Impact of Smart Metering Data Aggregation on Distribution System State Estimation

Qipeng Chen, Dritan Kaleshi, Zhong Fan and Simon Armour

***Abstract*—Pseudo MV/LV (Medium/Low Voltage) transformer loads are usually used as partial inputs to the Distribution System State Estimation (DSSE) in MV systems. Such pseudo load can be represented by the aggregation of Smart Metering (SM) data. This follows the government restriction that Distribution Network Operators (DNOs) can only use aggregated SM data. Therefore, we assess the subsequent performance of DSSE, which shows the impact of this restriction - it affects the voltage angle estimation significantly. The possibilities for improving the DSSE accuracy under this restriction are further studied. First, two strategies that can potentially relax this restriction's impact are studied: the correlations among pseudo loads' errors are taken into account in the DSSE process; a power loss estimation method is proposed. Second, the investments (i.e., either adding measurement devices or increasing the original devices' accuracy) for the satisfactory DSSE results are assessed. All these are for addressing DNOs' concerns on this restriction.**

***Index Terms*—distribution system state estimation, medium voltage power system, smart meter.**

I. Introduction

BY 2020, the majority of EU and US consumers will have their smart meters installed. A smart meter can measure the active and reactive power of loads in every 30 minutes [1]. This development will bring in huge amounts of data. On one hand, stakeholders have shown their firm beliefs in the power and usefulness of the data, but on the other hand, they are keen to find out how to make the most out of the data. Another concern about such Smart Metering (SM) data is its impact on consumer privacy, so the governments have proposed restrictions on stakeholders' SM data access. For Distribution Network Operators (DNOs), only aggregated SM data is allowed to be accessed, so no individuals' privacy can be identified. More details are published in the UK Data Access and Privacy Consultation Report [2] which has followed wider EU and international research related to SM data access and privacy [3], [4]. The above policy has also been extended to a wider range of remote access meters which have similar functionality as smart meters [5]. This restriction concerns stakeholders because it may limit the use of the data in fulfilling their obligations. DNOs' major obligation is to maintain distribution systems' reliability. In recent years, there has been significant interest in involving new types of resources (e.g. electrical vehicles and distributed renewable generations) which bring into these systems new dynamics and give rise to the necessity of accurate operating condition information.

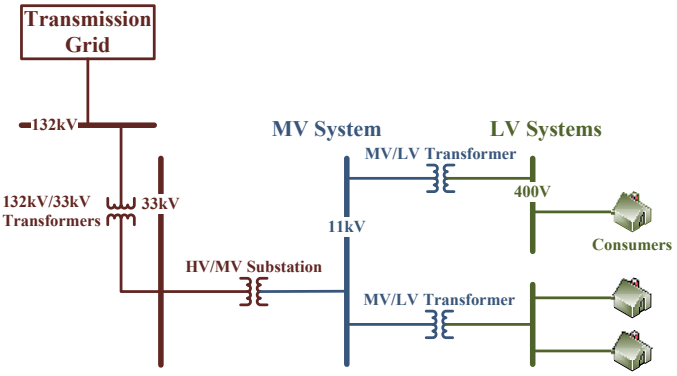


Fig. 1. The UK power distribution system topology.

Medium Voltage (MV) systems' conditions are hard to acquire due to the limited number of measurement devices installed. In Fig. 1 we show the topology of a UK power distribution system. It has High, Medium and Low Voltage (HV, MV and LV) systems. The MV system starts from a HV/MV substation and ends at multiple MV/LV transformers. In most of the countries, the installation of measurement devices is limited to the HV/MV substation but is rarely at MV/LV transformers - the number of MV/LV transformers is nearly two orders of magnitude bigger than the number of HV/MV substations [6]. However, a MV system's condition can be estimated by Distribution System State Estimation (DSSE) which can be facilitated by SM data. We thereby study the performance of the DSSE in MV systems under the data aggregation restriction.

A power system may be taken as a network with nodes and branches. State estimation in a power system can estimate the state of this system, where the state is commonly the set including all nodes' voltage magnitudes and phase angles. Based on the system state, the other electrical quantities such as nodes' loads and branches' power flows and current can be calculated. All of these quantities denote the operating condition of this system. In MV systems, substations' volt-

Manuscript received July 13, 2015; revised September 2, 2015, November 19, 2015 and March 9, 2016; accepted May 11, 2016; Paper no. TII-16-0136.

Q. Chen and S. Armour are with the Communication Systems and Networks Group, Department of Electrical and Electronic Engineering, University of Bristol, United Kingdom, e-mail: {Qipeng.Chen, Simon.Armour}@bristol.ac.uk.

D. Kaleshi is with Digital Catapult, London, United Kingdom, e-mail: Dritan.Kaleshi@cde.catapult.org.uk.

Z. Fan is with Toshiba Research Europe Limited, Telecommunications Research Laboratory, Bristol, United Kingdom e-mail: Zhong.Fan@toshiba-trel.com.

age magnitudes and power flows are usually measurable, in addition to which there may also be a limited number of transformers having measurement devices that produce similar types of measurements. All these are taken as partial inputs to DSSE. To execute DSSE, the pseudo measurements that denote the active and reactive power of loads (hereafter referred to as pseudo loads) at the other transformers without measurement devices are also required. The pseudo loads were modelled monthly or for even longer periods [7], which is applicable for a system under steady state operations but not for the more dynamical system of the near future. The use of SM data offers a new opportunity for more accurate load modeling.

Under the data aggregation restriction, the pseudo load at a transformer is the aggregation of loads of the consumers that are downstream from this transformer. Consequently, the power loss along a LV system is neglected. We evaluate the DSSE performance with such pseudo loads as partial inputs, which actually evaluates the impact of the LV system power loss on the MV system DSSE. Using such SM data aggregated pseudo loads for DSSE has been studied in [8] and [9], but both ignore the power loss impact. If this impact is insignificant, the concern about this data aggregation restriction would be addressed. Otherwise, more efforts may be required to relax this impact and to improve the DSSE accuracy. Therefore, this work also studies the possibilities to improve the DSSE accuracy under this restriction.

Consequently, two potential strategies are considered in this paper respectively. The first strategy defines the correlation between the errors of each pair of measurement variables then takes all such correlations into account in the state estimation process. As discussed, the state estimation in a power system can estimate the system state based on the available real and pseudo measurements. Prior to the state estimation process, a measurement variable is weighted by its degree of brief which is given by the variance of the errors of this variable. And to take into account the weights of all measurement variables in the state estimation process, the state estimation formula traditionally involves a Matrix of Measurement Error Covariances. In the state estimation related research, such as in [8] and [9], this matrix is usually assumed to be diagonal, which is under the assumption that the errors of any two measurement variables are not correlated. However, as described in [10] and [11], the errors of a pair of measurement variables may be correlated, and the state estimation accuracy could be improved by taking into account such correlations. For this case, that covariance matrix may not be assumed to be diagonal. The errors of the pseudo loads at different MV/LV transformers may have correlations, because: the power loss is a main cause to a LV system's pseudo loads' errors; the time-varied power loss of different LV systems may be correlated. In this paper, such correlations are defined and are taken into account in constructing the covariance matrix. This is the first strategy considered to be potential for relaxing the impact of the data aggregation restriction on DSSE in this paper. Alternatively, a method that can estimate the LV system power loss is proposed as the second strategy.

The performance of the DSSE in a MV system also depends on this system's real measurement configuration in terms of the total number and the accuracies of the involved measurement devices. Even if the DSSE accuracy with the above strategies is not satisfactory, DNOs can introduce more investments by either adding extra measurement devices or raising the original devices' accuracies to perform more accurate DSSE. And the data aggregation restriction may not be a big issue to DNOs, if affordable investments can lead to satisfactory DSSE results. Therefore, the performance of the DSSE in MV systems with different measurement configurations is assessed in this paper as well: first, it is assumed that only the HV/MV substation in a MV system has a measurement device, based on which the DSSE performance with respect to different accuracy settings of the measurement device is assessed; second, it is assumed that transformers can also have such devices, based on which the DSSE performance with increased numbers of additional measurements is evaluated.

In recent years, there has been significant interest in using the phasor measurements provided by Phasor Measurement Units (PMUs) to improve the performance of state estimation. However, as described in [6], PMUs are rarely used in power distribution systems due to their high financial costs. Therefore, in this study using PMUs is not taken as a strategy to improve the performance of the DSSE in MV systems. A study with respect to the cost-effectiveness to use PMUs for the DSSE in MV systems may be carried out in the future, which could be helpful for DNOs to determine the trade-off between the costs of these devices and the DSSE performance improvement. And the method to also involve PMUs' phasor measurements into the state estimation process has been introduced in [12].

In addition to the simulations above, one more simulation which also takes into account SM data's errors is also conducted. The above simulations consider the LV system power loss as the only cause to pseudo loads' errors. However, in a realistic scenario, a smart meter cannot produce 100% accurate measurements, and smart meters may not be perfectly synchronized. As a consequence, extra errors may be introduced into the pseudo loads, and the DSSE performance may be further affected.

