A Comparison of MV Distribution System State Estimation Methods using Field Data

Barry Hayes and Milan Prodanovic
IMDEA Energy Institute
Madrid, Spain
{barry.hayes, milan.prodanovic}@imdea.org

***Abstract*—This paper compares the performance of five different Distribution System State Estimation (DSSE) methods, using field data taken from a European MV distribution network. The performance of each method is assessed in terms of its solution accuracy, robustness to noise and input measurement uncertainty, and ability to identify bad data and network topology errors. The advantages and disadvantages of each approach are discussed with regard to their application to static state estimation in MV distribution systems, where the quantity and quality of available network measurements is typically low. The Weighted Least Squares (WLS) approach is by far the most widely-used method in this context. However, the results from this study demonstrate that Extended Kalman Filter (EKF) techniques have significant advantages, particularly in terms of their ability to handle various types of input data errors. The performance of robust solution methods for distribution system state estimation is also compared.**

***Index Terms*—State estimation, power distribution, smart grids, power system analysis computing, distributed energy management systems.**

I. Introduction

Distribution systems were traditionally designed as unidirectional links between transmission network bulk supply points and end-users, and were operated passively, with limited monitoring and control capabilities. The recent drive towards actively-managed distribution systems creates a requirement for a much improved level of visibility throughout the distribution network [1], [2]. This has led to increased interest in the development of Distribution System State Estimation (DSSE) techniques [3]-[10].

However, the main practical issue for implementing SE in distribution systems is the lack of available on-line network measurements. While monitoring and communications infrastructure in distribution networks is gradually improving as systems are modernised, the quantity and quality of available measurements is usually still very limited, especially when compared to transmission systems. The introduction of smart metering infrastructure provides detailed information on the energy consumption at each network node. However, using smart meter measurements as a direct input to DSSE is generally not possible due to low data rates, low data reliability, and a lack of time synchronisation. Despite these limitations, smart metering can benefit DSSE by providing much better load pseudo-measurements, and can also help to determine the network post-fault topology and state [5], [6]. Other advanced metering infrastructure, such as PMUs, can greatly improve observability, but these have not been installed in distribution systems in significant numbers to date. Hence, a number of authors have developed DSSEs which rely heavily on pseudo-measurements of power injections at each network node, along with a very limited number of real, on-line measurements at certain network locations (e.g. from SCADA or PMU units installed at primary substations or main feeder heads) [8]-[10].

Given this context, the presented paper applies five different SE methods to a typical European MV distribution network, with the aim of simulating and assessing the SE performance in a distribution network scenario where the number of real measurements is low, and the DSSE inputs are comprised mainly of high-variance load pseudo-measurements. The SEs are tested to compare their performance in terms of SE accuracy, noise performance, robustness, handling of bad input data and topological errors, and computation time. The paper is structured as follows: Section II briefly outlines the different SE solvers used; Section III describes the MV network case study and discusses the performance of each SE; and Section IV gives the discussion and conclusions.

II. Methodology

The most widely-used approach for DSSE in the literature is "static" SE, which applies a single scan of measurement input data to estimate the system state at the current point in time. These methods typically use Weighted Least Squares (WLS) or Weighted Least Average (WLAV) estimation, or a combination of these approaches [9], [11]. Several authors have applied "dynamic", or tracking SEs, based on Kalman filtering approaches at the distribution level, e.g. [6]. In addition, DSSE solvers employing robust statics methods have been used in [3], [9]. In this paper, five different SE solvers are applied to the MV distribution network case study: 1) Weighted Least Average (WLAV); 2) Weighted Least Squares (WLS); 3) Weighted Least Squares-Robust (WLS-R);

The authors kindly acknowledge the support of the European Commission projects on Marie Sklodowska-Curie researcher mobility action (FP7-PEOPLE-2013-COFUND), the SmartHG research project (FP7-ICT-2011-8, ICT-2011.6.1), and the Spanish Ministry of Economy and Competitiveness project RESmart (ENE2013-48690-C2-2-R).

