Analysis of State Estimation Accuracy in Distribution Networks Using Pseudo-Measurement Selection and Real Measurement Incorporation

Lilla Barancsuk, Bálint Hartmann

Department of Electric Power Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, H-1111 Budapest, Hungary
barancsuk.lilla@vik.bme.hu, harmann.balint@vik.bme.hu

Abstract: In this paper, we investigate computationally efficient methodologies for generating pseudo-measurements with the aim of selecting the set of pseudo-measurements that can achieve maximum state estimation accuracy for a given network. We analyze the impact of various information sources for residential and demand-response load pseudo-measurements. Additionally, we examine three meteorological approaches to approximate photovoltaic output. By preforming numerical experiments across four real Hungarian low-voltage supply areas, we present the scenarios that yield the maximum state estimation accuracy for each load, which can be conveniently incorporated into any industrial application.

Keywords: distribution system state estimation; pseudo-measurement; smart meter

1 Introduction

State Estimation (SE) is a key element of power system operation ever since it was first introduced by Schweppe et al. in 1969 [1]. While transmission system operation quickly embraced this computational method, the ongoing structural shifts in the power network [2] – including electrification, distributed and variable generation, and the integration of smart functionalities – have elevated SE into the focus of scientific research, yielding substantial advancements in the field. Dehghanpour et al. conducted an extensive review of the method encompassing mathematical problem formulation, pseudo-measurement applications, metering instrument placement, network topology challenges, impacts of renewable energy integration, and addressing cybersecurity concerns [3].

Recently, the application of SE to low-voltage (LV) distribution networks has gained traction, primarily due to the distinctive characteristics and behavior of such

networks. Typically, outage management and control options at the LV level inadequately address the requirements of distribution system operators (DSOs). Over the past three years, E.ON DSO Hungary has collaborated with academic researchers to support distributed operation and control, augmenting the capabilities of its existing LV Supervisory Control and Data Acquisition (SCADA) system. The main driver of this process is escalating power consumption trends and the imperative for capital-intensive network enhancements. Both aspects underscore the necessity of leveraging intelligent solutions (e.g., energy storage, voltage regulators, distributed generation control, demand response) to sustain cost-effective operation. Nonetheless, as the system becomes more complex, ensuring sufficient observability assumes paramount significance. Distribution system state estimation (DSSE) can provide the necessary information for operational software and long-term network planning, but only if the scarcity of measurement data is addressed properly [4].

In LV systems, observability criteria for SE are commonly fulfilled using pseudo-measurements, designed to complement the sparse spatial grid information. Existing literature introduces diverse pseudo-measurement generation techniques aimed at mitigating SE errors to the scale of metering unit offset. Most approaches rely on historical load and generation profiles to model temporal change in node parameters, while other solutions extract supplementary attributes and network features [3, 4]. Spatial and temporal dependencies inherent in pseudo-measurement techniques were tackled in [5, 6]. Metering unit limitations were addressed in [7], and the effect of local prosumers was examined in [8].

Recently, there has been a growing interest in the adoption of data-driven approaches for the generation of pseudo-measurements and state estimation, particularly neural networks [9]. These approaches have been gaining traction due to the potential of neural networks in recognizing complex patterns in the data to enhance the quality of pseudo-measurements. Nevertheless, it is worth noting that these techniques tend to be computationally expensive and need expert knowledge for proper implementation.

Given the resource requirements of such data driven approaches, a pressing need arises for simpler yet effective alternatives that accommodate recent trends of distribution networks [3]. As emphasized already a decade ago in [10], practical applications in industrial network management continue to prioritize robust and computationally efficient methods. To further this objective, in this paper, we focus on cost-effective and computationally efficient pseudo-measurement generation techniques.

The novelty of this paper lies in the following.

- Exploration of efficient techniques for enhancing Distribution System State Estimation (DSSE) accuracy through pseudo-measurement generation, prioritizing simplicity and resource-efficiency.

- Analysis of multiple information sources for generating pseudo-measurements for residential and demand-response loads, alongside meteorological methods for estimating photovoltaic output.
- Validation of the proposed approach via numerical experiments on real Hungarian LV networks, demonstrating its sufficiency in achieving high state estimation accuracy.

The remainder of the paper is organized as follows. Section 2 introduces the methods, including the pseudo-measurement generation techniques and the simulation scenarios. A case study involving four real low-voltage network models are presented in Section 3, and results are discussed in Section 4, alongside the conclusions drawn.

2 Methodology and Data

2.1 Distribution System State Estimation

State estimation (SE) is the process of deducing the state of a power system, i.e., the system's nodal phasor voltages or branch current phasors based on the power system's model and available measurements. This approach is pivotal for real-time monitoring and control of the network, enabling operators to make informed decisions. In LV networks, real-time data from measurement devices are scarce due to the high cost of metering units and the large number of network nodes. Using these limited measurements and a static physical representation of the network, the nodal voltages of the system are estimated using an iterative algorithm, such as the weighted least squares approach detailed in Section 2.1.1. The result of the estimation can then be used in network operation, fault detection and system issue identification, all are of high concern in today's network [11]. However, due to the computational intensity of state estimation, frequent real-time execution in industrial circumstances can be unfeasible, which further necessitates the utilization of low-complexity, efficient pseudo-measurement generation.

