Robust Medium-Voltage Distribution System State Estimation using Multi-Source Data

Junbo Zhao, *Member, IEEE*, Can Huang, *Senior Member, IEEE*, Lamine Mili, *Life Fellow, IEEE*, Yingchen Zhang *Senior Member, IEEE*, Liang Min, *Senior Member, IEEE*

***Abstract*—Due to the lack of sufficient online measurements for distribution system observability, pseudo-measurements from short-term load or distributed renewable energy resources (DERs) forecasting are used. However, the accuracy of them is low and thus significantly limits the performance of distribution system state estimation (DSSE). In this paper, a robust DSSE that integrates multi-source measurement data is proposed. Specifically, the historical low-voltage (LV) side smart meters are used to forecast load and DERs injections via the support vector machine (SVM) with optimally tuned parameters. By contrast, the online smart meters at LV side are utilized to derive equivalent power injections at the MV/LV transformers, yielding more accurate pseudo-measurements compared to the forecasted injections. Furthermore, to deal with bad data caused by communication loss, instrumental errors and cyber attacks, robust DSSE that relies on generalized maximum-likelihood (GM)-estimation criterion is developed. The projection statistics are developed to adjust the weights of each measurement, leading to better balance between pseudo- and real-time measurements. Numerical results conducted on modified IEEE 33-bus system with DG integration demonstrate the effectiveness and robustness of the proposed method.**

***Index Terms*—Distribution system state estimation, machine learning, pseudo measurement, real-time measurement, smart meter.**

I. Introduction

WITH the increasing integration of renewable energy, load characteristics become highly stochastic, thus posing a great challenge for power distribution systems monitoring and control. As one of the fundamental functions in the distribution management system (DMS), distribution system state estimation (DSSE) is able to provide reliable and accurate data for DMS applications, such as outage management, loss reduction, demand response, etc [1]–[3]. However, low measurement redundancy of distribution network often prevents the operator from estimating all state variables, i.e., the voltage magnitude and angle at each bus [4]. To improve the system observability, pseudo-measurements derived from historical data or other indicators are usually advocated [3]. But those pseudo-measurements have lower accuracy than real-time measurements, which would degrade the DSSE performance. On the other hand, given the limited budget on deploying new sensors, pseudo-measurements have to be used.

To model pseudo-measurements, several methods have been proposed in the literature. In [5], the errors of pseudo-measurements were assumed to be much greater than those of real-time measurements. In [6], various distribution functions (Weibull, normal, Erlang, and beta) have been applied to fit the group domestic loads. As a result, they proposed the beta distribution function. However, the beta distribution cannot be used for the weighted least squares (WLS) based DSSE. A similar research effort was reported by Ghosh *et al*. [7] in DSSE problem, where they validated various models such as normal, log-normal, and beta distribution through chi-square goodness of fit test. In [8], two methods to model pseudo-measurements were proposed. One was based on analyzing the correlation between real-time measurements taken at substations and load pseudo-measurements, while in another approach the accuracy of the load pseudo-measurements was improved through Gaussian Mixture Models (GMM). Simulation results showed that both methods had similar performance. The GMM was further investigated in [9], [10] to produce pseudo-measurements for DSSE. In [11], an artificial neural network (ANN) algorithm was adopted to generate pseudo-measurements to compensate the missing real-time measurements during contingent events. However, a prior knowledge of the real-time measurements is required. Thanks to the wide-area installations of smart meters at the LV side of the distribution systems [12], sufficient historical load and DERs data are available for better pseudo-measurements modelings [13]. In [14], the smart meters were used together with the online measurements for MV DSSE. The impacts of smart meter data aggregation on the MV DSSE was investigated in [15]. However, the data quality issues associated with smart meters have not been addressed and there is still a grand challenge of how to combine smart meter and traditional measurements effectively for improving state estimation accuracy.

In this paper, a robust DSSE for the MV network is proposed that integrates multi-source measurement data from smart meters, supervisory control and data acquisition (SCADA) and pseudo-measurements. In particular, the historical aggregated smart meter data are utilized by the machine learning method, i.e., support vector machine (SVM) for short-term load and DERs injection forecasting. These forecasts are taken as pseudo-measurements. On the other hand, the online smart meters at LV side are utilized to derive equivalent power injections at the MV/LV transformers. Upon the availability of three types of data, the projection statistics are developed to detect data quality issues and then adjust the weights of each measurement. Finally, the generalized maximum-likelihood (GM)-estimator is presented to filter out measurement errors and suppress the influences of bad data and model uncertainties.

