymbol{\sigma}_1^0)$
% Forward Pass - Kalman smoother
for t = 1 to T **do**
Calculate mean $(\boldsymbol{\mu}_t^t)$ and covariances $(\boldsymbol{\sigma}_t^t)$ respectively of the hidden state variable given observations $(\mathbf{z_1}, ..., \mathbf{z_t})$.
end for
Require: covariance across time T and $T-1$ $(\hat{\boldsymbol{\sigma}}_{T,T-1})$
% Backward Pass - Rauch recursions
for t = T to 2 **do**
Calculate mean $(\hat{\boldsymbol{\mu}}_{t-1})$ and covariance matrices $(\hat{\boldsymbol{\sigma}}_{t-1})$, $(\hat{\boldsymbol{\sigma}}_{t,t-1})$ of the full smoother estimates.
end for
return $\hat{\boldsymbol{\sigma}}_t, \hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\sigma}}_{t,t-1}$ for all t

Algorithm 2 Expectation-Maximisation method - Learning

Require: $\mathbf{z}, \epsilon$
initialize $\mathbf{A}, \mathbf{C}, \mathbf{\Gamma}_s, \mathbf{\Gamma}_o, \boldsymbol{\mu}_1^0, \boldsymbol{\sigma}_1^0$
while $\Delta L > \epsilon$ **do**
% E-Step
Run Kalman smoothing and update log-likelihood
initialize $\boldsymbol{\beta} \leftarrow 0, \boldsymbol{\delta} \leftarrow 0, \boldsymbol{\gamma} \leftarrow 0,$
for t = 1 to T **do**
Calculate intermediary variables(δ, β, γ)
end for
% M-Step
Calculate model parameters $(\mathbf{A}, \mathbf{C}, \boldsymbol{\Gamma_s}, \boldsymbol{\Gamma_o}, \boldsymbol{\mu}_1^0, \boldsymbol{\sigma}_1^0)$
Calculate change in log-likelihood (ΔL)
end while
return $\mathbf{A}, \mathbf{C}, \mathbf{\Gamma}_s, \mathbf{\Gamma}_o, \boldsymbol{\mu}_1^0, \boldsymbol{\sigma}_1^0$

To compensate for the delayed SM measurements, EM method is used to forecast the load during that time and interpolate the measurements to a common timestamp. The EM algorithm follows two steps: Expectation (E) step and Maximization (M) step. Starting from the initial parameter θ_0, the E and M steps are iteratively performed until the log likelihood ceases to increase.

$$E\ step: \mathbf{D}_{k+1} \leftarrow \arg\max_{\mathbf{D}} F(\mathbf{D}, \boldsymbol{\theta}_k) \tag{18}$$

$$M\ step: \boldsymbol{\theta}_{k+1} \leftarrow \arg\max_{\boldsymbol{\theta}} F(\mathbf{D}_{k+1}, \boldsymbol{\theta}) \tag{19}$$

where, $\mathbf{D}$ is the distribution of variables (mean and covariance of measurements), $\boldsymbol{\theta}$ denotes the parameters $(\mathbf{A}, \mathbf{C}, \boldsymbol{\Gamma_s}, \boldsymbol{\Gamma_o}, \boldsymbol{\mu}_0,$ and $\boldsymbol{\sigma}_0)$. The inference and system identification process is iteratively performed until the solution converges. Using the current model parameters and given observations, the states are estimated.

The flowchart shown in Fig. 1 explains the proposed method including the Projection Statistics based IR-WLS using hybrid measurements.

