

A Machine Learning Approach for Transformer Load Allocation in Power Distribution Networks

Kevin López-Eugenio , Pedro Torres-Bermeo , Jorge Buele , *Member, IEEE*, Nancy Velasco , and José Varela-Aldás , *Senior Member, IEEE*

Abstract—Accurate transformer load allocation is essential for distribution network planning and operation, particularly in systems without advanced metering infrastructure. This paper proposes a bottom-up data-driven methodology that integrates clustering of measured load curves, supervised classification, and energy-based scaling to estimate transformer-level demand while preserving temporal characteristics and energy consistency. The methodology was validated using real data from an Ecuadorian distribution feeder with measurements available at the feeder head and at an intermediate recloser. Power flow results were compared with the conventional load allocation methods available in CYMDIST in order to evaluate the accuracy of the proposed method against standard approaches. At the intermediate recloser, where downstream allocation accuracy is critical, the proposed method achieved the lowest error, with a mean absolute error of 32.23 kVA and a mean absolute percentage error of 6.60%, compared with the range of errors obtained with the CYMDIST methods, which reached up to 186.77 kVA and 33.05%, respectively. These results suggest that the proposed methodology provides a practical approach for improving load allocation in minimally monitored distribution networks.

Link to graphical and video abstracts, and to code:
<https://latam.ieee9.org/index.php/transactions/article/view/10346>

Index Terms—K-means, load allocation, machine learning, CYMDIST, power distribution networks, and SDG 7.

I. INTRODUCTION

THE increasing complexity of modern electric distribution systems, driven by the integration of distributed energy resources, rising non-residential loads, and the need for higher operational efficiency, highlights the importance of accurately estimating load profiles at the transformer-level. Traditional approaches based on aggregated information, such as monthly energy consumption (kWh) or installed equipment capacity (kVA), often lack the resolution required for realistic distribution network modeling and technical studies [1].

Consequently, data-driven methodologies have gained prominence, utilizing advanced clustering and machine learn-

ing techniques to derive normalized load curves that reflect consumer behavior and temporal demand characteristics [2]. These approaches are especially relevant for utilities lacking advanced metering infrastructure (AMI), making indirect estimation methods essential for accurate planning and operational decisions [1]. As a result, network studies may not adequately represent the real distribution of demand along the feeder.

This limitation becomes increasingly critical in the current context of growing distributed generation penetration, where more realistic feeder models are required for technical analyses such as power flow simulation, voltage regulation assessment, loss estimation, and operational planning. Therefore, there is a need for practical methodologies capable of estimating transformer-level load profiles using limited measurement data while preserving consistency with real feeder operating conditions [3].

A. State of the Art

Load estimation in distribution systems is essential for network planning and operation. Traditional approaches based solely on kWh or kVA values are insufficient for modern dynamic simulations, particularly in systems lacking AMI [1], [4]. Data-driven methods have therefore gained attention, including K-means clustering, Dynamic Time Warping (DTW), and Gaussian Mixture Models [2]. Hybrid frameworks combining clustering with supervised learning, such as LightGBM, have achieved accuracies above 95% in transformer load estimation for low-instrumentation networks [3].

Artificial neural networks (ANNs) have also proven effective for estimating transformer loading from historical or indirect data [5], [6], while multi-stage approaches integrating clustering, gradient boosting, and feature selection improve forecasting accuracy and computational efficiency [7]. In parallel, state estimation methods including Weighted Least Squares (WLS), Kalman filtering, Bayesian inference, and convex optimization support the modeling of partially observable networks [8], [9]. Fuzzy neural, case-based reasoning (CBR), and deep neural network approaches further improve estimation under limited instrumentation [10], [11].

Recent reviews also highlight advances in data-driven estimation, optimization, and transformer health management using clustering and PCA [9], [12]–[14]. Power flow modeling remains important for validating estimation accuracy [15], while non-intrusive and measurement-assisted approaches continue to improve transformer demand estimation in minimally monitored networks [16], [17].

The associate editor coordinating the review of this manuscript and approving it for publication was Martín Pedemonte (*Corresponding author: José Varela-Aldás*).

K. López-Eugenio, J. Buele, and José Varela-Aldás are with the Centro de Investigación en Mecatrónica y Sistemas Interactivos MIST, Facultad de Ingenierías, Universidad Tecnológica Indoamérica, Ecuador (e-mails: klopez23@indoamerica.edu.ec, jorgebuele@uti.edu.ec, and josevarela@uti.edu.ec).

P. Torres-Bermeo is with Empresa Eléctrica Ambato Regional Centro Norte S. A. (EEASA), Ecuador (e-mail: ptorres@eeasa.com.ec).

N. Velasco is with the Facultad de Informática y Electrónica, Escuela Superior Politécnica de Chimborazo, Riobamba 060155, Ecuador (e-mail: nancy.velasco@esepoch.edu.ec).

Despite these advances, an important gap remains in integrating transformer-level load estimation with real feeder operating conditions. Most existing approaches rely on top-down allocation rules based on installed capacity or energy consumption, which may not accurately reproduce the spatial distribution of demand and may require relaxation factors for power flow convergence. In this context, the present work uses the clustering and supervised classification stages previously reported in [3] as input to develop and validate a feeder-level bottom-up allocation methodology. The proposed framework combines transformer-level estimated profiles, customer-category disaggregation, feeder-head loss correction, and global consistency adjustment, enabling the transition from transformer demand estimation to realistic feeder-level load allocation validated with field measurements.

Thus, this paper is structured as follows: Section 2 describes the materials and methods; Section 3 presents the results and discussion; and Section 4 summarizes the conclusions and future research directions.

II. MATERIALS AND METHODS

This research has been conducted using data provided by an electrical distribution utility from Ecuador. Specifically, one primary feeder (0100020T06) was extracted from the Geographic Information System (GIS) database of the utility, selected due to the availability of both feeder head and intermediate recloser measurements. It is important to note that not all primary feeders in the utility's system are equipped with intermediate reclosers that record electrical variables. Therefore, feeder 0100020T06 was selected as a suitable case study, given the availability of high-resolution measurements at both points.

