# Distribution Grid State Estimation With Limited Actual and Pseudo Measurements

Steven de Jongh, *Graduate Student Member, IEEE*, Felicitas Mueller, *Graduate Student Member, IEEE*, Claudio A. Cañizares, *Fellow, IEEE*, Thomas Leibfried, *Senior Member, IEEE*, and Kankar Bhattacharya, *Fellow, IEEE*

***Abstract*—Methods for distribution system state estimation in Low Voltage (LV) distribution grids are discussed in this paper, for systems with a high penetration of Distributed Energy Resources (DERs) such as solar generators and heatpumps. The proposed methods are specifically designed for LV grids with sparse measurement availability, such as feeders with measurements only at the distribution transformer, as is typically the case in some European LV grids. For these cases, device locations, temporal data, and weather data are used in the proposed techniques to estimate variables at unmeasured grid nodes. The impact of smart meters is also investigated by simulating the impact of individual smart meter measurements on the estimation results. The proposed methods are based on time series disaggregation of transformer measurements, such as thermoelectrical demand, baseload, and solar generation, enabling improvements over existing Pseudo-Measurement (PM) generation techniques. Furthermore, the paper presents approaches for estimating voltages and currents in the feeder using both actual and PMs, based on classical estimation methods and interval estimation techniques for unmeasured variables. The results for an realistic German LV grid show that the proposed disaggregation step allows to significantly improve the results of the state estimation results over state-of-the-art methods.**

***Index Terms*—Disaggregation, distribution system state estimation, interval estimation, low voltage grids, pseudo measurements, smart meters.**

## GLOSSARY

| | |
|---|---|
| ASHP | Air-Source Heat Pump |
| DER | Distributed Energy Resources |
| DSO | Distribution System Operator |
| DSSE | Distribution System State Estimation |
| EM | Ensemble Model |
| EV | Electric Vehicle |
| GIS | Geographic Information System |
| GLD | Grid Load Disaggregation |
| GSHP | Ground-Source Heat Pump |
| HP | Heat Pump |
| LR | Linear Regression |
| LV | Low Voltage |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MARPE | Mean Absolute Rated Percentage Error |
| ML | Machine Learning |
| MLE | Maximum Likelihood Estimate |
| MSE | Mean Squared Error |
| MV | Medium Voltage |
| NN | Neural Network |
| PM | Pseudo-Measurement |
| PMU | Phasor Measurement Unit |
| PV | Photovoltaic |
| RT | Regression Tree |
| SE | State Estimation |
| SGD | Stochastic Gradient Descent |
| SLP | Standard Load Profile |
| SM | Smart Meter |
| UPM | Uniform PM Method |
| WLS | Weighted Least Squares |
| XGB | eXtreme Gradient Boosting |

Received 26 November 2024; revised 18 March 2025; accepted 21 May 2025. Date of publication 6 June 2025; date of current version 25 August 2025. This work was supported by MITACS, Karlsruhe Institute of Technology and University of Waterloo. Paper no. TSG-02055-2024. *(Corresponding author: Steven de Jongh.)*

Steven de Jongh, Felicitas Mueller, and Thomas Leibfried are with the Institute of Electric Energy Systems and High-Voltage Technology, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany (e-mail: steven.dejongh@kit.edu).

Claudio A. Cañizares and Kankar Bhattacharya are with the Electrical and Computer Engineering Department, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2025.3576697.

Digital Object Identifier 10.1109/TSG.2025.3576697

## I. INTRODUCTION

A DISTRIBUTION system state is defined as the vector of all voltages and currents at the nodes of the grid. Estimating this state in Low Voltage (LV) grids has become increasingly important due to the ongoing integration of Distributed Energy Resources (DERs), such as Photovoltaic (PV) generators, Heat Pumps (HPs), and Electric Vehicle (EV) chargers. LV grid planning has traditionally not taken the power associated with these DERs into account, thus requiring the monitoring of the grid state to prevent overloads, equipment damage, and power outages [1]. To avoid these issues, the targeted control of flexible grid components and the shifting of load and generation to time periods of lower demand is needed.

Regulations such as the revised version of §14a of the German Energy Industry Act (EnWG) provide Distribution System Operators (DSOs) with new requirements for grid monitoring to be aware of the actual state of the grid [2], specifically the nodal voltages and currents on cables, lines,

1949-3053 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

and transformers. However, DSOs in Germany and other European countries have access to only a limited number of spatially distributed measurements at DSO-owned LV grid equipment, such as transformers, cables and cable distributors, thus necessitating the development of specialized voltage and current estimation methods considering that a limited subset of assets is equipped with measurement devices. In this context, Distribution System State Estimation (DSSE) can be applied to estimate the node voltages and feeder currents from measurements, under the following challenges in the absence of a wide deployment of Smart Meters (SMs), as is the case of Germany [3]:

- *Sparse Measurements:* The number of measurements in LV grids is significantly lower compared to the transmission system level, leading to a scarce placement of meters in the grid. This may make the system unobservable, which prevents the direct application of existing estimation algorithms. Hence, methods for Pseudo-Measurement (PM) generation are necessary to allow systems state estimation.
- *Low Aggregation Level:* Measurements at the LV distribution transformer level involve profiles that are aggregated from individual buildings. Since profiles of individual buildings exhibit high variability due to the occupants' behavior, predicting their consumption is challenging.
- *Measurement Equipment Limitations:* The measurement equipment used in LV grids is typically measured voltages and current magnitudes, with voltage phase angles being rarely recorded. This prevents the calculation of voltage phasors and thus results in model simplifications being necessary in the state estimation process.

To ensure the feasibility of controlling active grid components under these challenges and in accordance with §14a [2], the development of DSSE methods for LV grids is a key topic of discussion among German regulators, distribution grid operators, and researchers. Initial studies propose standardized procedures [4], aimed at enabling LV DSSE under the aforementioned conditions and highlighting its relevance for electrical distribution networks to support the expansion of PV, heat pumps, and charging infrastructure. However, the LV DSSE state-of-the-art is at early stages, necessitating further development of specialized methods with high estimation accuracy.

In the scientific literature, various approaches have been proposed to address the aforementioned challenges. Thus, a fully measured LV grid is considered in [5], where measurements from SMs and transformers are used, and PMs are generated using average values of SM measurements, obtaining estimates using the Weighted Least Squares (WLS) method. In [6], different estimation methods for distribution systems are considered; however, all nodes are equipped with SMs, which does not represent a realistic case for typical LV grids in various jurisdictions with limited deployment of new generation of SMs with real-time measurement capabilities (e.g., Germany [7]).

In reality, most distribution grids lack full observability, meaning that there is not enough information to directly estimate the system state. A common approach to deal with non-observable distribution systems involves generating PMs at unmeasured or inadequately monitored nodes. An early method is described in [8], where PMs are derived from feeder measurements obtained at the distribution transformer, using proportional scaling factors, solving the DSSE problem using the WLS method. However, this approach applies simple scaling factors without accounting for individual consumer characteristics and devices, leading to high PM errors. Another approach to DSSE described in [9] uses power flow calculations to estimate voltages and currents in LV grids. This method relies on measurements from Medium Voltage/Low Voltage (MV/LV) transformers and cable distributors, and PMs are generated based on worst-case load current allocations, resulting in conservative estimates of bus voltages and power flows that do not consider the load models. Hence, the current paper proposes a methodology based on the training of detailed load models for state estimation purposes.

In one of the early works on DSSE described in [10], the authors apply the WLS method to solve the estimation problem using PMs based on industrial and residential consumption profiles. However, the model aggregates multiple buildings, reducing the spatial resolution significantly compared to the scenario analyzed here, where a building-level perspective in the context of extensive LV grids is considered. A more detailed investigation into consumption profiles for DSSE is found in [11], which explores both traditional Standard Load Profiles (SLPs) and profiles including renewable energy and thermal electrical heating systems, applying WLS for estimation. A general statistical model is used for devices such as PV, HPs, and EV chargers, without incorporating specific building or feeder measurement data, which is a notable limitation compared to the approaches presented in the current paper that are based on more precise models of the underlying components. However, unlike the PM profiles in [11] the approach proposed in this paper leverages real-time measurements by first disaggregating the power flows into the feeder. This allows the proposed methods to refine PM estimates and outperform profile-based approaches.

In [12], a method is proposed that assumes that a large amount of SM data from measured nodes can be used to extrapolate PMs for unmeasured nodes, contingent upon the readily available and extensive representative measurements, which is not yet accessible to DSOs with currently deployed first generation SMs. The Augmented Matrix Method (AMM), discussed in [13], solves a DSSE problem by using a PV model to generate PMs based on weather data and known module parameters, but it lacks a detailed model for other devices such as HPs and EV chargers, and assumes that residential consumption profiles can be sufficiently estimated from a limited number of measured profiles. The limitations of SLP-based PM generation for LV grids are analyzed in [14], where a method based on Markov Chains (MCs) is proposed for PM generation, with the WLS approach used to solve the observable system. Although this represents a considerable improvement over SLP-based approaches for PM generation, MC models have high data requirements. Furthermore, the inclusion of feeder measurements, Geographic Information System (GIS) data, and device-specific data in the training process as proposed here to improve the DSSE process is not considered in the existing literature.

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

In summary, all existing DSSE approaches exhibit one or more of the following shortcomings:

1) Need for a fully observable system, requiring that each LV node is equipped with a new generation SM [5], or a micro-Phasor Measurement Unit ($\mu$PMU) [15], which is a strong assumption that is rarely applicable to existing LV grids.
2) Heavy reliance on data that is not accessible to the DSO, as in the case of DSSE techniques based on Neural Networks (NNs) trained on unmeasured data [16], [17], which is only available in simulations.
3) The lack of use of other relevant data sources available to the DSO, such as GIS data, exogenous data like weather measurements and forecasts, and technical information about installed devices.

Therefore, this paper proposes a methodology that allows DSSE in LV distribution systems with a minimum amount of measurements, such as only one measurement at the transformer. To achieve this, a novel PM generation method based on Machine Learning (ML) techniques is presented, which allows the measured time series at the feeder to be decomposed into their respective components. Based on the latter, methods for DSSE are presented and compared in detail, investigating how an increasing penetration of SMs influences the results obtained from the proposed DSSE procedure. Furthermore, a method for estimating intervals of states is also proposed. Thus, the main contributions of the paper can be summarized as follows:

- Novel methods are proposed that allow to disaggregate transformer measurements into the underlying components for thermoelectric demand, baseload, and PV generation. These methods are based on ML algorithms developed for supervised and unsupervised time series disaggregation [18], [19], which are adapted here for the task of PM generation that is a central component of DSSE.
- A new approach for PM generation based on aggregated signals is proposed and compared to a state-of-the-art method for PM generation in LV grids, referred to as the Uniform Pseudo-Measurement (UPM) method, which is used in [4] and [9].
- The effect on DSSE of using PM, based on disaggregated signals and intervals, is investigated and compared to classical DSSE methods, for a realistic LV grid in Germany.

The rest of the paper is structured as follows: In Section II, the modeling of LV grids is reviewed, and an overview of typical measurement equipment in these grids is provided. A novel procedure for PM generation using an ML-based disaggregation algorithm is proposed in Section III, together with estimation methods for DSSE. The results of the presented methods for a realistic German LV grid are discussed in Section IV. Finally, in Section V the main conclusions of the presented work are highlighted, and a potential outlook is discussed.

## II. Models and Measurements Overview

### A. LV Grid Modeling

LV grids can be modeled as graphs connected to the MV grid through MV/LV transformers, and may have either radial or meshed topologies that can be subdivided into individual feeders that distribute power to various buildings in each feeder. The feeders can be further divided using cable distributors, which are nodes operated by the DSO and can be used to create arbitrary grid topologies. Fig. 1 illustrates a typical German LV grid, where a single MV/LV transformer can supply multiple buildings. In the example grid shown, the MV/LV transformer supplies 20 buildings through two cable distributors.