To our best knowledge, this is the first work that demonstrates the impact of the SM data aggregation restriction on DSSE, which addresses DNOs' concerns about this restriction. This paper is organized as follows. A DSSE technique, the design of the testing systems and the assessment criteria for DSSE results are described in Section II. In Section III, pseudo load errors' correlations are specified, and the LV system power loss estimation method is introduced. The simulation results are discussed in Section IV. Conclusions are drawn in Section V.

II. DSSE and Evaluation

A. Testing System Design

A test system has two levels with a MV system and multiple LV systems. The MV system's network topology is shown in Fig. 2, which is from [13] where the impedance is specified. It has one HV/MV substation (node 1) and 32

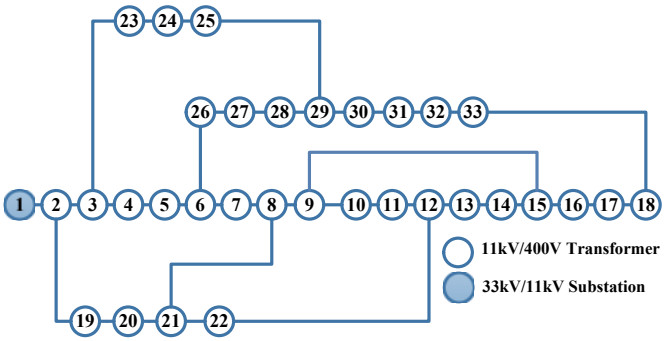


Fig. 2. The 33-node MV power distribution system (from [13]).

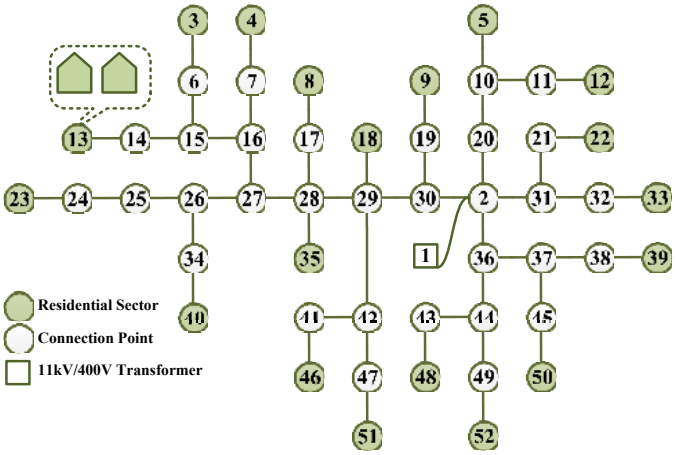


Fig. 3. An example of the LV power distribution system (from [14]).

MV/LV transformers (node 2-33). Each transformer is the starting point of a LV system. We model LV networks by ourselves but follow the method that is proposed in [14]. We model fifteen LV networks for sub-urban areas with different types of underground cables and $300kVA$ transformers. Fig. 3 shows an example, where node one denotes the MV/LV transformer. A filled node is a residential sector which is actually a group of households. Hereafter we call a modelled LV network a BLVN. We model eleven two-level test systems in total, with their names as $MV\&LV_1$, $MV\&LV_2$, ...and $MV\&LV_{11}$ respectively. For the first ten $MV\&LVs$, the LV networks in each $MV\&LV$ are the same, which is one of the first ten BLVNs. $MV\&LV_{11}$ simulates the real world scenario that the LV networks in a system are different. For this case, we allocate the first ten BLVNs to the system's 32 LV networks randomly. The remaining five BLVNs are used for power loss estimation.

The Electricity Customer Behavior Trial Database is made accessible to the public by the Commission for Energy Regulation (CER) [15]. We then allocate this database into each $MV\&LV$ system and calculate their true system conditions. This database records the half-hourly active power consumptions between 14-07-2009 and 31-12-2010 of 4225 residential consumers. We only use their records of the last 50 days. These consumers are with certain residential services. We thus divide these consumers into 32 groups according to their services. We simply assume that the power of a CER consumer during a 30-minute interval keeps constant, under which the original active power consumption data is multiplied by two to get the active power. The reactive power of a household is unknown, so we simulate reactive power records ourselves. We set for each group of households a constant power factor. As specified in [16], the minimum acceptable power factor for residential consumers is between 85% and 90%, so our simulated power factors are randomly derived from this range. Therefore, the reactive power can be calculated accordingly. Provided that the active power p and the power factor pf of a household is known, the reactive power is computed as $p\times\tan(\arccos(pf))$. For a $MV\&LV$, we allocate the 32 groups to the LV networks of this system one by one. For every LV network, we then randomly assign the corresponding group of households to all of the consumer sectors.

For a $MV\&LV$ with allocated SM data, the true time-series operating conditions of each LV system can be calculated. This is accomplished by the power flow analysis with the toolbox MATPOWER [17]. Therefore, the true active and reactive loads at MV/LV transformers are known. The power flow analysis is then used to calculate the true time-series operating conditions of the MV system. The secondary side voltage magnitudes of a MV/LV transformer and the substation are defined as 400V and 11kV respectively.

B. Weighted Least Squares State Estimation Technique

We use the Weighted Least Squares (WLS) State Estimation [18] in our study. In a power system, DSSE aims to identify an optimal set of nodal voltage magnitudes and phase angles such that the system's quantities calculated based on these voltage estimates have the minimum overall error against the provided measurements. This aim can be formulated as a WLS optimization problem:

$$\begin{aligned}\min_{\boldsymbol{x}} J(\boldsymbol{x}) &= \sum\nolimits_{j=1}^{Z_n}(z_j - h_j(\boldsymbol{x}))^2/R_{jj} \\ &= [\boldsymbol{z} - h(\boldsymbol{x})]^T R^{-1}[\boldsymbol{z} - h(\boldsymbol{x})]\end{aligned} \tag{1}$$

where $\boldsymbol{z}$ is the set of the provided measurements. Traditionally, it includes the measurements of nodes' voltage magnitudes and loads and branches' current magnitudes and power flows. j is the index of a measurement, and Z_n is the total number of measurements. $\boldsymbol{x}$ is the set of system state estimates that is given by $[\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{V}}]$, where $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{V}}$ are the sets of nodes' voltage angles and magnitudes respectively. $h_j(\boldsymbol{x})$ is a measurement function that relates $\boldsymbol{x}$ to the measurement z_j, so the residual $r_j = z_j - h_j(\boldsymbol{x})$ reflects the difference between the provided measurement and the measurement calculated from the estimated states. R is the covariance matrix of the errors of measurements, and it is given by Equation 2, where an element denotes the covariance between the errors of a pair of measurement variables. For example, $C_{1,2}$ means the covariance between the errors of z_1 and z_2. And $C_{1,2}$ can also be given by $\sigma_1\sigma_2\rho_{1,2}$, where σ_1 denotes the standard deviation of the errors of z_1, and $\rho_{1,2}$ denotes the correlation coefficient between the errors of z_1 and z_2. In most of the state estimation related research, the errors of measurement

variables are assumed to be uncorrelated, and R is assumed to be a diagonal matrix. Therefore, only the diagonal elements $C_{1,1}$, $C_{2,2}$, …, C_{Z_n-1,Z_n-1} and C_{Z_n,Z_n} are retained, where an element actually denotes the variance of a corresponding measurement's errors. For example, $C_{1,1}$ is given by $\sigma_1\sigma_1$. In Section IV we will first show the accuracy of the DSSE based on such diagonal R matrix. However, as described in [11], measurements' errors may be correlated, and taking into account their correlations could improve the DSSE accuracy. Therefore, in Section III we will show the process to specify the correlation coefficients of measurements' errors and to construct the corresponding R matrix. Then we will show the subsequent DSSE improvement in Section IV.

$$R = \begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,Z_n-1} & C_{1,Z_n} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,Z_n-1} & C_{2,Z_n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{Z_n-1,1} & C_{Z_n-1,2} & \cdots & C_{Z_n-1,Z_n-1} & C_{Z_n-1,Z_n} \\ C_{Z_n,1} & C_{Z_n,2} & \cdots & C_{Z_n,Z_n-1} & C_{Z_n,Z_n} \end{bmatrix} \tag{2}$$

Newton's method is applied, in order to derive the optimal solution of $\boldsymbol{x}$. It initializes all nodes' voltage magnitudes to 1 and phase angles to 0 and iteratively updates $\boldsymbol{x}$ through $\boldsymbol{x}^{k+1} = \Delta\boldsymbol{x}^{k+1} + \boldsymbol{x}^k$ in each iteration k, until $J(\boldsymbol{x})$ in (1) is minimized. The increment $\Delta\boldsymbol{x}^{k+1}$ is calculated through (3), where H is the measurement Jacobian, and $G(\boldsymbol{x}^k)$ is called the gain matrix which equals $H^T(\boldsymbol{x}^k)R^{-1}H(\boldsymbol{x}^k)$. For more details of state estimation see [19], [20].

$$[G(\boldsymbol{x}^k)]\Delta\boldsymbol{x}^{k+1} = H^T(\boldsymbol{x}^k)R^{-1}[\boldsymbol{z} - h(\boldsymbol{x}^k)] \tag{3}$$

C. DSSE in a MV System

In this part, the implementation of the DSSE in the MV system in Fig. 2 is introduced. First, the DSSE required real and pseudo measurements are introduced. Second, these measurements' errors are specified for constructing the DSSE required measurement error covariance matrix. As has been discussed, different measurements' errors are first assumed to be uncorrelated. Under this assumption, the R matrix is diagonal, where each of its elements is associated with the variance of the errors of a measurement variable. Hereafter, such an element is called a diagonal R element. Finally, this section ends with the introduction of the evaluation criteria for the DSSE results. To better illustrate the DSSE implementation, a common situation for the MV system in Fig. 2 is introduced as follows: first, only the substation has a measurement device which measures the voltage magnitude at node one and the active and reactive power flows from node one to two; second, the standard deviation of the relative errors (SDRE) of a substation measurement is assumed to be 0.5%.

