

Minimal Investment Installation Planning of Smart Meters for Load Balancing Using a Data-Driven Approach

Victor A. Jimenez*, Adrian Will

Grupo de Investigación en Tecnologías Informáticas Avanzadas/Facultad Regional Tucumán/Universidad Tecnológica Nacional, San Miguel de Tucumán 4000, Tucumán, Argentina.

* Corresponding author. Email: victoradrian.jimenez@frt.utn.edu.ar (V.A.J.)

Manuscript submitted April 8, 2026; accepted June 3, 2026; published June 18, 2026.

doi: 10.12720/sgce.15.1.45-61

Abstract: Reassigning the phase connections of selected single-phase customers is an effective strategy for mitigating load imbalance in low-voltage distribution networks. Computational methods designed to optimize phase configurations inevitably depend on individual consumption data and accurate knowledge of customer-to-transformer connectivity. However, during the early deployment stages of smart grids—characterized by low smart meter penetration—such granular information is frequently incomplete or entirely missing, rendering most data-driven approaches inapplicable. To address these constraints, this paper introduces a novel, cost-effective load-balancing framework optimized for environments with limited metering penetration. The proposed strategy minimizes initial capital expenditures by strategically identifying the absolute minimum number of smart meters required for deployment. Using data from these newly installed devices, the methodology simultaneously infers customer phase connections and substation topology. Furthermore, it pinpoints the optimal candidates for physical reallocation, maximizing unbalance mitigation while strictly minimizing the total number of physical switching operations. The framework is validated through simulations utilizing real-world consumption profiles from diverse residential sectors in Tucumán, Argentina. The results demonstrate that a substantial reduction in load imbalance is achievable by instrumenting only 30% of the customer base and rephasing up to 15% of connected loads.

Keywords: data-driven methods, load balancing, phase identification, phase switching

1. Introduction

Our technological world has become dependent on a continuous supply of electricity. The availability of new technologies, coupled with population and industrial growth, leads to a constant increase in energy demand and a need for very high service quality. The technical anomalies associated with this phenomenon frequently violate power quality standards and impose severe operational penalties on electric utilities [1]. Consequently, the Smart Grid paradigm [2] arose in response to these challenges, promoting a more reliable and efficient operation of modern power systems. By integrating information and communication technologies into the electrical infrastructure, Smart Grids provide real-time monitoring capabilities, enabling continuous access to the operational status of the grid at each network node. This data availability enables the implementation of many techniques to optimize the management of electricity distribution and related operations [3].

One feasible technique to implement, given the data available from Smart Grids, is *Load Balancing*. Unlike

medium- and high-voltage transmission networks, low-voltage distribution grids are difficult to configure and manage, leading to systemic phase imbalances. Usually, it comprises many single-phase customers consuming energy unequally, leading to an imbalance in the total load per phase [4]. This imbalance in load distribution constitutes a significant contributor to technical losses and, consequently, leads to substantial operational inefficiencies and management difficulties within the electrical system [5].

Load transfer is one of the most effective load-balancing techniques. This approach consists of reassigning customers to a different phase and distributing the total load evenly across the feeder. There are many data-driven methods for suggesting which customers should change and which new phases they should adopt. However, they require an Advanced Metering Infrastructure (AMI) that can provide accurate load consumption measurements. Moreover, most of them also need to know the substation and phase where the customers are connected. This information is not always available [6], but it can be estimated from load consumption measurements.

In Argentina, the implementation of Smart Grids is slow due to the reduced investment capacity of the electricity distribution utilities [7]. It is currently in the early stages, focused mainly on acquiring and testing smart meters and communication technologies [8]. The installation of new smart meters is done on a small scale, i.e., one neighborhood or isolated residential area at a time. Information about customers' phase connections is also often incomplete, outdated, or unknown [9]. This lack of information is caused by the addition of new customers, frequent changes due to repairs and maintenance, adaptation and redistribution of installed power, and insufficient process automation, among other reasons [10]. Even though regulatory norms exist and distribution companies try to follow them, distinguishing between the phases in the field is not easy. Under these conditions, applying existing load-balancing techniques is a real challenge.