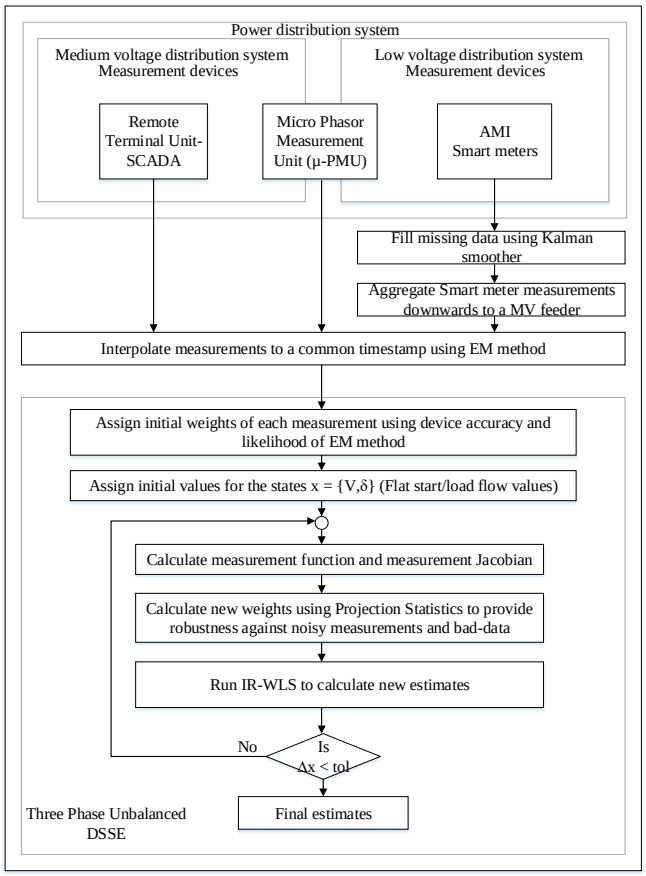


Fig. 1. FlowChart for Robust Unbalanced DSSE

V. Simulation Results and Discussion

Throughout this paper, the 2-norm error of DSSE is used to find the accuracy of the estimates of the three-phase unbalanced DSSE [22].

$$error = \sqrt{\sum_{i=1}^{N} |v_i^{est} - v_i^{act}|^2} \tag{20}$$

where N is the number of nodes (all single phase buses / feeders), v_i^{est} is the estimated voltage phasor in pu and v_i^{act} is the actual voltage phasor in pu. The simulations are performed in MATLAB environment using Intel Core i3 3.40 GHz CPU and 4GB RAM.

A. Effect of bad-data in WLS DSSE

The three-phase WLS DSSE, explained in Section II, is applied to IEEE 13, 37, and 123 test feeder systems [27]. The WLS-DSSE without any bad-data is run for over 1000 iterations considering random measurement noise. The time taken for the estimation process and the 2-norm error of the estimates are tabulated in Table I. From the results, it is observed that the estimate error is within limits for different factorisation methods.

TABLE I
PERFORMANCE OF WLS-DSSE WITHOUT ANY BAD-DATA AND CONSIDERING RANDOM MEASUREMENT NOISE

Factorisation methods	IEEE 13 bus system		IEEE 37 bus system		IEEE 123 bus system	
	Error	Time(sec)	Error	Time(sec)	Error	Time(sec)
Orthogonal WLS	0.00038	0.0065	0.0023	0.0498	0.0177	2.6125
Hybrid(QR) WLS	0.00038	0.0066	0.0023	0.0563	0.0178	2.6577
Peters and Wilkinson (LU) WLS	0.00038	0.0058	0.0023	0.0460	0.0177	2.5979

Though the WLS-DSSE works effectively for noiseless measurements, it can be seen that, even a small amount of noise or bad-data in measurements affects the accuracy of the estimates to a great extent. Since noisy measurements are more common in distribution systems, 10% random noise and atleast two outliers are added at a time to the measurements of random meters. The power flow measurements through short lines and the lower redundant measurements have high leverages. A good leverage measurement improves the accuracy of the estimates whereas the gross error in leverage measurements affects the accuracy of the estimates. An adversary tries to attack these high leverage measurements thereby creating maximum impact. The attack strategies provided in [4], [28] are used to simulate the bad-data on the measurements.