Under this common situation, the inputs to the MV system DSSE are those real measurements at the substation and

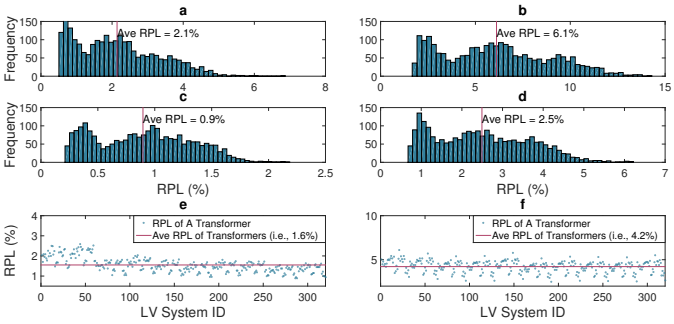


Fig. 4. LV system power loss statistics. a), the distribution of RPL of active power of the LV system with the highest maximum active RPL; b), the distribution of RPL of reactive power of the LV system with the highest maximum reactive RPL; c), the distribution of RPL of active power of the LV system with the lowest maximum active RPL; d), the distribution of RPL of reactive power of the LV system with the lowest maximum reactive RPL; e), the averaged RPL of active power of each of the 320 LV systems; f), the averaged RPL of reactive power of each of the 320 LV systems.

the active and reactive pseudo loads at MV/LV transformers, where the pseudo loads are given by:

$$\widetilde{P} = \sum\nolimits_{n=1}^{S_n} p_n \approx P - \Delta P \tag{4a}$$

$$\widetilde{Q} = \sum\nolimits_{n=1}^{S_n} q_n \approx Q - \Delta Q \tag{4b}$$

where $\widetilde{P}$ is the aggregated active load, p is the active power measurement from a smart meter, n is the index of a household downstream from this transformer, S_n is the total number of households, P is the true active power and ΔP is the LV system power loss. Q and q are for reactive power. These substation measurements and the transformer pseudo loads are the inputs to the MV system DSSE. The system state $\boldsymbol{x}$ of this system is $[\widetilde{\theta}_2, \ldots, \widetilde{\theta}_i, \ldots, \widetilde{\theta}_{33}, \widetilde{V}_1, \ldots, \widetilde{V}_i, \ldots, \widetilde{V}_{33}]$, where i is the index of a node. $\widetilde{\theta}_1$ is set as the reference angle which equals zero, so $\widetilde{\theta}_1$ is not involved in $\boldsymbol{x}$ and $\widetilde{\theta}_i$ is the angle difference between node one and i.

That 0.5% SDRE is used to introduce errors into the true electrical quantities (i.e., the power flow analysis calculated voltage magnitude and power flow) to simulate the real measurements at the substation. Provided that the value of a true electrical quantity at the substation is given, the corresponding real measurement is given by: $\widetilde{A} = A \times (1 + \omega)$, where A is a true electrical quantity value, $\widetilde{A}$ is the simulated real measurement, and ω is randomly generated by the Gaussian function $\mathcal{N}(0, 0.5\%)$. This 0.5% SDRE is also used to set this measurement's associated diagonal R element, which is given by: $(\widetilde{A} \times 0.5\%)^2$.

Specifying the SDRE for transformers' pseudo loads then defining the corresponding diagonal R elements for them is much more difficult. In the following, the error of the pseudo load is assessed, then the diagonal R elements for the pseudo load measurements are derived correspondingly. As described, the first ten $MV\&LVs$ have different LV networks. Consequently, we have 320 different LV systems for analyzing the LV system power loss and defining the corresponding R elements for pseudo loads. In graph a)-d) of Fig. 4 we show the distributions of the Relative Power Loss (RPL) of either

active or reactive pseudo load of four extreme LV systems that are of the highest (lowest) maximum active (reactive) RPL. The active and reactive RPL are given by $\frac{\Delta P}{P}$ and $\frac{\Delta Q}{Q}$ respectively, where P and Q denote the true active and reactive power at a transformer and a time point, and ΔP and ΔQ denote their respective power loss. We also highlight the averaged value of each graph. The RPL distributions are different among LV systems, which is led by their different network configurations (i.e., network topology, cable lengths and types and the number of residential consumers) and their amounts of power consumption. For example, the active RPL of the graph a) associated LV system is as high as 7%, but the maximum value of graph c) is only 2.5%. Such difference for the reactive power is more significant, which can be seen from graph b) and d).

The average RPL of a transformer's pseudo load may be taken to describe this pseudo load's SDRE in offline simulations. Provided that the average RPL of the active pseudo load and the reactive pseudo load of transformer b are given by α and β, the diagonal R elements for this transformer's active and reactive pseudo loads (i.e., $\widetilde{P}_t^b$ and $\widetilde{Q}_t^b$) at time t may be defined as: $(\widetilde{P}_t^b \times \alpha)^2$ and $(\widetilde{Q}_t^b \times \beta)^2$ respectively. However, this may not be applicable in a real-world scenario, because the true loads of this system are not known, and the averaged RPL of the pseudo load cannot be calculated. In other words, the power loss situation for a specific transformer is not known, so it is not available to specialize the diagonal R elements for this transformer. Nevertheless, DNOs may be able to install measurement devices at a limited number of their owned transformers. As a result, these transformers may be taken as samples, which allows DNOs to get a very general idea about transformers' power loss situation. Consequently, all transformers' active/reactive pseudo loads may be assigned the same diagonal R element.

Therefore, in this study transformers' active/reactive pseudo loads' diagonal R elements are assigned in a similar way to reflect this truth of difficulty. All transformers' active pseudo loads are assigned the same SDRE, and their reactive pseudo loads are assigned another SDRE. As shown in graph e), the averaged RPL for the active pseudo loads of different LV systems are shown. The mean value of this graph is highlighted. Graph f) is for the reactive power. Therefore, the SDRE assigned for active pseudo load and reactive pseudo load of all LV systems are these two graphs' mean values which are 1.6% and 4.2% respectively. For example, the corresponding diagonal R element for a transformer's active pseudo load could be given by: $(\widetilde{P}_t \times 1.6\%)^2$, where $\widetilde{P}_t$ denotes the active pseudo load of this transformer at time t.

As described, we also evaluate the investments of additional measurement devices. In our study, a measurement device is placed at a transformer when it is required. It measures the voltage magnitude and loads at this transformer and the active and reactive power flows into and out of this transformer. A 0.5% SDRE is also defined for such measurements for introducing errors and for defining the corresponding diagonal R elements.

D. DSSE Evaluation Metrics and Criteria

The assessment metrics are the Maximum Absolute Relative Voltage Magnitude Error ($|\text{RME}|_{max}$) and the Maximum Absolute Voltage Angle Error ($|\text{AE}|_{max}$). They are given by:

$$|RME|_{max} = \max_{1 \le i \le 33} |V_i - \widetilde{V}_i|/V_i \tag{5a}$$

$$|AE|_{max} = \max_{2 \le i \le 33} |\theta_i - \widetilde{\theta}_i| \tag{5b}$$

where i is the index of a node.

The criteria of both $|\text{RME}|_{max}$ and $|\text{AE}|_{max}$ depend on the subsequent DSSE application, so there are no fixed minimum acceptable thresholds for these two metrics. We therefore use those criteria as in previous references. In [21], the performance of DSSE is acceptable if the $|\text{RME}|_{max}$ and the maximum absolute relative voltage angle error of more than 95% of cases are under 1% and 5% respectively. In [22], when more than 95% of cases' $|\text{RME}|_{max}$ and $|\text{AE}|_{max}$ are under 0.7% and 0.7 crad the state estimation performance is satisfactory, where 1(crad)=0.01(rad). In [9], a 0.6% threshold of $|\text{RME}|_{max}$ is used to determine if the DSSE result is satisfactory.

For voltage magnitude, we assess the proportion of the cases whose $|\text{RME}|_{max}$ are smaller than 1%, and also assess the proportion of the cases whose $|\text{RME}|_{max}$ are smaller than 0.6%. And 95% is taken as the minimum acceptable threshold of the proportion. For voltage angle, we use $|AE|_{max}$ as the metric rather than the relative error of [21], because the true voltage angles of node two, three and four of the system in Fig. 2 are nearly zero for most of the time. This follows the recommendation in [23]. We find that the 0.7 crad of [22] is not to be trustworthy for our system. In our simulation, the criterion for $|\text{AE}|_{max}$ is defined as 0.07 crad. This is the maximum $|\text{AE}|_{max}$ for our initial simulation (i.e., only the substation has real measurements with a 0.5% SDRE and the pseudo load is by SM data aggregation without further modification). The 0.07 crad is not the criterion for determining if the DSSE results are satisfactory for real-world applications. It is used to assess if the power loss estimation, the error correlation specification, the increased substation measurement accuracy and the adding of measurement devices can make changes to DSSE.