2.1.1 The Weighted Least Squares Algorithm

The state estimation method most applied is the Weighted Least Squares (WLS) algorithm [3]. The primary objective of WLS is to minimize the weighted sum of the difference between the measurement vector z and the measurement function $h(x)$, which depends on the system's state x and the system model, $h(x)$:

$$J(x) = \sum_{i=1}^{n} w_i \left(z_i - h_i(x)\right)^2 \tag{1}$$

The weights (w_i) are inversely proportional to the measurement variances. This allows the calculation of the most likely state $\hat{x}$ through the minimization of $J(x)$:

$$\hat{x} = \underset{x}{\operatorname{argmin}} J(x) \tag{2}$$

DSSE employs measurements represented as expected values and standard deviations, where the former is used as the measurement, and the latter determines the weight of each measurement. Given that DSSE applications often involve nonlinear relations between the state (such as bus voltage phasor or branch current phasor) and measurements (including active/reactive powers and voltage/current values) as exemplified in 2.1.2, an iterative Newton method is employed to tackle this nonlinearity.

The DSSE algorithm used for the numerical experiments performed is based on the WLS approach and is customized to fit the datasets and IT systems used by E.ON DSO Hungary [12]. We obtained topological data from GIS models, and the modelling work was carried out in Python using the pandapower package [13].

2.1.2 The Physical Modelling of Power Systems

An electric power grid, comprised of buses and power lines, forms a graph where nodes (N) represent buses and edges represent lines. A line across nodes n_1 and n_2 can be described by its line series admittance: $y_{n_1,n_2} = g_{n_1,n_2} + jb_{n_1,n_2}$ and total shunt susceptance: $\mathrm{j}b^{s}_{n_1,n_2}$, following a nominal π line model. Based on Ohm's law the current flowing from node n_1 to node n_2 across the line is given by

$$\mathrm{i}_{\mathrm{n}_1,\mathrm{n}_2} = \left(\mathrm{y}_{\mathrm{n}_1,\mathrm{n}_2} + \mathrm{jb}^{\mathrm{s}}_{\mathrm{n}_1,\mathrm{n}_2}/2\right)\mathrm{u}_{\mathrm{n}_1,\mathrm{n}_2} - \mathrm{y}_{\mathrm{n}_1,\mathrm{n}_2}\,\mathrm{u}_{\mathrm{n}_1,\mathrm{n}_2}, \tag{3}$$

and according to Kirchhoff's current law, the injected current into node n_2 equals the sum of currents on the lines incident to node n_2: $i_n = \sum_{n_1 \in N_{n_2}} i_{n_1,n_2}$, where $N_{n_2} \subseteq N$ is the set of nodes directly connected to node n_2. This formulation can be represented as a matrix equation, stacking all line currents into a single vector: $\underline{i} = Y\underline{u}$, where Y is the bus admittance matrix commonly expressed as $Y = G + \mathrm{j}B$. Let $s_n := p_n + \mathrm{j}q_n = u_n\overline{i_n}$ express the complex power injected into node n. The active and reactive power injected into each node in the network can then be expressed by the so-called power flow equations with nodal voltages given in polar coordinates as $u_n = U_n\exp(\mathrm{j}\theta_n)$:

$$p_n = \sum_{n'=1}^{\mathrm{N}} U_n U_{n'}\left(G_{n,n'}\sin\theta_{n,n'} - B_{n,n'}\cos\theta_{n,n'}\right), \tag{4}$$

$$q_n = \sum_{n'=1}^{N} U_n U_{n'}\left(G_{n,n'}\cos\theta_{n,n'} + B_{n,n'}\sin\theta_{n,n'}\right), \tag{5}$$

where $G_{n,n'}$ and $B_{n,n'}$ are elements of the real and imaginary parts of the admittance matrix, G and B, respectively. Equations (3), (4) and (5) imply that all power system quantities can be expressed by nodal voltage amplitudes and angles: $\underline{U} = U_1, U_2, \ldots, U_n, \theta_1, \theta_2, \ldots, \theta_n$. Consequently, $\underline{U}$ serves as the system state or state vector. From equations (4) and (5), it can be concluded that power injections are

nonlinear functions of the state vector. To derive the state vector from the nonlinear relations, the approach introduced in (2) can be employed, substituting x in with the state vector, and replacing z with active and reactive power injection measurements and pseudo-measurements. The $h(x)$ model function in this formulation contains static network parameters, such as G and B. These parameters are computed from the physical attributes of the network, typically obtained from DSO information systems. The iterative approach facilitates solving the nonlinear model, resulting in the determination of the most probable system state as the resulting state vector [1].

2.2 Pseudo-measurement Generation

Due to the lack of real-time measurements state estimation must heavily rely on pseudo-measurements for equation (2) to be solvable. In DSSE, most of the information constitutes pseudo-measurements, which play a big role in affecting how accurate the estimated state is. The primary source of pseudo-measurements are synthetic load profiles (SLPs), also known as customer class curves heavily used in the literature [12]. These profiles, usually provided with a time resolution of 1-4 hours, often come with an anticipated maximum error of 50%. Another source of pseudo-measurements is historical load data. These data can be a basis for the modelling of load behavior, offering approximations of consumption leveraging probabilistic analysis [14]. In recent years, several data driven, and machine intelligence-based load modelling techniques were proposed, e.g., particle swarm optimization and genetic algorithm in [15] or a neural network in [9].

Most DSSE systems and pseudo-measurement generation approaches also rely on measured data collected from real meters in the network [7]. Many recent works investigate the effects of incorporating these measurements into the estimation scheme on the accuracy of the DSSE [16].

In this study, we formulate three distinct scenarios of pseudo-measurement generation. These scenarios are designed to assess estimation accuracy when applying various complexities of load modelling, particularly in the presence of smart meter (SM) data. The pseudo-measurement generation technique proposed in this paper integrates information from four distinct sources: synthetic load profiles (for both loads and photovoltaic generators), historical data (HD), actual smart meter measurements taken at the simulated network sites (SM), and photovoltaic (PV) production curves calculated using solar irradiance data or SLP. The specifics of these scenarios are detailed in subsequent sections.