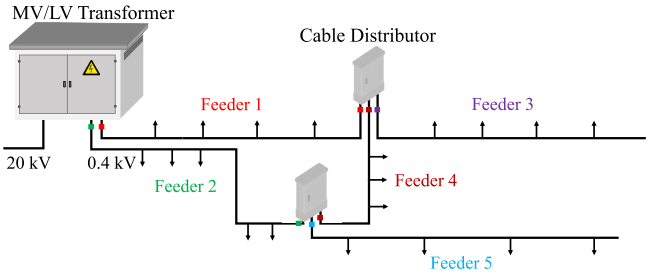



Fig. 1. Schematic of a 400 V LV grid with a 20 kV/400 V transformer supplying multiple buildings through two cable distributors and five feeders.

TABLE I
AVAILABLE GIS OBJECTS AND PARAMETERS

| GIS object | Parameters |
|---|---|
| Cable / Line | Maximum Current<br>Admittance |
| Transformer | Rated power<br>Admittance |
| Heat Pump | ID of household<br>Rated power |
| PV generator | ID of household<br>Rated power<br>Orientation |
| Consumer data | ID of household<br>Yearly energy consumption |

Using data from the DSO's GIS, which typically contains the data depicted in Table I, an electrical model can be derived for all feeders by modeling each section of each feeder using an equivalent $\pi$-circuit and choosing the respective parameters based on the type and length of each cable. Based on these models, DSSE can be applied to estimate all nodal voltages in the LV grid.

### B. Measurements in Distribution Grids

For DSSE, three typical types of measurements are considered in the LV grid:

- *Feeder Measurements:* These are physical measurements at the MV/LV transformer or LV cable distributors. These measurements include voltage magnitudes, active and reactive power flows flowing into the feeder, as well as currents.
- *SM Measurements:* These physical measurements are recorded at individual buildings and contain the voltage magnitude, as well as active and reactive power injections at the respective node in the LV grid.
- *PMs:* These are virtual estimates of active and reactive power injections at nodes that are not equipped with a

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

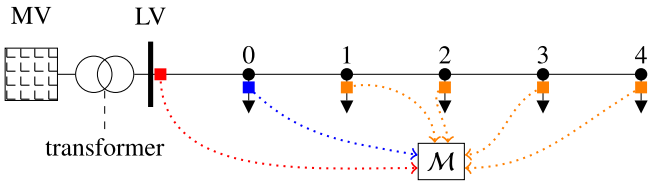



Fig. 2. Exemplary LV feeder measurements, where ■ depicts a feeder measurement, ■ corresponds to an SM measurement, and ■ illustrates a PM.

physical measurement device to ensure the LV grid is observable.

While the DSOs can install measurement devices at the MV/LV transformers and LV cable distributors, measurements from SMs in buildings may be sparsely distributed across the LV grid, which is the case of the smart meter rollout in various countries, particularly Germany. Consequently, varying levels of SM penetration must be considered in DSSE studies. In the case of a low ratio of measuring devices to grid nodes in a LV grid, the introduction of virtual PMs plays a crucial role to have an observable system that allows the application of estimation algorithms. Since the quality of PMs greatly affects the accuracy of the DSSE results, the generation of these synthetic measurements is discussed in detail in this paper.

Fig. 2 shows an example of a radial feeder that comprises all three types of these measurements considered here. In particular, the transformer is equipped with a feeder measurement, which measures the inflowing power into the feeder, as well as the voltage magnitude at the transformer. It is assumed that only one building is equipped with an SM, measuring both the voltage magnitude, and active and reactive power injections at this node. The rest of the buildings are considered not to have measurement equipment installed and thus must be estimated using PMs. All measurements are collected in the measurement set $\mathcal{M}$, which is used as input for the estimation algorithms discussed next.

## III. Distribution System State Estimation

### A. State Estimation Algorithm

The proposed DSSE algorithm operates under the assumption that a graphical model of the LV grid is available and that its topology and parameters remain static throughout the simulation period. This graphical model is represented as a graph $\mathcal{G} := \{\mathcal{N}, \mathcal{E}\}$, where $\mathcal{N} = \{1, 2, \ldots, n\}$ denotes the set of the $n$ nodes connected by $m$ edges $(k, l)$, with $k, l \in \mathcal{N}$ form the edge set $\mathcal{E} = \{(k_1, l_1), (k_2, l_2), \ldots, (k_m, l_m)\}$. Each edge corresponds to a transformer, switch, overhead line, or cable, characterized by its conductance $g$ and susceptance $b$, which are assumed known, depending on the length and type of the respective edge element. Given the graphical model $\mathcal{G}$ and the measurement set $\mathcal{M}$, the algorithm solves an estimation problem at each discrete time step $t$, with a chosen temporal resolution $\Delta t$. Thus, the DSSE algorithm solves a sequence of estimation problems over the simulation period $t_{\max}$.

The proposed DSSE procedure, illustrated in Fig. 3, can be divided into six steps:

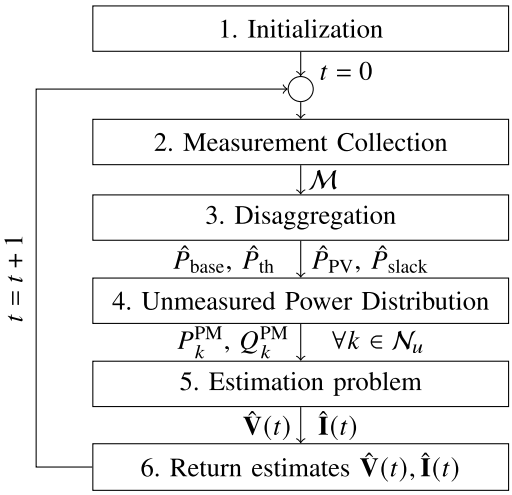



Fig. 3. Flowchart of the proposed multi-timestep DSSE algorithm.

1) *Initialization:* The graphical model $\mathcal{G}$ of the LV grid is parameterized and the arising types of measurements needed are identified at $t = 0$.
2) *Measurement Collection:* Measurements are collected into the measurement set $\mathcal{M}$, which includes physical measurements from feeders and SMs, as discussed in Section II-B.
3) *Disaggregation:* A novel Grid Load Disaggregation (GLD) algorithm is applied to disaggregate the feeder measured power $P_{\text{feed}}$ into PV power $\hat{P}_{\text{PV}}$, thermo-electrical demand $\hat{P}_{\text{th}}$, electrical baseload $\hat{P}_{\text{base}}$, and unassigned powers $\hat{P}_{\text{slack}}$ and $\hat{Q}_{\text{slack}}$ in the disagregation process, facilitating precise PM allocation, as discussed in Section III-B.
4) *Unmeasured Power Distribution:* The disaggregated feeder measurement components are used to distribute PM power injections $P_k^{\text{PM}}$ and $Q_k^{\text{PM}}$ across unmeasured nodes $k$ using various distribution algorithms presented in Section III-C.
5) *Estimation Problem:* The fully observable system, comprising physical measurements, PMs, and the graphical model $\mathcal{G}$, is used to solve the estimation problem discussed in Section III-D.
6) *Return Estimates:* The algorithm yields estimates of nodal voltage magnitudes $\hat{\mathbf{V}}(t)$ for all nodes, and estimates of line/transformer current magnitudes $\hat{\mathbf{I}}(t)$ flowing over the edges of the distribution grid graph for time step $t$. The algorithm then proceeds with the next estimation for $t + 1$.

By executing this algorithm, the vectors $\hat{\mathbf{V}} = [\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_n]$ and $\hat{\mathbf{I}} = [\hat{I}_1, \hat{I}_2, \ldots, \hat{I}_m]$ with voltage and current estimates at each simulated time step. These estimates are compared here against the simulated voltages $\mathbf{V}$ and currents $\mathbf{I}$ to assess the accuracy of the proposed methods, as well as comparing them with the existing state-of-the-art DSSE methodologies.

### B. Measurement Disaggregation

The objective of the proposed GLD approach is to utilize feeder measurements, which include the combined load and

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

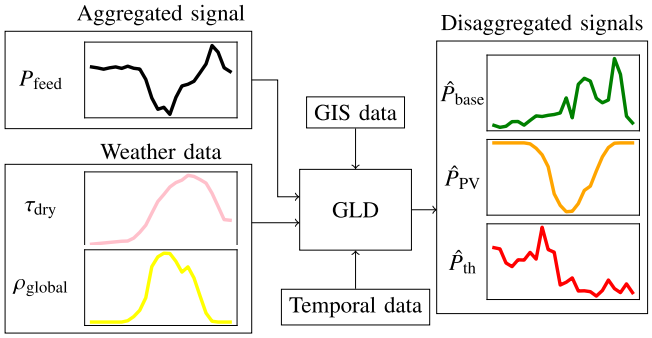



Fig. 4. Schematic of the proposed GLD methodology for disaggregation of LV grid measurements.

generation profiles from all downstream nodes, and disaggregate them into individual components associated with specific nodes in the LV grid based on the devices installed in these buildings. Fig. 4 illustrates the proposed GLD methodology, where the input data consists of the measured aggregated signal, encompassing the active and reactive power measurements at the feeder. Additionally, time series of dry-bulb temperatures $\tau_{\text{dry}}$ and global irradiance $\rho_{\text{global}}$ are fed into the GLD algorithm. Note that the DSO knows the installed devices and their corresponding grid nodes through grid connection requests and annual energy billing, which can be integrated into the GLD algorithm via GIS data. Furthermore, temporal data, which includes the current time step, time of day, time of year, and type of day, are provided to the algorithm. With these inputs, the GLD algorithm disaggregates the measured aggregated signal into three components: estimated electrical baseload power $\hat{P}_{\text{base}}$, representing the power consumption of appliances in all connected buildings; estimated power demand of thermoelectric devices $\hat{P}_{\text{th}}$, which accounts for heating and warm water usage; and estimated PV power $\hat{P}_{\text{PV}}$ generated by the solar panels installed on the buildings.

*1) Power Balance:* The power balance equation for active power in the LV grid can be expressed as follows:

$$P_{\text{feed}}(t) - P_{\text{th}}(t) - P_{\text{base}}(t) + P_{\text{PV}}(t) - P_{\text{loss}}(t) = 0 \quad (1)$$

where $P_{\text{feed}}(t)$ represents the total power measured at time step $t$ for all feeders supplying the LV grid, while $P_{\text{th}}(t)$ denotes the total power consumed by thermoelectric devices, such as HPs. The total baseload in the LV grid is captured by the variable $P_{\text{base}}(t)$, PV generation is represented by $P_{\text{PV}}(t)$, and the active power losses are modeled by the variable $P_{\text{loss}}(t)$. A similar balance equation can be defined for reactive power.

It is important to note that (1) must hold at every time step, assuming that all exchanged power is measured, and there are no additional power-generating or consuming components in the LV grid. Furthermore, the terms in (1) represent the aggregated active power of individual components, which can be attributed to the devices installed in the grid as follows:

$$P_{\text{feed}}(t) = \sum_{k \in \mathcal{F}} P_{\text{feed},k}(t) \quad (2)$$

$$P_{\text{th}}(t) = \sum_{k \in \mathcal{D}_{\text{th}}} P_{\text{th},k}(t) \quad (3)$$

$$P_{\text{PV}}(t) = \sum_{k \in \mathcal{D}_{\text{PV}}} P_{\text{PV},k}(t) \quad (4)$$

$$P_{\text{base}}(t) = \sum_{k \in \mathcal{D}_{\text{base}}} P_{\text{base},k}(t) \quad (5)$$

where $\mathcal{F}$ is the set of all feeders connected to the LV side of the transformers supplying the LV grid, typically consisting of one feeder in the common radial feeder configurations found across Europe. The sets $\mathcal{D}_{\text{th}}$, $\mathcal{D}_{\text{PV}}$, and $\mathcal{D}_{\text{base}}$ include all building equipment of the respective types: thermal, PV, and baseload.