This paper proposes a new strategy for installing a minimum required number of smart meters in new areas to enable data-driven load balancing with significantly reduced investment. It can be applied when replacing conventional meters with smart meters is desired. First, the number of smart meters required to obtain sufficient measurements for applying load-balancing techniques is analyzed. The analysis suggests the minimum number of meters to be installed in an uncovered area. After the smart meters are deployed, a phase identification procedure is applied to establish the association between customers and their corresponding substation phases by using the measurements collected from the installed devices. The phase identification method used in this work is based on Genetic Algorithms [9] because it provides an accurate response with only a few days of data, even in scenarios where smart meters are installed for only a subset of customers. Finally, the optimal phase configuration is obtained using a data-driven load-balancing method. The load balancing technique adopted in this study [11] aims to mitigate phase imbalance while minimizing the number of required switching operations, thereby reducing implementation costs. Combining the above methods for phase identification and load balancing produces accurate results with significantly less investment than other methods. The effectiveness of the method was assessed through simulations based on real consumption data collected from a residential neighborhood in Tucumán Province, Argentina.

The rest of this paper is organized as follows: Section presents a review of existing methods for phase identification and load balancing; Section 2 describes all the steps to balance the load from a transformer substation, including the acquisition of smart meters, phase identification and the phase switching; Section 3 presents the simulation results obtained from the implementation of the proposed methodology; Finally, Section 4 summarizes the main conclusions and discusses possible directions for future research.

2. Literature Review

This study restricts its scope to load-balancing strategies that rely exclusively on load transfer, deliberately avoiding approaches that necessitate specialized hardware for rephasing. Consequently, techniques utilizing reactive energy compensation—such as capacitor banks designed for power factor correction [12]—as well as solutions involving active power filters, current injection devices, or generation systems [13] are excluded. Such hardware-dependent methodologies incur high procurement costs and often fail to provide a sustainable solution when severe, long-term imbalances occur [14]. Similarly, online load-transfer techniques featuring remotely managed phase-switching devices [15] are omitted; while they successfully maintain continuous grid balance, the required capital expenditures for installing these switches remain prohibitive, particularly since standard commercial smart meters lack this functionality natively.

Conversely, offline load-transfer alternatives for load balancing are extensive, encompassing analytical formulations [16], combinatorial optimization [17], and fuzzy logic frameworks [18, 19]. Among these, heuristic optimization stands out as the most prevalent approach, utilizing objective functions tailored to specific constraints. For instance, Immune Algorithms have been deployed to minimize both neutral currents and reconnection costs [20], while Particle Swarm Optimization (PSO) has targeted phase current disparities [21]. Hybrid frameworks also exist, such as Fuzzy Evolutionary PSO [22], which integrates swarm intelligence with evolutionary mechanisms. Although Genetic Algorithms (GAs) are widely favored, conventional architectures still dominate the literature [23–26]. These formulations define their objective functions through diverse imbalance metrics, often penalized by the frequency of phase re-allocations. Ultimately, selecting the optimal algorithm depends strictly on the data available in the target scenario, as some models require phase-specific current readings, while others rely on voltage or power profiles. In this context, the method developed in [11] introduces a specialized class of GAs characterized by minimal data overhead, high operational flexibility, and the prevention of redundant phase changes. Due to its reliance on easily accessible inputs and its capacity to operate without universal customer metering, this specific approach is adopted in the present work.

A fundamental prerequisite for any of the aforementioned load-balancing techniques is accurate knowledge of each customer's phase connection, which can be derived from smart meter data. Certain phase identification techniques depend on hardware-centric solutions, such as Power Line Carrier (PLC) systems [27], GPS-synchronized timers [28], or high-precision phasor measurement units [29]. These options are disregarded here because their implementation requires significant financial investments, extensive field deployment time, and specialized personnel [30].

