

# Modular Load Estimation Framework for Low-Observability Distribution Networks: Crete Case Study

Georgios Falekas

*Hellenic Electricity Distribution  
Network Operator (HEDNO)  
Perrevou 20 & Kallirois 5  
Athens, Greece  
G.Falekas@deddie.gr*

Ilias Manitaris

*Hellenic Electricity Distribution  
Network Operator (HEDNO)  
Perrevou 20 & Kallirois 5  
Athens, Greece  
I.Manitaris@deddie.gr*

Efstratia Sarafoglou

*Hellenic Electricity Distribution  
Network Operator (HEDNO)  
Perrevou 20 & Kallirois 5  
Athens, Greece  
E.Sarafoglou@deddie.gr*

**Abstract**— This paper presents a lightweight, data-sparse methodology for estimating nodal loads in low-observability distribution networks. A modular approach is designed for settings where smart metering is unavailable and only feeder-level SCADA measurements exist, while allowing for increasing levels of truth points as available. Two estimation strategies are shown: a baseline “null-hypothesis” estimator, which assumes uniform loading, and a profile-based estimator, which uses temporal variations extracted from smart meters. A novel alpharithmic naming convention is proposed. The methodology is implemented on a real MV feeder in rural Crete, and its performance is validated against SCADA data. The framework is designed to comply with system operator workflow and allow for rapid estimator development. The results show that even under severe observability constraints, accurate voltage and power estimates can be achieved using this replicable approach.

**Keywords**— *Load Estimation, Low-Observability Networks, Distribution System Monitoring*

## I. INTRODUCTION

Electrical distribution network segments often suffer from low observability, having only sparse measurements at primary substations or feeder heads [1]. This lack of real-time data at downstream nodes complicates load estimation, which is crucial for planning, state estimation, and integration of renewables. Traditionally, network operators resort to pseudo-measurements derived from historical averages or standard load profiles (SLPs) to fill the data gap [2]. This research proposes a pragmatic and computationally lightweight framework for reconstructing load patterns at unmonitored nodes using only feeder-level SCADA data. Two estimators are developed: a “null-hypothesis” estimator that assumes uniform load allocation and nominal PV generation, and a profile-based estimator that uses pre-extracted smart meter load profiles to inform temporal behavior and heterogeneity. Both estimators are applied to a real MV feeder in rural Crete, allowing for comparative evaluation under realistic conditions.

The novelty of this work lies in the integration of these estimation strategies into a replicable framework that enables Distribution System Operators (DSOs) to improve observability in non-instrumented regions. Unlike existing methods that rely on smart meter or high-resolution AMI data, the proposed approach is designed to operate under minimal data availability, making it particularly suited for

transitional or remote grids. Moreover, the use of a Mediterranean AMI dataset adds geographic and behavioral relevance for application in Greek networks. The framework is designed to be fully modular and provide increasingly accurate estimates given progressively more information. Both modeling and estimator compartments can range from simplistic to complex.

The proposed forecasting tool has significant and direct value for DSO operations, planning, and long-term digital transformation, as well as for the increasingly prevalent Digital Twin and flexibility services studies. While national smart metering programs are advancing, many rural and islanded areas across Europe remain partially uninstrumented due to rollout priorities. In Greece, initial deployments are focused on high-density urban regions as well as industrial LV customers, with remote low-voltage (LV) areas—particularly small villages and islands with mostly residential customers such as those in Crete—yet to be fully integrated into AMI systems. As a result, LV observability in these regions remains limited to substation or feeder-level SCADA measurements at the MV level. The present methodology provides a practical, cost-effective, and replicable solution for achieving observability in these opaque regions of the grid, until the complete AMI project of the LV customers is achieved during the next few years.