The remainder of this paper is organized as follows. Section II describes the problem formulation. In Section III, the proposed robust DSSE is shown in detail. Section IV shows and discusses the numerical results and finally Section V concludes the paper.

II. Problem Formulation

For an distribution system with n-buses, the relationship between the measurement vector $\boldsymbol{z} \in \mathbb{R}^m$ and the state vector $\boldsymbol{x} \in \mathbb{R}^n$ (i.e., voltage magnitudes and angles) can be described as follows [16]:

$$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{h}(\cdot) \in \mathbb{R}^{m\times n}$ is the vector of measurement functions; $\boldsymbol{e} \in \mathbb{R}^m$ is a random measurement error vector assumed to be distributed with zero mean and covariance matrix $\boldsymbol{R}$. To obtain

This work was supported by the U.S. Department of Energy through Advanced Grid Modeling program. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 with IM release number LLNL-CONF-775662.

Junbo is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762; Lamine is with the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Falls Church, VA 22043, USA. Can and Liang are with the Lawrence Livermore National Laboratory, Livermore, CA, USA; Yingchen is with National Renewable Energy Laboratory, Golden, CO 80401 USA (e-mail: junbo@msstate.ece.edu, huang38@llnl.gov).

978-1-7281-3103-0/20/$31.00 ©2020 IEEE

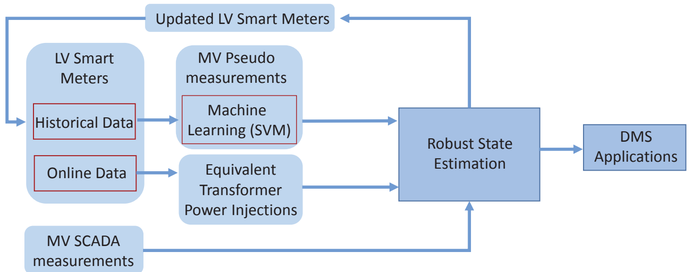

Fig. 1. Framework of the proposed robust DSSE.

state estimation, iterative WLS criterion [16] is adopted to solve the following objective function

$$\widehat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}} \left[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})\right]^T \boldsymbol{R}^{-1} \left[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})\right], \tag{2}$$

with the most basic form of iteration updates

$$\widehat{\boldsymbol{x}}_{k+1} = \widehat{\boldsymbol{x}}_k + \boldsymbol{\Delta}_k, k = 1, 2, ..., \tag{3}$$

$$\boldsymbol{\Delta}_k = (\boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{R}^{-1} (\boldsymbol{z} - \boldsymbol{h}(\widehat{\boldsymbol{x}}^k)), \tag{4}$$

where $\boldsymbol{H} = \partial \boldsymbol{h} / \partial \boldsymbol{x}|_{\boldsymbol{x}=\widehat{\boldsymbol{x}}^k}$ is the Jacobin matrix. The algorithm converges once $\boldsymbol{\Delta}_k$ becomes smaller than a pre-specified tolerance threshold.

Due to the limited number of real-time measurements in the distribution system, pseudo-measurements from short-term load forecasting are typically added to $\boldsymbol{z}$ for measurement redundancy improvement. However, their accuracies are much lower than the online SCADA measurements, leading to degraded DSSE performance. With the wide-area deployment of smart meters, its usage with SCADA measurements provides an alternative way of enhancing DSSE. It should be noted that smart meters are usually used for short-term load forecasting [13] to obtain pseudo-measurements. As the smart meters have better accuracy and spatial resolutions than the historical load data from load survey, modeling, etc., it indeed improves the quality of pseudo-measurements. However, the benefits of online smart meters are not fully explored. In addition, the smart meter readings may be subject to cyber attacks, electricity theft and communication issues, yielding incorrect values. To address these issues, a robust DSSE that integrates both historical and online smart meters with SCADA measurements is proposed.

III. Proposed Robust DSSE

The framework of the proposed robust DSSE is displayed in Fig. 1. It consists of four major blocks, namely the use of aggregated historical smart meters for generating MV pseudo-measurements via machine learning; the derivation of measurements from online smart meters for the MV/LV transformers; the robust state estimation using pseudo-measurements, derived online measurements from smart meters and real-time SCADA measurements, and the closed feedback loop for updating historical LV smart meters.