It is observed from Table-I and Table-II that the estimates are significantly affected by outliers and the time taken for DSSE also increases with bad-data, as the number of iterations required to converge increases. Therefore, statistical techniques are necessary to develop an unbalanced DSSE that is robust to noisy measurements and bad-data.

B. Robust IR-WLS and its response to bad-data

In this section, PS based IR-WLS, as explained in Section III-B, is used to detect and re-assign less weights to the bad measurements. The DSSE is run for more than 1000 iterations for each test system at different load profiles (as in Fig. 2 and Fig. 3). Random noise of 10% is added to the measurements and atleast two randomly selected measurement devices are attacked, using attack strategies provided in [13], [28]

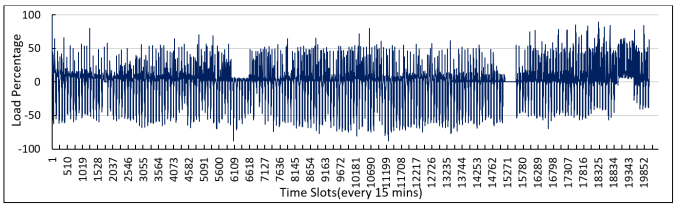


Fig. 2. Real power measurements from smart meter

The robustness of DSSE under bad-data can be observed using the accuracy of the estimates, as shown in Table-II. Though the time taken for IR-WLS is higher than that of WLS DSSE, it is faster when compared to other robust methods [11] [29]. Fig. 4 and Fig. 5 represent the DSSE accuracy in the presence of bad-data with and without IR-WLS.

Table-III depict the $\mathbf{PS}_i$, $\chi^2_{k,0.975}$ and the weight residual, for a particular set of measurements in IEEE 13 bus system. The measurement set contains good-leverage point on measurement index 3, bad-leverage point on measurement index 34, and outlier on measurement index 5. From the calculated weight residual, we can infer that that the PS based IR-WLS down-weights the bad-leverage points and outliers effectively.

This method involves computing the projection statistic of each point and compares with the statistical cutoff value to detect the presence of bad-data. It is observed that Projection Statistics based IR-WLS improves the accuracy of the estimates in the presence of bad-data. With LU factorization and without considering IR-WLS, the average estimation error