## II. RELATED WORK

Distribution system state estimation (DSSE) and load allocation traditionally rely on pseudo-measurements to compensate for scarce field data. These pseudo-measurements are often derived from standard load profiles or historical billing data [3]. However, standard profiles (often developed in different regions or for average consumers) may not capture local variability. State estimation proposals in distribution systems heavily rely on pseudo-measurements from SLPs, but inaccuracies in those profiles can lead to state estimation errors [4]. Recognizing this, more data-driven or adaptive approaches have been explored to generate pseudo-measurements [5]. The main challenge lies in representing the uncertainty and diversity of customer loads with limited information. In poorly observed networks, it is essential to incorporate any available knowledge, such as SCADA readings at the feeder, customer class information, or data from similar networks, to refine load estimates. DSOs can also leverage actual network structure and electrical characteristics. Clustering techniques on smart meter data have emerged as powerful tools to segment and categorize load patterns, yielding typical daily

---

The work presented is based on research conducted within the framework of the Horizon Europe European Commission project Crete Valley (Grant Agreement No. 101136139). The content of the paper is the sole responsibility of its authors and does not necessarily reflect the views of the EC.

profiles that can act as improved pseudo-measurements [6]. Meanwhile, techniques from evidence theory and Bayesian inference allow combining disparate sources with various uncertainty levels to arrive at a probabilistic estimate of unknown quantities. These advanced techniques have seen use in grid applications such as fault diagnosis and sensor fusion, but have not been widely applied to load allocation [7]. The aforementioned methods, while powerful, typically assume access to historical or AMI data for the specific network under study, or they rely on multiple inputs (e.g., customer classes, smart meter readings, or advanced proxy indicators) to inform load distribution. In practice, such inputs may not be available, especially in LV rural networks with minimal instrumentation.

In contrast, the present work adopts a pragmatic and replicable approach tailored to data-scarce conditions. Instead of relying on prior customer classification or advanced fusion techniques, two lightweight estimators are implemented, operating with limited inputs. The first is a baseline "null-hypothesis" estimator that assumes uniform loading and nominal PV generation, and calibrates assigned loads to match feeder SCADA measurements. The second is a profile-based estimator, which uses publicly available smart meter profiles from a Mediterranean dataset and scales them on their format basis to match observed feeder-level power. Both methods are evaluated on a real MV feeder in rural Crete. Rather than emulating fully instrumented state estimation frameworks, this study focuses on the development and preliminary validation of a modular, low-overhead methodology that leverages basic tools and publicly available data to generate plausible load estimations in unmonitored LV networks.

### III. METHODOLOGY

The methodology is designed to operate within the constraints of data privacy and proprietary grid asset protection. To address these challenges while enabling collaborative research and reproducible analysis, the approach incorporates a deliberate anonymization strategy that preserves the essential electrical characteristics of medium voltage networks while obscuring sensitive geographical and infrastructure details.

PandaPower is chosen as the framework environment, representing a strategic departure from DIgSILENT PowerFactory. PandaPower eliminates the substantial overhead typically associated with commercial power system analysis platforms, including licensing restrictions, complex data interfacing requirements, and rigid database structures that can impede freeform research and external collaboration.

The network is defined deliberately without geographical data to preserve anonymity and disallow matching. Installed transformer power is aggregated at each bus using a custom 's\_nom\_mva' variable. Line definitions include the length and type-material characteristics. This results in a graph-based representation which preserves only analysis relevant parameters while rendering the actual location and configuration of the distribution system indiscernible. Each bus serves as an aggregation point for total connected load, with additional future granularity, such as the LV network from the bus, treated as a new network. The graph-based representation also allows for graph-based search and/or optimization algorithms. Transformers are not included in the current implementation, but the framework allows for

their inclusion, if more accurate losses, effects or measurements are needed.

The framework employs a novel systematic alphanumeric naming convention specifically designed for radial distribution networks, which is proposed for standard adaptation. It is designed to provide inherent topological information while simultaneously anonymizing the network. During topology definition, the resultant self-evident "main trunk" buses are sequentially labeled as bus\_0, bus\_1, ..., bus\_N. Secondary branches are denoted using alternating alphanumeric extensions, where bus\_3 generates branches bus\_3a and bus\_3b, with subsequent sub-branches following as bus\_3a1, bus\_3a2, and further subdivisions as bus\_3a1a, bus\_3a1b, and so on. Lines are simply named after the connecting buses; for example, line\_3a1\_3a1a. Table I clearly showcases the proposed naming scheme.