A. Pseudo-Measurements Modeling

Although a lot of smart meters have been installed, they are typically deployed at the LV network. The smart meters are able to provide power consumption and voltage magnitude reading every 15min. Since the behaviors of each customers are full of randomness, it is challenging to predict load profiles at the LV network. However, if all customers along the downside of MV/LV transformers are aggregated together, the load profiles would be much smoother as seen at the MV level. Motivated by this, we propose to aggregate customers' power measurements so as to obtain the loading information at the MV/LV transformers. These historical data are recorded and will be used as the inputs for training the forecasting model via the SVM. Note that SVM is a widely used machine learning tool for data classification and regression [17]. In this paper, the support vector regression (SVR) is adopted for the load time series prediction. Specifically, given training data $(x_1, y_1), ..., (x_l, y_l)$, where x_i are inputs and y_i are their associated output values, SVR solves the following optimization problem:

$$\begin{aligned} f = & \min_{w,b,\xi,\xi^*} \tfrac{1}{2} w^T w + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \\ s.t. \quad & y_i - \left(w^T \phi(x_i) + b\right) \le \varepsilon + \xi_i \\ & \left(w^T \phi(x_i) + b\right) - y_i \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0, i = 1, ..., l \end{aligned} \tag{5}$$

where x_i is mapped to a higher dimensional space by the kernel function ϕ and will be set as Gaussian radial basis function, i.e., $\phi(x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)$, where σ is a tuning parameter; ξ_i is the upper training error (ξ_i^* is the lower) subject to the ε-insensitive tube $\left|y - \left(w^T \phi(x) + b\right)\right| \le \varepsilon$. C and ε are parameters that control the regression quality and the width of the tube. The constraints in (5) indicate that we should put most data (x_i, y_i) in the tube $\left|y - \left(w^T \phi(x) + b\right)\right| \le \varepsilon$. If x_i does not fall into the tube, there is an error ξ_i or ξ_i^* that should be minimized in the objective function.

The appealing feature of SVR is that it avoids under-fitting and over-fitting the training data by minimizing the training error $C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$ as well as the regularization term $\frac{1}{2} w^T w$. For traditional least square regression, ε is always set to 0 and the data is not mapped into higher dimensional spaces. Thus, SVR is a more general and flexible treatment on regression problems. As is well-known that the choices of the kernel parameter σ and the control parameter C would affect the prediction results, the intelligent stochastic searching method, i.e., particle swarm optimization is used to tune them.

Once the regression model is learnt from the historical data, the forecasted MV load/DERs real and reactive power injections for the ith bus can be obtained, i.e., $\tilde{P}_i^{MV}$ and $\tilde{Q}_i^{MV}$. The forecasting error covariance matrices are $\tilde{\boldsymbol{R}}_p$ and $\tilde{\boldsymbol{R}}_q$, respectively.

B. MV/LV Transformer Measurement Derivations

It is worth pointing out that the loads in similar geographical areas usually exhibit temporal and spatial correlations. These are captured by the SVR-based short-term load forecasting. However, besides the statistical information in the historical smart meters, there are system state information contained in the real-time smart meters that have not been used yet. By using the smart meter data aggregation strategy, the online real and reactive power consumptions at the LV side of the ith MV/LV transformers can be estimated, yielding $\widehat{P}_i^{LV}$ and $\widehat{Q}_i^{LV}$, respectively. Then, the estimated equivalent MV/LV transformer power injections $\widehat{P}_i^{MV}$ and $\widehat{Q}_i^{MV}$ can be derived via the following equations:

$$\widehat{P}_i^{MV} = \widehat{P}_i^{LV} + P_i^0 + P_i^s (\left|\widehat{S}_i^{LV}\right| / S_i^0)^2, \tag{6}$$

$$\widehat{Q}_i^{MV} = \widehat{Q}_i^{LV} + Q_i^0 + Q_i^s (\left|\widehat{S}_i^{LV}\right| / S_i^0)^2, \tag{7}$$

where $\widehat{S}_i^{LV}$ is the aggregated complex load power of LV customers measured by smart meters; P_i^0 and Q_i^0 are the active and reactive power of the transformer when there is no load connected, respectively; P_i^s and Q_i^s are the short-circuit active and reactive power of the transformers, respectively; S_i^0 is the rated apparent power of the transformer. As $\widehat{P}_i^{LV}$, $\widehat{Q}_i^{LV}$ and $\widehat{S}_i^{LV}$ contain errors, they will be propagated to the derived $\widehat{P}_i^{MV}$ and $\widehat{Q}_i^{MV}$. Furthermore, it is found from (6)-(7) that the errors are in nonlinear relationship. To this end, the unscented transformation-based method proposed in our previous work [18] is advocated to calculate the covariance matrices of $\widehat{P}_i^{MV}$ and $\widehat{Q}_i^{MV}$, yielding $\widehat{\boldsymbol{R}}_p$ and $\widehat{\boldsymbol{R}}_q$, respectively.