*2) Ensemble Model:* The proposed GLD models are designed to learn representations for the various components in (1) using the input data outlined in Fig. 4. This results in the following Ensemble Model (EM), which is a combination of multiple sub-models:

$$\mathbf{y}_{\text{agg}} = f_{\mathbf{w}}^{\text{agg}}(\mathbf{X}_{\text{PV}}, \mathbf{X}_{\text{th}}, \mathbf{X}_{\text{base}}) \quad (6)$$

$$= \underbrace{f_{\mathbf{w}_{\text{PV}}}(\mathbf{X}_{\text{PV}})}_{\mathbf{y}_{\text{PV}}} + \underbrace{f_{\mathbf{w}_{\text{th}}}(\mathbf{X}_{\text{th}})}_{\mathbf{y}_{\text{th}}} + \underbrace{f_{\mathbf{w}_{\text{base}}}(\mathbf{X}_{\text{base}})}_{\mathbf{y}_{\text{base}}} \quad (7)$$

where $\mathbf{y}_{\text{agg}} \in \mathbb{R}^{t_{\max} \times 1}$ represents a vector of the aggregated signal, i.e., the total power measured in the LV grid over a time period $t_{\max}$. The goal of the proposed GLD algorithm is to learn a model $f_{\mathbf{w}}^{\text{agg}}$ through a regression task that closely matches the observed data, using input data matrices $\mathbf{X}_{\text{PV}}$, $\mathbf{X}_{\text{th}}$, and $\mathbf{X}_{\text{base}}$, which contain various temporal and weather-related features for the three components. The weight vector $\mathbf{w} = \{\mathbf{w}_{\text{PV}}, \mathbf{w}_{\text{th}}, \mathbf{w}_{\text{base}}\}$ contains the trainable parameters. The aggregated model is composed of an ensemble of three individual sub-models $f_{w_i}$ corresponding to a component signal that needs to be disaggregated. Each component in the EM receives its own input matrix $\mathbf{X}_i$ and has a unique function approximator $f$ with weights $\mathbf{w}_i$. Although the disaggregated signals $\mathbf{y}_i$ are initially unknown, once the model is trained, the sub-models can be used to compute the disaggregated signals $\mathbf{y}_i$ by passing the respective input data through the function approximator of the sub-model.

*3) Ensemble Model Training:* To train the EM defined in (7), the data flow illustrated in Fig. 5 is applied. The input consists of three matrices $\mathbf{X}_{\text{base}}$, $\mathbf{X}_{\text{th}}$, and $\mathbf{X}_{\text{PV}}$, which contain historical exogenous variables over the period $t_{\max}$, such as temporal features and weather data. The number of features in these matrices $n_{\text{feats},i}$ may vary depending on the chosen method and the selected sub-models. Additionally, the aggregated measured power profile $\mathbf{y}_{\text{agg}}$ serves as the target variable and contains the measured power for each historical time step in $t_{\max}$.

The objective of the EM is to estimate a profile $\hat{\mathbf{y}}_{\text{agg}}$, which is the sum of the estimated vectors $\hat{\mathbf{y}}_{\text{base}}$, $\hat{\mathbf{y}}_{\text{th}}$, and $\hat{\mathbf{y}}_{\text{PV}}$. The prediction quality is then evaluated using a loss function `loss`, such as the Mean Squared Error (MSE) or Mean Absolute Error (MAE) of $\hat{\mathbf{y}}_{\text{agg}}$ with respect to $\mathbf{y}_{\text{agg}}$, allowing for the optimization of sub-model weights based on the Stochastic Gradient Descent (SGD) approach used here, which requires the gradients $\frac{\partial \mathbf{w}_{\text{base}}}{\partial \texttt{loss}}$, $\frac{\partial \mathbf{w}_{\text{th}}}{\partial \texttt{loss}}$, and $\frac{\partial \mathbf{w}_{\text{PV}}}{\partial \texttt{loss}}$.

It is important to note that the primary objective of the EM is not to obtain $\hat{\mathbf{y}}_{\text{agg}}$ itself, but rather the individual components $\hat{\mathbf{y}}_{\text{base}}$, $\hat{\mathbf{y}}_{\text{th}}$, and $\hat{\mathbf{y}}_{\text{PV}}$. However, since these components are

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

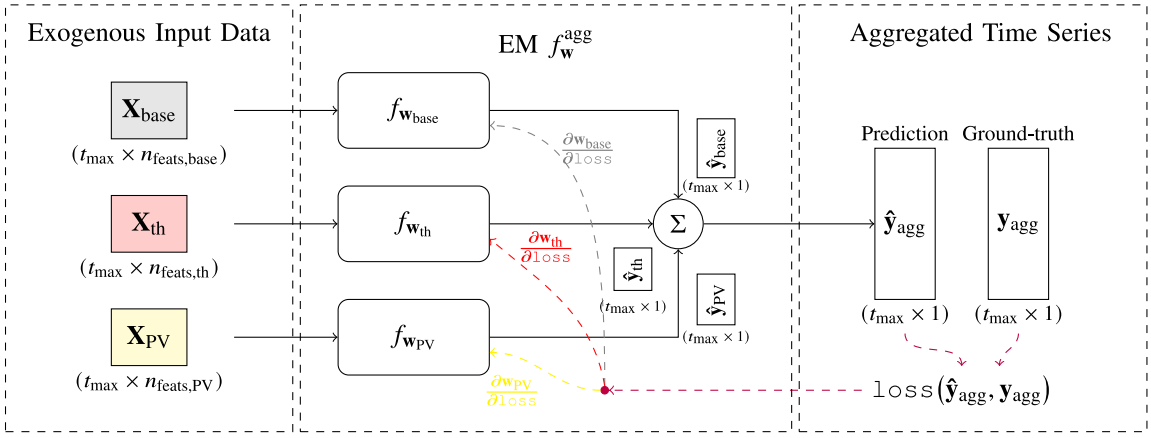



Fig. 5. Data flow in the EM training for GLD.

not directly measurable, training must be conducted indirectly through $\mathbf{y}_{\text{agg}}$, making the proposed approach semi-supervised. To ensure that the learned sub-models $f_{\mathbf{w}_{\text{base}}}, f_{\mathbf{w}_{\text{th}}}, f_{\mathbf{w}_{\text{PV}}}$ provide meaningful representations of the unobservable components, the following strategies can be applied when selecting function approximators for these sub-models:

- *Input Data Constraints:* The input data provided to each sub-model via the matrices $\mathbf{X}_i$ can be constrained to ensure that each sub-model learns only its respective component of the signal.
- *Model Constraints:* The function class of the approximators can be restricted to align with the true underlying models. This can be achieved by limiting sub-models to linear or piecewise-linear functions or by incorporating activation functions that constrain their output.

In general, the sub-models within the EM can be implemented using any parameterizable function approximator, such as Neural Networks (NNs), Regression Trees (RT), or Linear Regression (LR) models. The training procedure used here follows the algorithm presented in Algorithm 1. Thus, based on pre-defined hyperparameters, such as the number of training epochs, learning rate, and batch size, the iterative training loop aims at fitting the weights in all sub-models of the EM. This is achieved by applying the following steps to each batch of each epoch:

1) Extract the data of the current batch.
2) Use the input data of the batch for a forward pass through the EM.
3) Calculate the loss with respect to the true output.
4) Update the weights of the sub-models in the EM using SGD.

Note that in the special case where all sub-models in the EM can be represented using a linear equation, the training procedure can be simplified as follows:

$$\mathbf{w}_{\text{EM}} = \mathbf{X}^{\dagger}\, \mathbf{y}_{\text{agg}} \tag{8}$$

where the $\cdot^{\dagger}$ operator calculates the pseudo-inverse of a matrix and $\mathbf{X} \in \mathbb{R}^{(t_{\max} \times n_{\text{feats,base}} + n_{\text{feats,th}} + n_{\text{feats,PV}})}$ is the following sparse block diagonal matrix containing the features of all sub-models:

**Algorithm 1** Training Procedure for EM for GLD

**Inputs:**
$\mathcal{X} = \{\mathbf{X}_k \ \forall k \in n\}$: Input datasets for all sub-models.
$\mathcal{Y} = \{\mathbf{y}_k \ \forall k \in n\}$: Output dataset containing aggregated signal.
$f_{\mathbf{w}_k}$: Regression models for sub-models.
$\text{n}_{\text{epochs}}$: Number of training epochs.
$\alpha$: Learning rate.
batch_size: Batch size for mini-batches.
**Returns:**
$f_{\mathbf{w}}^{\text{agg}}$: Trained EM.
$f_{\mathbf{w}_k}$ $\forall k \in n$: Trained sub-models.

**for** epoch $\in \{1, \ldots, \text{n}_{\text{epochs}}\}$ **do**
  **for** batch $\in \{\mathcal{Y}, \mathcal{X}\}$ **do**
    $y, x \leftarrow$ batch — Extract data of batch.
    $\hat{y} \leftarrow f_{\mathbf{w}}^{\text{agg}}(x)$ — Forward pass.
    $l \leftarrow \texttt{loss}(\hat{y}, y)$ — Calculate loss.
    $\mathbf{w} \leftarrow \mathbf{w} + \alpha \cdot \nabla l$ — SGD step.
  **end for**
**end for**

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{\text{base}} & & \\ & \mathbf{X}_{\text{th}} & \\ & & \mathbf{X}_{\text{PV}} \end{bmatrix} \tag{9}$$

Observe that the proposed EM is highly flexible and supports various regression methods; hence, a limited selection of possible sub-models was used in the paper. The chosen sub-models used here are based on physical models of the underlying components, as well as a nonlinear regression approach known for its strong performance in time series predictions. The next sections provide an overview of the proposed sub-models for the EM used in the GLD task.

*4) Baseload Demand Sub-Models:* The following two sub-models for the baseload model are used and compared here:

$$\begin{cases} \hat{\mathbf{P}}_{\text{base}} = w_{\text{base}}\, \mathbf{x}_{\text{SLP}} & \text{(10a)} \\ w_{\text{base}}\, \mathbf{x}_{\text{reg,base}} & \text{(10b)} \end{cases}$$

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

where (10a) is based on a vector $\mathbf{x}_{\text{SLP}}$ that incorporates an SLP, as defined in [11], which is generated by using a methodology employed by DSOs in Germany [20], [21]. On the other hand, sub-model (10b) provides greater flexibility and allows for a decoupling between the input data used for the regression method to generate $\mathbf{x}_{\text{reg,base}}$, and the input data for the EM approach. The scalar weight $w_{\text{base}}$ in (10a) and (10b) is then learned within the EM dependent on the sub-model used, scaling the profiles to best fit the aggregated signal.