Alternative data-driven methodologies can operate effectively under these constrained conditions. These analytical methods are generally categorized by their primary input variable: voltage profiles or consumption metrics. Within voltage-based approaches, simple models evaluate linear correlation coefficients between adjacent consumers [31], whereas more advanced frameworks employ Harmonic Voltage Correlation [32]. Clustering techniques are also prominent, utilizing algorithms like K-Means [33], Maximum Spanning Trees [33], or Constraint-driven Hybrid Clustering to capture arbitrary cluster geometries [34–36]. Additionally, supervised models like Artificial Neural Networks have been explored [37], though they demand prior knowledge of certain phase connections to construct a valid training set. A shared limitation of these voltage-centric frameworks is their strict dependence on high-frequency, highly precise voltage time-series data.

Regarding consumption-driven strategies, early implementations utilized Integer Programming [38], which exhibits a high sensitivity to missing data or measurement noise. More recently, network topologies have been inferred through Principal Component Analysis combined with graph theory [10]. Similarly, the

authors in [30] leveraged linear correlation to analyze consumption variations between end-users and substation phases; unlike older methods, this framework remains robust even under low smart meter penetration. Although enhancements proposed in [39] subsequently boosted its reliability, both algorithms process consumers sequentially, causing errors to propagate and undermining estimation accuracy. To overcome these limitations, a recent approach [9] introduces a heuristic optimization framework based on Genetic Algorithms. Designed specifically for low-penetration environments and incomplete datasets, it operates efficiently using 15-minute average power intervals and requires fewer total samples. Given that its structural characteristics align precisely with the practical constraints defined in this research, this latter methodology is selected for our implementation.

3. Load Balancing from Scratch

3.1. Evaluation of the Load Imbalance

Before considering any load-balancing method, it is essential to quantify the existing imbalance, which requires continuous electricity consumption measurements from each phase at the transformer substation via a commercial meter sampling at regular 15-minute intervals. These measurements will be used to calculate the imbalance index and decide if it is necessary to balance the load in that substation.

Several criteria have been proposed to quantify load imbalance in three-phase electrical systems [25]. Among the most widely adopted metrics are the indices established by the IEC, IEEE, and ANSI standards [40], which rely on symmetrical component relationships derived from the mathematical decomposition of voltage or current phasors. Although these approaches provide highly accurate assessments, they require measuring both magnitudes and phase angles for each phasor [41]. To avoid these intensive data requirements, this study adopts the imbalance formulation defined in the NEMA MG-1 standard [42], adapted for current magnitudes. This indicator is expressed by Eq. (1), where $I_{max,t}$ represents the current magnitude of the most heavily loaded phase at time interval t , and $I_{avg,t}$ denotes the mean current across the three phases during the same period. These quantities are derived from 15-minute average active current profiles, aligning with the standard sampling intervals deployed in commercial energy metering infrastructure.

$$B_t = \frac{I_{max,t} - I_{avg,t}}{I_{avg,t}} \times 100 \quad (1)$$

A different indicator is required to evaluate load imbalance over an extended interval containing multiple measurements. In this study, the imbalance metric B_M introduced in [11] is adopted. This index is defined by Eq. (2), where M denotes the total number of samples and (B_t) corresponds to the imbalance value calculated at time (t). One of the main advantages of the B_M metric is that it preserves the same units as B_t , which facilitates its interpretation and comparison.

$$B_M = \sqrt{\frac{\sum_{t=1}^M B_t^2}{M}} \quad (2)$$

According to the NEMA MG-1 standard, the maximum permissible voltage imbalance is 1% [40], which is typically associated with a current imbalance ranging between 15% and 20% [41]. In the present study, phase current imbalance is reduced using the B_M metric, with 20% set as the desired limit.

The next step is to estimate how many smart meters must be installed to obtain sufficient information to identify customer phases and balance load on the three-phase network correctly.