By resorting to equations (6)-(7), many derived power measurements from online smart meters can be obtained. As the real-time smart meters better reflect the system operating conditions, the derived power measurements will yield better accuracy than the pseudo-measurements. Furthermore, thanks to the widespread installation of smart meters, a large number of such derived equivalent power measurements can be obtained. Thus, a system with higher measurement redundancy is expected allowing us to improve estimation accuracy as well as design robust estimators against bad data.

C. Robust DSSE and Historical LV Data Updating

In practice, the real-time measurements are always subject to various types of errors caused by many different reasons, such as communication issues, cyber attacks, instrumental errors, etc. If these errors are large, bad data occur. On the other hand, if there are changes on system operation conditions due to topology errors, sudden load/DERs injections, the pseudo-measurements from short-term forecasting that are close to the event locations are no longer reliable. Therefore, robust DSSE is required. In this paper, we develop the robust DSSE based on the generalized maximum-likelihood (GM)-estimation criterion. Its objective function is:

$$J(\boldsymbol{x}) = \sum_{i=1}^{m+2p+2q} \omega_i^2 \rho(r_{S_i}), \tag{8}$$

where p and q are the number of load/DERs buses; ω_i is the weight aimed at bounding the influence of bad data, including vertical outliers and bad leverage points; $\rho(\cdot)$ denotes the Huber cost function [19]; $r_{S_i} = r_i/\sigma_i\omega_i$ is the standardized residual; $r_i = z_i - h_i(\widehat{\boldsymbol{x}})$; σ_i is the standard deviation of ith measurement, which is the square-root of the diagonal element of $\overline{\boldsymbol{R}} = block[\boldsymbol{R}\ \tilde{\boldsymbol{R}}_p\ \tilde{\boldsymbol{R}}_q\ \widehat{\boldsymbol{R}}_p\ \widehat{\boldsymbol{R}}_q]$. $block$ means augmenting the matrices in a diagonal manner.

The calculation of ω_i is as follows: let the pseudo-measurement vector $\tilde{\boldsymbol{z}} = [(\tilde{\boldsymbol{P}}^{MV})^T\ (\tilde{\boldsymbol{Q}}^{MV})^T]^T$, the derived measurement vector $\widehat{\boldsymbol{z}} = [(\widehat{\boldsymbol{P}}^{MV})^T\ (\widehat{\boldsymbol{Q}}^{MV})^T]^T$ and the online SCADA measurement vector $\boldsymbol{z}_s$, yielding $\boldsymbol{z} = [\tilde{\boldsymbol{z}}^T\ \ \widehat{\boldsymbol{z}}^T\ \ \boldsymbol{z}_s^T]^T$. There are two bad data scenarios we need to deal with, namely the incorrect pseudo-measurements due to the change of system operating conditions and the incorrect online smart meters and SCADA measurements. If it is the former case, some of the forecasted measurements will not be consistent with the online measurements. To check that, the projection statistics (PSs) [19] are applied to the following innovation matrix

$$\boldsymbol{N} = [\tilde{\boldsymbol{z}} - \widehat{\boldsymbol{z}}\ \tilde{\boldsymbol{z}}' - \widehat{\boldsymbol{z}}'], \tag{9}$$

where $\tilde{\boldsymbol{z}}'$ and $\widehat{\boldsymbol{z}}'$ represent the forecasted measurements and online measurements at the previous time instants. Then, if the PS value of the ith element satisfies $PS_i > \chi^2_{2,0.975}$, it is declared as a bad data. The choice of threshold $\chi^2_{2,0.975}$ is because the PS values follow a chi-squares distribution [19]. After that, the weight of the ith pseudo-measurement is modified as follows:

$$\varpi_i = \min\left(1, \chi^2_{2,0.975}/PS_i^2\right), i = 1, ..., p+q \tag{10}$$

It should be noted that the incorrect online smart meters and SCADA measurements can happen as well. To deal with that, PSs are applied to the Jacobian matrix $[\boldsymbol{H}_{\widehat{\boldsymbol{z}}}\ \boldsymbol{H}_{\boldsymbol{z}_s}]$ associated to $\widehat{\boldsymbol{z}}$ and $\boldsymbol{z}_s$. If the PS value of the jth element satisfies $PS_j > \chi^2_{k,0.975}$, it is confirmed to be a bad data, where k is the number of non-zero entry for the jth row of $[\boldsymbol{H}_{\widehat{\boldsymbol{z}}}\ \boldsymbol{H}_{\boldsymbol{z}_s}]$. Note that the sparse version of PSs should be used here as most elements of the Jacobian matrix are zeros. After that, the weight associated with the jth measurement will be changed as follows:

$$\varpi_j = \min\left(1, \chi^2_{k,0.975}/PS_j^2\right), j = p+q+1, ..., m+2p+2q, \tag{11}$$