*5) Thermal Demand Sub-Model:* For the thermal device model, the following matrix containing the dry-bulb temperatures is defined first:

$$\mathbf{X}_\tau = \begin{bmatrix} \tau_{\text{dry}}(t) & \tau_{\text{dry}}(t-1) & \dots & \tau_{\text{dry}}(t-L) \\ \tau_{\text{dry}}(t+1) & \tau_{\text{dry}}(t) & \dots & \tau_{\text{dry}}(t-L-1) \\ \vdots & \vdots & \vdots & \vdots \\ \tau_{\text{dry}}(t_{\max}) & \tau_{\text{dry}}(t_{\max}-1) & \dots & \tau_{\text{dry}}(t_{\max}-L) \end{bmatrix} \tag{11}$$

where $L$ represents the number of lagged time steps considered, which enables modeling the thermal inertia of the buildings, and $\tau_{\text{dry}}(t)$ is the dry-bulb temperature measured at time step $t$. A simple linear sub-model can then be formulated as:

$$\hat{\mathbf{P}}_{\text{th}} = \left[\mathbf{X}_\tau \ \mathbf{1}_{t_{\max}\times 1}\right] \mathbf{w}_{\text{th}} \tag{12}$$

This linear model uses the lagged temperatures and an intercept as features, learning a weight vector $\mathbf{w}_{\text{th}}$ during the training phase.

A second model that is able to better capture the piecewise linear behavior often observed in space heating and warm water consumption extends (12) by introducing a piece-wise linear sub-model as follows:

$$\hat{\mathbf{P}}_{\text{th}} = \underbrace{\begin{bmatrix} \mathbf{X}_\tau^{\text{sh}} & \mathbf{0}_{n_{\text{sh}}\times L+1} & \mathbf{1}_{n_{\text{sh}}\times 1} & \mathbf{0}_{n_{\text{sh}}\times 1} \\ \mathbf{0}_{n_{\text{ww}}\times L+1} & \mathbf{X}_\tau^{\text{ww}} & \mathbf{0}_{n_{\text{ww}}\times 1} & \mathbf{1}_{n_{\text{ww}}\times 1} \end{bmatrix}}_{\mathbf{x}_{\text{th}}} \mathbf{w}_{\text{th}} \tag{13}$$

where $\mathbf{X}_\tau^{\text{sh}}$ and $\mathbf{X}_\tau^{\text{ww}}$ are used to capture the behavior in the space heating and warm water ranges, respectively. These matrices are derived by splitting the input data $\mathbf{X}_\tau$ into two parts based on a cutoff temperature. The learnable weight vector $\mathbf{w}_{\text{th}} = [a^{\text{sh}}\ a^{\text{ww}}\ b^{\text{sh}}\ b^{\text{ww}}]^\intercal$ contains the slopes $a$ and intercepts $b$ of the two linear functions.

Similar to the regression model described in (10b), the following regression model can also be utilized to generate input data for the thermal component in the EM:

$$\hat{\mathbf{P}}_{\text{th}} = w_{\text{th}}\ \mathbf{x}_{\text{reg,th}} \tag{14}$$

This is a simpler linear sub-model with a feature vector $\mathbf{x}_{\text{reg,th}}$ and a scalar weight $w_{\text{th}}$.

*6) PV Sub-Model:* Two distinct sub-models are utilized here for predicting PV generation. The first one employs LR, relying on the measured global irradiance as follows:

$$\hat{\mathbf{P}}_{\text{PV}} = w_{\text{PV}}\ \rho_{\text{global}} \tag{15}$$

where $\rho_{\text{global}}$ is a vector containing global irradiance measurements for all time steps $t_{\max}$, and the scalar weight $w_{\text{PV}}$ is learned during SGD or eXtreme Gradient Boosting (XGB) training, with no intercept included.

TABLE II
GLD MODELS CONSIDERED FOR PM GENERATION

| GLD Method | | Baseload Model | Thermal Model | PV Model |
|---|---|---|---|---|
| GLD1 | Sub-Model | (10a) | (13) | (15) |
| | Data | SLP | $\tau_{\text{dry}}$ | $\rho_{\text{global}}$ |
| GLD2 | Sub-Model | (10a) | (12) | (15) |
| | Data | SLP | $\tau_{\text{dry}}$ | $\rho_{\text{global}}$ |
| GLD3 | Sub-Model | (10b) | (14) | (16) |
| | Data | XGB | XGB | XGB |

Another investigated PV sub-model also based on regression techniques as in (10b) and (14) can be formulated as follows:

$$\hat{\mathbf{P}}_{\text{PV}} = w_{\text{PV}}\ \mathbf{x}_{\text{reg,PV}} \tag{16}$$

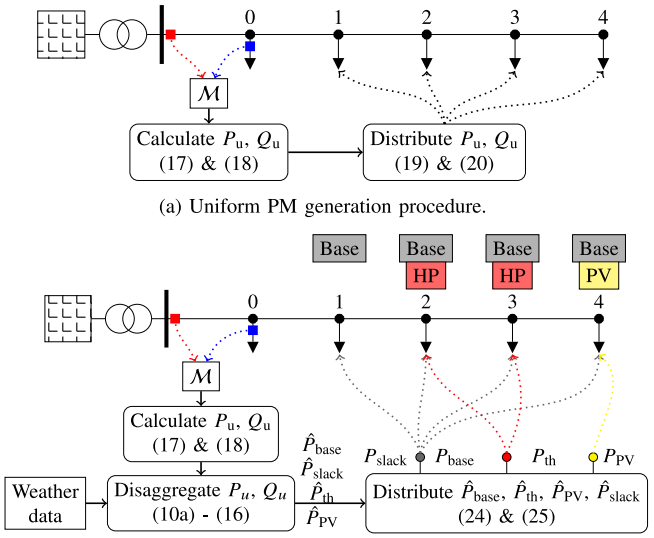
where $\mathbf{x}_{\text{reg,PV}}$ is a vector predicted by an external regression method, and the weight $w_{\text{PV}}$, which scales the prediction of the regression model, is learned during the training of the EM.

*7) Ensemble Model Configurations:* To evaluate the effectiveness of the proposed GLD method in comparison to traditional approaches for PM generation, three promising GLD configurations are analyzed, and their outcomes are compared to the current state-of-the-art PM generation method. Specifically, the three configurations presented in Table II, referred to as GLD1, GLD2, and GLD3, which vary in both the models employed and the input data required. Thus, the first two methods consider all possible modeling approaches requiring limited amount of data, while the third one regards models with high data requirements. In particular, the following input data is required for the three proposed approaches:

- *GLD1 and GLD2:* The input data, as per Fig. 5, consists of the SLP feature in (10a) with $n_{\text{feats,base}} = 1$, and the global irradiance feature in (15) with $n_{\text{feats,PV}} = 1$. GLD1 utilizes the piecewise-linear thermal features in (13) with $n_{\text{feats,th}} = 2 \cdot (L+1) + 2$ features, and GLD2 is based on the linear model (12) with $n_{\text{feats,th}} = L + 2$.
- *GLD3:* The input data for GLD3 consists of three vectors for each component, i.e., $n_{\text{feats,base}} = n_{\text{feats,PV}} = n_{\text{feats,th}} = 1$, representing the output vector of each XGB regression method $\mathbf{x}_{\text{reg,base}}$, $\mathbf{x}_{\text{reg,th}}$, and $\mathbf{x}_{\text{reg,PV}}$ that is pre-trained on the respective component.

Both GLD1 and GLD2 utilize linear models for baseload and PV generation. However, GLD1 incorporates a piecewise-linear model for thermal demand, whereas GLD2 applies a linear thermal device model. In contrast, GLD3 relies on input data generated by a different regression method and learns the weights of the corresponding outputs. Consequently, GLD3 requires the training of regression models on data accessible to the DSO, such as SM recorded data from a subset of buildings within the relevant LV grid. The distinct input data used for all three models ensures that each component is learned independently, preventing spillover between models. The eXtreme Gradient Boosting (XGB) regression method in [22] is employed for predicting the underlying models due to its characteristics as a fast and high-performing regression technique for time series forecasting in the power sector [23].

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.




(a) Uniform PM generation procedure.

(b) Grid Load Disaggregation PM generation procedure.

Fig. 6. PM generation with (a) UPM and (b) GLD methods for a sample 5-node radial feeder, with feeder measurements ■, and SM measurements ■.

### C. Pseudo-Measurement Generation

The generation of synthetic PM is essential in DSSE to ensure the observability of the system when combined with physical measurements, making it a crucial element of any DSSE algorithm. Fig. 6 illustrates the two types of methods compared in this paper.

*1) Uniform Pseudo-Measurement Method:* This method [4], [9] depicted in Fig. 6d, represents the current state-of-the-art approach in the literature. In this case, the unassigned power is first calculated as follows:

$$P_u(t) = \sum_{k\in\mathcal{F}} P_{\text{feeder},k}(t) - \sum_{k\in\mathcal{N}_m} P_k(t) \tag{17}$$

$$Q_u(t) = \sum_{k\in\mathcal{F}} Q_{\text{feeder},k}(t) - \sum_{k\in\mathcal{N}_m} Q_k(t) \tag{18}$$

where $P_u(t)$ and $Q_u(t)$ represent the unassigned active and reactive powers, respectively. These powers are determined by subtracting the power injection using all measured nodes, which are collected in the set $\mathcal{N}_m$, from the total measured power of the LV feeder. The remaining power injection accounts for the power consumed at unmeasured nodes, as well as any additional losses within the feeder. The unmeasured powers are then allocated across all unmeasured nodes as follows:

$$P_k^{\text{PM}}(t) = f_k P_u(t) \quad \forall k \in \mathcal{N}_u \tag{19}$$

$$Q_k^{\text{PM}}(t) = f_k Q_u(t) \quad \forall k \in \mathcal{N}_u \tag{20}$$

where $P_k^{\text{PM}}$ and $Q_k^{\text{PM}}$ represent the injection PMs for all unmeasured nodes, which are included in the set $\mathcal{N}_u$. The factors $f_k$ distribute a fraction of the unassigned power to the corresponding nodes, where the total sum is typically close to one when accounting for losses.

Various methods for determining $f_k$ can be employed, depending on the available data. Table III outlines different common distribution techniques based on GIS and billing

TABLE III
DISTRIBUTION METHODS FOR PM GENERATION

| Distribution Method | Formula $f_k$ | Input Data |
|---|---|---|
| Equal scaling | $\frac{P_u}{N_{\text{bldgs}}}$ | Number of buildings $N_{\text{bldgs}}$ |
| Rated Power scaling | $\frac{P_u P_k^{\text{rated}}}{\sum_k^{N_u} P_k^{\text{rated}}}$ | Rated power of buildings $P_k^{\text{rated}}$ |
| Energy scaling | $\frac{P_u E_k^{\text{year}}}{\sum_k^{N_u} E_k^{\text{year}}}$ | Energy of buildings $E_k^{\text{year}}$ |

data available to the DSO using the following three methods considered here: equal scaling, which assumes the same factor for all unmeasured buildings; rated power scaling, which considers the rated power of the loads in each building $P_k^{\text{rated}}$; and energy scaling, which is based on the annual energy consumption of each building $E_k^{\text{yearly}}$.

*2) Grid Load Disaggregation Pseudo-Measurements:* The PM generation method proposed here enhances the UPM method by incorporating a disaggregation step and modifying the distribution process, as depicted in Fig. 6b. Initially, the unassigned power is calculated using the formulas in (17) and (18). This is followed by a disaggregation step utilizing the methods outlined in Section III-B. The trained sub-models are then used to predict the power injections at the current time step $t$, resulting in the estimates $\hat{P}_{\text{th}}(t)$, $\hat{Q}_{\text{th}}(t)$, $\hat{P}_{\text{base}}(t)$, $\hat{Q}_{\text{base}}(t)$, $\hat{P}_{\text{PV}}(t)$, and $\hat{Q}_{\text{PV}}(t)$. The active and reactive slack power not accounted for by the sub-models is then calculated as follows:

$$\hat{P}_{\text{slack}}(t) = P_u(t) - \hat{P}_{\text{th}}(t) - \hat{P}_{\text{base}}(t) + \hat{P}_{\text{PV}}(t) - \hat{P}_{\text{loss}}(t) \tag{21}$$

$$\hat{Q}_{\text{slack}}(t) = Q_u(t) - \hat{Q}_{\text{th}}(t) - \hat{Q}_{\text{base}}(t) + \hat{Q}_{\text{PV}}(t) \tag{22}$$

where $\hat{P}_{\text{slack}}(t)$ and $\hat{Q}_{\text{slack}}(t)$ represent the residual between measured power and disaggregated signals.