2.2.1 Synthetic Load Profiles

SLPs used for the study were supplied by E.ON DSO Hungary [17], presenting a realistic picture of the electricity usage patterns of Hungarian LV consumers. These curves are traditionally used by the DSO for network planning and to predict electricity consumption throughout the year. This dataset contains consumer curves

for multiple load types, of which we use the curves for "controlled" (ripple- and radio ripple controlled domestic electric water heaters) and "PV" customers. The SLPs consist of consumption rates for the entire year of 2021 with a time resolution of 15 minutes. To produce the actual consumption values, these datasets are scaled, for each load they are multiplied with the average annual consumption (AACs) of the load to be modelled. Figure 1 shows standardized customer curves for residential and controlled loads.

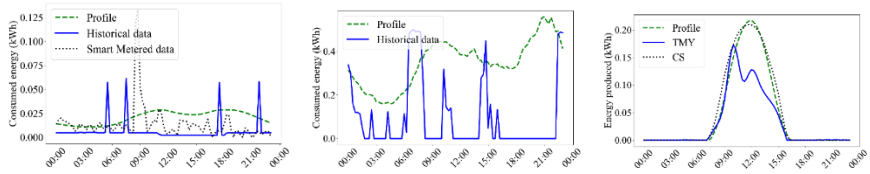

Figure 1

Various data sources for residential, demand-response consumption and PV production (presented in order) modelling for the day of 2021-01-01. Historical and Smart Metered datasets show high levels of intermittency compared to profile. The PV production is shown for a photovoltaic plant of 10 kWp. As seen in the right figure, the DSO profile is very similar to the clear sky model.

2.2.2 Historical Dataset

The historical dataset we used was collected from a different Hungarian distribution grid than the ones we studied in this paper. This dataset includes many consumption patterns that reflect the typical electricity use in Hungary. It's made up of consumption profiles from 334 residential users and 69 controlled users, each consisting of 15-minute active power consumption values for the whole course of year 2020 with different AACs for each set. To align the curves from the different sources, the historical set is truncated to start with the same day (Friday) as 2021.

2.2.3 Smart Metered Data

We had access to Smart Meter (SM) measurements from the year 2021 for each area. These measurements included quarter-hour readings of both active and reactive energy consumption. The measurements were linked to specific locations in the networks using load IDs. The quantity of available SM measurements for each supply area is outlined in Table 5. Smart metered measurements for a residential load are shown in Figure 1.

2.2.4 Photovoltaic Output

As outlined in section 2.3, one of the scenarios in the pseudo-measurement dataset incorporates the output of PV generators. PV energy production is known for its volatility, causing substantial fluctuations and imbalances in the power system, thus

emphasizing the importance of accurate PV modeling [18, 19, 20]. While modelling photovoltaic output might come with significant inaccuracies, using approximations remains valuable, particularly when considering the broader impact on the network rather than focusing on a specific timestamp. Since all simulated supply areas encompass multiple PV generators (as indicated in Table 5), modelling the effect of PV output is crucial for achieving good estimates [21].

The current body of literature offers various methods for estimating PV generation, including the utilization of weather models [22], probabilistic approaches [23], machine learning and data-driven techniques [24], among others of which irradiation estimation stand out as the most robust means of PV output approximation due to their reliance on historical weather patterns.

Irradiation estimates share a common feature in that they rely on historical data, often spanning several years, to construct solar irradiation profiles and, consequently, typical PV output patterns. It's important to note that while this standardized approach is well-suited for analyzing aggregated PV generation, it may not perform optimally when applied to individual PV installations. Nonetheless, they offer a consistent basis for comparative analysis, allowing us to assess the relative impact of PV generation on distribution networks. Due to their widespread usage and their applicability across locations, we included clear-sky (CS) and typical meteorological year (TMY) models alongside SLPs for PV output calculations.

A CS model is an estimate of the amount of sunlight reaching the Earth's surface under idealized clear sky conditions [25]. This model is useful for approximating the maximum potential solar energy that can be harvested.

We also use a TMY irradiance dataset [26]. TMY data provide a realistic representation of weather conditions based on historical data for a specific location. TMY data are valuable for simulating the actual performance of PV systems under real-world conditions.

For accessing CS and TMY irradiance data, we draw from the European Commission's open meteorological database known as the Photovoltaic Geographical Information System [25]. PV production estimates are shown in Figure 1.

Table 1
Data source utilization across scenarios

	SLP	**SM**	**HD**	**PV**
Scenario 1 (residential consumers)				
SC1a		✓	✓	
SC1b		✓	✓	
Scenario 2 (demand response)				
SC2a	✓	✓	✓	
SC2b	✓	✓	✓	

Scenario 3 (distributed generation)				
SC3a	✓	✓	✓	
SC3b		✓	✓	✓
SC3c		✓	✓	✓

2.3 Simulation Scenarios

To analyze the accuracy of the DSSE regarding various load types three scenarios are devised, each of them focusing on a typical configuration and/or a challenge with respect to DSSE. In each scenario, the networks contain varying types of loads and/or PV sources. In these scenarios, we have matched each of the loads with a time series from one of the data sources, considering relevant combinations as detailed in Table 2. The scenarios are detailed in the following sections.

2.3.1 Scenario 1: Considering only Residential Consumers

Scenario 1 considers only household customers connected to the network, excluding other consumer types. Residential consumers are using a random dataset from the HD, matching their Average Annual Consumptions (AACs). Table 3 summarizes the pseudo-measurement sources. For this scenario, two sub-cases were investigated.

Table 2

Pseudo-measurement sources of the different load types in Scenario 1

Load type	SC1a		SC1b	
	Reference	**Simulation**	**Reference**	**Simulation**
SM	SM	HD	SM	SM
Non-SM residential	HD	HD	HD	HD
Controlled	-	-	-	-
PV	-	-	-	-
# of weeks	52	52	26	26

In the unmonitored case (SC1a), each residential load is associated with a randomly selected dataset from the HD that matches the AACs of the load to be modelled, following the approach described in [12].