Once the weights of all measurements are determined, the next step is to solve (8), whose minimum should satisfy the following equation:

$$\frac{\partial J(\boldsymbol{x})}{\partial \boldsymbol{x}} = \sum_{i=1}^{m+2p+2q} -\frac{\boldsymbol{c}_i \omega_i}{\sigma_i} \psi(r_{S_i}) = \boldsymbol{0}, \tag{12}$$

where $\boldsymbol{c}_i^T$ is the ith column vector of the Jacobian matrix $\boldsymbol{H}$; $\psi(r_{S_i}) = \partial\rho(r_{S_i})/\partial r_{S_i}$. We multiply and divide both sides of (12) by r_{S_i}, yielding

$$\boldsymbol{H}^T \overline{\boldsymbol{R}}^{-1} \boldsymbol{Q} (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})) = \boldsymbol{0}, \tag{13}$$

where $q(r_{S_i}) = \psi(r_{S_i})/r_{S_i}$ and $\boldsymbol{Q} = \mathrm{diag}(q(r_{S_i}))$. By taking the first-order Taylor series expansion of $\boldsymbol{h}(\boldsymbol{x})$ about $\widehat{\boldsymbol{x}}^\ell$ and using the iteratively reweighted least squares (IRLS) algorithm, we get the following iterative form:

$$\Delta\widehat{\boldsymbol{x}}^{(\ell)} = \left(\boldsymbol{H}^T \overline{\boldsymbol{R}}^{-1} \boldsymbol{Q}^{(\ell)} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^T \overline{\boldsymbol{R}}^{-1} \boldsymbol{Q}^{(\ell)} (\boldsymbol{z} - \boldsymbol{h}(\widehat{\boldsymbol{x}}^\ell)), \tag{14}$$

where ℓ is the iteration counter. The algorithm converges if $\|\Delta\widehat{\boldsymbol{x}}^{(\ell)}\|_\infty \leq 10^{-3}$.

Once the state estimates are obtained, they can be substituted into the measurement function to estimate the derived measurements at the MV/LV transformers. Based on the relationship between the derived measurements and online smart meters, more accurate smart meters can be estimated. They are stored and will be used for regression model training by SVM in the next time. Therefore, better pseudo-measurements can be expected.

IV. Numerical Results

To validate the performance of the proposed robust DSSE, extensive simulations are carried out on the IEEE 33-bus system, whose distribution line data and loads can be found in [20]. 8 distributed generations (DGs) are integrated into this system and they are modeled as power injections. Their locations are at buses 5, 9, 12, 15, 18, 22, 29 and 33, respectively. Each DG produces 80% real and reactive power of the load it is connected to. The DSSE is updated every 15 min, which is consistent with the smart meter scan rate. The SCADA measurements are updated every 5 min and only those SCADA measurements that have same time stamps as the smart meters are used. The online SCADA measurements include the voltage magnitude at bus 1, the real and reactive power injections at buses 4, 5, 8, 11, 14, 17, 20, 23, 26, 29, and 32, and the real and reactive power flows on lines 2-3, 2-19, 6-7, 6-26, 10-11, 14-15, 16-17, 17-18, 24-25, 27-28, 30-31 and 32-33. The true operating conditions of the distribution system, i.e., voltage magnitudes and angles are obtained from the load flow calculations based on the backward/forward sweep method. All buses except for the reference bus 1, the buses 8, 9, 12, 15, 21, 22, 25, 29 and 33 have MV/LV transformers. The smart meters at the LV side are aggregated for short-term load forecasting and the derivation of equivalent transformer injections. For the SCADA measurements, the added noise of voltage magnitude measurements is assumed to be Gaussian white noise with a 0.5% standard deviation, while those of the power injection and power flow measurements are assumed to be Gaussian white noise with a 1% standard deviation. For the LV smart meters, the added noise is assumed to be Gaussian white noise with a 3% standard deviation. All the tests are performed on a PC with Intel Core i5, 2.50 GHz, 8GB of RAM.