Although the exact losses in the LV grid at a given time step are unknown, the fact that the losses can be approximated in practice by a quadratic relationship of the grid power flows, since node voltages vary in a narrow range, can be used to estimate these losses. Thus, assuming that the feeder parameters, such as, topology of the grid and the admittances of the lines, cables, and transformers, are known and fixed over the simulation period, the losses can be approximated with the following quadratic function:

$$\hat{P}_{\text{loss}}(t) \approx a_{\text{loss}} \sum_{k\in\mathcal{F}} P_{\text{feeder},k}^2(t) + b_{\text{loss}} \sum_{k\in\mathcal{F}} P_{\text{feeder},k} + c_{\text{loss}} \tag{23}$$

where $a_{\text{loss}}$, $b_{\text{loss}}$, and $c_{\text{loss}}$ are unknown parameters that can be obtained by running Monte-Carlo power flow simulations using different random distributions of the LV feeder power over all connected households, and considering a strict quadratic relationship, as expected in power grids, i.e., $b_{\text{loss}} = 0$ and $c_{\text{loss}} = 0$.

Once the components are computed, the PMs are allocated to the nodes where the respective device types are present, as identified from GIS data. This involves the following three steps, using one of the distribution methods from Table III:

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

1) *Baseload and Slack Distribution:* Distribute the sum of the estimated total baseload $\hat{P}_{\text{base}}$ and the total slack $\hat{P}_{\text{slack}}$ across all unmeasured buildings.
2) *Thermoelectric Power Distribution:* Distribute the total estimated demand $\hat{P}_{\text{th}}$ across all buildings from thermoelectric devices.
3) *PV Power Distribution:* Distribute the estimated PV power $\hat{P}_{\text{PV}}$ across all unmeasured buildings equipped with a PV generator.

The application of Steps 1 through 3, in any order, enables the calculation of PMs for each unmeasured node as follows:

$$P_k^{\text{PM}} = \hat{P}_{\text{base},k} + \hat{P}_{\text{slack},k} + \hat{P}_{\text{th},k} - \hat{P}_{\text{PV},k} \quad \forall k \in \mathcal{N}_u \tag{24}$$

$$Q_k^{\text{PM}} = \hat{Q}_{\text{base},k} + \hat{Q}_{\text{slack},k} + \hat{Q}_{\text{th},k} - \hat{Q}_{\text{PV},k} \quad \forall k \in \mathcal{N}_u \tag{25}$$

In these equations, the PM estimated powers $P_k^{\text{PM}}$ and $Q_k^{\text{PM}}$ for unmeasured nodes $k \in \mathcal{N}_u$ are zero for nodes where no equipment of the respective types are installed, which are assumed known to the DSO.

### D. Minimal Measurement Estimators

After generating PMs for all unmeasured nodes using the UPM approach described in (19) and (20), or by applying the proposed GLD models outlined in (24) and (25), the system becomes fully observable, enabling the estimation process to proceed. The measurements for the current time step are concatenated as follows:

$$\mathbf{z}_{\text{feed}} = \left[V_{\text{feed},k}\ P_{\text{feed},k}\ Q_{\text{feed},k} \quad \forall k \in \mathcal{F}\right] \tag{26}$$

$$\mathbf{z}_{\text{SM}} = \left[V_k\ P_k\ Q_k \quad \forall k \in \mathcal{N}_m\right] \tag{27}$$

$$\Rightarrow \mathbf{z}_{\text{phys}} = \left[\mathbf{z}_{\text{feed}}\ \mathbf{z}_{\text{SM}}\right] \tag{28}$$

where the measurements from all monitored feeders and SM in the LV grid are combined, including measurements of nodal voltage magnitudes, as well as active and reactive power flows. The following overall measurement vector is obtained by adding the generated PMs:

$$\mathbf{z}_{\text{PM}} = \left[P_k^{\text{PM}}\ Q_k^{\text{PM}} \quad \forall k \in \mathcal{N}_u\right] \tag{29}$$

$$\Rightarrow \mathbf{z} = \left[\mathbf{z}_{\text{phys}}\ \mathbf{z}_{\text{PM}}\right] \tag{30}$$

where the measurement vector $\mathbf{z} \in \mathbb{R}^{|\mathcal{M}|}$ is the concatenation of physical measurements $\mathbf{z}_{\text{phys}} \in \mathbb{R}^{|\mathcal{M}_{\text{phys}}|}$ for a set of actual measurements $\mathcal{M}_{\text{phys}}$ and virtual measurements $\mathbf{z}_{\text{PM}} \in \mathbb{R}^{|\mathcal{M}_{\text{PM}}|}$ for a set of PMs $\mathcal{M}_{\text{PM}}$. Note that the total number of measurements must be at least $2|\mathcal{N}| + 1$ to ensure that the system is observable and a solution can be computed. Hence, depending on the available feeder and SM measurements in $\mathcal{M}_{\text{phys}}$, PMs for active and reactive power must be added at all unmeasured nodes in $\mathcal{N}_u$ to meet this requirement. For example, for the feeder in Fig. 2, which consists of six nodes, five household nodes, and one slack node at the transformer, there must be at least $2 \times 6 + 1 = 13$ measurements. Hence, the physical measurements at the feeder and the SM at household 0 are supplemented by adding PMs at the households at nodes 1, 2, 3 and 4, resulting in a total of $\mathcal{M} = 14$ measurements in $\mathbf{z}$ thus making the system observable, as illustrated in Table IV.

TABLE IV
FIG. 2 MEASUREMENT SETS AND VARIABLES

| Measurement Type | $\mathcal{M}$ | Variables | No. Total |
|---|---|---|---|
| Feeder | $\{0\}$ | $V_{\text{feed},0}$, $P_{\text{feed},0}$, $Q_{\text{feed},0}$ | 3 |
| SM | $\{0\}$ | $V_0$, $P_0$, $Q_0$ | 3 |
| PM | $\{1, 2, 3, 4\}$ | $P_1^{\text{PM}}$, $Q_1^{\text{PM}}$, $P_2^{\text{PM}}$, $Q_2^{\text{PM}}$ $P_3^{\text{PM}}$, $Q_3^{\text{PM}}$, $P_4^{\text{PM}}$, $Q_4^{\text{PM}}$ | 8 |
| | | | $\lvert\mathcal{M}\rvert = 14$ |

An estimate of the state can then be obtained by solving the following optimization problem:

$$\min_{\tilde{\mathbf{V}}} \quad J\left(\tilde{\mathbf{V}}\right) = \mathbf{r}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{r} \tag{31}$$

$$\text{s.t.} \quad \mathbf{r} = \mathbf{z} - h\left(\tilde{\mathbf{V}}\right)$$

where the objective seeks to minimize the squared residuals in the residual vector $\mathbf{r}$, which represent the difference between the measurement vector $\mathbf{z}$, and the measurement functions $h$ applied to the state vector $\tilde{\mathbf{V}}$, which consists of the complex nodal voltages. The diagonal matrix $\mathbf{R}$ contains the inverse measurement variances, serving as weighting factors for the individual measurements. The measurement functions $h$ contain, depending on the type of measurement, linear and nonlinear functions describing the power flows in the grid. Consequently, an iterative method, such as the common WLS method is necessary to solve the SE optimization problem.

It should be noted that, although the solution of (31) only yields complex voltage values, the line/transformer currents flowing through the edges of the graph of the LV grid can then be readily obtained by solving the following equation:

$$\tilde{I}_i = \tilde{Y}_{kl}\left(\tilde{V}_k - \tilde{V}_l\right) \tag{32}$$

where $i \in \mathcal{E}$, and $\tilde{Y}_{kl}$ is the assumed known line/transformer admittance of the edge element $i$. Therefore, the proposed method determines limit violations in DSSE by calculating the magnitude of the complex LV voltages and currents, as highlighted in Fig. 3.

*1) Estimating State Intervals:* Solving the optimization problem (31) yields a Maximum Likelihood Estimate (MLE), which contains the phasor voltage estimates $\hat{\mathbf{V}}$ for all nodes in the system and represents the most probable state for the measured data, assuming the errors follow a normal distribution. However, in practical applications, these estimates can be inaccurate due to measurement errors and, more critically, errors in predicted PMs. Such inaccuracies may result in the failure to detect overloading conditions in the LV grid. Therefore, it is important to estimate the possible ranges of voltages in the grid.

To determine the voltage ranges, Steps 4 and 5 in Fig. 3 can be modified by considering an error in the predicted PMs as follows:

$$\hat{P}_k^{\text{PM}} = P_k^{\text{PM}} + pdf(0, \sigma_k)\ \forall k \in \mathcal{N}_u \tag{33}$$

where $\hat{P}_k^{\text{PM}}$ represents the simulated PM including an error, and *pdf* denotes a normal probability density function with zero mean and $\sigma_k$ standard deviation. This results in a Monte-Carlo approach where the uncertainty in the PM estimates

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

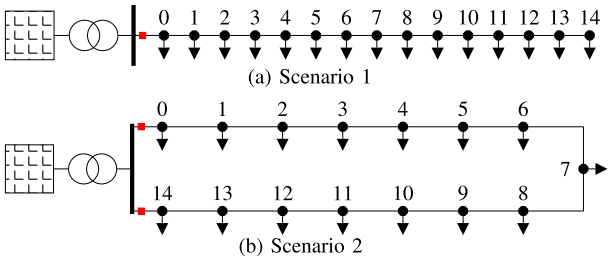



Fig. 7. Considered LV feeder scenarios.

is taken into account by randomly sampling around the estimated values, which leads to a new PM value at each node, resulting in deviating estimates of voltages and currents when solving (31) with the sampled values.

To estimate state intervals, the following steps are performed:

1) Compute point-estimates for the PMs $P_k^{\text{PM}}$ at all unmeasured nodes $k \in \mathcal{N}_u$.
2) Generate noisy estimates for the PMs using (33), where a predefined uncertainty serves as the standard deviation $\sigma_k$.
3) Solve the DSSE optimization problem described in (31) for the noisy estimates, yielding estimates for all grid nodes, lines, and transformers.

By repeating steps 2 and 3 for $n_{\text{samples}}$ times, the resulting voltage and current estimates can be expressed as a discrete distribution of possible grid states at a given time step. Since sampling the normal distribution in (33) and solving the optimization problem in (31) is computationally efficient, with execution times in the sub-second range, tens or even hundreds of samples for typical distribution LV feeder sizes can be quickly calculated.

## IV. Results

### A. Test System and Measurements

To assess the effectiveness of the proposed DSSE approach, two realistic LV feeders based on an actual LV grid in the Karlsruhe region, are studied. This represents a common grid configuration in Germany, where each transformer services about 100 residential buildings through multiple LV feeders, with 10-20 buildings per feeder. Thus, the proposed methodology can be readily extended to an entire LV grid by applying it in parallel to each feeder connected to the MV/LV transformer. Fig. 7 illustrates the two grid configuration considered here. Scenario 1 represents the single-phase representation of a radial LV feeder supplying 15 homes, with a nominal voltage of 400 V, and Scenario 2, the same homes are assumed to be supplied by a meshed grid topology. The 15 residential households are connected using NAYY 4x150 SE and NAYY 4x50 SE underground cables. Each home was simulated for a full year with a time resolution of $\Delta t = 1$ h.