III. DSSE Improvement Strategies

Two strategies which may be able to relax the impact of the data aggregation restriction on the DSSE in MV systems are studied in this paper, as discussed. In this section, these two strategies are introduced in sequence. The process to define the correlation between the errors of a pair of measurement variables is first described. The proposed power loss estimation method is then introduced.

A. Error Correlation Specification

The errors of the pseudo loads of different LV systems may be correlated. More specifically, there may exist correlations between two MV/LV transformers' active pseudo loads' errors and between their reactive pseudo loads' errors. Besides, the

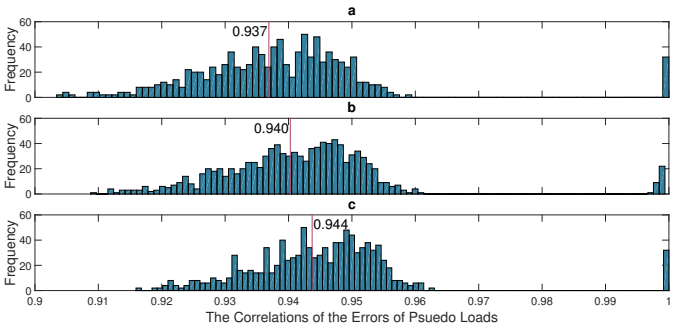


Fig. 5. The correlations of the errors of pseudo loads. a), the correlations of the errors of active pseudo loads; b), the correlations between the errors of active pseudo loads and the errors of reactive pseudo loads; c), the correlations of the errors of reactive pseudo loads.

errors of active pseudo loads and the errors of reactive pseudo loads at the same transformer may be correlated. And there may also be correlations between the errors of active pseudo loads at a transformer and the errors of reactive pseudo loads at another transformer. In the following, such pseudo load error correlations are specified, so a full error covariance matrix rather than the diagonal R matrix could be constructed. Besides, in this study, the real measurements are simulated by randomly introducing errors into the true electrical quantities, so any pair of real measurement variables' errors are assumed to be uncorrelated. The errors of a real measurement variable and the errors of a pseudo measurement variable are assumed to be uncorrelated as well.

The time series active/reactive pseudo loads at a node in the MV system that is shown in Fig. 2 could be obtained by SM data aggregation. However, the true loads at this node are not known in a realistic scenario, because this node rarely has a measurement device, as discussed. Therefore, it is impossible to compute the true correlation coefficient between the errors of two pseudo load variables. Nevertheless, it is reasonable to assume that DNOs may be able to install measurement devices at a limited number of their owned transformers. And this assumption is the same as that has been used in Section II-C when the diagonal R elements for pseudo loads are defined. As a consequence, an averaged pseudo load error correlation may be obtained by analyzing the measurements of the sampled transformers. And this averaged value may be generally taken as the correlation coefficient between the errors of any pair of pseudo load variables in the MV system in Fig. 2.

As discussed, we have 320 different test LV systems, based on which we obtain the averaged correlation coefficient values. First, for any pair of LV systems, we calculate the correlation coefficient between their active pseudo loads' errors. And we show the distribution of such correlation coefficients of all pairs in graph a) of Fig. 5. The correlation coefficient distribution for reactive pseudo load is shown in graph c). Besides, for any pair of LV systems, we also compute the correlation coefficient between the errors of reactive pseudo load of one system and the errors of active pseudo load of the other system. Graph b) shows such correlation distribution. The averaged value in a graph has been highlighted, and we use these values to model the full error covariance matrix. As shown in this figure, we assume that the correlation coefficient between any two transformers' active pseudo loads' errors is 0.937, the correlation coefficient between the errors of active pseudo load of a transformer and the errors of reactive pseudo load of another transformer is 0.940, and the correlation coefficient between any two transformers' reactive pseudo loads' errors is 0.944. Note that in graph b) there are a few bars which are close to 1. And this is the distribution when the correlation coefficient is between the active pseudo loads' errors and the reactive pseudo loads' errors at the same transformer. We simply assume such correlation coefficient to be 1. Additionally, we assume that the correlation coefficient of the errors of two real measurement variables is 0, and the correlation coefficient between the errors of a real measurement variable and the errors of a pseudo load variable is 0. Finally, any element in the full error covariance matrix could be calculated by taking into account measurement errors' correlations.

The following is an example to compute the covariance between the active pseudo loads' errors of transformer b_1 and the reactive pseudo loads' errors of transformer b_2 at time t, where the pseudo active load of b_1 at t is $\widetilde{P}_t^{b_1}$, and the reactive pseudo load of b_2 at t is $\widetilde{Q}_t^{b_2}$. First, the standard deviation of the errors of the active pseudo load at b_1 could be given by: $\widetilde{P}_t^{b_1} \times 1.6\%$, and the standard deviation of the errors of the reactive pseudo load at b_2 could be given by: $\widetilde{Q}_t^{b_2} \times 4.2\%$ (as discussed in Section II-C). Then as shown in graph b) of Fig. 5, the averaged correlation coefficient between these two kinds of errors is 0.94. Therefore, the corresponding covariance between these two kinds of errors could be given by: $\widetilde{P}_t^{b_1} \times 1.6\% \times \widetilde{Q}_t^{b_2} \times 4.2\% \times 0.94$.

B. Power Loss Estimation

In a relatively large area (e.g., a country), there should be LV systems where their MV/LV transformers are equipped with measurement devices, so the power loss of these systems are known. The general idea of our power loss estimation method is to learn the power loss from these systems then apply these statistics to represent the power loss for the LV systems without transformer measurement devices. For a LV system without a transformer measurement device, the time-series aggregated active and reactive power of loads (i.e., pseudo loads) are the only known information. Therefore, we use such limited information for power loss estimation and pseudo load modification. We first describe the power loss estimation in a LV system with a measurement device at the transformer. We then explain how this method can be made applicable to a system without such a measurement device. The voltage magnitude of a MV/LV transformer's secondary side commonly remains in a narrow range (i.e., plus or minus 2% or 5% of the rated value). Besides, we assume that the network configuration of a LV system is not changeable. Therefore in a LV system the power loss is mostly affected by all consumers' total power consumption. In graph a) of Fig. 6, we show the loss of active power in accordance with the aggregated active load in a randomly simulated LV system. We can see that they are highly correlated. The correlation between

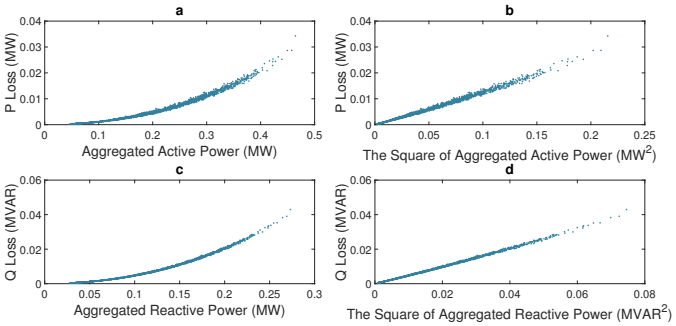


Fig. 6. Correlations between the power loss and the aggregated power in a LV system. a), correlation for active power; b), correlation between power loss and the square of the aggregated active power. c), correlation for reactive power; d), correlation between power loss and the square of the aggregated reactive power.

the square of the aggregated active load and the active power loss becomes linear, which can be seen in graph b). The same applies to the reactive power as well, which can be seen in graph c) and d). Provided that a LV system's transformer has a measurement device, the historical time-series true loads and aggregated loads of this system are known, so is the time-series power loss. Therefore, we can learn how the power loss correlates with the aggregated loads and use such information to estimate the power loss when a new value of the aggregated load is known. A simple method is as follows. Provided that the active and reactive power loss ΔP_t and ΔQ_t at time t are to be estimated, the aggregated loads $\tilde{P}_t$ and $\tilde{Q}_t$ are known, we find the pair of $\tilde{P}_{t'}$ and $\tilde{Q}_{t'}$ of a previous time point t' where $\tilde{P}_{t'}$ and $\tilde{Q}_{t'}$ is the nearest pair to $\tilde{P}_t$ and $\tilde{Q}_t$. ΔP_t is then given by $\Delta P_t \approx \Delta P_{t'} = P_{t'} - \tilde{P}_{t'}$, where $P_{t'}$ is the true active load at t'. The modified pseudo active load is $\hat{P}_t = \tilde{P}_t + \Delta P_t$. ΔQ_t and $\hat{Q}_t$ are estimated in the same way.

This method is not applicable to a LV system that does not have a transformer measurement device because both of the time-series true loads and the time-series power loss are unknown. However, we modify the above method such that we learn the power loss from the LV systems with measurement devices then use such information to estimate the power loss of the other systems without such devices. It also involves the procedure to define the diagonal R element for the modified pseudo load. This method is shown in Fig. 7.