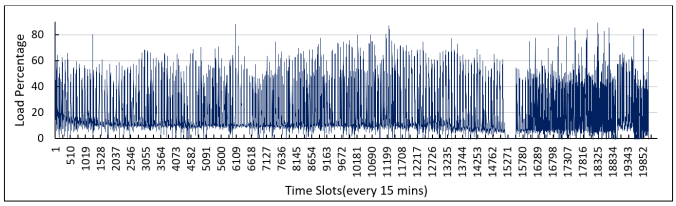


Fig. 3. Reactive power measurements from smart meter

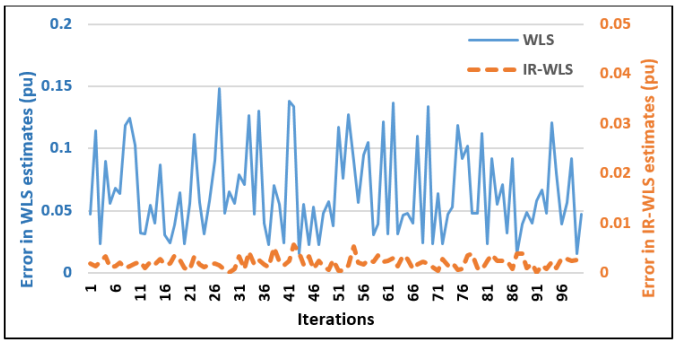


Fig. 4. WLS vs. IR-WLS in the presence of bad-data for IEEE 37 Node Feeder

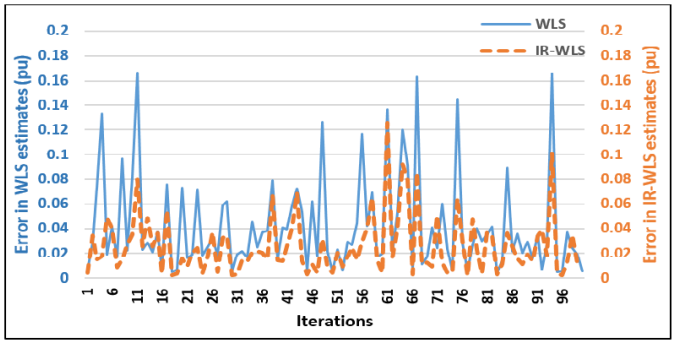


Fig. 5. WLS vs. IR-WLS in the presence of bad-data for IEEE 123 Node Feeder

TABLE II
PERFORMANCE OF THE PROPOSED DSSE

Test Case	IEEE 13 bus system		IEEE 37 bus system		IEEE 123 bus system	
	Error	Time(sec)	Error	Time(sec)	Error	Time(sec)
WLS-DSSE without bad-data	0.00038	0.0058	0.0023	0.0600	0.0177	2.5979
WLS-DSSE with bad-data	0.0956	0.0349	0.0863	0.5582	0.0389	4.6336
Proposed IR-WLS DSSE with bad-data	0.0031	0.0440	0.0041	0.9674	0.0217	38.2071

TABLE III
PROJECTION STATISTICS FOR THE BAD-DATA MEASUREMENTS

Measurement Index	$\chi^2_{k,0.975}$ value	$\mathbf{PS}_i$ value	Weight residual	Bad data
3	31.5264	80.0334	0.72	No
4	11.1433	8.8581	0.97	No
5	36.7807	14.4796	81.11	Yes
6	36.7807	11.1176	0.98	No
33	11.1433	10.2089	1.05	No
34	31.5264	93.2642	59.24	Yes

was 0.0389pu in IEEE 123 bus system whereas IR-WLS improved the DSSE accuracy and the error reduced to 0.0217 pu. Similarly, the 2-norm error of the estimates of the IEEE 13 bus system reduced from 0.0956 to 0.0031 and that of IEEE 37 bus system reduced from 0.0863 to 0.0041.

C. Hybrid Measurements and Delay Compensation

In this paper, multiple measurement devices such as SCADA and SMs are considered. The DSSE is run for MV distribution system considering SCADA at few MV nodes. For nodes that doesnot contain RTU/SCADA/μPMUs, aggregated SM measurements considering the LV system losses, as explained in Section IV-A, are used. Smart meters are considered to be available in all nodes of the LV distribution system and the aggregated value of the SMs considering LV system losses is calculated using power flow or LV DSSE. Alternatively, if the distribution transformer has SM installed, then the corresponding measurements can be considered. Use of SMs not only increases the accuracy of the estimates but also improves the observability of the MV DSSE.

Since the SM measurements arrive at different times, the measurements are less reliable due to varying loads. The missing SM measurements are filled using Kalman smoother. To compensate for the variable loads in distribution system, the proposed method also considers Expectation-Maximisation based interpolation for delay compensation of the SM measurements, along with adjusting the variance of measurement devices, as proposed in [22].

It is observed that for the real-time measurements from SMs, the EM method converges, and the log-likelihood and convergence plots are shown in Fig. 6. Matlab's financial toolbox app is used for the simulation studies.

The state estimator is run for the interpolated SM load measurements by EM method, and the output (estimates) of the state estimator are compared with linearized loads. Measurement variances are adjusted based on delay, as in [22]. This estimator also includes the robust PS based IR-WLS. The state estimator is run for load profiles of IIT Kanpur smart grid [30] over six months (shown in Fig. 2 and Fig. 3). The