(5) Extended Kalman Filter (EKF); 5) Extended Kalman Filter-Robust (EKF-R). The performance of each SE is assessed based on estimation accuracy, robustness to noise and gross input data errors, ability to detect network topology errors, and computation time. SEs 1-5 are outlined very briefly below. For a more detailed description of each technique, the reader is referred to [1], [3], [7]-[9].

1) Weighted Least Average (WLAV) Estimator: The distribution network state is expressed as the vector **x** containing the voltage magnitudes and voltage angles at each node in the system. The input measurement vector **z** comprises of measurements of power/current injections at system buses, measurements of active/reactive power flows in system branches, pseudo-measurements of network quantities, or any combination of the above. This forms a set of over-determined, non-linear equations, where **h(x)** are the power flow functions corresponding to each measurement in **z**, and **e** is the vector of measurement errors. The Jacobian matrix $\mathbf{H} = \delta \mathbf{h}(\mathbf{x}) / \delta \mathbf{x}$ relates **z** to **x** and is composed of the partial derivatives of the measurements as a function of the system state. The SE objective function is given by:

$$min_x \qquad \mathbf{J}_{\mathbf{WLAV}}(\mathbf{x}) = \bar{\mathbf{W}}(abs[\mathbf{z} - \mathbf{h}(\mathbf{x})]) \qquad (1)$$

where $\bar{\mathbf{W}}$ is the input measurement weight matrix with each weight set according to the inverse of the variance of its corresponding measurement. The minimisation in (1) is solved using the Newton-Raphson method. The presence of bad data in the system measurement data set can be detected by applying statistical tests to the objective function $\mathbf{J}(\hat{\mathbf{x}})$, and to the normalised residual vector given by:

$$\mathbf{r} = \mathbf{z} - \mathbf{h}(\hat{\mathbf{x}}) \qquad (2)$$

The residual error vector is normalised by $\mathbf{r_n} = \rho_{jj}^{-1}\mathbf{r}$, where ρ_{jj} is the diagonal of the covariance matrix:

$$\mathbf{C_r} = \bar{\mathbf{W}}^{-1} - \mathbf{H}(\hat{\mathbf{x}})\, \mathbf{G}^{-1}\, \mathbf{H^T}(\hat{\mathbf{x}}) \qquad (3)$$

2) Weighted Least Squares (WLS) Estimator: The WLS estimator is the most widely-approach for static SE. The method is the same as the WLAV approach (1)-(3), but with the quadratic (least-squares) minimisation function:

$$min_x \qquad \mathbf{J}_{\mathbf{WLS}}(\mathbf{x}) = \bar{\mathbf{W}}[\mathbf{z} - \mathbf{h}(\mathbf{x})]^2 \qquad (4)$$

3) Weighted Least Squares-Robust (WLS-R) Estimator: In the presence of presence of gross input data errors (outlier values), conventional SE approaches can have computational issues resulting in the minimisation becoming insoluble. The WLS estimator is particularly sensitive to outlier points, since the larger the residual, the larger an influence it has on the quadratic objective function. In order to improve the robustness of the SE in the presence of bad data, "robust" estimators were proposed in [3], [8], [11], which work by re-assign the weights of $\bar{\mathbf{W}}$ as the SE minimisation is solved, so that the weights of suspected bad data (outlier) points are reduced in the calculation. In this paper, the equivalent weight function proposed in [8] is applied to create a "robust" WLS estimator. The equivalent weights are re-calculated at each SE iteration as follows:

$$\overline{w_i} = \begin{cases} w_i, & D_{i'} \le k_0 \\ w_i \dfrac{k_0}{D_{i'}} \left(\dfrac{k_0 - D_{i'}}{k_1 - k_0} \right)^2, & k_0 < D_{i'} \le k_1 \\ 0, & D_{i'} > k_1 \end{cases} \qquad (5)$$

$$D_{i'} = \frac{|r_i|}{\alpha(1 - z_i)\; med_j \left| r_j - med_i r_i \right|} \qquad (6)$$

where i , $j = 1, 2, ..., N$, $\alpha = 1.438$, and *med* is the median, and the variables k_0 and k_1 are calculated based on the influence function from robust statistics theory as described in [8]. The objective is to reduce the influence of input data points with extreme values which can cause the estimator to break down.