Provided that two LV systems have measurement devices at transformers, we call them the training and validating systems respectively. There is another one without such device, which is called the testing system. They are distinguished by the 'L', 'V' and 'E' superscripts. Fig. 7 shows how to use system 'L' to estimate the power loss (ΔP_t^E, ΔQ_t^E) and to get the modified pseudo loads ($\hat{P}_t^E$, $\hat{Q}_t^E$) of system 'E' at time t. It also shows how to use system 'V' to define the modified pseudo loads' errors' standard deviations (std_$\hat{P}_t^E$, std_$\hat{Q}_t^E$). The inputs include the aggregated loads of system 'E' at t ($\tilde{P}_t^E$, $\tilde{Q}_t^E$), the time-series aggregated loads ($\tilde{\mathbf{P}}^{\mathbf{L}}$, $\tilde{\mathbf{Q}}^{\mathbf{L}}$, $\tilde{\mathbf{P}}^{\mathbf{V}}$, $\tilde{\mathbf{Q}}^{\mathbf{V}}$) and true loads ($\mathbf{P}^{\mathbf{L}}$, $\mathbf{Q}^{\mathbf{L}}$, $\mathbf{P}^{\mathbf{V}}$, $\mathbf{Q}^{\mathbf{V}}$) of systems 'L' and 'V', where a time-series variable $\mathbf{P}$ is given by $\{P_1, \ldots, P_b, \ldots, P_t\}^T$ (T means matrix transpose).

To get (ΔP_t^E, ΔQ_t^E) given ($\tilde{P}_t^E$, $\tilde{Q}_t^E$), we aim to find the time index t' such that among all ($\tilde{\mathbf{P}}^{\mathbf{L}}$, $\tilde{\mathbf{Q}}^{\mathbf{L}}$), ($\tilde{P}_{t'}^L$, $\tilde{Q}_{t'}^L$) is the pair that is nearest to ($\tilde{P}_t^E$, $\tilde{Q}_t^E$). We then take ($\Delta P_{t'}^L$, $\Delta Q_{t'}^L$) as (ΔP_t^E, ΔQ_t^E). The pseudo loads are then modified as $\hat{P}_t^E = \tilde{P}_t^E + \Delta P_t^E$, $\hat{Q}_t^E = \tilde{Q}_t^E + \Delta Q_t^E$.

(P_t^E, Q_t^E) is not known, so we cannot calculate the errors of ($\hat{P}_t^E$, $\hat{Q}_t^E$) directly. We therefore assess how accurate if we use system 'L' to modify the pseudo load for system 'V'. We first find 100 pairs of ($\tilde{P}^V$, $\tilde{Q}^V$) from ($\tilde{\mathbf{P}}^{\mathbf{V}}$, $\tilde{\mathbf{Q}}^{\mathbf{V}}$) such that they are the nearest to ($\tilde{P}_t^E$, $\tilde{Q}_t^E$). For each pair of ($\tilde{P}^V$, $\tilde{Q}^V$), we use system 'L' to get the corresponding ($\hat{P}^V$, $\hat{Q}^V$) and record their errors. The standard deviations of the errors of the 100 pairs of ($\hat{P}^V$, $\hat{Q}^V$) are taken as the standard deviations (std_$\hat{P}_t^E$, std_$\hat{Q}_t^E$) for DSSE. For example, the R element of DSSE for the measurement $\hat{P}_t^E$ is (std_$\hat{P}_t^E$)^2. Note that when the power loss estimation is involved in the DSSE process, the error covariance matrix is assumed to be diagonal. As discussed, the error covariance matrix may be constructed as a full matrix rather than a diagonal matrix when measurements' errors are correlated. We have found that pseudo loads' errors are correlated due to the power loss in LV systems. However, with the above power loss estimation method, the power loss could be estimated and taken into account when modelling the

Inputs: $\tilde{P}_t^E$, $\tilde{Q}_t^E$, $\tilde{\mathbf{P}}^{\mathbf{L}}$, $\tilde{\mathbf{Q}}^{\mathbf{L}}$, $\tilde{\mathbf{P}}^{\mathbf{V}}$, $\tilde{\mathbf{Q}}^{\mathbf{V}}$, $\mathbf{P}^{\mathbf{L}}$, $\mathbf{Q}^{\mathbf{L}}$, $\mathbf{P}^{\mathbf{V}}$, $\mathbf{Q}^{\mathbf{V}}$
Outputs: ΔP_t^E, ΔQ_t^E, $\hat{P}_t^E$, $\hat{Q}_t^E$, std_$\hat{P}_t^E$, std_$\hat{Q}_t^E$
procedure PQ_ESTI
    t' =K_Nearest($\{\tilde{P}_t^E, \tilde{Q}_t^E\}$, $\{\tilde{\mathbf{P}}^{\mathbf{L}}, \tilde{\mathbf{Q}}^{\mathbf{L}}\}$, 1)
    $\Delta P_t^E = P_{t'}^L - \tilde{P}_{t'}^L$
    $\Delta Q_t^E = Q_{t'}^L - \tilde{Q}_{t'}^L$
    $\hat{P}_t^E = \tilde{P}_t^E + \Delta P_t^E$
    $\hat{Q}_t^E = \tilde{Q}_t^E + \Delta Q_t^E$
    $\mathbf{K}$ =K_Nearest($\{\tilde{P}_t^E, \tilde{Q}_t^E\}$, $\{\tilde{\mathbf{P}}^{\mathbf{V}}, \tilde{\mathbf{Q}}^{\mathbf{V}}\}$, 100)
    where $\mathbf{K} = \{K_1, \ldots, K_a, \ldots, K_{100}\}$
    for all K_a of $\mathbf{K}$ **do**
        K_Nearest($\{\tilde{P}_{K_a}^V, \tilde{Q}_{K_a}^V\}$, $\{\tilde{\mathbf{P}}^{\mathbf{L}}, \tilde{\mathbf{Q}}^{\mathbf{L}}\}$, 1)
        calculate $\Delta P_{K_a}^V$, $\Delta Q_{K_a}^V$, $\hat{P}_{K_a}^V$, $\hat{Q}_{K_a}^V$ as above
        Err_$P_{K_a}^V = P_{K_a}^V - \hat{P}_{K_a}^V$
        Err_$Q_{K_a}^V = Q_{K_a}^V - \hat{Q}_{K_a}^V$
    end for
    std_$\hat{P}_t^E$=std(Err_$P_{K_a}^V$), $a \in [1, 100]$
    std_$\hat{Q}_t^E$=std(Err_$Q_{K_a}^V$), $a \in [1, 100]$
end procedure
$\mathbf{x} = \{x^1, \ldots, x^i, \ldots, x^m\}$
$\mathbf{Y} = \{\mathbf{y}^1, \ldots, \mathbf{y}^i, \ldots, \mathbf{y}^m\}$
$\mathbf{y}^i = \{y_1^i, \ldots, y_j^i, \ldots, y_n^i\}^T$
procedure K_NEAREST($\mathbf{x}$, $\mathbf{Y}$, N_k)
    for all $i \in [1, m]$ **do**
        normalize $\{x^i;\mathbf{y}^i\}$ through $\{x^i;\mathbf{y}^i\}/\max(\{x^i;\mathbf{y}^i\})$
        get $\{\underline{x}^i;\underline{\mathbf{y}}^i\}$
    end for
    for all $j \in [1, n]$ **do**
        calculate Dist$_j$=$\sum_i^m (\underline{x}^i - \underline{y}_j^i)^2$
    end for
    return the indexes of the top N_k biggest Dist$_j$.
end procedure

Fig. 7. The method for power loss estimation.

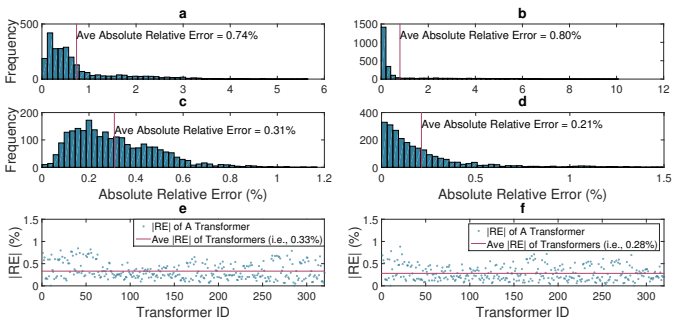


Fig. 8. Power loss estimation performance. a) the |RE| distribution of active power of system 'A' by learning from system 'C'; c) the |RE| distribution of active power of system 'C' by learning from system 'A'; b) the |RE| distribution of reactive power of system 'B' by learning from system 'D'; d) the |RE| distribution of reactive power of system 'D' by learning from system 'B'; e), the averaged active power |RE| for each of the 320 LV systems when learning from the two systems 'A' and 'C'; e), the averaged reactive power |RE| for each of the 320 LV systems when learning from 'B' and 'D';

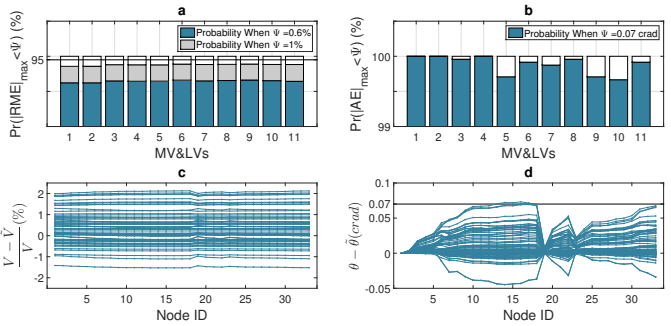


Fig. 9. The performance of DSSE for the common situation. a) 11 $MV\&LV$ systems' voltage magnitude estimates' accuracies; b) 11 $MV\&LV$ systems' voltage angle estimates' accuracies; c) relative errors of voltage magnitude of all nodes and time points of $MV\&LV_{11}$; d) errors of voltage angle of all nodes and time points of $MV\&LV_{11}$.