To achieve the objective of estimating the load associated with each distribution transformer within the network, the methodology has been organized into five main stages, as shown in Fig. 1. The first stage corresponds to data acquisition, which integrates GIS information, energy consumption records from customers of the 0100020T06 feeder, and measured transformer load curves. The second stage involves the segmentation of load curves, where clustering techniques are applied to determine representative daily load profiles. In the third stage, transformer power load curves are estimated by assigning representative profiles through supervised classification and scaling them according to the monthly energy consumption of each transformer. The fourth stage consists of load allocation and power flow simulation, where the estimated demands are integrated into CYMDIST, a distribution system analysis software developed by EATON and part of the CYME suite. This tool, licensed by the utility, is widely used for steady-state and dynamic analyses of distribution networks. In this study, modeling was performed using CYMDIST version 9.4 Rev 2, incorporating load allocation and power flow modules for feeder modeling and analysis [18]. CYMDIST is appropriate in this context due to its integrated allocation methods and detailed modeling capabilities for medium- and low-voltage distribution systems, as well as its widespread use in Ecuador as a standard tool for technical studies in

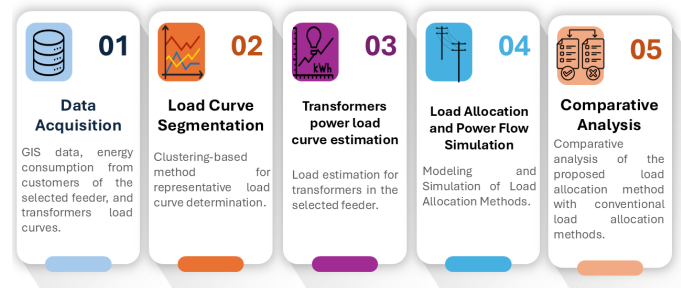


Fig. 1. Methodology for load allocation in a distribution feeder.



Fig. 2. GIS representation of feeder 0100020T06, including the feeder-head circuit breaker and intermediate recloser.

distribution utilities. Finally, the fifth stage corresponds to the comparative analysis, in which the proposed load allocation methodology is evaluated against conventional methods to assess its performance and reliability.

A. Data Acquisition

The input data used in this study come from two primary sources. First, load curves were collected from a representative sample of distribution transformers equipped with measurement devices, with a temporal resolution of 10 minutes. Second, the distribution network topology of the selected feeder was extracted from the utility's GIS.

In addition, monthly energy consumption records from all customers connected to this feeder were considered to apply the load allocation methods available in the CYMDIST software. The selected feeder also includes an intermediate recloser, which provides real-time power measurements. Fig. 2 illustrates the selected feeder in the GIS software. These measurements were used to validate and compare the performance of the proposed load allocation methodology against conventional approaches. It is important to clarify that the intermediate recloser used as a validation point does not correspond to the midpoint of the feeder in terms of physical distance, but rather to a location associated with approximately half of the downstream feeder load. Its position was not selected for the purposes of this study; instead, it reflects the location of monitoring equipment already installed in the actual distribution network.

TABLE I
SUMMARY OF TRANSFORMERS CONNECTED TO THE
SELECTED FEEDER

Type and Phases	Quantity	Capacity (kVA)	Monthly kWh	Clients
Single-phase A	71	1,235.0	184,745.26	2,713
Single-phase B	40	762.5	94,306.96	1,068
Single-phase C	57	895.0	148,486.99	2,063
Three-phase ABC	23	955.0	247,648.47	1,946
Total	191	3,847.5	675,187.69	7,790

Table I summarizes the characteristics of the transformers connected to the feeder, classified by type, phase configuration, capacity, and associated energy consumption.

B. Load Curve Segmentation

Measured transformer load curves were normalized and grouped into representative daily profiles using K-means clustering with Dynamic Time Warping (DTW) as the similarity metric. This unsupervised segmentation identifies typical demand patterns across customer categories, which serve as reference clusters for subsequent allocation. The clustering procedure is identical to that reported in [3], and its robustness has been demonstrated with cross-correlation and temporal consistency indices above 0.95.

C. Transformer Power Load-Curve Estimation and Peak-Load Estimation

To extend representative load curves to unmeasured transformers, a Light Gradient Boosting Machine (LightGBM) classifier was trained using the measured sample as labeled data, as shown in [3]. The independent variables include transformer rated capacity, customer composition, connection type (single-, two-, or three-phase), and geographic location. Each unmeasured transformer is assigned to the most suitable cluster obtained in the previous stage.

Once a cluster is assigned, the normalized daily curve is scaled according to the total monthly energy consumption of the transformer's connected customers, ensuring energy conservation. This process yields the estimated maximum power demand ($P_{\max}$) for each distribution transformer, as described in [3] and expressed by:

$$P_{\max} = \frac{\text{kWh}_{\text{connected_load}}}{\Delta t \cdot t \cdot \sum P_i} \quad (1)$$

In this equation, $P_{\max}$ denotes the estimated maximum load power (in kW) of the transformer. The term $\text{kWh}_{\text{connected_load}}$ refers to the total monthly energy consumption of all customers connected to the transformer, including public lighting. The parameter Δt represents the time resolution of the normalized load curve (10 minutes, or equivalently 1/6 of an hour), while t is the number of days in the analyzed month. Finally, $\sum P_i$ is the sum of the instantaneous values of the normalized daily load curve assigned to the transformer. This formulation is based on the principle of energy conservation, ensuring that the total area under the scaled load curve (expressed in kWh) matches the actual monthly energy consumption recorded for the transformer.

By solving for $P_{\max}$, the normalized curve is transformed into a realistic load profile aligned with the transformer's expected demand. For load allocation analysis, the transformer's power value is extracted under three scenarios: minimum, average, and maximum demand.

D. Load Allocation Methodologies

1) *Proposed Method*: The distribution system operated by the utility was modeled using CYMDIST. The network topology was exported from the GIS of the utility in ASCII format and subsequently imported into CYMDIST, ensuring consistency through a full data migration.

The modeled network spans from the primary substation feeders down to the distribution transformers. In the modeling environment used by the utility, the low-voltage network is not explicitly represented; instead, the demand of all customers downstream is aggregated at the distribution transformer node. Consequently, the estimated transformer demand (P_T) obtained in the previous section cannot be directly assigned to a single lumped load but must be disaggregated among the customer categories connected to each transformer, namely: residential, commercial, industrial, and public lighting.

The total transformer demand P_T (kW) estimated from the scaling procedure is given by:

$$P_T = \sum_{c \in C} P_c + P_{PL} \quad (2)$$

where P_c is the aggregated demand of customer category c (residential, commercial, or industrial), P_{PL} is the demand from public lighting, and C is the set of customer categories connected to the transformer.

Since the coincident peak demand of the feeder occurs at night, the load allocation under peak conditions was performed by subtracting the known contributions of P_{PL} from the total transformer demand and proportionally distributing the remainder among the customer categories:

$$P_c = \alpha_c \cdot (P_T - P_{PL}), \quad \forall c \in C \quad (3)$$

where α_c is the allocation coefficient for category c , defined according to the relative proportion of customers associated with that category. For minimum demand conditions, the same process was applied using the corresponding minimum load curve. For average demand allocation, public lighting was not considered, and the remaining demand was proportionally distributed only among residential, commercial, and industrial customers.