In the measured case (SC1b), residential loads without SM measurement sets are paired with random residential datasets from the HD, similarly to SC1a. However, loads with corresponding SM data are modelled using their SM dataset.

2.3.2 Scenario 2: Demand Response Loads

Recent research suggests that actual load profiles may deviate significantly from standard residential Synthetic Load Profiles (SLPs) [27] due to the presence of different consumption patterns in modern distribution grids. Demand response enabled loads (DRELs), primarily controlled loads like domestic water heaters in Hungarian distribution networks, exhibit behavior that significantly differs from conventional residential consumers [28]. Consequently, residential customer class curves are inadequate for approximating their consumption patterns [17].

Since all modelled networks include a substantial number of controlled loads (as shown in Table 5), this scenario enhances SE by incorporating controlled pseudo-measurements. Previously overlooked demand response loads in the network are reintroduced, and their power consumption is modelled using two different approaches. Further details of the scenario can be found in Table 3.

Table 3
Pseudo-measurement sources of the different load types in Scenario 2

Load type	Reference	SC2a	SC2b
SM	SM	SM	SM
Non-SM residential	HD	HD	HD
Controlled	HD	SLP	HD
PV	-	-	-
# of weeks	26	26	26

In SC2a, all controlled loads in the network use the same power consumption profile from SLPs, which is scaled to match the load's Average Annual Consumption (AAC). In SC2b, all controlled loads are paired with randomly selected controlled datasets from the HD, following a similar approach as described in [12].

2.3.3 Scenario 3: Distributed Generation

In recent years, the growing adoption of renewable energy sources, particularly photovoltaic (PV) generation, has led to their widespread installation in distribution networks. Consequently, the accurate and reliable estimation of PV output has emerged as a critical concern for effective network operation, control, and monitoring [29]. As mentioned in Section 2.2.4, there are multiple established methods in the literature for estimating PV output based on weather patterns. Scenario 3 systematically explores the approaches presented in 2.2.4 to assess the effect of these estimations on SE accuracy. The details of the scenarios are shown in Table 4.

In case of SC3a all PV producers in the networks use the same power generation profile derived from SLPs and the PV panels' reference peak power. In case of SC3b all PV producers in the networks use the same power generation profile

derived from clear sky irradiance and the PV panels' reference peak power. While in case of SC3c all PV producers in the networks use the same power generation profile obtained from TMY irradiance and the PV panels' reference peak power.

Table 4
Pseudo-measurement sources of the different load and generation types in Scenario 3

Load type	**Reference**	**SC3a**	**SC3b**	**SC3c**
SM	SM	SM	SM	SM
Non-SM residential	HD	HD	HD	HD
Controlled	-	-	-	-
PV	Same as the scenario	SLP	CS	TMY
Num. of weeks	26	26	26	26

3 Case Study

3.1 Real Low-voltage Networks

The low-voltage networks used for the case study are situated in the western part of Hungary, encompassing MV/LV transformers and low-voltage supply circuits. These areas are denoted by specific letters: A, B, C, and D, each comprising 2, 10, 5, and 4 circuits, respectively. Layouts of the networks are presented in Figures 7-10 in the Appendix. The layout and load characteristics of these areas align with typical Hungarian transformer regions. They were deliberately selected to include diverse load types, such as residential, and controlled loads, alongside high PV penetration, ensuring a representative cross-section for the scenarios. We obtained electric parameters from the GIS system of E.ON, adopting an approach that strikes a balance between minimizing human intervention and achieving precise estimations. This method allows for certain approximations while facilitating automated topology creation, as exemplified in [13].

Table 5
Number of loads by type in the networks modelled for the case study

Supply area	**A**	**B**	**C**	**D**
Type of data available				
PV	8	27	9	3
SM	7	23	23	10
SM and controlled	2	1	18	5
SM and PV	5	19	7	3
Residential	76	408	169	185
Controlled	24	79	82	54
Total	**78**	**418**	**171**	**187**

3.2 Dataset Preparation

3.2.1 Reference Dataset

For each scenario, we generate a distinct set of pseudo-measurement data, which is explained in detail in Sections 2.3.1-2.3.3. Additionally, we create a reference dataset for benchmarking purposes. The reference dataset is established by executing a load-flow algorithm using input data from a distinct but analogous timeframe compared to the simulated data. The input data for the load-flow algorithm in each scenario can be found in Tables 2-5 under the "Reference" column. This method essentially constructs a virtual scenario that represents a likely network state. We employ this virtual scenario to compare and assess the outcomes of the scenarios. The load-flow simulation is executed separately for each scenario, case, and network, producing distinct reference datasets for each case.

In each instance (except for SC1a), we divide the datasets into 52 weeks. Odd-numbered weeks are employed as input for the estimation, while even-numbered weeks serve as input for the load-flow simulation, the results of which are considered the reference dataset. This approach yields 26 weeks' worth of simulation data for most scenarios, and 52 weeks' worth of data for SC1a, where benchmarking is conducted for the entire year.

3.2.2 Estimating Uncertainty

As established in Section 2.1, SE deduces the network state by assessing the reliability of the input. Measurements with lower uncertainty levels are considered more reliable and are given higher importance in the calculation. Thus, realistic calculation of uncertainties is essential for informative modelling.