A. Case 1: Steady-State Operation Condition

In this scenario, the load and DG profiles with low degrees of stochasticity are used. Therefore, the system is operating under steady-state conditions. The historical smart meter data that contain

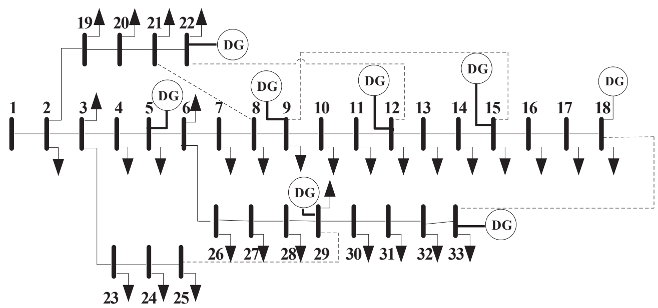

Fig. 2. Diagram of the IEEE 33-bus system with DERs.

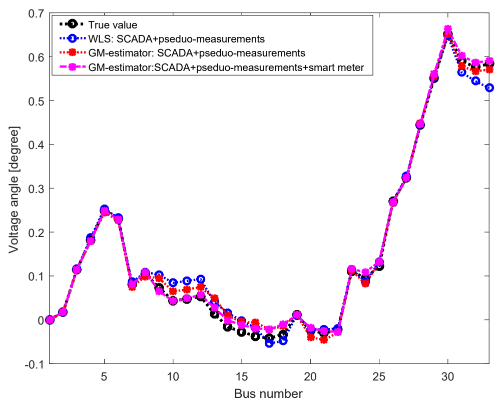

Fig. 3. Case 1: estimated bus voltage angles under normal operation condition.

100 points for each MV/LV transformers are utilized for short-term load forecasting while the 101th measurements are used for testing. With the SVR method tuned by the PSO, the root mean square error of load forecasting is about 5.2%. Therefore, the standard deviation of the added noise for pseudo-measurement is set as 5% with some uncertainties as we could not get the true error. Only the estimated bus voltage angles are shown in Fig. 3 due to the space limitation. It can be found from the figure that as the pseudo-measurements are of reasonably good accuracy, all methods are able to achieve high statistical efficiency of estimating the voltage angles. The buses with relatively large errors are those that are close to DGs. This is expected as the randomness and uncertainties of DGs cause difficulties in getting accurate short-term forecasts. On the other hand, since the online smart meters have better accuracy than the pseudo-measurements, the derived equivalent power injections at the MV/LV transformers are also better. As a result, the proposed robust DSSE achieves better results than the other two methods.

B. Case 2: High Degree of Stochasticity of DGs

The test conditions are similar to Case 1 except for that the DGs are of high degree of stochasticity. In particular, the uncertainties of the DGs are assumed to be 20% simulated by Gaussian distribution. By using the PSO-tuned SVR method, the root mean square error of load forecasting is about 22.5%. To this end, the standard deviation of the added noise for pseudo-measurement is set as 20%. Figs. 4-5 display the estimated voltage magnitudes and angles of all buses, respectively. As expected, due to the high degree of stochasticity of DGs, those load forecasts that are close to DG buses have high uncertainties, yielding poor pseudo-measurements. These uncertainties are further propagated to the state estimates, causing large biases. Due to the lack of robustness, the traditional WLS method obtains the worst state estimates. By contrast, with the projection statistics as well as the GM-estimation criterion, the uncertainties of pseudo-

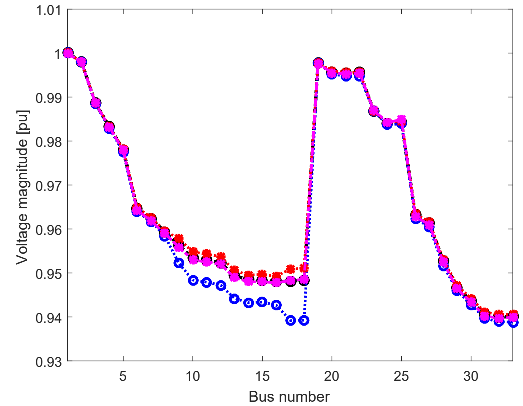

Fig. 4. Case 2: estimated bus voltage magnitudes with high degree of stochasticity of DGs.