TABLE I. HIERARCHICAL NAMING SCHEME

| Level                  | Pattern                | Example             |
|------------------------|------------------------|---------------------|
| Main trunk             | bus_[N]                | bus_0, bus_1, bus_2 |
| 1 <sup>st</sup> branch | bus_[N][A]             | bus_2a, bus_2b      |
| 2 <sup>nd</sup> branch | bus_[N][A][N]          | bus_2a1, bus_2a2    |
| 3 <sup>rd</sup> branch | bus_[N][A][N][A]       | bus_2a1a, bus_2a1b  |
| Lines                  | line_[source]_[target] | line_2a1_2a1a       |

The proposed hierarchical alphanumerical naming structure provides two distinct analytical advantages. First, the convention enables complete network topology reconstruction from bus names alone, eliminating the need for graphical visualization or adjacency matrices to understand network connectivity. Any intended user can deduce the full radial structure directly through rule knowledge, facilitating model sharing and collaborative analysis without supplementary documentation.

The second advantage addresses a key limitation in interpreting power flow directionality. In networks with distributed generation, net power measurements can mask the underlying contributions of load and generation, especially when bidirectional flows result in near-zero or negative net values. The hierarchical naming convention mitigates this by enabling automatic identification of upstream and downstream relationships. While the main drawback is long identifiers in highly branched networks, this is offset by the benefits of interpretability and automated topology recognition, particularly in research contexts. For visualization, name length can be managed via a reference dictionary.

The framework implements a categorical bus assignment system, allowing for high-level differentiated load and generation estimation algorithms. In the current implementation, buses are assigned to one of four commonly used categories across literature and industry: residential (RES), commercial (COM), industrial (IND), and generation (GEN), through a custom column. Any type and number of categories can be assigned, and a bus can have a fraction of each category to allow for diversity. In the currently investigated network, buses are assigned to single categories based on the real topology. Figure 1 showcases the graph of the radial line explored in this work.

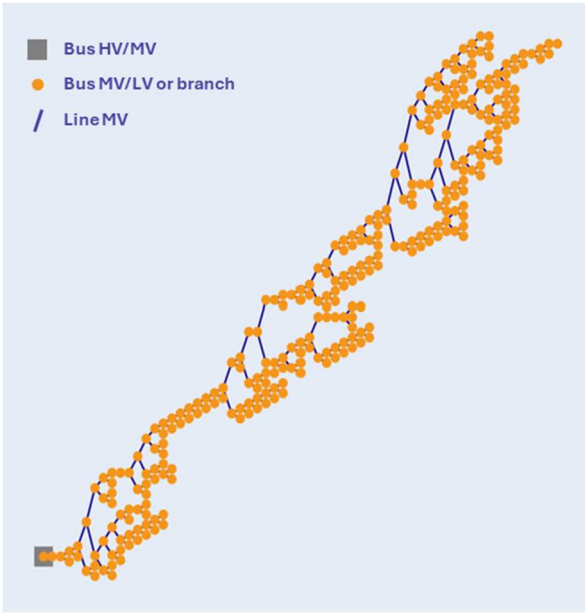


Fig. 1. Graph representation of the investigated Cretan MV line

From the starting point of the MV line at the HV/MV substation, the radial distribution network is depicted down to the MV/LV connection points. Buses are assigned to each connection point of the network, such as the interchange with the LV grid as well as the branches within the MV distribution network.

The framework integrates external load and generation profiles through structured CSV datasets that capture temporal consumption and production patterns. To construct these profiles, the publicly available GoEner smart meter dataset [8] was used, which contains high-resolution hourly consumption data from a diverse pool of residential, commercial, and industrial consumers across Spain. The dataset is enriched with metadata including contracted tariff codes and economic activity classifications, enabling the categorization of users according to typical DSO definitions. The Spanish context is characterized by Mediterranean climatic conditions and usage behaviors and offers strong regional similarity to the island of Crete, making the dataset particularly well-suited for this application.

After parsing and cleaning over 16,000 user time series, daily load curves were extracted and normalized by total daily energy. This converts each daily curve into a profile whose mean column sums to 1.0 over 24 hours, effectively isolating the shape of consumption from its magnitude. All of the other columns are scaled relatively to the mean. As a result, each profile represents only the temporal distribution of demand, independent of the user's absolute energy usage. Users were classified into three consumer categories: residential, commercial and industrial, based on tariff and economic activity classification metadata. Each normalized daily curve was then assigned to a group according to consumer category, day of the week, and season (winter, spring, summer, or autumn, determined by calendar month), yielding a structured taxonomy of 84 possible profile combinations (3 categories times 7 days times 4 seasons), plus one PV profile per season. For each of these groups, a representative 24-hour load shape was extracted by averaging all normalized curves within the category. In Figure 2, the SCADA data has also been similarly formatted.