By assuming a Gaussian error distribution, we calculate the standard deviation of the datasets by fitting a Gaussian curve to the distribution of differences between the pseudo-measurements and the reference datasets. This curve fitting process is conducted for both absolute and relative differences across all data sources. The resulting Gaussian standard deviation is regarded as the measurement uncertainty. To gain a comprehensive understanding of the inherent uncertainties in the sources, we collect error distribution data across all networks and throughout the entire year for each source. Table 6 presents the standard deviations of data sources for both active (P) and reactive power (Q) pseudo-measurements. It's worth noting that relative uncertainties between active and reactive measurements exhibit a strong correlation within a given source. As expected, smart metered measurements and PV SLPs emerge as the most reliable sources.

Table 6
Uncertainties of the data sources

Data source	Relative [%]		Absolute [kW, kvar]	
	P	Q	P	Q
HD residential	86.86	81.26	1.245	0.252
SM	74.02	90.41	0.466	0.133
HD controlled	141.74	141.72	0.433	0.142
SLP controlled	119.14	119.10	0.448	0.147
SLP PV	61.88	68.58	0.316	0.102
CS PV	158.26	147.11	0.966	0.224
TMY PV	54.88	61.45	0.358	0.113

3.3 Simulation Process Overview

The scenarios are assessed by means of numerical simulation. The entire course of a year in 15-minute resolution is evaluated for every network-scenario combination.

The process is as follows (according to the flowchart in Figure 2):

1) Selection of the scenario and the network to be evaluated.
2) Building of the digital representation of the selected a network from network parameters provided by E.ON.
3) Generation of measurements and pseudo-measurements spanning the entire year.
4) Execution of the state estimation and the load-flow analysis, which is responsible for generating the reference dataset.
5) Analysis of convergence of the estimation.
6) Comparison of the output from both algorithms to calculate error metrics.
7) Saving the results as output files.

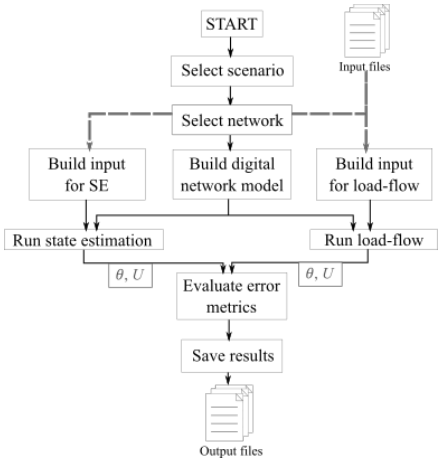

Figure 2
Steps of the simulation and evaluation process

4 Results and Discussion

4.1 Error Metrics

To analyze the results, we assessed various error metrics across all areas and scenarios. Error metrics were calculated in Python via the numpy library. First, the relative and absolute error of voltage magnitudes (Δu) and the absolute phase error ($\Delta\theta$) for each node of the network (i) and for each time step (t) were calculated as follows.

Voltage magnitude error:

$$\Delta u_{\mathrm{ABS}}(t,i) = \tilde{U}(t,i) - U_{\mathrm{REF}}(t,i) \tag{6}$$

Phase error:

$$\Delta\theta(t,i) = \tilde{\theta}(t,i) - \theta_{\mathrm{REF}}(t,i) \tag{7}$$

where $\tilde{U}(t,i)$ and $\tilde{\theta}(t,i)$ are the voltage magnitude and phase provided by the SE and $U_{\mathrm{REF}}(t,i)$ and $\theta_{REF}(t,i)$ are the reference voltage magnitude and phase at node i for timestep t, respectively. The following aggregated error metrics are assessed for all the above errors:

Root mean square error (RMS):

$$e_{\mathrm{RMS}} = \sqrt{\frac{\sum_{t=1}^{T}\sum_{i=1}^{N} y(t,i)^2}{NT}}. \tag{8}$$

Mean absolute error (MAE):

$$e_{\mathrm{MAE}} = \frac{\sum_{\mathrm{t=1}}^{T}\sum_{\mathrm{i=1}}^{N}|y(t,i)|}{NT}. \tag{9}$$

Mean relative error (MRE):

$$e_{\mathrm{MRE}} = \frac{\sum_{\mathrm{t=1}}^{T}\sum_{\mathrm{i=1}}^{N}\frac{u(t,i)}{U_{REF}(\mathrm{t,i})}}{NT}, \tag{10}$$

where $y(t,i)$ represents the specific error quantity (voltage magnitude or phase error). Mean relative error is calculated solely for voltage magnitudes, as it is not meaningful with respect to the voltage angles. The error is calculated for time step t and at node i. The calculations of the overall error metrics encompass a total of T time periods and encompass an evaluation of N nodes within the networks (the number of network nodes is specified in Table 5). T equals 17472 for all scenarios except for SC1a where T=34944. Error metrics were obtained by averaging over all nodes and time steps of the analyzed networks. Comparisons were drawn between the scenarios and network topologies based on the error metrics. The results of the experiments are discussed in the following section. To facilitate clarity and precision, the overall error metrics in the subsequent section are denoted by specifying both the type of metric employed and the physical quantity being assessed, as exemplified in Table 7.

Table 7
Evaluated error metrics and their notations

Error metric	Physical quantity	
	u (Voltage)	θ (Phase)
RMS	Δu_{RMS}	$\Delta\theta_{\mathrm{RMS}}$
MAE	Δu_{MAE}	$\Delta\theta_{MAE}$
MRE	Δu_{MRE}	--

4.2 Discussion of the Scenarios

4.2.1 Scenario 1

In Scenario 1 (Figure 3), a significant difference can be observed between the cases SC1a and SC1b. In the unmonitored case (SC1a) without SM, the state estimation exhibits a larger error. In contrast, when SM measurements are used (SC1b), the results show improved accuracy. The error tends to correlate with the network's size, with the largest error observed at the site with the highest number of nodes, B. This phenomenon is attributed to the error's propensity to increase with greater distances from the network's feed-in point. Location D, where both MRE and RMS for voltage magnitudes are the largest across the networks is an exception,

suggesting that the small number of SMs (10) compared to the number of network nodes (187) contributes to the low observability of this location.