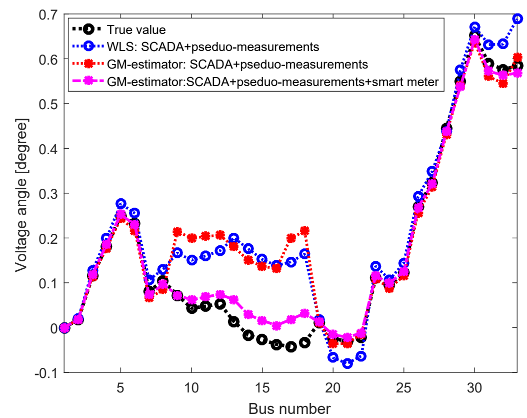

Fig. 5. Case 2: estimated bus voltage angles with high degree of stochasticity of DGs.

measurements can be mitigated slightly, yielding better estimates than the WLS. Note that the distribution system does not have good level of measurement redundancy and therefore the robustness of the estimator is limited. This can be validated by comparing the results with our proposed robust DSSE. The smart meters are used to derive equivalent power injections at the MV/LV transformers. These measurements are of less than 5% uncertainties, which are much more accurate than the pseudo-measurements. By combining them with the real-time SCADA measurements, we can find that the influences of the DG uncertainties have been significantly mitigated. The errors of the estimates are acceptable. Note that the accuracy of the robust DSSE depends heavily on the quality of smart meters and SCADA measurements. In fact, the low quality pseudo-measurements have cause negative impacts to the results. However, we still have to use them for helping us detect and identify bad data. There is a trade-off between robustness and measurement redundancy improvement by the pseudo-measurements.

C. Case 3: Occurrence of Bad Data

Due to communication issues and sudden DG injection changes, the power injection measurements P_9, P_{11} and P_{16} are corrupted with gross errors. In particular, the received measurements are 4 times of their original values. Note that P_9, P_{11} and P_{16} represent the forecasted value, the online SCADA measurement and the derived measurement from smart meters, respectively. The purpose is to cover the scenario that all three types of measurements can have bad data. Figs. 6-7 show the estimated voltage magnitudes and angles of all

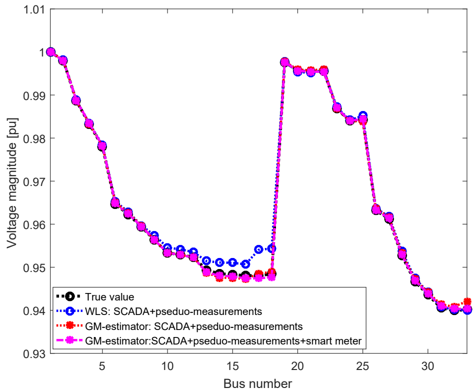

Fig. 6. Case 3: estimated bus voltage magnitudes when measurements P_9, P_{11} and P_{16} are corrupted with gross errors.

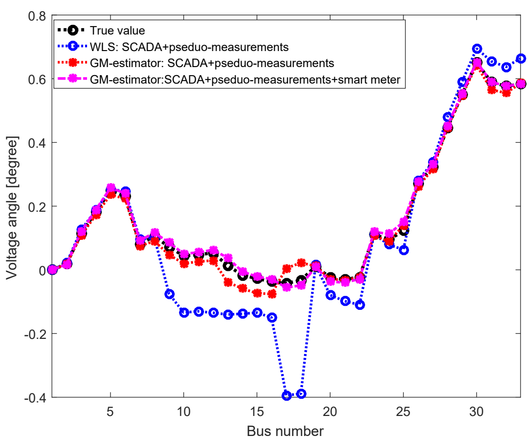

Fig. 7. Case 3: estimated bus voltage angles when measurements P_9, P_{11} and P_{16} are corrupted with gross errors.

buses in the presence of bad data, respectively. It can be observed from the results that as WLS lacks of robustness to bad data, its estimation results are significantly biased. It is interesting to note that unlike the transmission system, where the voltage angles have much stronger correlation with power injections than that of the voltage magnitudes, both the voltage magnitudes and angles have high correlations with the power injections in the distribution system. This is due to the high R/X ratio. As a result, we can find from the figures that because of incorrect power injections, both the estimated voltage magnitudes and angles have large errors. By contrast, thanks to the robustness, the GM-estimator that uses pseudo-measurements is able to suppress the influence of bad data. The results are still biased as the pseudo-measurements are not of high quality, which restrict the capability of the GM-estimator. On the other hand, the use of derived measurements from smart meters at the MV/LV transformers allows the GM-estimator greatly bound the negative effects of bad data, yielding slightly biased state estimates.