3.2. Recommended Number of Smart Meters to Install

Load imbalance may arise from several factors, including: (i) differences in consumption levels among

customers (since certain users demand significantly more energy than others), (ii) an unequal distribution of customers connected to each phase (where consumption profiles are similar, but some phases support a larger number of connections), or (iii) a combination of both. The second reason is consistent with residential neighborhoods and apartment buildings since residential consumers within the same socio-economic sector typically exhibit statistically similar or homogeneous baseline consumption profiles. However, it is not clear which phase provides energy to each customer. For this reason, this work focuses on this particular case to determine in advance (without customer measurements) how many smart meters should be installed.

In its simplest formulation, the load imbalance at a distribution transformer substation can be mitigated by transferring selected customers from the most heavily loaded phase to the remaining phases. Determining which customers should be reassigned requires access to their electricity consumption profiles, meaning that those users must be equipped with smart meters. Therefore, it is necessary to focus on customers with measurements available and connected to the most loaded phase. The probability of identifying these customers within the set of users connected to the substation depends on the number of smart meters installed by the substation that provides energy. If there are few smart meters available, the probability of identifying suitable customers is low. On the contrary, it grows rapidly as more meters are installed.

To illustrate the proposed approach, consider a simplified electrical distribution system composed of $N = 200$ single-phase customers supplied by a transformer substation, as depicted in Fig. 1. In this scenario, customers are randomly distributed across phases, but more are in phase A. This imbalance in the number of customers per phase causes the load imbalance. Consider the following case:

- (40%) customers in phase A,
- (30%) customers in phase B,
- (30%) customers in phase C.

Electricity consumption is recorded at fixed time intervals using 15-minute average samples for only $n \leq N$ customers. Consequently, consumption data are available for only a subset of the users connected to the distribution network (hereafter referred to as Measured Customers). Assuming that all nodes exhibit similar consumption patterns, the resulting imbalance metric B_M is approximately equal to 20%.

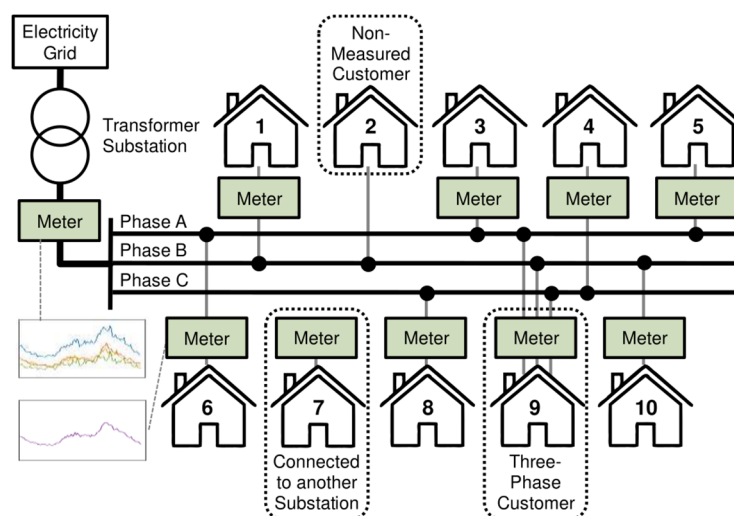


Fig. 1. Classification of low-voltage distribution network customers analyzed for load balancing.

In this example, $k = 13$ customers should be evenly moved from phase A to phases B and C. After the rephasing process, the grid must have an equal number of nodes in each phase. If $n = 20$ of the 200

customers (only 10%) are randomly selected to install smart meters, the probability that 13 of them are connected to the most loaded phase is 0.02. Instead, if $n = 40$ customers are randomly selected to get smart meters (20% of the total), the probability that 13 of them are connected to the most loaded phase rises to 0.9.