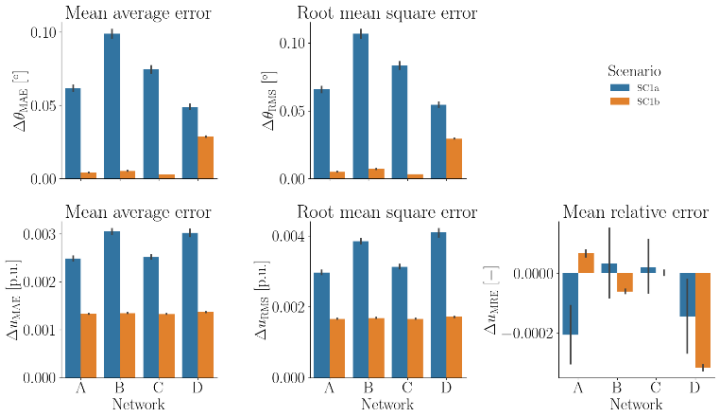

Figure 3
State estimation error for Scenario 1

4.2.2 Scenario 2

The two demand response scenarios exhibit similar performance (Figure 4), with slightly higher errors compared to Scenario 1 due to the introduction of uncertainty from demand response load modeling. Controlled consumers with intermittent consumption patterns pose a challenge to estimation. The error correlates with network size. As seen in Figure 6, SC2b produces slightly smaller errors, except for location C where the controlled load penetration is the highest. Historical data or SLPs for demand response modeling have minimal impact on results, while sophisticated approaches like price-based approximations may improve accuracy [16], however, using readily available SLPs still provides satisfactory results, aligning with previous findings on demand response loads [16].

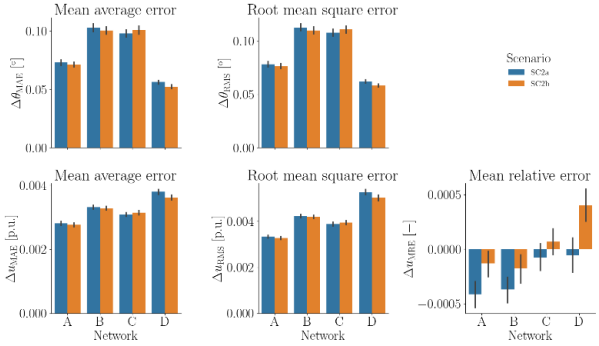

Figure 4
State estimation errors across each network for Scenario 2

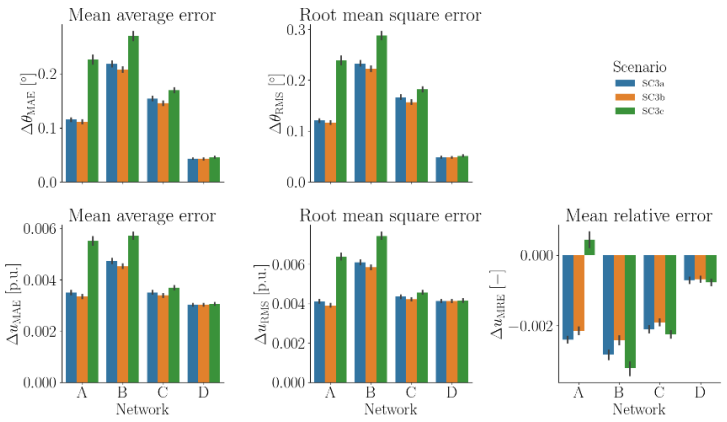

Figure 5
State estimation error for Scenario 3

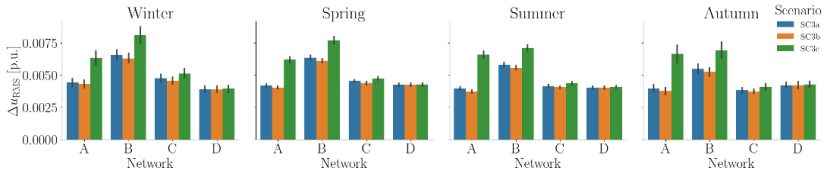

Figure 6
State estimation RMS error for Scenario 3 across the seasons. For locations with lower PV penetration, like C and D, the error remains stable across the year, however for higher PV production locations, the error is highly dependent on the season.

4.2.3 Scenario 3

When distributed generation is incorporated into the measurement set (as shown in Figure 5), measurement errors are the highest compared to the other scenarios. As for previous scenarios, the highest error occurs at site B. Two key factors contribute to this outcome. Firstly, site B is the largest among the considered sites, and as demonstrated in previous scenarios, error rates tend to increase with network size. Additionally, this network hosts a significant number of PV plants, with 19 out of 23 PV feed-in points equipped with smart meters. Consequently, when comparing actual PV output (represented by SM reference data) with approximations, the error is significantly amplified.

This analysis emphasizes the limitations of current approximations in capturing PV consumer behavior, highlighting the importance of smart metering for PV output monitoring in distribution networks. Moreover, it indicates that the TMY approximation (SC3b) is less accurate compared to other scenarios due to higher uncertainty in the TMY dataset. Particularly, errors are more pronounced in

locations with high PV penetration (e.g., locations A and B), suggesting that usng clear sky values for irradiation changes yields more precise state estimation outcomes than relying on the varying irradiance based on the TMY model.