4) Extended Kalman Filter (EKF): The EKF approach is formulated by the dynamic model [12]:

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{g}_k + \mathbf{w}_k \qquad (7)$$

where $\mathbf{F}_k$ is the state transition matrix, $\mathbf{g}_k$ is a vector representing the trend behaviour of the state trajectory, and $\mathbf{w}_k$ represents the process noise, modelled as Gaussian white noise with zero mean and covariance matrix **C**. At each time step *k*, the Jacobian matrix $\mathbf{H}_k$ is evaluated with the current predicted states, and the EKF equations, where the recursions are given by:

$$\mathbf{x}_{k+1} = \mathbf{x}_{k+1} + \mathbf{K}_{k+1}[\mathbf{z}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1})] \qquad (8)$$

where $\mathbf{x}$ is the *estimated* state vector, $\mathbf{x}$ is the *predicted* state vector, and $\mathbf{K}_{k+1} = \mathbf{W}_{k+1}\, \mathbf{H^T}_{k+1}\, \mathbf{C^{-1}}$. The state transition matrix $\mathbf{F}$ and vector $\mathbf{g}$ are updated at each iteration using the Holt-Winters method outlined in [13]. The "innovations" at each time step are defined as:

$$\mathbf{v}_{k+1} = \mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1} \qquad (9)$$

where $\hat{\mathbf{z}}_{k+1} = \mathbf{h}(\hat{\mathbf{x}}_{k+1})$, $\mathbf{v}$ is approximately a white Gaussian process with zero mean and covariance $\mathbf{C}_v$ is given by:

$$\mathbf{C}_v = \bar{\mathbf{W}}_{k+1} + \mathbf{H}_{k+1}\, \mathbf{M}_{k+1}\, \mathbf{H}^T{}_{k+1} \qquad (10)$$

where $\mathbf{M}_{k+1}$ is the covariance, and innovations are normalised according to $\mathbf{v_n}(i) = |\mathbf{v_n}| / \boldsymbol{\sigma}_\mathbf{n}(i)$, where $\boldsymbol{\sigma}_\mathbf{n}(i)$ is the standard deviation of the *i*-th iteration [14].

5) Extended Kalman Filter-Robust Estimator (EKF-R): This SE uses the EKF equations described in (7)-(10), but in this case the input measurement weight matrix $\mathbf{W}_{k+1}$ is adjusted according to (5)-(6), in order to reduce the influence of outlier values in the measurement inputs.

III. SE Performance Tests using Field Data

The MV distribution network used in this paper is taken from the European Commission project "SmartHG" [15]. This network is a suburban/rural 10kV system with a weakly-meshed structure. A reduced version of the network schematic is shown in Fig. 1. The network has a peak demand of 3.2 MW, which is made up primarily of suburban/rural residential customers (77% of the total annual demand), with the remaining demand comprising factory, district heating and street lighting loads. There are around 1,600 customers located at 46 MV nodes, where each MV node corresponds to a secondary transformer substation (10:0.4 kV).

On-line measurements of active and reactive power consumption were available at the primary (50:10 kV) transformer. At the MV demand nodes, there are no direct, on-line measurements of voltage or active/reactive power measurements available (this is typical of MV distribution networks, where measurement redundancies are very low due to economic and technical reasons). However, smart meters are installed at all end-users in the LV network, and the hourly consumption/production data for each MV node were available in the form of aggregated smart meter measurements. The case study network has a European configuration where the entire MV system is in three phases, and the connected LV networks are well-balanced across the three phases. Therefore it is assumed in this paper that the MV distribution system is balanced, or at least that the phase imbalance is not critical. In order to apply SEs 1-5 to LV networks, or MV networks with single and two-phase laterals, the methodology described in Section II should be modified to include the 3-phase power flow equations.