This allocation process was applied to all 191 distribution transformers connected to the selected feeder. Due to the repetitive nature of the procedure, the methodology was automated through Python scripts, ensuring reproducibility and computational efficiency.

Following the approach reported in [3], where an overall accuracy of approximately 95% was achieved in the estimation of the maximum transformer load, an additional adjustment step was introduced to ensure global consistency between the modeled load allocation and the real feeder head measurements. However, it must be noted that the real feeder

head measurements inherently include the technical losses of the feeder. Therefore, it is necessary to estimate these losses through a CYMDIST power flow simulation in order to isolate the actual load component of the feeder demand.

Let P_{meas} denote the total active power measured at the feeder head, and P_{loss} the active power losses of the feeder estimated by CYMDIST. The effective measured load that should be used as reference for the adjustment is then:

$$P_{feederload} = P_{meas} - P_{loss} \quad (4)$$

On the other hand, let P_{est} correspond to the summation of all P_T values, where P_T represents the estimated demand of each distribution transformer connected to the feeder. To reconcile both values, an adjustment factor β was defined as:

$$\beta = \frac{P_{feederload}}{P_{est}} \quad (5)$$

The factor β was applied uniformly to all transformer demands, producing the adjusted active power:

$$P_T^{adj} = \beta \cdot P_T \quad (6)$$

where P_T^{adj} is the adjusted transformer demand. This procedure guarantees that the aggregated demand of the modeled transformers matches the measured feeder demand after excluding technical losses, thereby avoiding systematic bias in subsequent power flow simulations. Once adjusted, active and reactive loads were integrated into CYMDIST, and power flow analyses were performed to evaluate demand propagation up to the intermediate feeder recloser.

Fig. 3 presents the flow diagram that summarizes the proposed load allocation methodology. The process begins with the estimation of transformer load curves derived from clustering and supervised classification. Next, the feeder losses (P_{loss}) are estimated through a single CYMDIST power flow simulation using the CYMDIST kWh method as the preliminary load allocation approach. These losses are subtracted from the feeder head measurement (P_{meas}) to determine the net feeder load ($P_{feederload}$). Based on this value, the adjustment coefficient β is computed, which is uniformly applied to all transformer demands, producing the adjusted demand P_T^{adj} and ensuring global consistency with the feeder head measurement excluding losses. The adjusted transformer loads are then integrated into the CYMDIST model, where power flow simulations are executed to evaluate demand propagation along the feeder, with validation against real measurements obtained from both the feeder head circuit breaker and the intermediate recloser.

2) *CYMDIST Load Allocation Methods*: To assign active and reactive power to concentrated loads along the distribution network system, CYMDIST provides a load allocation module which works with four built-in approaches: the Connected kVA, kWh, Actual kVA, and REA methods. Each method adopts a different strategy for allocating demand based on available data, as briefly described below along with their respective mathematical formulations [18].

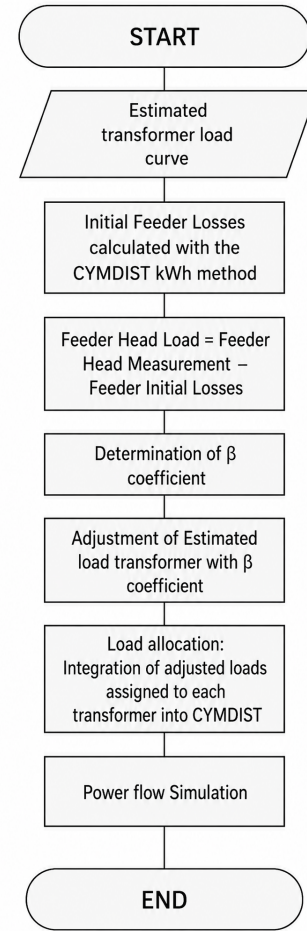


Fig. 3. GIS representation of feeder 0100020T06, including the feeder-head circuit breaker and intermediate recloser.

a) *Connected kVA CYMDIST Method*: This method distributes the feeder demand according to the rated apparent power of the transformers connected to each section, weighted by the corresponding load factor. It is commonly applied when direct measurements at individual sections are unavailable and nominal transformer capacities are used as the basis for proportional load allocation.

The total weighted connected capacity of feeder k is defined as:

$$T_{kVA,k} = \sum_s S_{rated,s,k} LF_s \quad (7)$$

The active and reactive power allocated to section s are then calculated as:

$$P_{s,k}^{alloc} = P_k^{dem} \frac{S_{rated,s,k} LF_s}{T_{kVA,k}} \quad (8)$$

$$Q_{s,k}^{alloc} = Q_k^{dem} \frac{S_{rated,s,k} LF_s}{T_{kVA,k}} \quad (9)$$

where: $S_{rated,s,k}$ is the nameplate rated apparent power (kVA) of the transformer at section s of feeder k ; LF_s is the load factor applied to section s ; $T_{kVA,k}$ is the total weighted connected capacity of feeder k ; P_k^{dem} and Q_k^{dem} are the total

active and reactive power demands measured at the feeder head; and $P_{s,k}^{\text{alloc}}$ and $Q_{s,k}^{\text{alloc}}$ are the active and reactive powers allocated to section s of feeder k .

b) kWh CYMDIST Method: In this approach, load allocation is based on the monthly billed energy (kWh) per section and phase. The energy contribution is adjusted by the corresponding load factor, and demand is proportionally assigned according to each section's relative energy consumption.

The total weighted monthly energy consumption of feeder k is calculated as:

$$TkWh(k) = \sum_s kWh(s, k) LoadFactor(s) \quad (10)$$

Based on this total, the active power demand is proportionally allocated to each section as:

$$PAlloc(s, k) = Pdem(k) \cdot \left[\frac{kWh(s, k) LoadFactor(s)}{TkWh(k)} \right] \quad (11)$$

Similarly, the reactive power demand allocated to each section is given by:

$$QAlloc(s, k) = Qdem(k) \cdot \left[\frac{kWh(s, k) LoadFactor(s)}{TkWh(k)} \right] \quad (12)$$

Where: $kWh(s, k)$ is the monthly billed energy (kWh) of section s in feeder k ; $LoadFactor(s)$ is the load factor applied to section s ; $TkWh(k)$ is the total billed energy for feeder k ; $Pdem(k)$ and $Qdem(k)$ are the total active and reactive power demands at the feeder head; $PAlloc(s, k)$ and $QAlloc(s, k)$ represent the allocated active and reactive powers for section s in feeder k .