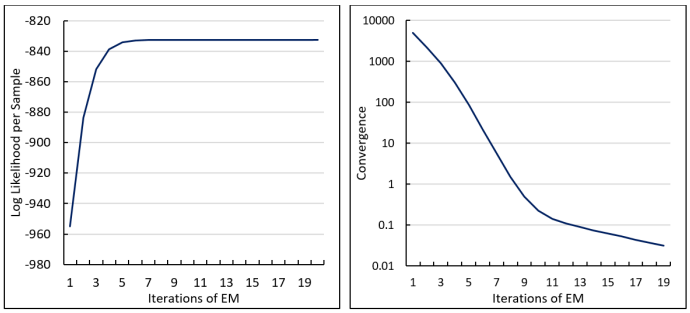


Fig. 6. Convergence and likelihood plots of EM method

algorithm considering hybrid measurements is run for more than 20000 iterations with different load profiles and random noise to test the accuracy of the proposed method. From these results, it is observed from Table IV, that the average estimates accuracy is better in EM-based measurement interpolation than the existing methods available in the literature [22]. The accuracy of the DSSE are verified in IEEE 13, 37, and 123 bus distribution systems considering measurement delay in real-time SM data.

TABLE IV
PERFORMANCE OF IR-WLS DSSE CONSIDERING BAD-DATA IN MEASUREMENTS AND RANDOM MEASUREMENT NOISE

Case	Average error in DSSE estimates		
	IEEE 13 bus	IEEE 37 bus	IEEE 123 bus
Ideal DSSE, Considering synchronised measurements	6.517 e-3	0.0823	0.0512
Traditional DSSE	12.230 e-3	0.2211	0.0658
DSSE in [22]	10.513 e-3	0.2115	0.0522
Proposed robust DSSE based on EM based load forecast	8.718 e-3	0.1563	0.0504
Percentage improvement of DSSE w.r.t. method in [22] and proposed DSSE (in brackets)	14.04 % (28.72 %)	4.34 % (29.31 %)	20.66 % (23.41 %)

VI. CONCLUSION

This paper addresses some of the key challenges in using SM measurements in DSSE. The proposed estimator is robust against bad-data, noisy measurements, and missing and delayed SM measurements. Since SM alone cannot make a distribution system observable, this paper considers hybrid

measurement sources and proposed a robust DSSE method. From the simulation results, it is observed that the projection statistics based IR-WLS is effective against multiple outliers and it requires little modification to the existing state estimators. The Kalman smoothing and EM-based interpolation provide better accuracy and robustness to the MV DSSE than the existing SM delay compensation methods available in the literature.

VII. Acknowledgment

The authors would like to thank the Ministry of Power, Govt. of India, and IIT Kanpur for providing financial support to carry out this research work with project no. MP&IIT/EE/2014297. This work is also partially funded by UK-India JUICE project with project numbers DST/EE/2017092 and EP/P003605/1.

References

[1] F. Ahmad, A. Rasool, E. Ozsoy, S. Rajasekar, A. Sabanovic, and M. Elitaş, "Distribution system state estimation-a step towards smart grid," *Renewable and Sustainable Energy Reviews*, 2017.

[2] A. Bose and K. A. Clements, "Real-time modeling of power networks," *Proceedings of the IEEE*, vol. 75, no. 12, pp. 1607–1622, Dec 1987.

[3] A. Garces, "A linear three-phase load flow for power distribution systems," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 827–828, 2016.

[4] A. Majumdar and B. C. Pal, "Bad data detection in the context of leverage point attacks in modern power networks," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2042–2054, May 2018.

[5] K. Li, "State estimation for power distribution system and measurement impacts," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 911–916, 1996.

[6] L. Holten, A. Gjelsvik, S. Aam, F. F. Wu, and W. H. E. Liu, "Comparison of different methods for state estimation," *IEEE Transactions on Power Systems*, vol. 3, no. 4, pp. 1798–1806, Nov 1988.

[7] R. Singh, B. Pal, and R. Jabr, "Choice of estimator for distribution system state estimation," *IET Generation, Transmission & Distribution*, vol. 3, no. 7, pp. 666–678, 2009.

[8] A. Monticelli, C. A. F. Murari, and F. F. Wu, "A hybrid state estimator: Solving normal equations by orthogonal transformations," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 12, pp. 3460–3468, Dec 1985.