Conclusions

This study investigates various techniques for generating pseudo-measurements to optimize DSSE accuracy, prioritizing resource-efficient approaches, particularly in networks with significant PV integration, validated using four real low-voltage networks. Results show that distributed generation significantly impacts the accuracy of state estimation, especially at larger sites with numerous PV units, emphasizing the need for increasing the observability by using smart metering. Comparing different estimation approaches for PV output underscores that CS and TMY models have their limitations. Moreover, our study demonstrated that demand response consumption introduces a level of uncertainty due to these loads' intermittent consumption patterns. In conclusion, improving the accuracy of SE in distribution networks necessitates a multi-faceted approach. Smart metering, demand response modelling, and selecting the most accurate model for PV output, are all critical for efficient network management with increasing renewable energy integration.

Acknowledgement

This research was funded by the Hungarian Academy of Sciences and E.ON Hungary grant number LENDÜLET_2019-111. All authors are with the MTA-BME Lendület FASTER Research Group, Department of Electric Power Engineering, Budapest University of Technology and Economics, 1111 Budapest, Hungary. (email: barancsuk.lilla@vik.bme.hu)

References

[1] F. Schweppe and J. Wildes, "Power System Static-State Estimation, Part I: Exact Model," IEEE Transactions on Power Apparatus and Systems, Vol. PAS-89, No. 1, pp. 120-125, Jan. 1970, doi: 10.1109/tpas.1970.292678

[2] P. Kádár, "Pros and cons of the renewable energy application," ACTA POLYTECHNICA HUNGARICA, Vol. 11, No. 4, pp. 211-224, Jan. 2014, doi: 10.12700/aph.25.04.2014.04.14

[3] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A Survey on State Estimation Techniques and Challenges in Smart Distribution Systems," IEEE Transactions on Smart Grid, Vol. 10, No. 2, pp. 2312-2322, Mar. 2019, doi: 10.1109/tsg.2018.2870600

[4] A. Primadianto, and L. Chan-Nan, "A review on distribution system state estimation," IEEE Transactions on Power Systems, Vol. 32, No. 5, pp. 3875-3883, Dec. 2016, doi: 10.1109/TPWRS.2016.2632156

[5] M. Pau, P. A. Pegoraro, A. Monti, Ferdinanda Ponci, and S. Sulis, "Impact of Current and Power Measurements on Distribution System State

Estimation Uncertainty," IEEE Transactions on Instrumentation and Measurement, Vol. 68, No. 10, pp. 3992-4002, Oct. 2019, doi: 10.1109/tim.2018.2883844

[6] M. Shafiei, G. Nourbakhsh, A. Arefi, G. Ledwich, and H. Pezeshki, "Single iteration conditional based DSSE considering spatial and temporal correlation," International Journal of Electrical Power & Energy Systems, Vol. 107, pp. 644-655, May 2019, doi: 10.1016/j.ijepes.2018.12.025

[7] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich and E. Vaahedi, "Distribution System State Estimation Based on Nonsynchronized Smart Meters," in IEEE Transactions on Smart Grid, Vol. 6, No. 6, pp. 2919-2928, Nov. 2015, doi: 10.1109/TSG.2015.2429640

[8] G. Cheng, S. Song, Y. Lin, Q. Huang, X. Lin, and F. Wang, "Enhanced state estimation and bad data identification in active power distribution networks using photovoltaic power forecasting," Electric Power Systems Research, Vol. 177, p. 105974, Dec. 2019, doi: 10.1016/j.epsr.2019.105974

[9] Y. Wang, J. Gu, L. Yuan, "Distribution network state estimation based on attention-enhanced recurrent neural network pseudo-measurement modeling," Protection and Control of Modern Power Systems, Vol. 8, No. 31, pp. 1-16, Jul. 2023, doi: 10.1186/s41601-023-00306-w

[10] P. Kádár, "Application of Optimization Techniques in the Power System Control," ACTA POLYTECHNICA HUNGARICA, Vol. 10, No. 5, Sep. 2013, doi: 10.12700/aph.10.05.2013.5.14

[11] W. Yu, J. Wang, H. Peng, J. Ming, C. Tao, T. Wang, "Fault diagnosis of power systems using fuzzy reasoning spiking neural P systems with interval-valued fuzzy numbers", Romanian Journal of Information Science and Technology, Vol. 1, No. 20, pp. 5-17, 2017

[12] G. B. Békési, L. Barancsuk, I. Táczi, and B. Hartmann, "Benchmarking Various Pseudo-Measurement Data Generation Techniques in a Low-Voltage State Estimation Pilot Environment," Applied Sciences, Vol. 12, No. 6, p. 3187, Mar. 2022, doi: 10.3390/app12063187

[13] L. Thurner et al., "Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems," IEEE Transactions on Power Systems, Vol. 33, No. 6, pp. 6510-6521, Nov. 2018, doi: 10.1109/tpwrs.2018.2829021

[14] P. Kapler, "Forecasting of Residential Power Consumer Load Profiles Using a Type-2 Fuzzy Inference System," ACTA POLYTECHNICA HUNGARICA, Vol. 19, No. 9, pp. 201-219, Jan. 2022, doi: 10.12700/aph.19.9.2022.9.11

[15] R. Singh, B. C. Pal and R. A. Jabr, "Statistical Representation of Distribution System Loads Using Gaussian Mixture Model," IEEE Transactions on

Power Systems, Vol. 25, No. 1, pp. 29-37, Feb. 2010, doi: 10.1109/TPWRS.2009.2030271

[16] J. Liu, R. Singh and B. C. Pal, "Distribution System State Estimation With High Penetration of Demand Response Enabled Loads," IEEE Transactions on Power Systems, Vol. 36, No. 4, pp. 3093-3104, Jul. 2021, doi: 10.1109/TPWRS.2020.3047269