c) Actual kVA Method: This method is conceptually similar to the Connected kVA Method but uses actual kVA measurements (e.g., from monitoring equipment) instead of nameplate ratings. As a result, it provides a more accurate representation of actual operating conditions. The total weighted actual apparent power of feeder k is calculated as:

$$TkVA(k) = \sum_s Actual_{kVA}(s, k) LoadFactor(s) \quad (13)$$

Based on this total, the active power demand is proportionally allocated to each section as:

$$PAlloc(s, k) = Pdem(k) \cdot \left[\frac{Actual_{kVA}(s, k) LoadFactor(s)}{TkVA(k)} \right] \quad (14)$$

Similarly, the reactive power demand allocated to each section is given by:

$$QAlloc(s, k) = Qdem(k) \cdot \left[\frac{Actual_{kVA}(s, k) LoadFactor(s)}{TkVA(k)} \right] \quad (15)$$

Where: $Actual_{kVA}(s, k)$ is the measured apparent power (kVA) at section s in feeder k ; $LoadFactor(s)$ is the load factor applied to section s ; $TkVA(k)$ is the total actual apparent power in feeder k ; $Pdem(k)$ and $Qdem(k)$ are the measured active and reactive power demands at the feeder head; $PAlloc(s, k)$ and $QAlloc(s, k)$ are the active and reactive powers allocated to section s in feeder k .

d) REA Method: The REA (Rural Electrification Administration) method is an empirical approach originally developed for rural electric cooperatives. It considers both the number of customers and their billed energy, using a nonlinear formula to better represent the diversity and nonlinearity of rural loads.

First, the customer diversity factor for section s of feeder k is calculated as:

$$A(s, k) = C(s, k) \left(1 - 0.4 C(s, k) + 0.4 \sqrt{C^2(s, k) + 40} \right) \quad (16)$$

The nonlinear scaling factor based on the average monthly energy consumption per customer is then calculated as:

$$B(s, k) = 0.005925 \cdot \left[\frac{kWh(s, k)}{C(s, k)} \right]^{0.885} \quad (17)$$

The REA allocation index for each section is obtained as:

$$REA(s, k) = A(s, k) \cdot B(s, k) \quad (18)$$

The total REA allocation index for feeder k is calculated as:

$$TREA(k) = \sum_s REA(s, k) \quad (19)$$

Based on this total, the active power demand is proportionally allocated to each section as:

$$PAlloc(s, k) = Pdem(k) \cdot \left[\frac{REA(s, k) LoadFactor(s)}{TREA(k)} \right] \quad (20)$$

Similarly, the reactive power demand allocated to each section is given by:

$$QAlloc(s, k) = Qdem(k) \cdot \left[\frac{REA(s, k) LoadFactor(s)}{TREA(k)} \right] \quad (21)$$

Where: $C(s, k)$ is the number of customers connected at section s of feeder k ; $kWh(s, k)$ is the monthly energy consumption of section s ; $A(s, k)$ is the diversity adjustment factor used in the REA method; $B(s, k)$ is the nonlinear scaling factor derived from the average billed energy per customer; $REA(s, k)$ is the empirical allocation index for section s ; $TREA(k)$ is the total allocation index for feeder k ; $Pdem(k)$ and $Qdem(k)$ are the total active and reactive power demands at the feeder head; $PAlloc(s, k)$ and $QAlloc(s, k)$ are the active and reactive powers allocated to section s in feeder k .

Each of these allocation strategies provides different levels of accuracy and data requirements. Their implementation within CYMDIST enhances the load modeling process, supporting realistic power flow results and improving the reliability of transformer-level demand estimations used in this study.

E. Comparative Analysis of Load Allocation Methods

To evaluate the accuracy and robustness of the proposed load allocation methodology, a comparative framework was designed based on error metrics commonly used in power systems and forecasting applications. In this study, two primary indicators are employed: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

The MAE measures the average magnitude of the absolute differences between estimated and actual values of transformer demand, without considering the direction of the error. It is defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (22)$$

where y_i denotes the actual demand value and $\hat{y}_i$ the corresponding estimated value for the i -th observation.

Complementarily, the MAPE expresses this error as a percentage of the actual value, allowing for a normalized comparison across transformers of different capacities. It is calculated as:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (23)$$

In this study, MAE and MAPE were applied to apparent power (S) (kVA), using the measured values at the feeder head and at the intermediate recloser as reference. The metrics were calculated for three representative operating conditions: maximum, average, and minimum demand. Therefore, the analysis provides a pointwise comparison between the proposed method and the conventional CYMDIST load allocation approaches, namely Connected kVA, kWh, Actual kVA, and REA, but it should not be interpreted as a full time-series validation of the estimated load curves.

III. RESULTS AND DISCUSSION

A. Load Curve Segmentation

To characterize the load profiles of distribution transformers, a sample of 208 metered units was analyzed from the total installed transformers across the entire system, as described in [3]. The measurement records for the sampled transformers were collected over a continuous period of seven days.

The load profiles from the monitored transformers were introduced into the K-means clustering model with Dynamic Time Warping (DTW) as the distance metric. Using the elbow method as the evaluation criterion to determine the optimal number of clusters, eight representative daily load curves were identified. These clusters were validated in [3] and describe the characteristic energy usage patterns of the various customer types analyzed in the distribution network.

B. Transformer Peak-Load Estimation

Based on the representative normalized load profiles identified in the clustering stage, each unmonitored distribution transformer in the selected feeder was assigned to one of the eight clusters using a supervised classification model implemented with the Light Gradient Boosting Machine (LightGBM) algorithm, with the hyperparameter configuration explained in [3]. Once the most appropriate representative load curve was assigned to each unmeasured transformer, the corresponding normalized daily profile was scaled to reflect its actual monthly energy consumption.

The maximum load power of each transformer was estimated using Equation (1).

Fig. 4 shows the comparison between the estimated load curve obtained with the proposed methodology and the measured feeder load curve for feeder 0100020T06. The results indicate that the proposed scaling procedure reproduces the temporal behavior of demand with high accuracy, achieving a peak demand estimation of 1,457.36 kW compared to the measured 1,543.71 kW, corresponding to a peak demand accuracy of 94.41%. The measured curve corresponds to a real operating day (maximum feeder demand), while the estimated curve is obtained from the aggregation of representative transformer load profiles. Fig. 5 shows the per-unit (p.u.) comparison, highlighting the similar temporal behavior between measured and estimated curves. The maximum demand day is used

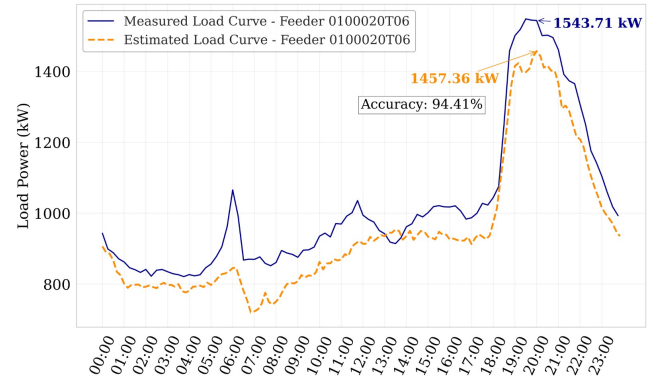


Fig. 4. Comparison between measured and estimated feeder load curves (Feeder 0100020T06).