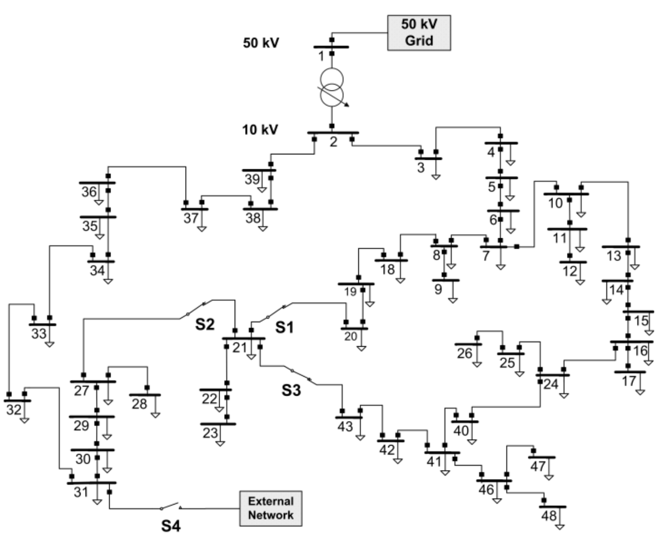

Figure 1. Schematic diagram of MV distribution network.

A. SE Solution Accuracy

The SE methods 1-5 outlined in Section II were applied to recorded data from the MV network. In order to create a set of reference values for the network state, i.e. the "true" values of voltage magnitude and angle throughout the network at each time step, the load flow was solved using a full set of aggregated smart meter measurements at each MV node as the network power injections. This data was pre-processed, removing any bad data and filling in any missing data points with load estimates, as outlined in [3], [4]. Bus 1 was used as the reference with $V = 1.0$ and $\delta = 0$. The "true" network state vector at each time step was compared with the values obtained from SEs 1-5. Real measurements were assumed to have measurement uncertainty of 1%, and pseudo-measurements of the real/reactive power injections at each MV node were applied with an uncertainty of 10% around the mean. Fig. 2 shows a sample of the calculated V and δ values at Bus 48 (the MV node with the greatest electrical distance from the primary substation, and the largest estimation error). The mean and maximum voltage magnitude and angle errors, are given in Table I, where all errors are calculated as Mean Absolute Percentage Errors (MAPEs). The results demonstrate that all SEs 1-5 reach a similar solution, which is expected since the level of uncertainty in the pseudo-measurements is low (10% in this case).

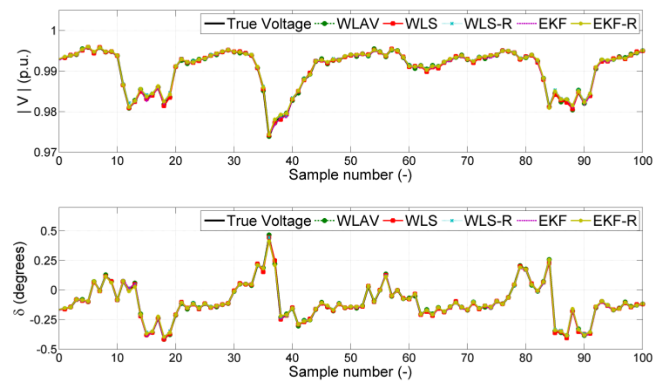

Figure 2. Sample of SE solutions for Bus 48 voltage magnitude and angle.

TABLE I. Mean and Maximum Errors for Bus 48 Voltage Magntitude and Angle Calculated Using SEs 1-5

MAPE	WLAV	WLS	WLS-R	EKF	EKF-R
|V| mean (%)	0.153	0.157	0.166	0.156	0.141
|V| max (%)	0.729	0.805	1.186	0.950	0.954
δ max (%)	0.979	0.963	0.905	0.980	0.915
δ mean (%)	9.998	8.958	8.026	9.808	8.183

B. Noise Performance Tests

In order to demonstrate the performance of SEs 1-5 in situations where there is a high level of uncertainty in the pseudo-measurements at each MV node[1], measurement uncertainty is modelled by adding Gaussian white noise to each MV node active/reactive power injection. The noise is increased incrementally at each MV node from 0% to 50%, and the MAPE of V and δ across the entire network is calculated for each case. The results are summarised in Fig. 3, where it is shown that all SEs 1-5 demonstrate similar performance while the level of uncertainty in the input real and reactive power injections remains below 25%. In the case of the voltage angle, δ, WLS-R and EKF-R demonstrate a slightly better performance while uncertainty is in the 0%-

[1] This may be the case, for instance, if on-line measurements of the power injections are not available, and estimated load profiles based on historical smart meter data are used instead.