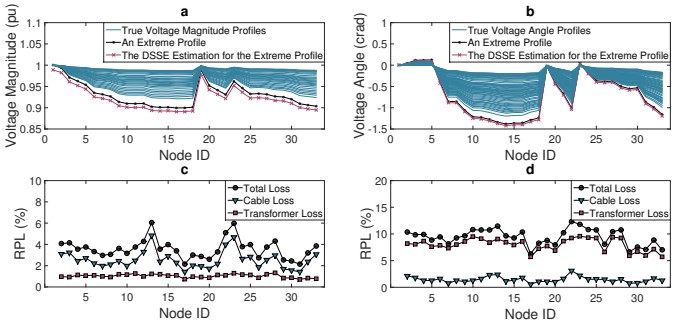


Fig. 10. The condition of the MV system of $MV\&LV_{11}$. a) voltage magnitude profiles; b) voltage angle profiles; c) active power loss in 32 LV systems; d) reactive power loss in 32 LV systems.

pseudo load. For this case, the errors of a pseudo load variable are not caused by the power loss anymore but are introduced by the power loss estimation method. Therefore, with the use of this method, the pseudo load errors are not correlated.

The above uses a single 'L' system and a single 'V' system. Actually, multiple training and validating systems can be used. We simply take multiple systems' data as the data from one system. For example, $\mathbf{P^L}$=$\{\mathbf{P^{L_1}},\mathbf{P^{L_2}}\}$=$\{P_1^{L_1}, \ldots, P_b^{L_1}, \ldots, P_t^{L_1}, P_1^{L_2}, \ldots, P_b^{L_2}, \ldots, P_t^{L_2}\}^T$, where '$L_1$' and '$L_2$' are two training systems.

It is obvious that the similarity between the chosen training systems and the testing system will determine how accurate the estimation of the power loss will be. The similarity between the chosen validating systems and the testing system will determine how well the defined (std_$\hat{P}_t^E$, std_$\hat{Q}_t^E$) is. Therefore, some initial network configuration information of the testing system helps choose better training and validating systems. However, such information is not taken into account in our study. In the following, we show the accuracy of the pseudo load after power loss estimation. We highlight the cases when the chosen training systems are extremely different from the testing system. In the simulation section, we will show that the defined error standard deviation for the modified pseudo load is also appropriate.

The active or reactive power loss of four extreme systems have been shown in Fig. 4. For the sake of simplicity, they are called system 'A', 'B', 'C' and 'D' in accordance with the graph ID. Let us recall that 'A' and 'C' have the highest and lowest maximum active power loss respectively among all 320 systems, and 'B' and 'D' are extreme for their reactive power loss. We use system 'C' to modify the pseudo active load of system 'A' then swap these two. The distributions of the Absolute Relative Error (|RE|) of their modified pseudo loads are shown in graph a) and c) of Fig. 8. We do the same for system 'B' and 'D' for their reactive power, where the results are shown in graph b) and d). As we can see, our method can lead to obvious improvement in terms of the pseudo load accuracy even for these extreme cases when the training and testing systems are significantly different. For example, the averaged RPL(|RE|) of system 'A' before and after the power loss estimation are 2.1% and 0.7% respectively. Furthermore, we use system 'A' and 'C' as the training systems to modify the pseudo active load of all 320 systems. The averaged |RE| of different systems are plotted in graph e), and we highlight the mean value which is only 0.33%. Graph f) is for the reactive power when system 'B' and 'D' are used as training systems, where the mean value is 0.28%. These two values are 1.6% and 4.2% respectively before the power loss estimation, which is shown in graph e) and f) of Fig. 4.

IV. Experiments

We first show the performance of the MV system DSSE for the common situation: only the substation has a measurement device; a real measurement's SDRE is 0.5%; the pseudo load is modelled by SM data aggregation without power loss estimation; and the measurement error covariance matrix is assumed to be diagonal. We then show the improved DSSE accuracy with each of the previously discussed strategies. Followed by this, we show the performance of DSSE with different accuracies of substation measurements or different numbers of additional measurement devices.

A. DSSE under the Common Situation

For assessing the performance of the DSSE under the common situation, we execute DSSE for each $MV\&LV$ for each one of the 2400 (i.e., 50 (days) $\times$ 48) time points. The results are shown in Fig. 9. We find that all test systems show similar results. First, the probabilities when $|\text{RME}|_{max}$ is smaller than 0.6% and 1% are around 65% and 88% respectively, which is shown in graph a). This is not satisfactory because we expect that 95% of cases' $|\text{RME}|_{max}$ can be smaller than 1% or even 0.6%. This is caused by the low accuracy of either the measurements at the substation or the pseudo transformer loads. Second, the $|\text{AE}|_{max}$ of more than 99.5% of cases are smaller than 0.07 crad, which is shown in graph b). The 0.07 crad is the maximum angle error of system $MV\&LV_{11}$, which can be seen in graph d) where the voltage angle errors of all nodes of a time point are shown as a line. We also show the details of the voltage magnitude relative errors in graph c). Because all $MV\&LV$ systems show similar results, we only use $MV\&LV_{11}$ for the following simulations. And the 0.07 crad is taken as the criterion for assessing the improvement of DSSE.

To better understand the results, we also show the condition of $MV\&LV_{11}$ for the time point where the maximum $|\text{RME}|_{max}$ occurs. In graph a) of Fig. 10, the voltage magnitudes of all nodes at different time points are shown with green lines. The profile of the time of the maximum $|\text{RME}|_{max}$ and its estimates are highlighted. In graph b), the profiles for voltage angles are shown. In graph c) and d) the active and reactive power loss for the 32 LV systems for that time point is shown. We detail the system total loss, the loss along LV cables and the power absorbed by transformers. As we can see, in a LV system the active power loss is mostly caused by cables, but the reactive power loss is mostly caused by the impedance of the transformer.

B. Improving the DSSE Performance

We first estimate LV systems' power loss and modify their pseudo loads for DSSE. As described, fifteen LV networks that are called BLVNs are modelled, where ten of them are used for modelling $MV\&LV$ systems. For power loss estimation we use three of the remaining systems as training systems and the other two as validating systems. The modified pseudo loads and the pseudo loads from SM data aggregation are compared in terms of their accuracies. We show such comparison results for active and reactive power in graph a) and d) of Fig. 11 respectively. In each graph, the averaged $|RE|$ of all nodes for each kind of pseudo loads are shown. We can see that the $|RE|$ of the modified pseudo loads are significantly lower for both active and reactive power.

This power loss estimation method can also define the standard deviations of the errors of the modified pseudo loads. In graph b) we demonstrate in a LV system of $MV\&LV_{11}$ how our defined and the true errors of the modified pseudo active load appear in pairs along time. As we can see, there is clear correlation between both. In other words, our defined error standard deviation can effectively reflect the error of the modified pseudo load. To go beyond this, we show the

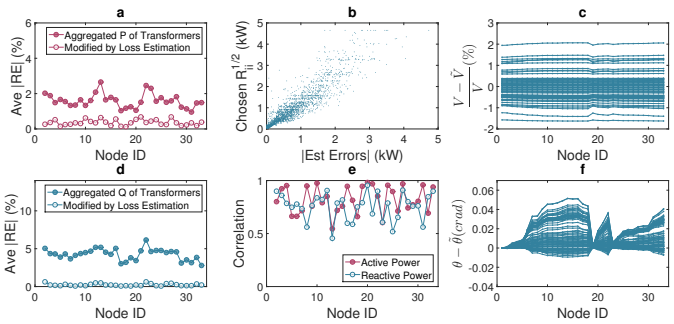


Fig. 11. DSSE with LV system power loss estimation. a) nodes' averaged $|RE|$ of pseudo active load; d) nodes' averaged $|RE|$ of pseudo reactive load; b) correlation between the defined and true accuracy of the modified pseudo active load of a LV system; e) the correlation coefficients for all LV systems; c) relative errors of voltage magnitude of all nodes and time points of $MV\&LV_{11}$; f) errors of voltage angle of all nodes and time points of $MV\&LV_{11}$.