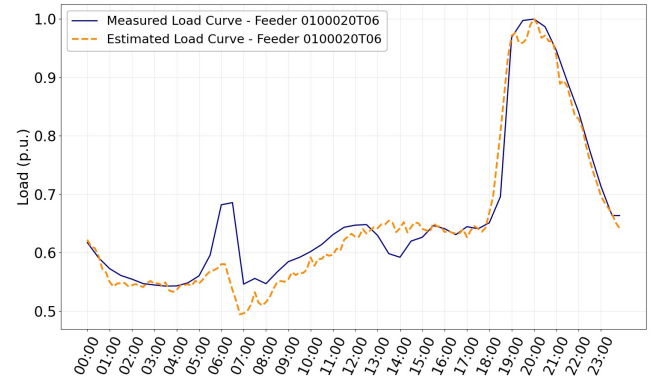


Fig. 5. Comparison between measured and estimated feeder load curves (Feeder 0100020T06) in p.u.

because it represents the most critical operating condition for system evaluation.

C. Load Allocation and Power Flow Simulation

1) *Proposed Method Simulation:* Fig. 6 shows the modeled feeder identified by code 0100020T06, which was selected for this study. Two monitoring points are indicated in the figure: the feeder head circuit breaker and an intermediate feeder recloser. These devices provide power measurements, which were used to validate the results of the power flow simulations.

Table II summarizes the reference values of apparent power measured at the feeder head. These measurements correspond to the maximum, average, and minimum demand conditions selected for the evaluation of load allocation methodologies. The subsequent comparative analysis of allocation strategies was performed using these operating points as boundary conditions, ensuring that the results reflect realistic system behavior across different loading scenarios.

Table III summarizes the real measurements recorded at the intermediate feeder recloser for the three demand scenarios analyzed: maximum, average, and minimum. These records represent the second validation point in addition to the feeder head measurements, allowing the evaluation of how each load allocation method reproduces not only the aggregated demand at the head but also the power flows along the feeder.



Fig. 6. GIS representation of feeder 0100020T06, including the feeder-head circuit breaker and intermediate recloser.

TABLE II
MEASURED FEEDER HEAD DEMAND USED AS REFERENCE
FOR LOAD ALLOCATION ANALYSIS

Demand Scenario	Date and Time	Active Power (kW)	Apparent Power (kVA)
Maximum	22/04/2025 20:00	1,543.71	1,576.72
Average	22/04/2025 16:15	1,020.88	1,094.98
Minimum	22/04/2025 03:45	821.29	849.49

TABLE III
MEASURED RECLOSER DEMAND USED AS REFERENCE
FOR LOAD ALLOCATION VALIDATION

Demand Scenario	Date and Time	Active Power (kW)	Apparent Power (kVA)
Maximum	22/04/2025 20:00	802.0	826.8
Average	22/04/2025 16:15	462.0	502.2
Minimum	22/04/2025 03:45	461.0	465.7

Measured values indicate that, under maximum demand, the recloser registered an apparent power of 826.8 kVA. Under average demand conditions, the apparent power decreased to 502.2 kVA. During minimum demand, the measured apparent power was 465.7 kVA, indicating a load predominantly composed of active power.

Following the procedure outlined in Fig. 3, once the proposed load allocation was completed for each demand scenario, power flow simulations were performed in the CYMDIST software. The simulations were executed using the Newton–Raphson unbalanced method, with convergence and scaling parameters defined to ensure stability and reproducibility across all operating conditions. Table IV summarizes the configuration parameters applied. A voltage tolerance of 0.1% and a maximum of 60 iterations were used as convergence criteria. All simulations started from nominal operating conditions (1.0 p.u.) and considered global load scaling factors of 100% for both active and reactive power.

The results of the power flow simulations obtained using the proposed load allocation methodology are summarized in Table V. At the feeder head, the methodology accurately

TABLE IV
POWER FLOW SIMULATION PARAMETERS USED IN
CYMDIST

Parameter	Value
Calculation Method	Unbalanced Newton–Raphson
Voltage Tolerance	0.1%
Maximum Iterations	60
Initial Condition	1.0 p.u.
Load Scaling Factors	P : 100%, Q : 100%

TABLE V
POWER FLOW RESULTS OBTAINED WITH THE PROPOSED
METHOD AT THE FEEDER HEAD AND THE INTERMEDIATE
FEEDER RECLOSER UNDER MAXIMUM, AVERAGE, AND
MINIMUM DEMAND CONDITIONS

Demand Scenario	Feeder Head		Intermediate Recloser	
	P (kW)	S (kVA)	P (kW)	S (kVA)
Maximum	1,545.09	1,578.51	811	827
Average	1,019.51	1,092.99	524	562
Minimum	819.91	847.94	415	429

reproduces the measured demand under all operating conditions. Under maximum demand, the estimated apparent power reaches 1,578.51 kVA, showing excellent agreement with the reference value. For the average demand scenario, the method yields 1,092.99 kVA, while under minimum demand the estimated value is 847.94 kVA.

Furthermore, at the intermediate feeder recloser, the proposed method also provides accurate estimations of downstream loading conditions. In particular, under maximum demand, the apparent power is 827 kVA, closely matching the measured value. Likewise, for average demand, the estimated value is 562 kVA, while for minimum demand the result is 429 kVA. Overall, the close agreement observed at both validation points demonstrates the robustness of the proposed methodology in reproducing real operating conditions in distribution feeders lacking smart metering infrastructure.

2) *CYMDIST Load Allocation Methods Simulation*: Table VI summarizes the power flow simulation parameters applied in CYMDIST for each allocation method. All simulations were carried out using the unbalanced Newton–Raphson method. The CYMDIST kVA, CYMDIST kWh, and CYMDIST Actual kVA methods converged under the default settings, using a voltage tolerance of 0.1%, a maximum of 60 iterations, and an initial condition of 1.0 p.u. By contrast, the CYMDIST REA method required constraint relaxation, with the voltage tolerance increased up to 6%, to achieve numerical convergence.