30% range. However, above 25% uncertainty, the performance of the robust solution methods deteriorates. The iterative re-calculation of the measurement weight matrix $\bar{\mathbf{W}}$ described by (5)-(6) fails when power injection uncertainty is above 30%, since the algorithm begins to consider noisy values in the input data set as "outliers", and attempts to reduce the influence of these variables in $\bar{\mathbf{W}}$, leading to a loss of valuable information from the input data set.

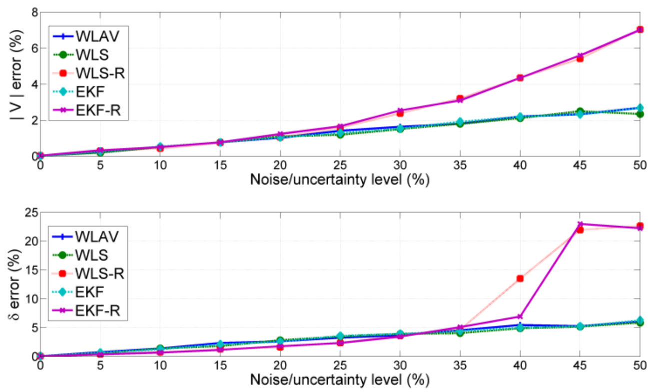

Figure 3. Voltage magnitude and voltage angle estimation error results with various levels of noise applied to all input measurements.

C. Robustness Test

In order to test SE "robustness", i.e. the ability to reach a solution when multiple errors and/or outliers are present in the input measurement data set, gross input data errors ($>5\sigma$) were applied at multiple MV nodes until the breakdown point of each SE solver was reached. The conventional SE solvers (WLAV, WLS and EKF) reach a breakdown point at 9-10 gross input data errors. At this point it was not possible for the SE algorithm to reach a correct solution due to poor conditioning of the Jacobian matrix **H**. The "robust" SEs (WLS-R and EKF-R) are able to reduce the influence of outlier values and reach a correct SE solution with voltage magnitude and angle errors of less than 10%. The breakdown point of the robust methods only occurs at > 40 input errors, at which point measurement data redundancy becomes insufficient to detect errors, Table II.

TABLE II. BREAKDOWN POINT OF EACH SE 1-5

Breakdown Point	**WLAV**	**WLS**	**WLS-R**	**EKF**	**EKF-R**
No. errors	9	9	>40	10	>40

D. Bad Data Identification

For identification of bad data in the MV network measurements, the "tracking" or dynamic SE methods tested (EKF and EKF-R) have advantages over the "static" SE approaches (WLAV, WLS and WLS-R) in their ability to clearly discriminate bad input measurements. Similar observations have been made for transmission-level SE [14]. In WLS-based SE methods[2], bad input data in one network measurement typically results in high residual errors not only in the corresponding node, but also at multiple nearby nodes, causing a "smearing effect" in the residual error vector. In EKF-based SEs, bad input data are detected as unexpected values in the innovation vector (9)-(10), which can provide more accurate error identification. In order to demonstrate bad data detection, gross input data errors ($>5\sigma$) are added at two random MV nodes in the test case MV network, Fig. 4.

For the WLS-based SE, there is an obvious smearing effect, where the bad input data result in multiple high residuals. This makes the input data errors difficult to distinguish clearly from measurement noise, and can cause problems for bad data identification, particularly if there are multiple simultaneous errors. The bad measurements are much more easily identifiable in the innovation vector (EKF-based SE), Fig. 4.

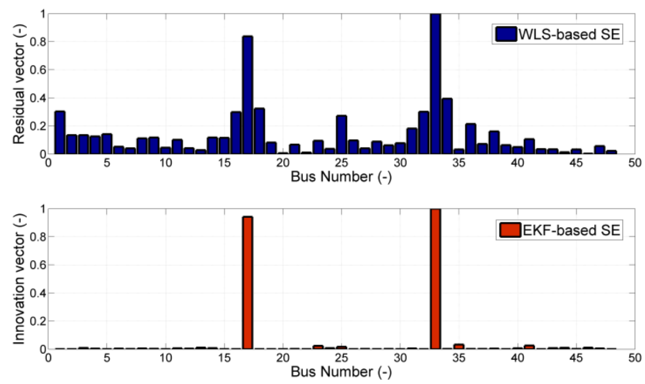

Figure 4. Sample of outputs where the same input data error is tested using "static" (WLS-based) and "tracking" (EKF-based) SE approaches (only the power injection error vectors are shown).