In general, for a network with N customers in total where n smart meters will be installed on randomly chosen customers, the probability of choosing at least k customers connected to the most loaded phase is given by Eq. (3), where N_A , N_B , and N_C denote the number of customers connected to each phase, with phase A corresponding to the most heavily loaded phase.

$$prob(k) = \sum_{i=0}^{n-k} \frac{\binom{N_A}{n+i} \binom{N_B+N_C}{n-k-i}}{\binom{N}{n}}, k < n, N_A \geq N_B \geq N_C, N = N_A + N_B + N_C \quad (3)$$

Fig. 2 shows how this probability varies with different numbers of measured customers (n) for three cases: 40%, 45%, and 50% of customers supplied by the most heavily loaded phase (13, 23, and 33 extra customers, respectively). The dotted lines indicate the points beyond which there are enough measured customers in the most loaded phase to fully correct the load balancing problem (probability greater than or equal to 0.9). Table 1 shows a detailed list of the analyzed cases. It can be seen that an imbalance of up to 35% can be corrected by installing meters for only 30% of customers. The number of smart meters should be adequate to guarantee, with high probability, that the measured subset includes a sufficient number of customers connected to the most heavily loaded phase to effectively mitigate the imbalance condition.

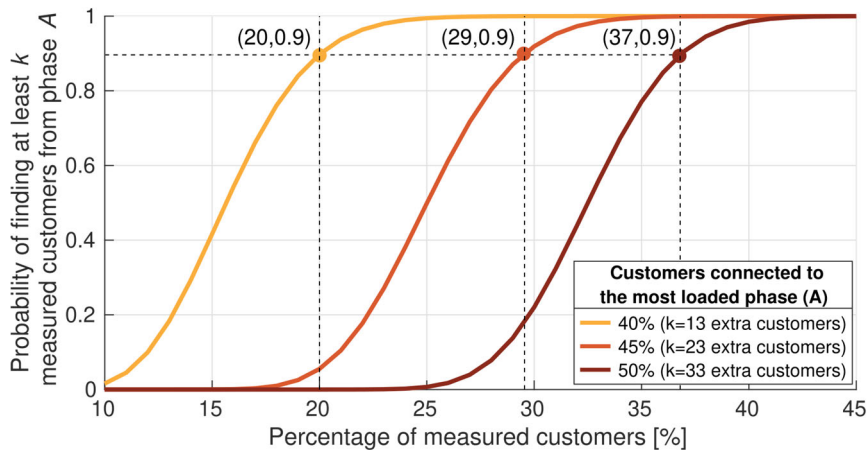


Fig. 2. Probability of identifying a sufficient number of measured customers connected to the most heavily loaded phase to achieve load balancing when customers are selected at random (for $n = 200$ customers in total).

Table 1. The Percentage of Measured Customers Needed to Achieve at Least a 90% Probability of Balancing the Load Correctly for 200 Customers in Total

Customers at the most loaded phase	Equivalent load imbalance (B_M metric)	Customers to be switched	Measured customers			
70	35%	5%	Slight	3	2%	7%
75	38%	10%	Mild	8	4%	15%
80	40%	20%	Moderate	13	7%	20%
90	45%	35%	Severe	23	12%	30%
100	50%	50%	Very Severe	33	17%	37%
120	60%	80%	Extreme	53	27%	48%

Although this analysis simplified the problem, it helps us determine in advance the number of meters to be installed in the neighborhood under study. Once the meters are installed, the next step is to initiate an appropriate metering period to apply data-driven load-balancing methods. So, in the first place, it is essential to determine the substation and phase connection for each customer using some phase identification method. The following section describes how to use these new measurements to estimate the substation and phase assignment associated with each customer using one of the existing methods based on data analysis.

3.3. Substation and Phase Identification

In this study, the phase identification approach presented in [9] is adopted for phase identification. It enables the detection of customers directly connected to a given transformer substation and, in that case, identifies which phase supplies them (all in a single method). Compared to other methods, it provides good accuracy with little data, even when only a few customers are measured or some data is missing. The proposed approach relies on a modified Niching-type Genetic Algorithm known as *Deterministic Crowding* [43]. It has certain advantages over traditional Genetic Algorithm architectures: it preserves population diversity and prevents premature convergence. Also, it has a few parameters that are easy to adjust, allowing us to obtain multiple solutions and thus determine the reliability of the response by analyzing its uniqueness [44].