Table VII summarizes the power flow results obtained with the CYMDIST load allocation methods at both the feeder head and the intermediate feeder recloser under maximum, average, and minimum demand conditions. At the feeder head, the apparent power values obtained with the kVA, kWh, and Actual kVA methods are very close to the measured reference values, whereas the CYMDIST REA method shows larger deviations due to the required relaxation for convergence. Under maximum demand, the feeder head apparent power ranges

TABLE VI
COMPARISON OF POWER FLOW SIMULATION PARAMETERS
USED IN CYMDIST FOR EACH ALLOCATION METHOD

Parameter	CYMDIST Methods			
	kVA	kWh	Actual kVA	REA
Voltage Tolerance	0.1%	0.1%	0.1%	6%*
Maximum Iterations	60	60	60	60
Initial Condition	1.0 p.u.	1.0 p.u.	1.0 p.u.	1.0 p.u.
Restriction Handling	Not applied	Not applied	Not applied	Applied

* For the REA method, constraint relaxation was required to achieve convergence, increasing the voltage tolerance up to 6%.

TABLE VII
POWER FLOW RESULTS OF CYMDIST LOAD ALLOCATION
METHODS AT THE FEEDER HEAD AND THE INTERMEDIATE
FEEDER RECLOSER UNDER MAXIMUM, AVERAGE, AND
MINIMUM DEMAND CONDITIONS

CYMDIST Methods	Feeder Head S (kVA)			Intermediate Recloser S (kVA)		
	Max.	Ave.	Min.	Max.	Ave.	Min.
kVA	1,576.49	1,093.28	848.55	1,013	740	602
kWh	1,576.53	1,093.41	848.61	884	680	574
REA	1,641.62	1,148.05	897.02	822	617	513
Actual kVA	1,576.55	1,093.56	848.75	828	635	538

from 1,576.49 kVA to 1,641.62 kVA. Under average demand, it varies from 1,093.28 kVA to 1,148.05 kVA, while under minimum demand it ranges from 848.55 kVA to 897.02 kVA.

In contrast, at the intermediate feeder recloser, the differences among methods become more evident, revealing substantial variation in the downstream load distribution produced by each allocation approach. Under maximum demand, the apparent power values range from 822 kVA to 1,013 kVA. For average demand, the results range from 617 kVA to 740 kVA, while under minimum demand the apparent power varies from 513 kVA to 602 kVA. These results confirm that, although several methods closely reproduce the feeder head measurements, their downstream behavior differs significantly.

D. Comparative Analysis

The accuracy of each load allocation method was quantified using MAE and MAPE computed on apparent power S (kVA), taking as reference the measured values reported in Tables II and III. The error metrics were aggregated across the three operating scenarios: maximum, average, and minimum demand.

Fig. 7 presents the comparison of MAE values for the analyzed methods at the feeder head and the intermediate recloser. At the feeder head, the CYMDIST kVA, CYMDIST kWh, and CYMDIST Actual kVA methods show the lowest MAE values, all below 1 kVA, because these approaches are directly anchored to the feeder head measurement. The proposed method also maintains a very low error at this point, with an MAE of 1.78 kVA. By contrast, the CYMDIST REA method exhibits a substantially higher MAE of 55.17 kVA, which is consistent with the relaxation required for convergence. At the intermediate recloser, the differences among methods become much more evident. In this downstream validation point, the proposed method achieves the lowest MAE, equal to 32.23 kVA, followed by CYMDIST REA, CYMDIST Actual kVA, CYMDIST kWh, and CYMDIST kVA.

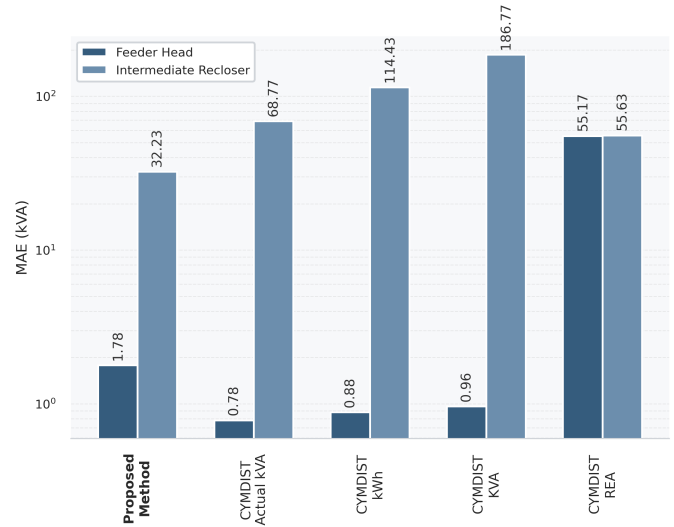


Fig. 7. Comparison of MAE for the analyzed methods at the feeder head and the intermediate recloser.

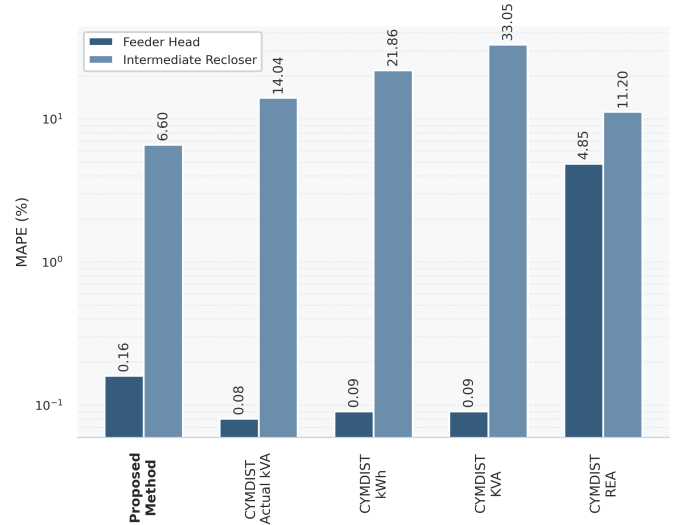


Fig. 8. Comparison of MAPE for the analyzed methods at the feeder head and the intermediate recloser.

Fig. 8 shows the corresponding MAPE comparison. At the feeder head, the CYMDIST kVA, CYMDIST kWh, and CYMDIST Actual kVA methods present errors below 0.10%, while the proposed method also preserves a very low deviation, with a MAPE of 0.16%. Again, the CYMDIST REA method shows a larger discrepancy at this location, reaching 4.85%. At the intermediate recloser, where the allocation must correctly reproduce the downstream distribution of load, the proposed method achieves the lowest MAPE, equal to 6.60%. The remaining methods show progressively larger errors, with CYMDIST REA reaching 11.20%, Actual kVA 14.04%, kWh 21.86%, and kVA 33.05%.

Overall, the comparative results indicate that methods directly tied to the feeder head measurement reproduce the head condition with very low error, except for CYMDIST REA, which required relaxed convergence criteria. However,

the downstream validation at the intermediate recloser reveals the greatest fidelity for the proposed method, which consistently achieves the lowest MAE and MAPE values among all evaluated approaches. This confirms that the proposed methodology provides a more realistic allocation of demand along the feeder.