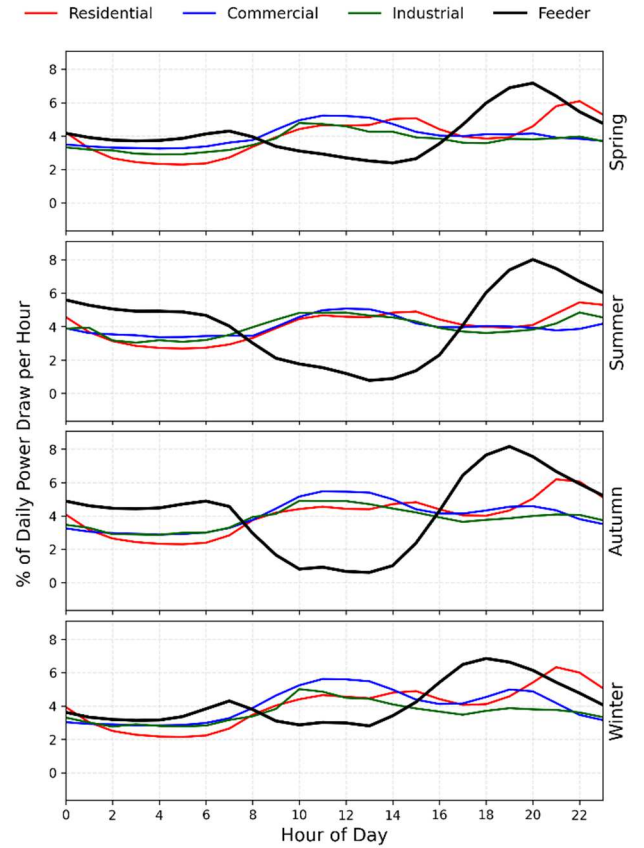


Fig. 2. Seasonal representations for the load profiles and the SCADA feeder data.

The final profiles include not only mean hourly values but also statistical descriptors such as standard deviation, minimum, and maximum load at each hour. These profiles form a comprehensive library of typical daily consumption patterns, suitable for data-scarce environments. It is important to note that the midday dip occurs due to high distributed generation, which reduces metered consumption.

The framework incorporates real-time SCADA measurements as reference points for estimation calibration and validation, with timestamp-synchronized net power flow measurements matched against simulated results to enable optimization and error quantification. Utilizing the directionality enabled by the naming convention, the comparison bus that corresponds to the real metering point can be assigned at will. Power flow calculations utilize the Newton-Raphson solver to determine network states for each estimated assignment individually for each timestamp. This approach creates a continuous validation cycle, where estimated loads are propagated through the network model to produce simulated power flows that can be directly compared against available SCADA measurements, allowing for iterations at will. More than one validation points can be defined if available.

The highly modular architecture is centered around seamless incorporation of abstract estimator logic code that respects minimal rules. An estimator factory class serves as the central registry, maintaining a dictionary of available estimator types (currently: manual, null-hypothesis, profile) for runtime calls. All estimators inherit from the base estimator abstract class, which establishes a standardized interface through mandatory methods. The base class handles

common functionality such as bus categorization, network cloning for safe manipulation, and parameter management. This inheritance structure ensures consistent data handling and parameter management across different estimation approaches while allowing complete freedom in the mathematical formulation of load assignment algorithms, as well as network parameters.

The simulation engine operates independently of estimator implementation details through well-defined interfaces and rigorous debugging and error checks. A main simulation code orchestrates the complete workflow by instantiating the selected estimator, managing timestamp iteration, and coordinating between network operations and results collection. A network manager provides a standardized abstraction layer for power flow operations, including load and generation assignment, power flow execution, and results extraction, effectively decoupling the estimation algorithms from the underlying standard PandaPower implementation details. Data handling modularity is achieved through a data handler class, which manages profile loading, SCADA data integration, and network model construction independently of simulation logic. This separation enables different data sources and formats to be incorporated without affecting the estimation or simulation components. The results manager operates similarly, providing standardized collection, storage, and export functionality that remains consistent regardless of the chosen estimation algorithm, providing baseline results. Estimator code can request custom results for specific analysis.