[17] E.ON, "DSO policy exclusively available on the website," Budapest, 2023. [Online] Available: https://www.eon.hu/hu/rolunk/vallalatcsoport/kozlemenyek/szabalyzatok-jogszabalyok/aram/eon-eszakdunantuliaramhalozati/elosztoi-szabalyzat-kizarolag-honlapon-publikalando-mellekletei.html, Accessed: 2023.10.10

[18] N. N. Son, L. T. Vinh, “Parameter Estimation of Photovoltaic Model, Using Balancing Composite Motion Optimization,” ACTA POLYTECHNICA HUNGARICA, Vol. 19, No. 11, pp. 27-46, Nov. 2022, doi: 10.127/APH.19.11.2022.11.2

[19] E. Manea, E. Budianu, M. Purica, C. Podaru, A. Popescu, C. Parvulescu, F. Babarada, “Silicon solar cells parameters optimization by adequate surface processing techniques,” Romanian Journal of Information Science and Technology, Vol. 11, No. 4, pp. 337-345, Jan. 2008

[20] M. Tekin, M. Sekkeli, “Active-Reactive Power Control of Grid-Tied Solar Energy Systems, Using Multi-Parameter Band Control for Grid Voltage Regulation,” ACTA POLYTECHNICA HUNGARICA, Vol. 17, No. 3, pp. 147-164, Jan. 2020, doi: 10.12700/APH.17.3.2020.3.8

[21] H. Z. Margossian, "Iterative State Estimation With Weight Tuning and Pseudo-Measurement Generation," IEEE Systems Journal, Vol. 15, No. 3, pp. 3165-3172, Sept. 2021, doi: 10.1109/JSYST.2021.3060072

[22] K. Yasunami, "An Estimation Method of Photovoltaics Power Output Utilizing Integral Power Consumption and Solar Radiation Intensity," IEEJ Transactions on Electronics, Information and Systems, Vol. 143, No. 2, pp. 117-124, 2023, doi: 10.1541/ieejeiss.143.117

[23] N. Mahdavi, D. Weeraddana, and Y. Guo, “Probabilistic estimation of PV generation at customer and distribution feeder levels using net-demand data,” IEEE Transactions on Smart Grid, Vol. 14, No. 3, pp. 1974-1984., Sep. 2022, doi: 10.1109/TSG.2022.3208606

[24] Y. Yuan, K. Dehghanpour, F. Bu and Z. Wang, "A Probabilistic Data-Driven Method for Photovoltaic Pseudo-Measurement Generation in Distribution Systems," 2019 IEEE Power & Energy Society General Meeting (PESGM), Atlanta, GA, USA, 2019, pp. 1-5, doi: 10.1109/PESGM40551.2019.8974026.

[25] T. Huld, R. Müller, A. Gambardella, "A new solar radiation database for estimating PV performance in Europe and Africa," Solar Energy, Vol. 86, No. 6, pp. 1803-1815, Jun. 2012, doi: 10.1016/j.solener.2012.03.006

[26] S. Wilcox, W. Marion, "Users Manual for TMY3 Data Sets (Revised)," United States: N. p., 2008, doi: 10.2172/928611

[27] A. Angioni, T. Schlösser, F. Ponci and A. Monti, "Impact of Pseudo-Measurements From New Power Profiles on State Estimation in Low-Voltage Grids," in IEEE Transactions on Instrumentation and Measurement, Vol. 65, No. 1, pp. 70-77, Jan. 2016, doi: 10.1109/TIM.2015.2454673

[28] N. Koltsaklis, J. Knápek, "The Role of Flexibility Resources in the Energy Transition," ACTA POLYTECHNICA HUNGARICA, Vol. 20, No. 11, pp. 137-158, 2023, doi: 10.12700/APH.20.11.2023.11.9

[29] M. Tekin and M. Sekkeli, "Active-Reactive Power Control of Grid-Tied Solar Energy Systems, Using Multi-Parameter Band Control for Grid Voltage Regulation," ACTA POLYTECHNICA HUNGARICA, Vol. 17, No. 3, pp. 147-164, 2020, doi: 10.12700/aph.17.3.2020.3.8

Appendix

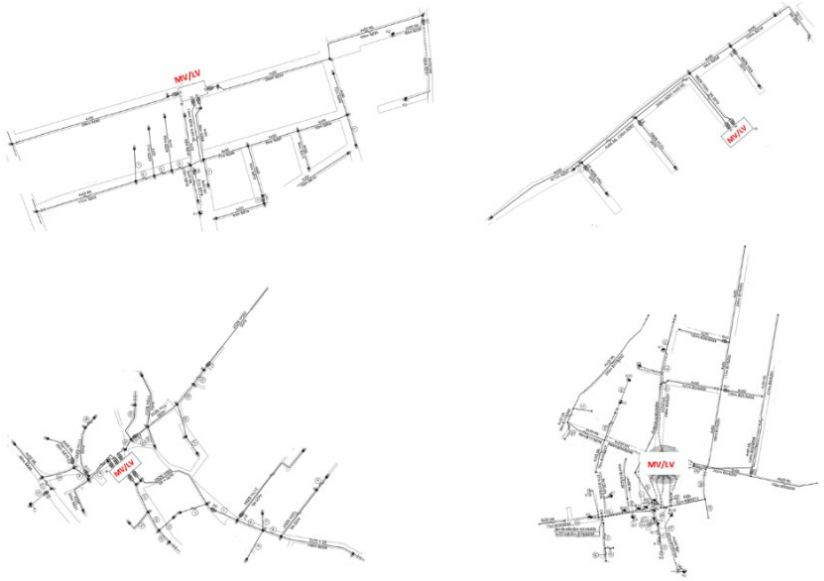

Figure 7

Layouts of supply areas. Order from left to right, top to bottom: D, A, C, B. Electric parameters and topology were extracted from the E.ON GIS system. Topology layout visualizations are the generated using the Python-pyplot library with manually added details.