E. Discussion

The results demonstrate that a bottom-up allocation strategy based on unsupervised segmentation and supervised assignment can reproduce feeder operating conditions with high fidelity even in the absence of AMI. At the feeder head, all CYMDIST methods anchored to the measured boundary condition closely matched the reference, achieving MAPE values below 0.10% for CYMDIST kVA, CYMDIST kWh, and CYMDIST Actual kVA, as shown in Fig. 8. The proposed approach, which does not directly impose the feeder head measurement during the initial estimation, still achieved a very low deviation of 0.16% MAPE once the global consistency adjustment was applied. This confirms that the cluster assignment combined with energy-based scaling described in Section II-C preserves the aggregate load level while retaining temporal realism, as illustrated in Fig. 4.

By contrast, the downstream validation at the intermediate recloser, shown in Figs. 7 and 8, provides direct evidence that the gap identified in the literature has been effectively addressed. While conventional CYMDIST allocation methods achieved low errors at the feeder head, their downstream performance was significantly poorer, with MAPE values of 33.05% (kVA), 21.86% (kWh), and 14.04% (Actual kVA), indicating their limited capability to reproduce the real distribution of demand along the feeder. In contrast, the proposed methodology achieved the lowest error at the intermediate recloser, with MAE equal to 32.23 kVA and MAPE equal to 6.60%, without requiring relaxation factors. This represents a substantial reduction in downstream error compared with all conventional approaches and demonstrates the ability of the proposed method to reproduce the downstream load distribution accurately under real operating conditions.

In the current implementation, the feeder technical losses were estimated through a CYMDIST power flow simulation after assigning the estimated transformer loads to the georeferenced network model. These losses were then used to correct the feeder-head demand measurement and to compute the global adjustment factor β . Therefore, the loss compensation step should be interpreted as a single CYMDIST-based correction rather than as an iterative procedure. Since technical losses depend on the final load allocation, a potential circular dependency may exist. This limitation is acknowledged in the present study. As future work, an iterative scheme will be implemented in which transformer loads are first assigned using a preliminary value of β ; then, a CYMDIST power flow simulation is executed to obtain the feeder technical losses; subsequently, β is recalculated using the feeder-head demand corrected by the updated losses; and the process is repeated until the relative variation of either β or the total feeder losses between two consecutive iterations is lower than a predefined tolerance.

The behavior of the REA method deserves special attention. Although REA can sometimes emulate diversity effects in rural feeders, in this study it required constraint relaxation to achieve numerical convergence in the power flow simulations, with voltage tolerance increased up to 6%, as indicated in Table VI. While this made it possible to include the method in the comparative analysis, the relaxation altered the feeder head boundary condition and produced apparent power values farther from the real operating point, as summarized in Table VII. This resulted in a feeder head error of 4.85% MAPE and a downstream mismatch of 11.20% MAPE, as shown in Fig. 8. From a practical and engineering perspective, this result remains relevant because REA is one of the conventional load allocation methods available in CYMDIST, and its numerical difficulties illustrate a limitation that distribution planners may encounter when applying traditional allocation procedures. Therefore, its inclusion in the comparison is useful not only as a benchmark, but also as evidence that methods requiring relaxed convergence criteria may compromise the physical realism and consistency of feeder modeling, particularly in planning studies.

The proposed method shows strong practical value. It is data-efficient, requiring only short-term transformer metering and monthly energy data; computationally efficient through LightGBM classification and simple scaling; and readily automatable with Python scripts for all 191 transformers. Most importantly, it yields the smallest downstream error among the tested alternatives, which is the relevant metric when the goal is to run feeder studies such as thermal loading, voltage profiles, and device settings without pervasive instrumentation.

IV. CONCLUSIONS

This work presented a data-driven methodology for transformer-level load allocation in distribution networks without advanced metering infrastructure. The framework integrates clustering of measured load curves, supervised classification using LightGBM, energy-based scaling, and a global consistency adjustment to reproduce real feeder demand under limited measurement availability.

Results showed minimal deviation at the feeder head (MAPE = 0.16%) and the lowest downstream error at the intermediate recloser (MAE = 32.23 kVA; MAPE = 6.60%) for the analyzed case study, confirming the effectiveness of the bottom-up approach compared with traditional top-down proportional rules. The method can support feeder studies related to voltage regulation, loss estimation, and load distribution analysis under limited instrumentation.

Moreover, the framework is designed to operate within standard simulation environments and can be implemented in practical distribution studies where feeder-head measurements and transformer data are available. However, the proposed methodology was validated using one real distribution feeder and one intermediate measurement point associated with a recloser. In addition, the error metrics were calculated only for three representative operating scenarios: maximum, average, and minimum demand. Therefore, the results should not be interpreted as a general demonstration of robustness or

scalability, nor as a full time-series validation of the estimated load curves. Future work will include additional feeders, different network topologies, diverse load compositions, multiple intermediate measurement points, and hourly error evaluation over a 24-hour period, depending on the availability of synchronized measurements.