Configuration management operates through a centralized system that allows runtime selection of estimator types, hyperparameter adjustment, and simulation settings without code modification. This configuration-driven approach enables systematic algorithm comparison and parameter sensitivity analysis, supporting both research applications and operational deployment scenarios. The modular architecture ultimately facilitates collaborative development, algorithm benchmarking, and incremental enhancement of individual components without system-wide modifications.

#### IV. ESTIMATORS AND RESULTS

The case study focuses on a radial MV feeder located in a low-density, rural area in Crete. This line is representative of areas well outside the scope of Greece's initial wave of installations of the national smart metering rollout, which prioritizes urban centers and mainland regions. However, it is an interesting line for network study, considering its distributed generation and touristic nature. The MV section of the network is modeled in PandaPower, following the schema as described in III-Methodology. Available SCADA data include substation-level active power, current, voltage, and power factor measurements, sampled at hourly resolution for a full year. No LV metering is available at customer or transformer level. LV loads are aggregated at the MV level.

##### A. Null Hypothesis Estimator

The Null Hypothesis (NH) Estimator implements an empirical calibration methodology that directly measures the relationship between load scaling and reference bus power measurements, rather than relying on theoretical

approximations of consumer behavior. Unlike traditional approaches that assume uniform load scaling with predetermined power factors, this estimator employs a two-stage measurement process to establish the power change coefficient that quantifies how reference bus power responds to system-wide load variations.

The estimation process begins by calculating time-dependent photovoltaic generation profiles based on seasonal and hourly solar irradiance data from the Photovoltaic Geographical Information System (PVGIS). The dataset corresponds to the entire region of Crete as one, providing irradiance values at an hourly temporal resolution, which matches the estimator granularity. Minimal load conditions (1% of nominal installed capacity) are initially applied to all load buses alongside the computed PV generation, establishing a baseline reference bus power measurement that reflects generation-only operation. A power change coefficient is then derived from a configurable load injection test at 10%, with the resulting change in SCADA power used to calculate the ratio of power change to injected load. This coefficient captures network-level effects, including impedance, topology, and interactions with distributed generation that influence how loads are reflected at the measurement point. The current version of the estimator uses one measurement per timestamp; additional points can be incorporated to address non-linearities at will. After obtaining the coefficient, the total system load required to match SCADA measurements is computed by dividing the difference between target and baseline power by the coefficient. The corresponding load scaling factor is then determined. This method inherently accommodates networks with significant distributed generation and reverse power flows, as the empirically derived coefficient accounts for bidirectional exchange effects.

$$c_p = \frac{P_{ref,test} - P_{ref,1\%}}{\sum L_{test}} \quad (1)$$

$$L_{req} = \frac{P_{SCADA} - P_{ref,1\%}}{c_p} \quad (2)$$

$$k = k_{test} \frac{L_{req}}{\sum L_{test}} \quad (3)$$

$$P_i = k \cdot S_{nom,i} \cdot \cos\phi \quad (4)$$

The power change coefficient  $c_p$  represents the sensitivity of the reference bus power to total system load variation. Here,  $P_{ref,test}$  and  $P_{ref,1\%}$  denote the feeder-level active power measured at the reference bus under the test and 1% loading conditions, respectively. The numerator thus reflects the observed power change due to the controlled load injection. The denominator  $\sum L_{test}$  is the total load added during the test case. This coefficient effectively captures the influence of network impedance, distributed generation backfeed, and topological attenuation on how added loads manifest at the measurement point. The required system load  $L_{req}$  is then calculated by dividing the difference between the SCADA-measured feeder power  $P_{SCADA}$  and the baseline power at 1% loading  $P_{ref,1\%}$ . This determines the total load that must be injected into the network in order to match the SCADA measurement, based

on the empirically measured load sensitivity. The final load scaling factor  $k$  is computed by proportionally adjusting the test-case factor  $k_{test}$  (default 0.10, corresponding to 10% of nominal capacity) based on the ratio between the required and test load. The per-bus load  $P_i$  is assigned using the computed scaling factor  $k$ , scaled by the bus's nominal apparent power  $S_{nom,i}$  of bus  $i$  and a power factor  $\cos\phi, k$ , ensuring accurate real power estimation.