E. Handling of Network Topology Errors

It is assumed in the above analysis that the topology of the network is known *a priori*, before carrying out the SE. However, topological errors can cause problems for DSSE, e.g. if switch statuses in the network are not accessible to the operator due to a fault, or if they are incorrect (i.e. a switch status is showing closed when it is actually open, or vice versa). In such scenarios, the network model used for the SE will be incorrect. Accordingly, several authors have applied an event-triggered approach to network model identification in distribution systems [5], [6]. If a topology error is suspected, e.g. due to a switching operation or a network fault, a network topology identification algorithm is run in order to identify the true network model from several possible network configurations.

In this paper, the Recursive Bayesian Algorithm (RBA) for network model identification proposed in [5] was applied to the MV test case network in Fig. 1. It was found that the choice of estimator has an influence on the network model identification, and that the EKF-based methods showed a slightly faster convergence on the correct network topology. A sample of the results of the event-triggered RBA analysis is shown in Fig. 5, where the objective was to identify the true network model from four possibilities:

[2] The term "WLS-based SE" is used below to refer to SEs 1-3, and "EKF-based SE" is used to refer to SEs 4-5.

- **M1**: all switches S1-S3 closed.
- **M2**: S1 open, S2 and S3 closed.
- **M3**: S2 open, S1 and S3 closed.
- **M4**: S3 open, S1 and S2 closed.

Each model M1-M4 was assigned the same initial probability (0.25), and the correct network topology model (M1) is identified, since its probability trends to unity following each RBA iteration, while the probabilities of models M2-M4 trend to zero.

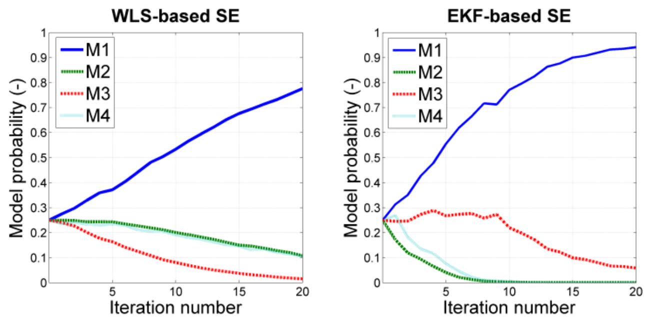

Figure 5. Network model identification using event-triggered RBA method.

F. *Computation Times*

The computation times for each of the 5 SE solvers are given in Table III, where the code was implemented in MatLab. The computational effort required for the EKF-based SEs are larger due to the added SE prediction step, and the "robust" methods required slightly longer solution times, due to the iterative re-calculation of measurement weights.

TABLE III. AVERAGE COMPUTATION TIME FOR ONE FULL MV NETWORK SOLUTION

WLAV	**WLS**	**WLS-R**	**EKF**	**EKF-R**
57.5 ms	59.0 ms	62.4 ms	68.2 ms	71.2 ms

IV. DISCUSSION AND CONCLUSIONS

This paper compared five different SE solution methods, through simulation using field data recorded from an MV distribution network. The performance of each SE 1-5 was compared considering various aspects: SE accuracy, noise/uncertainty, robustness, computation times, and handling of bad data. All SEs 1-5 reached similar solutions while the uncertainty in the input measurements was low (less than 25%). The "robust" SEs, WLS-R and EKF-R, had higher breakdown points, and were able to reach a correct solution even in the presence of outlier values which caused the conventional SEs to fail. However, there was trade-off between robustness and noise performance; with the "robust" methods showing larger errors when input measurement uncertainties are greater than 25%, Section III-B. The analysis found significant differences between the EKF-based and WLS-based solvers in detection and processing of bad input data. It was demonstrated that the EKF-based approaches allowed better discrimination of bad data, since the innovation vector does not produce the "smearing effect" that occurs in the WLS-based SE error residuals. EKF-based approaches also showed better performance in the network topology identification problem, with slightly faster convergence on the correct model.