REFERENCES

- [1] E. Báez, R. Duarte, E. Buzarquis, and P. Gardel, "Load curves estimation of distribution transformers using rna," in *2018 IEEE Biennial Congress of Argentina (ARGENCON)*, 2018, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/ARGENCON.2018.8646235>
- [2] G. Rouwhorst, E. M. S. Duque, P. H. Nguyen, and H. Slootweg, "Improving clustering-based forecasting of aggregated distribution transformer loadings with gradient boosting and feature selection," *IEEE Access*, vol. 10, pp. 443–455, 2022. [Online]. Available: <https://doi.org/10.1109/ACCESS.2021.3137870>
- [3] P. Torres-Bermeo, J. Varela-Aldás, K. López-Eugenio, N. Velasco, and G. Palacios-Navarro, "Optimization of electric transformer operation through load estimation based on the k-means algorithm," *Energies*, vol. 18, no. 14, p. 3755, 2025. [Online]. Available: <https://doi.org/10.3390/en18143755>
- [4] H. Ren, N. N. Schulz, V. Krishnan, and Y. Zhang, "Online static load model estimation in distribution systems," in *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*, 2019, pp. 153–158. [Online]. Available: <https://doi.org/10.1109/ISIE.2019.8781530>
- [5] Y. Tan, C. Liao, Y. Li, Y. Cao, M. Shahidehpour, and C. Chen, "A linear power flow model for balanced distribution network with droop-controlled dstatcom and voltage controlled dg," *International Journal of Electrical Power & Energy Systems*, vol. 117, p. 105665, 2020. [Online]. Available: <https://doi.org/10.1016/j.ijepes.2019.105665>
- [6] J. Jardini, C. Tahan, S. Ahn, and E. Ferrari, "Distribution transformer loading evaluation based on load profiles measurements," *IEEE Transactions on Power Delivery*, vol. 12, no. 4, pp. 1766–1770, 1997. [Online]. Available: <https://doi.org/10.1109/61.634203>
- [7] P. F. Austnes, C. García-Pareja, F. Nobile, and M. Paolone, "Probabilistic load forecasting of distribution power systems based on empirical copulas," *Sustainable Energy, Grids and Networks*, vol. 42, p. 101708, 2025. [Online]. Available: <https://doi.org/10.1016/j.segan.2025.101708>
- [8] W. Yan, G. Xu, X. Zhang, and R. Zhao, "A forward/backward robust state estimation algorithm for radial and simple loop distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 163, p. 110366, 2024. [Online]. Available: <https://doi.org/10.1016/j.ijepes.2024.110366>
- [9] A. Primadianto and C.-N. Lu, "A review on distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875–3883, 2017. [Online]. Available: <https://doi.org/10.1109/TPWRS.2016.2632156>
- [10] J. Wu and Y. Yu, "Cbr-based load estimation for distribution networks," in *MELECON 2006 - 2006 IEEE Mediterranean Electrotechnical Conference*, 2006, pp. 952–955. [Online]. Available: <https://doi.org/10.1109/MELCON.2006.1653256>
- [11] A. Bragantini, A. Sumper, V. Gadelha, and S. Galceran-Arellano, "Neural network-based three-phase state estimation for unobservable low voltage grids," *International Journal of Electrical Power & Energy Systems*, vol. 167, p. 110568, 2025. [Online]. Available: <https://doi.org/10.1016/j.ijepes.2025.110568>
- [12] M. Fotopoulou, S. Petridis, I. Karachalios, and D. Rakopoulos, "A review on distribution system state estimation algorithms," *Applied Sciences*, vol. 12, no. 21, 2022. [Online]. Available: <https://doi.org/10.3390/app122111073>
- [13] E. Mulliez, M. A. Sanz-Bobi, A. Sanchez, P. Mazidi, A. Gonzalez, and R. Bachiller, "Life characterization of power distribution transformers using clustering techniques," *PHM Society European Conference*, vol. 4, no. 1, 2018. [Online]. Available: <https://doi.org/10.36001/phme.2018.v4i1.129>
- [14] A. Seppälä, *Load research and load estimation in electricity distribution*, 1996. [Online]. Available: <https://publications.vtt.fi/pdf/publications/1996/P289.pdf>
- [15] M. Mahmoodi, S. M. N. R.A., L. Blackhall, and P. Scott, "A comparison on power flow models for optimal power flow studies in integrated medium-low voltage unbalanced distribution systems," *Sustainable Energy, Grids and Networks*, vol. 38, p. 101339, 2024. [Online]. Available: <https://doi.org/10.1016/j.segan.2024.101339>
- [16] B. Taheri, M. Sedighzadeh, M. R. Nasiri, and A. S. Fini, "A non-intrusive load monitoring algorithm based on real-time feature extraction and deep learning model," *International Journal of Electrical Power & Energy Systems*, vol. 168, p. 110637, 2025. [Online]. Available: <https://doi.org/10.1016/j.ijepes.2025.110637>
- [17] W. Kersting and W. Phillips, "Load allocation based upon automatic meter readings," in *2008 IEEE/PES Transmission and Distribution Conference and Exposition*. IEEE, 2008, pp. 1–7. [Online]. Available: <https://doi.org/10.1109/TDC.2008.4517057>
- [18] EATON, "Cymdist – distribution system analysis software package," 2024, accessed: January 2025. [Online]. Available: <https://www.eaton.com/us/en-us/products/utility-grid-solutions/software-modules/distribution-system-analysis-package-CYMDIST.html>



Kevin López-Eugenio received his Engineering degree in Electrical Engineering in 2022 from Escuela Politécnica Nacional, Quito, Ecuador, and a Master's degree in Big Data and Data Science in 2025 from Universidad Tecnológica Indoamérica, Ambato, Ecuador. He is currently working in the Planning Department of a distribution utility in Ecuador. His areas of interest include energy efficiency, distribution system planning, machine learning, artificial intelligence applied to power systems, renewable energies, and smart grids.



Pedro Torres-Bermeo received his Engineering degree in Electrical Engineering from Universidad de Cuenca, Cuenca, Ecuador, and a Master's degree in Big Data and Data Science in 2025 from Universidad Tecnológica Indoamérica, Ambato, Ecuador. He currently works as a chief in the Planning Department and Studies Section at Empresa Eléctrica Ambato Regional Centro Norte S.A. (EEASA). His areas of interest include energy efficiency, distribution system planning, machine learning, artificial intelligence applied to power systems, renewable energies, and smart grids.



Jorge Buele (Member, IEEE) received his B.S. degree in electronic engineering with a focus on instrumentation from Universidad de las Fuerzas Armadas (ESPE), Latacunga, Ecuador, in 2018, and the M.S. degree in integrated management systems from Universidad Internacional de La Rioja, Spain, in 2021. He is currently pursuing the Ph.D. degree in electronic engineering with the University of Zaragoza, Zaragoza, Spain. He has been a full-time Professor at Universidad Tecnológica Indoamérica, Ambato, Ecuador, where he coordinates research projects related to virtual reality. His research interests include automation, extended reality environments, and process control. Mr. Buele is a member of the IEEE EMBS.



Nancy Velasco received her M.S. degree in Automation and Robotics from Universidad Politécnica de Madrid, España, in 2015, and the M.S. degree in Mathematics and Computing from Universidad Internacional de la Rioja, España, in 2018. She is currently a Professor of the Faculty of Computer Science and Electronics with Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador. She is a coauthor of 25 scientific articles in the fields of electronics and mathematics. Prof. Velasco was a Committee Member of the IEEE Ecuador Technical Chapters Meeting in 2016 and 2018. She received recognition for advising winning projects at science fairs in 2018 and 2019 and recognition for her research work from Universidad de las Fuerzas Armadas (ESPE), Sangolquí, Ecuador, from 2016 to 2021.



José Varela-Aldás (Senior Member, IEEE) is an Industrial Engineer from Universidad Técnica de Ambato, holds a Master's degree in Control Systems from Escuela Superior Politécnica de Chimborazo, and a Ph.D. in Electronic Engineering from the University of Zaragoza. He is currently an Associate Professor at Universidad Tecnológica Indoamérica, where he serves as Director of Research. In 2023, he was recognized as the Best Young Researcher by IEEE Ecuador. In 2024, he served as the President of the IEEE Ecuador Robotics and Automation Society.

In 2025, he founded IEEE PUMABOT, the first battlebot league in Ecuador. His research interests include control systems, robotics, IoT, machine learning, and virtual reality.