To handle edge cases in high-PV networks, where reverse flows dominate or the power change coefficient is unreliable ( $<1\%$  change in power between tests or  $<1\%$  load needed), the algorithm applies contextual logic based on power flow directionality. For  $P_{ref,1\%} < 0$  and  $P_{SCADA} < 0$ , the relative magnitudes determine the response: if  $P_{SCADA} > P_{ref,1\%}$ , a moderate scaling factor of  $k = 0.2$  is applied; otherwise, a minimal scaling factor of  $k = 0.05$  is assigned, suggesting potential measurement discrepancies or PV underestimation. When  $P_{ref,1\%} < 0$  and  $P_{SCADA} \geq 0$ , a higher scaling factor of  $k = 0.3$  is applied, acknowledging that significant load injection is required to overcome the generation surplus and achieve net positive power flow. The default fallback for indeterminate cases is  $k = 0.1$ , as a conservative moderate load assumption.  $k_{test}$  is fully configurable. Most distribution networks are over-dimensioned to accommodate growth and ensure stability, so even moderate scaling can yield high aggregate loads, which remain consistent with design practices. The NH estimator thus serves as a fast, empirical baseline for comparative evaluation, with one-step calibration and computational complexity of  $O(2n)$ , where  $n$  is the number of time steps, resulting in a very fast estimate. Figure 3 showcases the NH estimator results in the profile format.

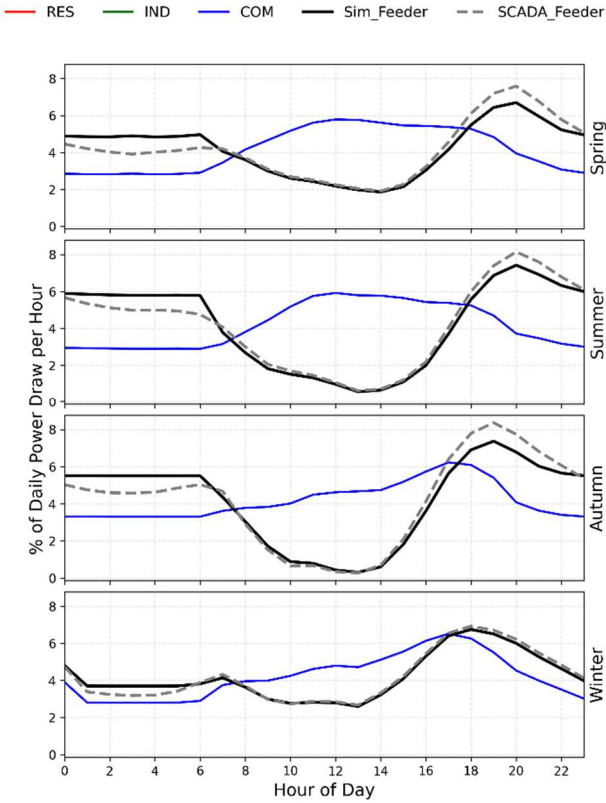


Fig. 3. Seasonal representations for the NH-estimated load profiles, the simulated net power reference and the SCADA feeder data.

## B. Profile-based estimator

The Profile Based (PB) Estimator addresses fundamental limitations of uniform scaling approaches by implementing asymmetric interpolation around the statistical mean of load profile data through individual timestamp optimization. This estimation incorporates the categorized temporal demand and generation profiles. The statistical mean is used as the natural reference point rather than forcing arithmetic symmetry around the min-max range.  $P(\lambda)$  is the estimated factor in terms of percentage of total daily power usage at this hour, with min, mean and max derived from the profiles. Total daily power is calculated from SCADA and split in the day using this factor.

$$P(\lambda) = \begin{cases} P_{min} + (\lambda + 1)(P_{mean} - P_{min}), & \lambda \in [-1, 0] \\ P_{mean} + \lambda(P_{max} - P_{mean}), & \lambda \in (0, 1] \end{cases} \quad (5)$$