Each individual in the Genetic Algorithm represents a solution to the problem, determining the phase number for each customer. This codification allows us to place the problem as a classification problem. Each solution is represented with a vector of integer coordinates $S = (s_1, s_2, \dots, s_n)$. The element $s_i \in \{0, 1, 2, 3\}$ represents the phase number ($A = 1, B = 2, \text{ or } C = 3$) proposed for customer i , or indicates with 0 when it is not directly connected to the substation (i.e., when it is connected to another substation). The vector size, n , is equal to the total number of *Measured Customers*. Fig. 3 illustrates the solution for the networks presented in Fig. 1 with 10 customers in total. In this example, 8 customers are connected through single-phase service lines, 1 customer has a three-phase connection, and the remaining customer is supplied by a different transformer substation. Only single-phase customers are considered in the algorithm since load balancing will be performed by changing the phase of this type of customer.

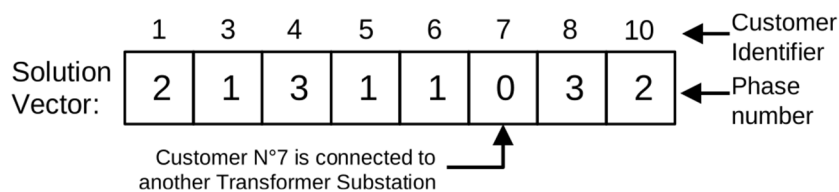


Fig. 3. Solution representation of the phase identification problem in genetic algorithms.

Niching Algorithms require the definition of a distance measure in the solution space (D) to compare solutions. In this work, the Hamming distance is adopted, which is defined as the number of elements whose values differ between two solutions.

The fitness function is based on a linear correlation analysis of variations in electricity consumption profiles in order to assess the quality of the candidate solutions. It was designed with the understanding that the correlation between customers' measurements and the measurements corresponding to the actual connected phase is greater than the correlation obtained with data from the remaining phases [9]. In addition, using variations of load consumption improves the phase identification process [30]. The analysis was performed using 15-minute average active power measurements collected for each customer and for each phase at the transformer substation. The calculation of the fitness value for each solution consists of

applying the following steps:

1. Group the customers' consumption profiles by the phase number indicated in solution S , creating one group per phase and an additional group for customers connected to another substation transformer.
2. Estimate the total load consumption for each group in solution S by summing the individual load profiles for each customer. This step uses Eq. (4), where $h_i(t)$ is the measured load consumption from customer i at time t , n is the number of measured customers processed by the Genetic Algorithm, and S is the vector solution.

$$p_j^S(t) = \sum_{i=1, s_i=j}^n h_i(t), \forall j \in \{1, 2, 3\} \quad (4)$$

3. Calculate the variations in the estimated load demand obtained from the aggregation of the individual consumption profiles by applying Eq. (5).

$$Vp_j^S(t) = p_j^S(t) - p_j^S(t - 1), \forall j \in \{1, 2, 3\} \quad (5)$$

4. In the same way, use Eq. (6) to calculate the variations of the load consumption from each phase, where $p_j(t)$ is the measured value of load consumption from the substation's phase j at time t .

$$Vp_j(t) = p_j(t) - p_j(t - 1), \forall j \in \{1, 2, 3\} \quad (6)$$

5. Perform a linear correlation analysis between the variations of the combined load consumption from customers and the variations of the measured load consumption from each phase. This analysis yields three Pearson's correlation coefficients: R_1 , R_2 , and R_3 .
6. Calculate the fitness value using Eq. (7), which combines and normalizes the three correlation values.

$$f(R_1, R_2, R_3) = \frac{1}{3} \sum_{j=1}^3 \frac{1}{1 + \left\| \frac{R_j - 1}{\lambda} \right\|^2} \quad (7)$$

The parameter λ controls the shape of the function and defines the transition region between good and bad solutions. Preliminary tests indicated an adequate value of 0.3.