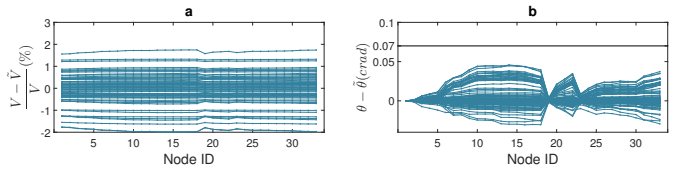


Fig. 12. DSSE with error correlation specification. a) relative errors of voltage magnitude of all nodes and time points of $MV\&LV_{11}$; b) errors of voltage angle of all nodes and time points of $MV\&LV_{11}$.

coefficients of such correlation for all 32 LV systems for both active and reactive power in graph e). More than 75% of cases' coefficients are higher than 0.7. The subsequent DSSE performance is shown in graph c) and f). By comparing Fig. 9 and 11 we find significant improvement on the voltage angle estimates, where the maximum $|\text{AE}|_{max}$ changes from 0.07 to 0.05 which is reduced by 30%. However, there is not an obvious change of $|\text{RME}|_{max}$ though the quality of the pseudo load has been improved significantly. It seems like the data aggregation restriction only shows clear impact on the angle but not magnitude estimates' accuracies.

In Fig. 12 we show the performance of the DSSE when the correlations among pseudo load errors are taken into account. As shown, with this strategy the maximum $|\text{AE}|_{max}$ equals 0.046 crad which is even slightly lower than the maximum $|\text{AE}|_{max}$ with the power loss estimation strategy. It is also found that the resultant $|\text{RME}|_{max}$ with these two strategies are similar. Therefore, with either strategy, the DSSE voltage angle estimation results could be improved significantly.

To further improve the accuracies of the voltage magnitude estimates, real measurements' accuracies should be increased. We therefore assess the DSSE performance with different real measurement accuracies. We assume that the SDRE of real measurements decreases from 1% to 0.1% with a 0.1% decrement. For each SDRE setting the performance of the DSSE of the following three situations are evaluated: the original situation - without error correlation specification and without power loss estimation; the situation with error correlation specification; and the situation with power loss estimation.

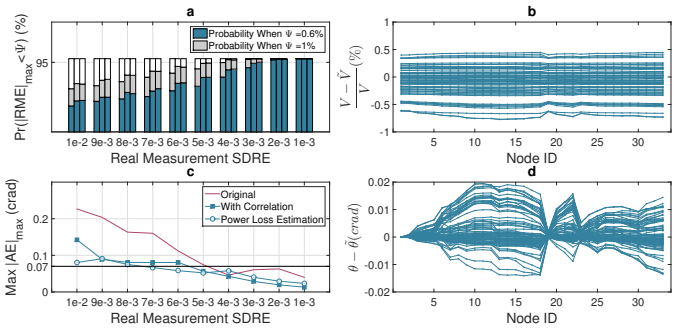


Fig. 13. DSSE performance with different real measurement accuracies. a) voltage magnitude estimates' accuracies with different real measurement accuracies (the bars from the left to the right of a x-axis unit denote the original situation, the situation with error correlation specification and the situation with power loss estimation); c) voltage angle estimates' accuracies with different real measurement accuracies; b) relative errors of voltage magnitude of all nodes and time points of $MV\&LV_{11}$ when real measurements' SDRE are 0.1%; d) errors of voltage angle of all nodes and time points of $MV\&LV_{11}$ when real measurements' SDRE are 0.1%.

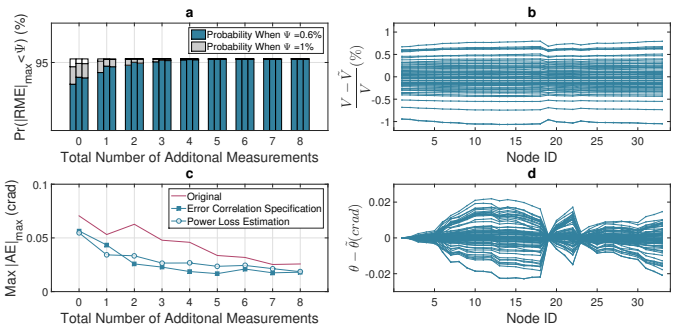


Fig. 14. DSSE performance with different numbers of additional real measurement devices. a) voltage magnitude estimates' accuracies with different numbers of devices (the bars from the left to the right of a x-axis unit denote the original case, the case with error correlation specification and the case with power loss estimation); c) voltage angle estimates' accuracies with different numbers of devices; b) relative errors of voltage magnitude of all nodes and time points of $MV\&LV_{11}$ with three additional real measurement devices; d) errors of voltage angle of all nodes and time points of $MV\&LV_{11}$ with three additional real measurement devices.

The results are shown in Fig. 13. As shown in graph a), to make sure 95% of cases' $|\text{RME}|_{max}$ to be lower than 1% and 0.6%, real measurements' SDRE should be increased to 0.4% and 0.2% respectively for all three situations. And to ensure all cases' $|\text{RME}|_{max}$ to be lower than 0.6% the real measurements' SDRE should be 0.1%. The maximum $|\text{AE}|_{max}$ of different real measurement accuracies can be seen in graph c). As can be seen, with the decrease of SDRE, the results of the DSSE of all three situations are improved significantly. Besides, regardless of the SDRE setting, the results of the situations with either power loss estimation or error correlation specification are much better than the results of the original situation. Furthermore, when SDRE is relatively small (i.e., smaller than 0.5%), the DSSE results of the situation with error correlation specification are slightly better than the situation with the other strategy. The maximum $|\text{AE}|_{max}$ of the situation with error correlation specification for the 0.4%, 0.2% and 0.1% real measurement SDRE settings are 0.043, 0.020 and 0.013 crad respectively which are reduced by 39%, 71% and 81% relative to the 0.07 crad. For the situation with error correlation specification, the details of the DSSE results of the 0.2% real measurement SDRE setting are shown in graph b) and d).

Increasing the real measurement accuracy to a certain level, like the 0.1% SDRE setting, may be technically difficult, so DNOs may consider the adding of additional measurement devices as an alternative for the DSSE improvement. We thereby also assess the impact of the number of additional measurement devices on DSSE. We add such devices to $MV\&LV_{11}$ one after another. A device is placed at a transformer that is chosen by the meter placement method described in [24]. It first identifies the top five branches with the biggest voltage error variances. The branch voltage is a complex value with both magnitude and angle, which is the difference between the voltages of two connected nodes. It then identifies one transformer from the nodes that are connected by these five branches such that this transformer has the biggest power flow error variance. This transformer will be the place for placing the measurement device.

Let us recall that this device measures voltage magnitude and active and reactive loads at the transformer and active and reactive power flows into and out of this transformer. The SDRE of these measurements is assumed to be 0.5%. Same as the above simulation, for all of the three DSSE situations, we show their results with respect to different numbers of additional measurement devices in Fig. 14. As shown in graph a), for 95% of cases' maximum $|\text{RME}|_{max}$ to be lower than 1% or 0.6%, only one or three more measurement devices are required for all these three situations. Graph b) shows the results for voltage angle. The results are similar to the results demonstrated in Fig. 13: with the increase of the number of additional measurement devices, the DSSE voltage angle estimation accuracy could be significantly improved; the results of the original situations are always significantly worse than the results of the other situations; the DSSE results of the situation with the error correlation specification strategy are better than the situation with the power loss estimation strategy. When one or three additional devices are installed, the maximum $|\text{AE}|_{max}$ for the situation with error correlation specification is 0.043 or 0.023 crad which is reduced by 38% or 67% relative to the 0.07 crad. Details of the DSSE results of this situation with three additional measurement devices are shown in graph b) and d).

C. DSSE with Smart Meter Errors

Finally, we introduce errors into those CER active power consumption records and our simulated reactive power measurements by adding errors according to an additive Gaussian function. We then aggregate the modified SM data to get pseudo transformer loads and proceed with DSSE. $MV\&LV_{11}$ is used for this evaluation. This evaluation is under the common situation described in II-C - only the substation has real measurements (i.e., voltage magnitude and active and reactive power flow measurements) that are with a 0.5% SDRE. In this simulation, we only consider the original

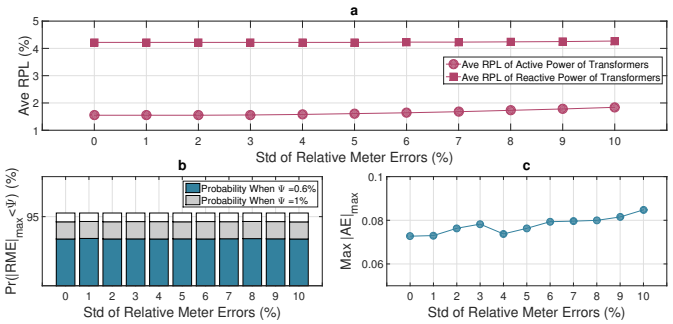


Fig. 15. DSSE performance when the smart meter data is with different levels of errors. a) pseudo loads' relative errors; b) relative errors of voltage magnitude; c) errors of voltage angle.

situation that is without error correlation specification and power loss estimation.

The errors are introduced as follows. We assume SM data's relative errors follow Gaussian distribution $\mathcal{N}(0,\ \eta)$, then we use the Gaussian function to introduce errors for all time points and smart meters. We consider eleven separate situations with different η that are from 0% to 10% with a 1% increment.