This approach addresses non-linear characteristics inherent in energy profile data, where typical residential profiles exhibit highly skewed patterns with mean values around 0.03 pu while maximum values reach 0.09 pu (of total daily load at each hour). Bounded scalar optimization is used to find the optimal  $\lambda$  value for each timestamp. Absolute error between simulated and SCADA power at the reference bus is minimized, starting from an initial guess of  $\lambda = 0$ , using Brent's algorithm within  $[-1, 1]$ . For each candidate  $\lambda$  selected by the algorithm, the estimator builds load and generation profiles using  $P(\lambda)$ , runs power flow analysis, and calculates the error. Non-convergent power flows return a large penalty ( $10^6$ ) to avoid infeasible solutions, and iterations are limited to 50. Figure 4 shows the results in the same format as previously for comparison.

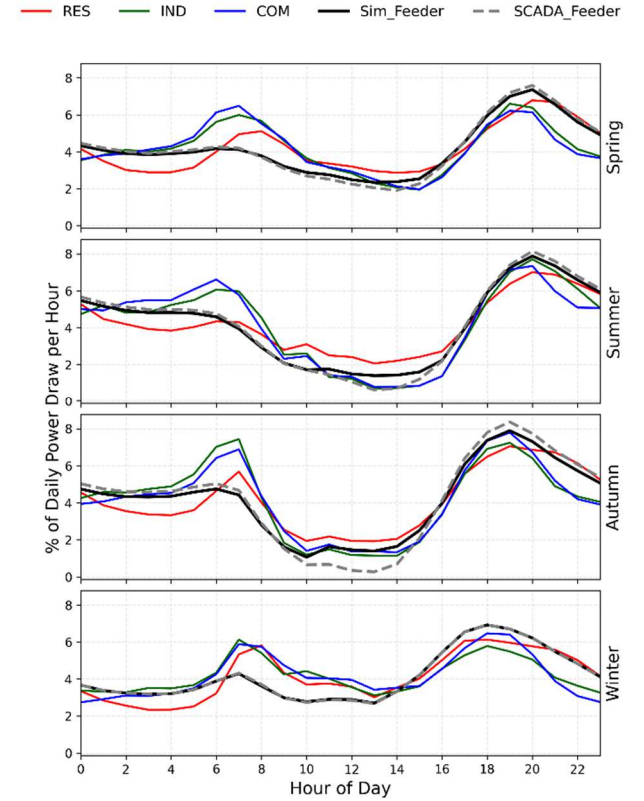


Fig. 4. Seasonal representations for the PB-estimated load profiles, the simulated net power reference and the SCADA feeder data.

For PV generation profiles, the estimator applies asymmetric interpolation while incorporating bounded constraints to prevent unrealistic generation estimates during transition periods. Hours 5-9 and 15-16 are subject to constraint at twice the mean value, addressing high variance ratios observed in cross-seasonal PV profile analysis where transition hours exhibit Max/Mean ratios exceeding 12:1.

Both estimators are very simplistic in the sense that they utilize only the net power flow at the reference bus, disregarding voltage and current. NH has no granularity, while PB differentiates between the categories. More advanced, stochastic estimators are required for fully representative scenarios. Despite this, both estimators achieve acceptable realism, and serve as a comparative basis for more granular estimators that consider more variables. Interestingly, NH identifies the midday peak and the night loads despite the reduced net power flow, while PB better adapts to the temporal variations as expected. We can surmise that a complex estimator is required to produce fully realistic load profiles. The purpose of this work is to establish a proper estimator development framework.

An important insight derived during development is the impracticality of Mean Absolute Percentage Error (MAPE) as the error metric. In similar cases where net power flow navigates through zero, PE is unstable and can explode despite small Absolute Error (AE). Both estimators can serve as a baseline to compare complex ones, as increasing layers of logic easily worsen results as seen during development. The purpose of the modular framework is to verify and open-source it for further research. Estimator metrics are displayed in Table II.