In this paper, a detailed comparison of the performance of WLS and EKF-based approaches to the static DSSE problem was carried out using field data from an existing MV distribution system. In general, EKF-based solvers have only been applied in distribution networks where high-resolution measurements are available, e.g. from PMUs. In distribution system applications where measurement data is very limited, WLS-based approaches have been used, e.g. [3]-[5], [7]-[9], [11]. However, the results in this paper suggest that EKF-based approaches have significant advantages over WLS-based methods, even in networks where the overall quantity and quality of measurements is poor. These points have been demonstrated using field data from an existing MV distribution network with a European configuration. In addition, this paper provided a comparison of the performance of "robust" DSSE solvers to conventional approaches, which has not been published before. Future work will carry out this analysis on other distribution system case studies, including MV/LV networks with significant phase imbalances.

REFERENCES

[1] B.P. Hayes and M. Prodanovic, "State estimation techniques for electric power distribution systems," presented at *Proc. UKSim-AMSS European Modelling Symposium (EMS 2014)*, Oct. 2014, Pisa, Italy.

[2] B.P. Hayes et al, "Optimal power flow for maximizing network benefits from demand-side management", *IEEE Trans. Power Systems*, vol. 29, no. 4, pp.1739-1747, Jul. 2014.

[3] B.P. Hayes and M. Prodanovic, "Short-term operational planning and state estimation in power distribution networks," presented at *CIRED Electricity Distribution Workshop*, Jun. 2014, Rome, Italy.

[4] B.P. Hayes, J.K. Gruber, and M. Prodanovic, "A closed-loop state estimation tool for MV network monitoring and," *IEEE Trans. Smart Grid*, published online Dec. 2014, DOI: 10.1109/TSG.2014.2378035

[5] R. Singh et al., "A recursive Bayesian approach for identification of network configuration changes in distribution system state estimation," *IEEE Trans. Power Systems*, vol. 25, no. 3, pp. 1329–1336, Aug. 2010.

[6] J. Huang, V. Gupta, and Y.-F. Huang, "Electric grid state estimators for distribution systems with microgrids," in *Proc. 2012 46th Annual Conf.in Information Sciences and Systems (CISS)*, pp. 1–6.

[7] I. Dzafic et al., "Real time estimation of loads in radial and unsymmetrical three-phase distribution networks," *IEEE Trans. Power Systems*, vol. 28, no. 4, pp. 4839–4848, Nov. 2013.

[8] E. Manitsas et al., "Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling," *IEEE Trans. Power Systems*, vol. 27, no. 4, pp. 1888–1896, Nov. 2012.

[9] J. Wu, Y. He, and N. Jenkins, "A robust state estimator for medium voltage distribution networks," *IEEE Trans. Power Systems*, vol. 28, no. 2, pp. 1008–1016, May 2013.

[10] S. Alam, B. Natarajan, and A. Pahwa, "Distribution grid state estimation from compressed measurements," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1631–1642, Jul. 2014.

[11] R. Singh, B.C. Pal, and R.A. Jabr, "Choice of estimator for distribution system state estimation," *IET Gen., Trans. and Dist.*, vol. 3, no. 7, pp. 666-678, Jul. 2009.

[12] A. Monticelli, *State Estimation in Power Systems: A Generalized Approach*. Norwell, MA: Kluwer, 1999, pp. 283-311.

[13] A. da Silva et al., "State forecasting in electric power systems," *IET Gen., Trans. and Dist.*, vol. 130, no. 5, pp. 237–244, Sep. 1983.

[14] M. D. C. Filho and J. De Souza, "Forecasting-aided state estimation: Parts I & II," *IEEE Trans. Power Systems*, vol. 24, no. 4, pp. 1667-1677, Nov 2009.

[15] European Commission SmartHG project website. [Online]. Available: http://smarthg.di.uniroma.it/