The accuracies of the subsequent pseudo transformer loads are first assessed. A point in graph a) of Fig. 15 shows the average RPL of either pseudo load (i.e., active or reactive load) over multiple transformers and time points. With the increase of η, the RPL increases, while the increment is small. When $\eta = 0\%$, the SM data is 100% accurate, and the RPL is caused by LV system power loss only. Therefore, compared with the power loss, the SM data errors show less impact on pseudo loads' accuracy.

As discussed in II-C, we have used 1.6% and 4.2% as the SDRE of active and reactive pseudo loads for getting the corresponding diagonal R elements for DSSE, when $\eta = 0\%$. For the DSSE with other η values, we also use the same two SDRE settings. Graph b) shows the DSSE results on voltage magnitude. As we see, by increasing η, the voltage magnitude accuracy is not changed. For voltage angle, we only show the maximum $|AE|_{max}$ over time for each η, where the results are shown in graph c). With the increase of η, the maximum $|AE|_{max}$ increases, but the results are in a narrow range which is between 0.072 and 0.085 crad. The results show that the SM data errors have insignificant impact on DSSE.

V. Conclusions

The SM data can benefit various grid applications, but the data access may be restricted for reasons, e.g. for the concern of consumer privacy. It is worthwhile to study if there exists conflict between the data access and the functional requirements. Therefore, in this work the impact of the SM data aggregation restriction on the DSSE performance in terms of the voltage magnitude and angle estimation accuracies has been assessed. Our results show that this restriction has insignificant impact on voltage magnitude estimation. Though the voltage magnitude estimates are not satisfactory under the common situation, it is mainly caused by the low substation measurement accuracy rather than this restriction. However, this restriction shows significant impact on voltage angle estimation.

Therefore, one should either relax this restriction or provide affordable solutions to improve the voltage angle estimation accuracy. With regard to affordable solution, the following two strategies have been considered. The first is to take into account pseudo load error correlations in the state estimation process, and the other is to estimate and involve the LV system power loss. The results show that both strategies can improve the voltage angle estimation accuracy significantly. Additionally, we have evaluated the DSSE performance with additional investments in terms of either increasing the accuracy of real measurements or adding more measurement devices. Our results show that a few investments can lead to significant DSSE improvement, so from the DSSE perspective it is not necessary to relax this data aggregation restriction. DNOs' concerns on the SM data aggregation restriction have been addressed and it has been seen that these problems can be solved satisfactorily.

Acknowledgment

The authors wish to take this opportunity to thank the Commission for Energy Regulation (CER) for supplying the Electricity Customer Behavior Trial database. This work was made possible by a grant from Toshiba Research Europe Limited and EPSRC Dorothy Hodgkin Scholarship.

References

[1] "Smart metering implementation programme smart metering equipment technical specifications version 1.1," Department of Energy & Climate Change, 3 Whitehall Place, London, Tech. Rep., 2014.

[2] "Smart metering implementation programme data access and privacy - government response to consultation," Department of Energy & Climate Change, 3 Whitehall Place, London, Tech. Rep., December 2012.

[3] "Opinion 12/2011 on smart metering," Article 29 Data Protection Working Party, B-1049 Brussels, Belgium, Tech. Rep., April 2011.

[4] "Data access and privacy issues related to smart grid technologies," Department of Energy (DOE), Washington DC, Tech. Rep., October 2010.

[5] "Decision on extending the smart meter framework for data access and privacy to remote access meters," Ofgem, 9 Millbank, London, Tech. Rep., February 2015.

[6] D. Della Giustina, M. Pau, P. Pegoraro, F. Ponci, and S. Sulis, "Electrical distribution system state estimation: measurement issues and challenges," *Instrumentation Measurement Magazine, IEEE*, vol. 17, no. 6, pp. 36–42, December 2014.

[7] Y. Chen, P. Luh, C. Guan, Y. Zhao, L. Michel, M. Coolbeth, P. Friedland, and S. Rourke, "Short-term load forecasting: Similar day-based wavelet neural networks," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 322–330, Feb 2010.

[8] M. Baran and T. McDermott, "Distribution system state estimation using ami data," in *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, March 2009, pp. 1–3.

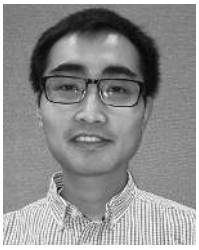
[9] J. Wu, Y. He, and N. Jenkins, "A robust state estimator for medium voltage distribution networks," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1008–1016, May 2013.

[10] C. Muscas, M. Pau, P. A. Pegoraro, and S. Sulis, "Impact of input data correlation on distribution system state estimation," in *Applied Measurements for Power Systems (AMPS), 2013 IEEE International Workshop on*, Sept 2013, pp. 114–119.

[11] ——, "Effects of measurements and pseudomeasurements correlation in distribution system state estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 12, pp. 2813–2823, Dec 2014.

[12] J. Zhu and A. Abur, "Effect of phasor measurements on the choice of reference bus for state estimation," in *Power Engineering Society General Meeting, 2007. IEEE*, June 2007, pp. 1–5.

[13] M. Baran and F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Trans. Power Del.*, vol. 4, no. 2, pp. 1401–1407, Apr 1989.
[14] I. Hernando-Gil, B. Hayes, A. Collin, and S. Djokic, "Distribution network equivalents for reliability analysis. part 1: Aggregation methodology," in *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*, Oct 2013, pp. 1–5.
[15] "Smart Metering Electricity Customer Behaviour Trials (CBTs)," Commission for Energy Regulation (CER), online, http://www.ucd.ie/issda/data/commissionforenergyregulationcer/, accessed: 01/10/2015.
[16] "Power factor correction evaluation," Norman Disney & Young, 1 Chandos Street, ST LEONARDS, Tech. Rep., September 2002.
[17] R. Zimmerman, C. Murillo-Sanchez, and R. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb 2011.
[18] F. Schweppe and E. Handschin, "Static state estimation in electric power systems," *Proceedings of the IEEE*, vol. 62, no. 7, pp. 972 – 982, July 1974.
[19] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. New York: CRC Press, 2004.
[20] A. Monticelli, *State Estimation in Electric Power Systems: A Generalized Approach*, ser. Power Electronics and Power Systems. Springer US, 1999.
[21] R. Singh, B. Pal, and R. Vinter, "Measurement placement in distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 668–675, May 2009.
[22] J. Liu, F. Ponci, A. Monti, C. Muscas, P. Pegoraro, and S. Sulis, "Optimal meter placement for robust measurement systems in active distribution grids," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 5, pp. 1096–1105, May 2014.
[23] P. Vide, F. Barbosa, and J. Carvalho, "Performance metrics for evaluation of a mixed measurement based state estimator," in *ELEKTRO, 2012*, May 2012, pp. 274–279.
[24] Y. Xiang, P. Ribeiro, and J. Cobben, "Optimization of state-estimator-based operation framework including measurement placement for medium voltage distribution grid," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2929–2937, Nov 2014.



Qipeng Chen received the B.Sc. degree in computer science from the University of Ulster, Ulster, U.K., in 2010, and the M.Sc. degree in advanced computing from the University of Bristol, Bristol, U.K., in 2011, where he is currently pursuing the Ph.D. degree with the Electrical and Electronic Engineering Department. His research interests include smart grid, machine learning, and data mining.

Dritan Kaleshi received the Ph.D. degree from the University of Bristol, and the Dipl.Ing. (Hons.) Degree in electronics from the Polytechnic University of Tirana, Albania. He was a Senior Lecturer of Communication Networks with the University of Bristol until 2015. He is currently a Visiting Research Fellow with the University of Bristol, and a 5G Fellow with Digital Catapult, London, U.K. He has authored over 60 papers, edited two international standards, and holds three patents. His research interests are in future networking architectures and protocols (5G and beyond), large scale loosely-coupled distributed systems design, modelling and performance evaluation, and data interoperability for sensor/actuator systems (IoT). He represents the U.K. in various international standardisation bodies (ISO/IEC, CEN/CENELEC) in areas related to IoT, home electronic systems, and smart grid.

Zhong Fan received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, China, and the Ph.D. degree in telecommunication networks from Durham University, U.K. He was a Research Fellow with Cambridge University, a Lecturer with Birmingham University, and a Researcher with Marconi Laboratories, Cambridge. He is currently a Chief Research Fellow with Toshiba Research Europe, Bristol, U.K. His research interests are wireless networks, IP networks, M2M, and smart grid communications. He received a BT Short-Term Fellowship for his work at BT Laboratories.

Simon Armour received the B.Eng. degree from the University of Bath, Bath, U.K., in 1996, and the Ph.D. degree from the University of Bristol, Bristol, U.K., in 2001. He has been a member of the Academic Staff with the University of Bristol, since 2001, where he has been a Senior Lecturer since 2007. He has authored or co-authored over 100 papers in the field of baseband PHY layer and MAC layer techniques for wireless communications with a particular focus on orthogonal frequency-division multiplexing, coding, multiple input/multiple output, and cross-layer multiuser radio resource management strategies. He has investigated such techniques in general terms, as well as specific applications to 4G, 5G and WLAN.