TABLE II. COMPARISON OF ERROR STATISTICS BETWEEN ESTIMATORS

| Metric       | Estimator       | MAPE(%) | Mean Absolute Error | Absolute Error Range |
|--------------|-----------------|---------|---------------------|----------------------|
| Power (MW)   | Null-Hypothesis | 10.53   | 0.1295              | 0.0000-0.7953        |
|              | Profile-Based   | 0.80    | 0.0032              | 0.0000-1.6541        |
| Voltage (kV) | Null-Hypothesis | 0.70    | 0.1475              | 0.000-0.6404         |
|              | Profile-Based   | 0.70    | 0.1475              | 0.000-0.6404         |
| Current (A)  | Null-Hypothesis | 20.38   | 7.350               | 0.0011-29.1170       |
|              | Profile-Based   | 8.11    | 2.9372              | 0.0004-42.6227       |

## V. CONCLUSIONS

This paper presents and validates two low-overhead methodologies for estimating LV loads in the absence of smart meters. The null-hypothesis estimator relies on conservative assumptions and empirical calibration through controlled load injections, while the profile-based estimator leverages realistic daily load shapes extracted from a publicly available Mediterranean smart meter dataset. Both

methods require minimal data and are computationally efficient.

Applied to a real feeder, both estimators demonstrated accurate reconstruction of active power and voltage values with minimal deviation from SCADA references. These results confirm the viability of lightweight, modular tools for enhancing situational awareness in data-scarce LV networks. The estimators' computational efficiency and low data requirements make them well-suited for deployment in legacy systems where full AMI coverage is not yet feasible. Importantly, the methodology offers more than just short-term estimation capabilities. By tracking estimation discrepancies across the network, it provides operators with insight into where smart metering infrastructure would have the greatest impact. In this way, the framework doubles as a decision-support tool for targeted instrumentation planning.

The modular architecture also supports scalability and adaptation: additional estimator types, refined profile libraries, and more granular SCADA validation points can be easily integrated in future iterations. As DSOs continue their digital transformation, tools such as these can facilitate the transition from traditional distribution operation to more intelligent, data-driven grid management.

## REFERENCES

- [1] N. Feng, Y. Du, and Y. Ding, "A Heuristic-Search-Based Topology Identification and Parameter Estimation Method in Low Voltage Distribution Grids With Low Observability," *IEEE Trans Smart Grid*, vol. 15, no. 6, pp. 5826–5839, Nov. 2024, doi: 10.1109/TSG.2024.3435069.
- [2] K. Dehghanpour, Y. Yuan, Z. Wang, and F. Bu, "A Game-Theoretic Data-Driven Approach for Pseudo-Measurement Generation in Distribution System State Estimation," *IEEE Trans Smart Grid*, vol. 10, no. 6, pp. 5942–5951, Nov. 2019, doi: 10.1109/TSG.2019.2893818.
- [3] B. Wang, Y. Ju, S. Hu, X. Feng, and J. Jiang, "Considering Topology Recognition for Pseudo-Measurement Generation in Low-Voltage Distribution Networks," in *2024 Second International Conference on Cyber-Energy Systems and Intelligent Energy (ICCSIE)*, IEEE, May 2024, pp. 1–6. doi: 10.1109/ICCSIE61360.2024.10698001.
- [4] A. Angioni, T. Schlosser, 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, doi: 10.1109/TIM.2015.2454673.
- [5] D. Xu, Z. Wu, J. Xu, and Q. Hu, "A pseudo measurement modeling based forecasting aided state estimation framework for distribution network," *International Journal of Electrical Power & Energy Systems*, vol. 160, p. 110116, Sep. 2024, doi: 10.1016/j.ijepes.2024.110116.
- [6] W. Zhang, M. Cheng, Q. Xiang, and Q. Li, "Enhancing Short-Term Load Forecasting Through K-Shape Clustering and Deep Learning Integration," *IEEE Access*, vol. 13, pp. 30817–30832, 2025, doi: 10.1109/ACCESS.2025.3541574.
- [7] S. Calderwood, K. McAreavey, W. Liu, and J. Hong, "Context-dependent combination of sensor information in Dempster–Shafer theory for BDI," *Knowl Inf Syst*, vol. 51, no. 1, pp. 259–285, Apr. 2017, doi: 10.1007/s10115-016-0978-0.
- [8] C. Quesada, L. Astigarraga, C. Merveille, and C. E. Borges, "An electricity smart meter dataset of Spanish households: insights into consumption patterns," *Sci Data*, vol. 11, no. 1, p. 59, Jan. 2024, doi: 10.1038/s41597-023-02846-0.