The electrical profiles of the homes in Scenario 1 and 2 are simulated using the `BEopt` software [24], which models each home based on weather data specific to Germany. Each home is modeled according to the specifications in Table V, accounting for various factors such as heating technology, i.e., Air-Source Heat Pump (ASHP), Ground-Source Heat Pump (GSHP), or hybrid systems using both gas and electricity; PV capacity and orientation; home footprint for various insulation levels; and the number of occupants with different usage patterns. The variety of factors considered allow to represent the diversity of buildings typically connected to LV grids.

TABLE V
LV GRID HOUSE CHARACTERISTICS

| ID | HP | PV | Footprint | Occupants |
|---|---|---|---|---|
| 1 | - | - | 197 m$^2$ | 4 |
| 2 | ASHP | 10 kWp; S | 114 m$^2$ | 2 |
| 3 | ASHP | - | 223 m$^2$ | 5 |
| 4 | - | 8 kWp; SW | 134 m$^2$ | 3 |
| 5 | - | 6 kWp; S | 251 m$^2$ | 2 |
| 6 | GSHP | - | 130 m$^2$ | 3 |
| 7 | - | 10 kWp; SE | 201 m$^2$ | 1 |
| 8 | hybrid | - | 204 m$^2$ | 2 |
| 9 | - | - | 86 m$^2$ | 2 |
| 10 | ASHP | 10 kWp; S | 264 m$^2$ | 5 |
| 11 | hybrid | - | 187 m$^2$ | 2 |
| 12 | GSHP | 4 kWp; SE | 104 m$^2$ | 1 |
| 13 | hybrid | 4 kWp; SSW | 239 m$^2$ | 3 |
| 14 | hybrid | - | 197 m$^2$ | 6 |
| 15 | ASHP | 4.5 kWp; S | 174 m$^2$ | 2 |

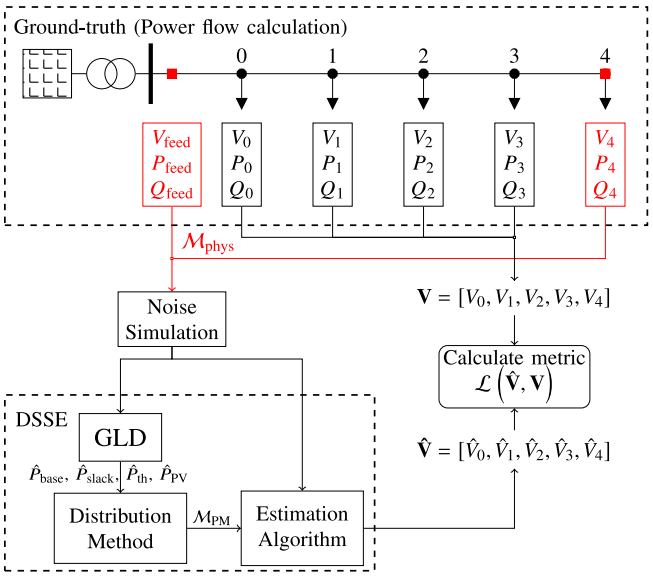



Fig. 8. Evaluation procedure for the accuracy of DSSE algorithms using power flow calculation as ground-truth, with assumed physical measurements marked as ■.

For each simulated time step, power flow calculations are performed using the `pandapower` package [25], which are considered as the ground-truth data for evaluation purposes, since actual measurements are not available. Based on this generated data, measurements are simulated by adding noise to the obtained quantities for nodes at which measurement equipment is assumed to be available in the investigated scenario. Fig. 8 illustrates the procedure used for the generation of ground-truth data for one of the feeders and for house ID 4. Thus, the assumed measured quantities are collected in $\mathcal{M}_{\text{phys}}$ and noise is added based on the specified accuracies for the measurement devices. The noisy measurements are then

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

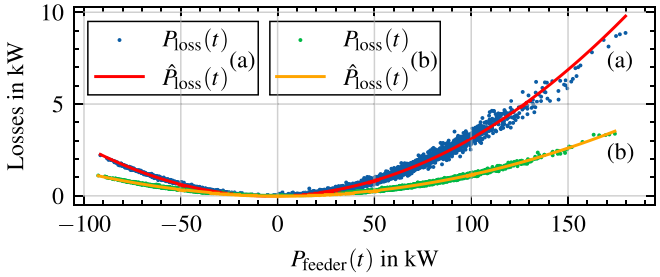



Fig. 9. Losses $P_{\text{loss}}(t)$ for the feeder in (a) Scenario 1 and (b) Scenario 2 with respect to feeder power flows for ground-truth calculations for one year of data and quadratic $\hat{P}_{\text{loss}}(t)$ as per (23).

passed onto the DSSE algorithm, where they are disaggregated and PMs are distributed based on the disaggregated components. The estimation algorithm receives the generated PMs $\mathcal{M}_{\text{PM}}$ and the noisy physical measurements and estimates the system state. Finally, a metric $\mathcal{L}$ is applied to compute the quality of the estimates with respect to the underlying ground-truth.

The computation of the fitted parameter $a_{\text{loss}}$ in (23) from ground-truth values for both LV scenarios is illustrated in Fig. 9. This yields $a_{\text{loss}} = 2.92 \times 10^{-4}$ for Scenario 1, and $a_{\text{loss}} = 1.19 \times 10^{-4}$ for Scenario 2.

Depending on the measurement setup being evaluated, PMs are generated for the DSSE procedure. It is assumed that physical measurements are subject to normally distributed errors, corresponding to an accuracy class of 0.5S, while PMs are assumed to have higher uncertainty, characterized by high standard deviations.

Three metrics are used to quantify the quality of PMs and DSSE estimates defined as follows:

$$\text{MAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{\sum_{t=1}^{t_{\max}} |y(t) - \hat{y(t)}|}{t_{\max}} \tag{34}$$

$$\text{MAPE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{100}{t_{\max}} \sum_{t=1}^{t_{\max}} \left| \frac{y(t) - \hat{y}(t)}{y(t)} \right| \tag{35}$$

$$\text{MARPE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{100}{t_{\max}} \sum_{t=1}^{t_{\max}} \left| \frac{y(t) - \hat{y}(t)}{y_{\text{rated}}} \right| \tag{36}$$

where the MAE calculates the mean error over of two vectors over all time steps $t_{\max}$. The Mean Absolute Percentage Error (MAPE) is used to calculate relative errors between two vectors. The metric Mean Absolute Rated Percentage Error (MARPE) normalizes the absolute error with respect to the rated power of the respective household $y_{\text{rated}}$ under consideration.

### B. Disaggregation and Pseudo-Measurements

The accuracy of the predicted PMs is crucial in DSSE. Therefore, the quality of the proposed disaggregation method and the predicted PMs errors are analyzed first.

Unlike the UPM method, the GLD approach involves a disaggregation step to estimate the signal components. Fig. 10 depicts the disaggregation results using GLD1, GLD2, and GLD3 for the first week of data. The upper plot shows the

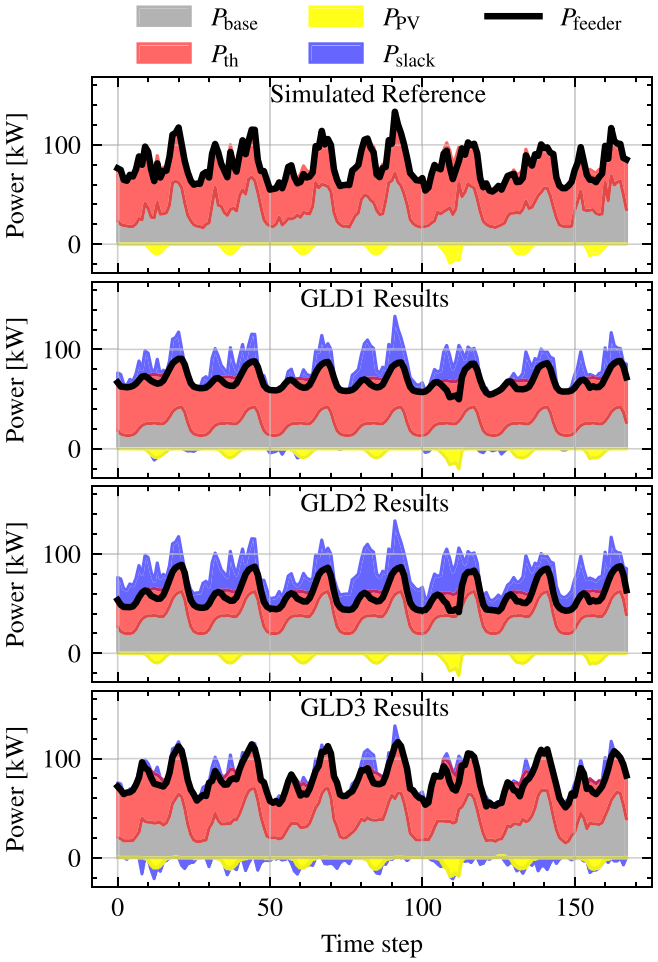
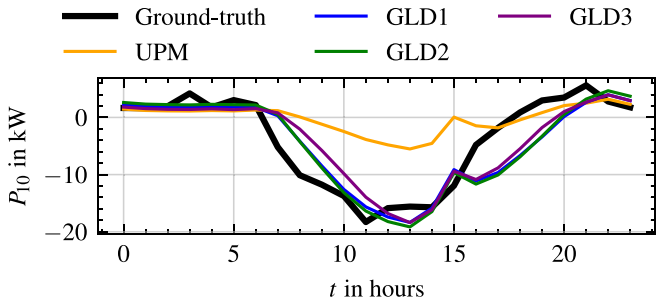



Fig. 10. Disaggregated components for one week of data for GLD1, GLD2, and GLD3.

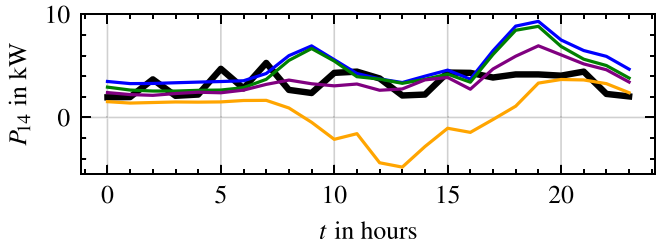
simulated components of the signal contributing to the feeder loading, used as the reference. In contrast, the lower plots display the components resulting from the disaggregation process for all three disaggregation methods. The blue area highlights the difference between measured and predicted power, i.e., the slack power, which is distributed across all unmeasured nodes. Note that the results for GLD3 capture the simulated components best, as highlighted by a relatively small slack power, as well as the daily variations, due to higher flexibility of the underlying nonlinear regression models. In contrast, GLD2 captures the PV power best, since it is constrained by a linear dependency between irradiance and $P_{\text{PV}}$, which restricts PV powers to be only negative. On the other hand, for GLD3, PV powers are not assumed to be only negative, thus occasionally resulting in small erroneous positive PV power predictions at night. Observe also that the prediction of the thermal component is consistently under-estimated when using GLD2, showing the limitation of using a linear model to predict the nonlinear thermal demand. Conversely, GLD1 shows better disaggregation results for thermal power, demonstrating the advantages of using a piecewise-linear thermal model.

The effects of different disaggregation techniques on the obtained PMs are analyzed and compared to ground-truth data

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.




(a) House ID 10 for day 175.




(b) House ID 14 for day 122.

Fig. 11. Ground-truth power injections and PMs using different disaggregation techniques.