Finally, for all three cases, the computing times of the proposed robust DSSE are less than 0.4s while that of the WLS is 0.1s. It is expected that WLS is faster than the robust DSSE but it lacks of robustness to bad data and yields lower accuracy. It should be noted that the computing time of each method is affected by the size of the distribution system. Thus, for very large-scale distribution system, a distributed implementation or parallel computing techniques are useful. We will investigate that in our future work.

V. Conclusion

This paper presents a robust DSSE for the MV network with the consideration of multiple data sources, including SCADA measurements, historical and online smart meters. Specifically, the historical aggregated smart meters are used by the SVR for short-term load forecasting to generate pseudo-measurements. The feedback loop updating of smart meters from the robust estimation is also presented to enhance the quality of the historical data. To further improve the system measurement redundancy with online measurements, we utilize the real-time smart meters at LV side to derive equivalent power injections at the MV/LV transformers. The projection statistics are extended and used together with the GM-estimator to detect and suppress the influence of bad data and model uncertainties. Numerical results conducted on modified IEEE 33-bus system demonstrate the effectiveness and robustness of the proposed method.

References

[1] Y. Xu, C. Huang, X. Chen, L. Mili, C. H. Tong, M. Korkali, and L. Min, "Response-surface-based Bayesian inference for power system dynamic parameter estimation," *IEEE Trans. Smart Grid*, in Press.

[2] B. P. Hayes, J. K. Gruber, and M. Prodanovic, "A closed-loop state estimation tool for MV network monitoring and operation," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2116-2125, 2015.

[3] A. Primadianto, C. Lu, "A review on distribution system state estimation," *IEEE Trans. Power Systs.*, vol. 32, no. 5, pp. 3875-3883, 2017.

[4] M. E. Baran, "Challenges in state estimation on distribution system," *IEEE Power Engineering Society Summer Meeting*, pp. 1-5, 2001.

[5] C. N. Lu, J. H. Teng, W.-H. E. Liu, "Distribution state estimation," *IEEE Trans. Power Systs.*, vol. 10, no. 1, pp 229-240, 1995.

[6] R. Herman, J. J. Kritzinger, "The statistical description of grouped domestic electrical load currents," *Elect. Power Syst. Res.*, vol. 27, pp. 43-48, 1993.

[7] A. K. Ghosh, D. L. Lubkeman, M. J. Downey, R. H. Jones, "Distribution circuit state estimation using a probabilistic approach," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 45-51, Feb. 1997.

[8] E. Manitsas, R. Singh, B. Pal, G. Strbac, "Modelling of pseudo-measurements for distribution system state estimation," In *IET-CIRED SmartGrids for Distribution*, pp. 1-4, 2008.

[9] R. Singh, *et. al*, "Distribution system state estimation through Gaussian mixture model of the load as pseudo-measurement," *IET Generation, Transmission and Distribution*, vol. 4, no. 1, pp. 50-59, 2010.

[10] R. Singh, B. C. Pal, R. A. Jabr, "Statistical representation of distribution system loads using Gaussian mixture model," *IEEE Trans. Power Systs.*, vol. 25, no. 1, pp. 29-37, 2010.

[11] M. B. Do Coutto Filho, J. C. S. de Souza, M. T. Schilling, "Generating high quality pseudo-measurements to keep state estimation capability," *IEEE Power Tech*, 2007.

[12] K. Dehghanpour, Z. Wang, J. Wang, et al, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2312-2322, 2019.

[13] I. Dzafic, R. A. Jabr, I. Huseinagi, "Multi-phase state esti-mation featuring industrial-grade distribution network models," *IEEE Trans. Smart Grid*, vol. 8, no.2, pp. 609-618, 2017.

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

[15] Q. Chen, D. Kaleshi, Z. Fan, et al, "Impact of smart metering data aggregation on distribution system state estimation," *IEEE Trans. Ind. Informat.*, vol. 12, no. 4, pp. 1426-1437, 2016.

[16] A. Abur, A. Gomez Exposito, *Power System State Estimation: Theory and Implementation*. Marcel Dekker, Inc.: New York, 2004.

[17] B. Scholkopf, A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[18] J. B. Zhao, *et. al*, "A robust state estimation framework considering measurement correlations and imperfect synchronization," *IEEE Trans. Power Systs.*, vol. 33, no. 4, pp. 4604-4613, 2018.

[19] J. B. Zhao, L. Mili, R. C. Pires, "Statistical and numerical robust state estimator for heavily loaded power systems," *IEEE Trans. Power Systs.*, vol. 33, no. 6, 6904-6914, 2018.

[20] M. E. Baran, F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balacing," *IEEE Trans. Power Del.*, vol. 4, no. 2, 1401-1407, 1989.