[9] F. F. Wu, W. H. E. Liu, L. Holten, L. Gjelsvik, and S. Aam, "Observability analysis and bad data processing for state estimation using hachtel's augmented matrix method," *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 604–611, May 1988.

[10] L. Mili, M. Cheniae, N. Vichare, and P. J. Rousseeuw, "Robust state estimation based on projection statistics [of power systems]," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 1118–1127, 1996.

[11] J. Zhao, G. Zhang, Z. Y. Dong, and M. L. Scala, "Robust forecasting aided power system state estimation considering state correlations," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2658–2666, July 2018.

[12] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[13] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1636–1646, May 2018.

[14] A. Gomez-Exposito and A. Abur, *Power system state estimation: theory and implementation*. CRC press, 2004.

[15] C. A. Fantin, M. R. C. Castillo, B. E. B. de Carvalho, and J. B. A. London, "Using pseudo and virtual measurements in distribution system state estimation," in *2014 IEEE PES Transmission Distribution Conference and Exposition - Latin America (PES T D-LA)*, Sept 2014, pp. 1–6.

[16] E. Manitsas, R. Singh, B. Pal, and G. Strbac, "Modelling of pseudo-measurements for distribution system state estimation," in *CIRED Seminar 2008: SmartGrids for Distribution*, June 2008, pp. 1–4.

[17] F. Ni, P. H. Nguyen, J. F. G. Cobben, H. E. van den Brom, and D. Zhao, "Uncertainty analysis of aggregated smart meter data for state estimation," in *2016 IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, Sept 2016, pp. 1–6.

[18] T. Xygkis, G. Karlis, I. Siderakis, and G. Korres, "Use of near real-time and delayed smart meter data for distribution system load and state estimation," *IET Conference Proceedings*, January 2014.

[19] A. Von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2014, pp. 1–5.

[20] A. Al-Wakeel, J. Wu, and N. Jenkins, "State estimation of medium voltage distribution networks using smart meter measurements," *Applied energy*, vol. 184, pp. 207–218, 2016.

[21] A. Alimardani, S. Zadkhast, J. Jatskevich, and E. Vaahedi, "Using smart meters in state estimation of distribution networks," in *2014 IEEE PES General Meeting — Conference Exposition*, July 2014, pp. 1–5.

[22] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2919–2928, 2015.

[23] M. Wandera, B. Jonasson, J. Benoit *et al.*, "Cybersecurity considerations for electrical distribution systems," Accessed September 20, 2018. [Online]. Available: http://www.eaton.com/ecm/groups/public/@pub/@eaton/@corp/documents/content/pct_1603172.pdf

[24] T. Sukumara, S. Sudarsan, J. Starck, and T. R. Vittor, "Cyber security–security strategy for distribution management system and security architecture considerations," *CIRED-Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 2653–2656, 2017.

[25] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyberphysical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, Jan 2012.

[26] S. Roweis and Z. Ghahramani, "A unifying review of linear gaussian models," *Neural computation*, vol. 11, no. 2, pp. 305–345, 1999.

[27] [Online]. Available: http://sites.ieee.org/pes-testfeeders/resources/ (accessed September 20, 2018).

[28] S. Tan, W.-Z. Song, M. Stewart, and L. Long, "Lpattack: Leverage point attacks against state estimation in smart grid." in *Global Communications Conference (GLOBECOM)*, 2014 IEEE, pp. 643–648.

[29] J. Zhao, L. Mili, and R. C. Pires, "Statistical and numerical robust state estimator for heavily loaded power systems," *IEEE Transactions on Power Systems*, pp. 1–1, 2018.

[30] S. Sahoo, S. K. Singh, S. S. Inamdar, A. Sharma, I. Husain, A. Yadav, S. Chakrabarti, V. K. Tiwari, S. Srivastava, R. Singh, S. Anand, K. Rajawat, A. Roy, and P. Sensarma, "Development of an R&D platform for smart city projects in the Indian context," *CIGRE 2018 Session, Paris, France.*, Aug. 2018.