TABLE VI
PM MAE IN KW FOR DIFFERENT DISTRIBUTION TECHNIQUES

| Scaling | UPM | GLD1 | GLD2 | GLD3 |
|---|---|---|---|---|
| Equal | **2.01** | 1.39 | 1.49 | 1.30 |
| Rated Power | 2.02 | 1.23 | 1.38 | 1.15 |
| Energy | 2.51 | **1.07** | **1.20** | **0.88** |

in Fig. 11. Note in Fig. 11a that the proposed GLD methods provide a more accurate estimation of PV infeed compared to the UPM method. The UPM approach distributes total power uniformly across all buildings in the feeder, disregarding the presence of a PV generator installed on building ID 10, with Fig. 11b revealing that PV power is assigned to buildings without PV installations, significantly affecting the accuracy of the PMs obtained using this methodology. Both plots show that GLD1 and GLD2 effectively capture the overall behavior of the buildings, despite their reliance on SLP-based load profiles, which limit their ability to reflect the stochastic nature of household loads. In contrast, GLD3 is shown not to be constrained by SLP-based baseload models, allowing it to generate PM estimates that more accurately represent actual household load behavior.

As shown in Table III, different distribution techniques can be applied during PM generation, affecting the quality of the predicted PMs. To evaluate the resulting errors in the absence of SMs, the MAE between the predicted PMs and actual active power for one year of data for the UPM, GLD1, GLD2, and GLD3 methods, along with three different distribution methods, are summarized in Table VI. The best setting for each method is highlighted in bold, showing that the lowest PM error for UPM is achieved using equal scaling, while the GLD methods yield the lowest errors with energy scaling. Moreover, it is evident that all GLD methods significantly

TABLE VII
STATISTICAL ANALYSIS OF PM FROM CONSIDERED GENERATION TECHNIQUES

| Method | Distr. | $\mu_{\lvert\text{MARPE}\rvert}$ | $\sigma_{\lvert\text{MARPE}\rvert}$ | $\mu_{\lvert\text{MAPE}\rvert}$ | $\sigma_{\lvert\text{MAPE}\rvert}$ |
|---|---|---|---|---|---|
| UPM | Equal | 12.27% | 13.11% | 66.65% | 81.75% |
| | Rated Power | 12.51% | 12.92% | 66.30% | 80.41% |
| | Yearly Energy | 15.06% | 16.08% | 83.03% | 88.77% |
| GLD1 | Equal | 9.58% | 10.26% | 53.81% | 63.88% |
| | Rated Power | 9.23% | 8.54% | 56.52% | 60.21% |
| | Yearly Energy | 7.45% | 7.07% | 46.98% | 55.12% |
| GLD2 | Equal | 9.56% | 10.55% | 53.37% | 64.95% |
| | Rated Power | 8.83% | 8.71% | 54.14% | 61.05% |
| | Yearly Energy | 7.43% | 7.48% | 46.90% | 55.05% |
| GLD3 | Equal | 8.26% | 10.06% | 39.15% | 52.04% |
| | Rated Power | 7.44% | 8.25% | 39.58% | 49.32% |
| | Yearly Energy | **5.52%** | **6.69%** | **27.99%** | **36.33%** |

reduce errors compared to the state-of-the-art UPM method, with GLD3 yielding the most accurate results. This demonstrates the benefits of incorporating the disaggregation step in the PM generation process.

A detailed statistical analysis of the resulting PM errors for all households is shown in Table VII. For this purpose, two metrics are compared: the MAPE in (35) and the MARPE in (36). Note that (35) can cause very large errors for small values of the true variable $y(t)$, which is why values around zero are excluded from the analysis. Table VII displays the mean $\mu$ and standard deviation $\sigma$ of the two metrics. Similar to Table VI, a clear trend of lower mean errors can be observed for the GLD methods, with GLD3 showing the lowest overall mean errors in both metrics. Moreover, the standard deviation for these methods is significantly lower compared to UPM and is in the range from 6.69% to 10.55% for the MARPE. As expected, the MAPE metric gives higher errors due to the normalization with respect to the actual value $y(t)$, which leads to very high relative errors for small actual values. Overall, the standard deviations are well below the standard deviations for PMs reported in [11], which are stated to be above 50% for the investigated scenario. This highlights the importance of incorporating feeder measurements and a disaggregation step into the proposed PM generation algorithm.

Since the entries in the $\mathbf{R}$ matrix in (31) are defined as the inverse variances $1/\sigma^2$ of the respective measurements, the accuracy class, which is assumed to be $3\sigma = 0.5\%$, as previously mentioned, is used for physical measurements. On the other hand, for PMs, the $\mathbf{R}$ entries are chosen assuming a very poor accuracy of $3\sigma = 50\%$, since these estimates are considered to have high errors; thus, exact $\sigma$ values are not essential for the DSSE to operate effectively.

The exemplary error distribution using the MARPE metric for household 6 over the simulation time span is illustrated in Figure 13, showing the histograms for all methods using their respective best distribution technique from Table VII. It can be observed that the distribution of GLD1, GLD2, and GLD3 better resembles a normal distribution and leads to smaller outliers. In contrast, the histogram of PM errors obtained with UPM is characterized by a higher skewness and high positive outliers caused by PV power that is incorrectly assigned to this household, despite the household not having a PV generator installed.

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

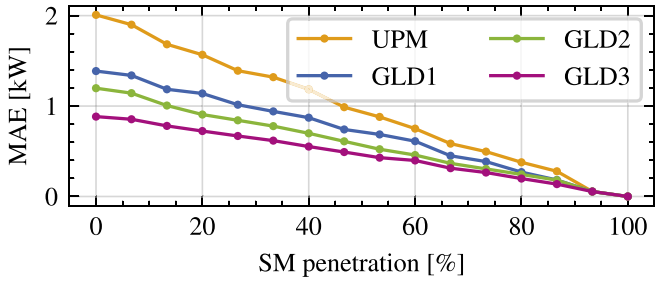



Fig. 12. Error of PM generation techniques for radial feeder and varying SM penetrations for UPM and proposed disaggregation techniques.

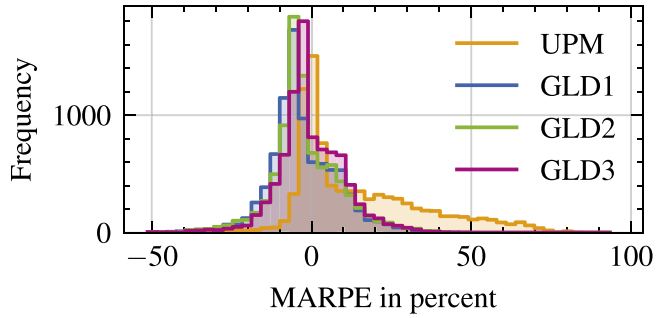



Fig. 13. Histogram of MARPE for household 6 for all methods using the best distribution technique.

Finally, the impact of varying the number of available measurements from SMs is examined by scaling the SM percentage from 0% to 100%, starting with the node closest to the transformer in Scenario 1. Fig. 12 presents the resulting MAE for UPM, GLD1, GLD2, and GLD3, using their respective best distribution methods. As expected, the error decreases with increasing SM penetration, reaching zero when every node is equipped with a metering device. Note that the proposed GLD1, GLD2, and GLD3 exhibit significantly lower errors compared to UPM, with an approximately linear reduction in errors.

### C. State Estimation

The PMs derived from the UPM, GLD1, GLD2, and GLD3 methods are utilized to solve the estimation problem across all time steps of the simulated year for both Scenario 1 and Scenario 2 using various SM configurations. Fig. 14 depicts the resulting voltage estimates for all methods at a single time step $t = 3421$ h, which corresponds to a sunny midday hour with the highest PV generation for the year. Observe that for the radial grid of Scenario 1, a noticeable increase in voltage along the feeder is observed, leading to the highest voltage errors at the end of the feeder, with GLD3 achieving the lowest errors and UPM showing the biggest deviation from the "true" simulated voltages. Conversely, the results for the meshed feeder in Scenario 2 shows the highest voltage occurring towards the electrical middle at Node 10. Similar to Scenario 1, the GLD-based methods provide more accurate estimates compared to UPM. The average relative voltage errors in Scenario 1 and 2 are 0.050% and 0.015%, respectively, which are significantly lower than the 0.62% error reported in [11]. However, a direct comparison is not possible due to differences in the grid and measurement configurations considered in both studies.

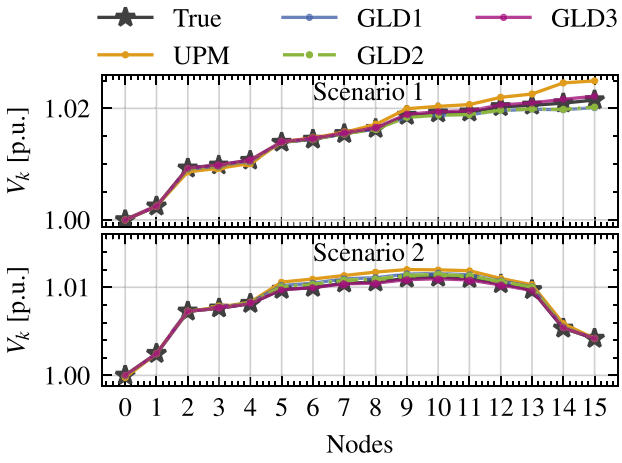



Fig. 14. Voltage estimates for $t = 3421$ h.

TABLE VIII
MAPE % REDUCTION WITH RESPECT TO UPM WITH NO SM FOR SCENARIO 1

| SM Node | UPM | GLD1 | GLD2 | GLD3 |
|---|---|---|---|---|
| {} | 0.00 | 31.18 | 36.32 | **42.68** |
| {7} | 11.63 | 45.65 | 47.10 | **57.06** |
| {14} | 42.50 | 72.27 | 70.30 | **73.42** |
| {7, 14} | 44.85 | 74.42 | 72.74 | **75.69** |

TABLE IX
NUMBER OF VOLTAGE LIMIT VIOLATIONS FOR SCENARIO 1 WITH GLD3 INTERVAL ESTIMATION OF ±1 KW

| SM Node | $\hat{V}(t) > 1.05$ p.u. | $\hat{V}(t) < 0.95$ p.u. |
|---|---|---|
| {} | 21 | 5716 |
| {7} | 2 | 3913 |
| {14} | 0 | 3063 |
| {7, 14} | 0 | 2383 |
| $V(t)$ violations | 0 | 2033 |

An analysis of four different SM configurations for Scenario 1 yields to the results in Table VIII, which shows the percentage reduction in MAPE with respect to the UPM method without SMs. Note that, first, the GLD-based methods yield significantly lower voltage errors than the UPM method, with GLD3 achieving the lowest errors, which are 42.68% lower in the absence of SM measurements. Second, adding SM consistently reduces errors across all methods, leading to a 44.85% reduction when SMs are installed at Nodes 7 and 14, corresponding to the middle and end of the feeder, in the UPM method. As expected, the placement of SMs significantly influences the DSSE results, as installing a SM at Node 7 results in smaller reductions compared to installing them at the end of the feeder at Node 14, making end-of-the-feeder installations of SMs the preferred choice.

Table IX presents the results when employing GLD3 for DSSE for interval estimation, as proposed in Section III-D1 for the more extreme case of Scenario 1, where the radial structure makes it more challenging to predict voltages towards the end of the feeder. To explore the number of missed detections of voltage bound violations, a typical voltage bound of ±5% is considered, thus ensuring that violations occur in the simulation. The number of actual voltage violations, as

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

well as estimated voltage violations assuming a PM error of $\pm 1$ kW, are simulated using the proposed interval estimation method. Regardless of the SM setup, no voltage violations of 1.05 p.u. are present in the "true" value $V(t)$, whereas 2033 violations of the lower bound 0.95 p.u. occur. The DSSE method detects 5716 possible violations when no SMs are installed, which is more than double the number of actual violations. However, this number can be reduced to 2383, with only 353 faulty detections, when SMs are installed at Nodes 7 and 14. These results underscore that while interval estimation methods can be effective in detecting potential voltage violations in sparsely measured LV grids, missed detections are likely.

### D. Discussion

The results of the proposed novel DSSE approach demonstrate significant improvements over existing state-of-the-art methods, obtaining valuable insights for PM generation in LV grids with a very limited number of measurements, and offering a promising avenue for future advancements in this field. However, it is important to highlight the following limitations of the presented study:

- Even though the analysis is validated with a simulated realistic system, the performance in practical applications may differ due to factors not considered in the simulation such as corrupted measurements and unbalancing.
- The proposed DSSE method assumes a static system with unchanging grid topology and known electrical parameters and load characteristics. However, in practice, these factors can vary dynamically. Hence, algorithms for topology identification and parameter estimation (e.g., [26]), and data validation from GIS may be needed in the case of dynamically changing grid conditions.
- The simulation focuses on a balanced LV grid, but in practice, LV grids often exhibit unbalanced conditions. Nevertheless, the proposed techniques could be readily extended to handle unbalanced systems by disaggregating signals from each phase and incorporating a three-phase grid model into the DSSE optimization problem, resulting in a 3-fold increase in the size of the problem. This should not be an issue when using today's efficient and robust optimization solvers, given the number of nodes of a typical European LV grid, particularly in Germany, as previously discussed. For example, for the sample balanced grid used here, the DSSE problem contains about 30 state variables and can be solved in less than 1 s.
- Although the impact of noise in the data was considered in the presented simulations, corrupt and erroneous data quality issues were not examined. The latter can be addressed using robust data processing techniques, including outlier detection and filtering, which may be necessary for ensuring the accuracy and reliability of the DSSE results in practice.

## V. Conclusion

Methods that disaggregate feeder measurements into components for thermal, baseload, and photovoltaic power were introduced for PM generation in the context of LV grid DSSE, particularly for systems with limited SM coverage or, in some extreme case, no measurements from LV nodes. Unlike state-of-the-art approaches, the proposed methods incorporate a disaggregation step based on exogenous GIS and, weather and temporal data to reduce negative impact of limited measurement availability in LV grids, significantly enhancing predictive accuracy in PM generation compared to existing technologies. Three different configurations of the novel GLD methods were proposed and compared with existing approaches from the literature. The accuracy of PM generation and its impact on DSSE results were examined using two realistic German LV feeders. The results demonstrate that using the GLD methods leads to a significant reduction in errors for both PM generation and DSSE results. The study also explored the influence of SM penetration and installation location on DSSE results, revealing that the GLD methods consistently produced the best estimation results across all settings, with the number and placement of SMs significantly affecting estimation accuracy.

Future research could explore more components and sub-models for the disaggregation step. Furthermore, the application of interval estimates via affine arithmetic could be used to obtain tighter estimated intervals.

## Acknowledgment

The A.I. based language models DeepL and OpenAI GPT-4o were used for text editing and other language-related tasks.

## References

[1] "Netzwirtschaft 2050—Szenarien für deutsche verteilnetze," Deloitte GmbH. 2021. Accessed: Oct. 11, 2024. [Online]. Available: https://www.deloitte.com/de/de/Industries/power-utilities-renewables/perspectives/netzwirtschaft-2050-zukunftsszenarien-deutsche-verteilnetze.html

[2] "EnWG §14 a—Gesetz über die elektrizitäts- und gasversorgung (energiewirtschaftsgesetz—EnWG) § 14a netzorientierte steuerung von steuerbaren verbrauchseinrichtungen und steuerbaren netzanschlüssen; festlegungskompetenzen." 2023. Accessed: Oct. 11, 2024. [Online]. Available: https://www.gesetze-im-internet.de/enwg_2005/__14a.html

[3] (ACER Eur. Union Agency Cooper. Energy Reg., Ljubljana, Slovenia). *2024 Market Monitoring Report*. 2024. Accessed: Oct. 11, 2024. [Online]. Available: https://www.acer.europa.eu/monitoring/MMR

[4] "Standardisiertes vorgehen für die Durchführung der netzzustandsermittlung auf basis von echtzeit-messwerten in der niederspannung zur einhaltung von mindestanforderungen an deren Sensitivität und spezifität." VDE FNN. Jan. 2025. [Online]. Available: https://www.bundesnetzagentur.de/DE/Beschlusskammern/1_GZ/BK6-GZ/2022/BK6-22-300/Mitteilung/Mitteilung_3/VDE_FNN_Empfehlung_zu_Tenorziffer_2e.pdf?__blob=publicationFile&v=1

[5] Y. Guo, Y. Zhang, and A. K. Srivastava, "Data-quality aware state estimation in three-phase unbalanced active distribution system," in *Proc. IAS Annu. Meeting*, 2018, pp. 1–7.

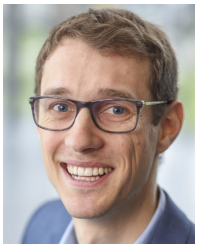
[6] E. B. Alzate, M. Bueno-López, J. Xie, and K. Strunz, "Distribution system state estimation to support coordinated voltage-control strategies by using smart meters," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 5198–5207, Nov. 2019.

[7] "O. reg. 425/06: Criteria and requirements for meters and metering equipment, systems and technology." Electricity Act 1998. 2017. Accessed: Oct. 11, 2024. [Online]. Available: https://www.ontario.ca/laws/regulation/060425

[8] T. Xygkis, N. Manousakis, G. Korres, and N. Hatziargyriou, "Distribution grid state estimation using load pseudomeasurements and topology identification techniques," in *Proc. 23rd CIRED*, 2015, pp. 15–18.

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.

[9] N. Neusel-Lange, C. Oerter, and M. Zdrallek, "State identification and automatic control of smart low voltage grids," in *Proc. 3rd IEEE PES ISGT Eur.*, 2012, pp. 1–6.
[10] K. Li, "State estimation for power distribution system and measurement impacts," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 911–916, May 1996.
[11] A. Angioni, T. Schlösser, F. Ponci, and A. Monti, "Impact of pseudo-measurements from new power profiles on state estimation in low-voltage grids," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 70–77, Jan. 2016.
[12] G. Cavraro, J. Comden, E. Dall'Anese, and A. Bernstein, "Real-time distribution system state estimation with asynchronous measurements," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3813–3822, Sep. 2022.
[13] D. Wäresch, *Entwicklung eines Verfahrens zur Dreiphasigen Zustandsschätzung in Vermaschten Niederspannungsnetzen* (Forschungsberichte des Lehrstuhls für Energiesysteme und Energiemanagement), vol. 4. Aachen, Germany: Shaker Verlag, 2018,
[14] D. Groß et al., "Evaluation of a three-phase distribution system state estimation for operational use in a real medium voltage grid," in *Proc. IEEE PES ISGT Europe*, 2019, pp. 1–5.
[15] S. de Jongh, F. Gielnik, F. Mueller, L. Schmit, M. Suriyah, and T. Leibfried, "Physics-informed geometric deep learning for inference tasks in power systems," *Electr. Power Syst. Res.*, vol. 211, Oct. 2022, Art. no. 108362.
[16] J.-H. Menke, N. Bornhorst, and M. Braun, "Distribution system monitoring for smart power grids with distributed generation using artificial neural networks," *Int. J. Electr. Power Energy Syst.*, vol. 113, pp. 472–480, Dec. 2019.
[17] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4347–4356, Nov. 2020.
[18] J. Kolter, S. Batra, and A. Ng, "Energy disaggregation via discriminative sparse coding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 1153–1161.
[19] K. Pan et al., "An unsupervised data-driven approach for behind-the-meter photovoltaic power generation disaggregation," *Appl. Energy*, vol. 309, Mar. 2022, Art. no. 118450.
[20] E. Proedrou, "A comprehensive review of residential electricity load profile models," *IEEE Access*, vol. 9, pp. 12114–12133, 2021.
[21] C. Stegner, O. Glaß, and T. Beikircher, "Comparing smart metered, residential power demand with standard load profiles," *Sustain. Energy, Grids Netw.*, vol. 20, Dec. 2019, Art. no. 100248.
[22] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794.
[23] J. Huang, M. Algahtani, and S. Kaewunruen, "Energy forecasting in a public building: A benchmarking analysis on long short-term memory (LSTM), support vector regression (SVR), and extreme gradient boosting (XGBoost) networks," *Appl. Sci.*, vol. 12, no. 19, p. 9788, 2022.
[24] C. Christensen, S. Horowitz, T. Givler, A. Courtney, and G. Barker, "*BEopt: Software for Identifying Optimal Building Designs on the Path to Zero Net Energy*, Nat. Renew. Energy Lab., Golden, CO, USA, 2005.
[25] L. Thurner et al., "Pandapower—An open source python tool for convenient modeling, analysis and optimization of electric power systems," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6510–6521, Nov. 2018.
[26] S. de Jongh, F. Mueller, F. Osterberg, C. A. Canizares, T. Leibfried, and K. Bhattacharya, "Data-driven topology and parameter identification in distribution systems with limited measurements," *IEEE Trans. Power Del.*, vol. 40, no. 1, pp. 249–260, Feb. 2025.



**Steven de Jongh** (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree in electrical engineering, KIT, where he is currently a Research Associate. His research interests include distribution system modelling, machine learning approaches for power systems and power system estimation and identification.

**Felicitas Mueller** (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree in electrical engineering, KIT, where she is a Research Associate. Her research interests include distribution grid modeling, energy management systems, and optimization of power systems.

**Claudio A. Cañizares** (Fellow, IEEE) is a University Professor, the Hydro One Chair, and the WISE Executive Director of the University of Waterloo, where he has been since 1993. His highly-cited research focuses on modeling, simulation, computation, stability, control, and optimization of power and energy systems. He has received the 2017 IEEE PES Outstanding Educator Award, the 2016 IEEE Canada Electric Power Medal, and multiple awards and recognitions from PES technical committees. He is a former EIC of IEEE TRANSACTIONS ON SMART GRID and the Director of the IEEE and PES Boards; and a Fellow of the Royal Society of Canada, and the Canadian Academy of Engineering.

**Thomas Leibfried** (Senior Member, IEEE) received the Dipl.-Ing. and Dr.-Ing. degrees from the University of Stuttgart, Stuttgart, Germany, in 1990 and 1996, respectively. From 1996 to 2002, he was with Siemens AG, working in the power transformer business in various technical and management positions. In 2002, he joined the Karlsruhe Institute of Technology, Karlsruhe, Germany, as the Head of the Institute of Electric Energy Systems and High-Voltage Technology. From 2014 to 2018, he served as the dean of his faculty. In 2024, he was elected as member of the review board "Electrical Engineering" of the German Research Foundation from 2024 until 2028. He is the author of numerous technical papers dealing with power system equipment, HVDC, and power flow control and management strategies. He is a member of VDE and CIGRE.

**Kankar Bhattacharya** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the Indian Institute of Technology, New Delhi, India, in 1993. He was with the Faculty of Indira Gandhi Institute of Development Research, Mumbai, India, from 1993 to 1998, and also with the Department of Electric Power Engineering, Chalmers University of Technology, Gothenburg, Sweden, from 1998 to 2002. In 2003, he joined the Electrical and Computer Engineering Department, University of Waterloo, Waterloo, ON, Canada, where he is currently a Professor and the Chair. His current research interests include power system economics and operational aspects. He is a Registered Professional Engineer in the province of Ontario.

Authorized licensed use limited to: IEEE Editors-in-Chief. Downloaded on August 24,2025 at 21:23:40 UTC from IEEE Xplore. Restrictions apply.