The algorithm uses two genetic operators: a Crossover Operator and a Local Search Operator designed for the phase identification problem. *Uniform Crossover* was used because it enables effective search-space exploration and is suitable for discrete codings. It randomly combines the coordinates of each parent with probability 0.5. On the other hand, the Local Search Operator is used as a common mutation operator, which modifies a single coordinate of a randomly selected solution for each generation to improve the algorithm's convergence. A complete description of the operator can be found in [9]. Finally, the stopping criterion counts the number of consecutive generations without finding a better solution than the previous one. 100 consecutive generations are enough to ensure solutions with good precision.

An important consideration when implementing the adopted phase identification step within the broader load-balancing method is determining the data window required to ensure high topological accuracy. This requirement inherently depends on smart meter penetration rates and data availability [30]. Fig. 4 shows the minimum measurement duration (in days) required by the adopted methodology to achieve a phase identification accuracy exceeding 95% across different network scales, as reported in [9]. In the context of the current load-balancing experiments, Fig. 4 serves as a reference to define the input datasets. For instance, in a scenario with 300 total customers and a low meter penetration of 30%, the phase identification algorithm requires a 51-day measurement window to guarantee topological reliability. Conversely, as meter penetration increases to 60% or 90%, the experimental data requirements are dynamically reduced to 21 and 7 days, respectively. This performance evaluation is critical because the

subsequent load-balancing optimization phase will operate with a similar time window. Consequently, Fig. 4 allows for a precise quantification of the identification accuracy achieved within this specific timeframe, thereby ensuring that subsequent phase-swapping operations are executed based on highly reliable topological data.

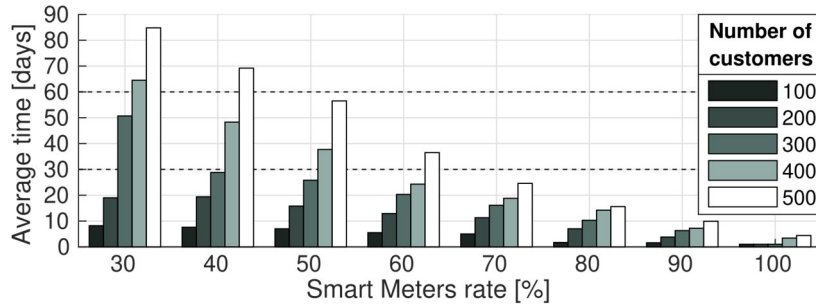


Fig. 4. Average time required to achieve 95% accuracy for electrical networks of different sizes and varying smart meter penetration ratios [9].

After the substation and phase identification, the next step is to analyze which customers should be switched to balance the system.

3.4. Load Balancing at the Substation

The selected load-balancing method addresses the problem as a heuristic optimization due to its complexity [11]. The method is based on a Genetic Algorithm, which is particularly well-suited to this type of problem and finds good solutions in a reasonable time. Deterministic Crowding without mutations is the algorithm used due to its multiple advantages mentioned before, even though the main objective is not to find more than one solution to the problem.

The solutions are represented as vectors with n integer coordinates. Each coordinate corresponds to a customer taken in an arbitrary but fixed order. It is assigned a value from $\{1, 2, 3\}$ to indicate the phase A, B, or C to which the customer should be connected to the feeder (only single-phase customers are considered). It is very similar to the coding used in the phase identification step, but the value 0 is not considered in this case.

The problem has two opposite objectives: minimize the imbalance and the number of customers who must be reassigned. The imbalance is measured using the B_M index in Eq. (2), which is calculated from phase-current estimates obtained from the substation transformer. These estimations are carried out by subtracting the individual customer's load from the initial phase's consumption profile, then adding it to the new phase determined by the solution. The current is then obtained from this estimated load per phase. On the other hand, the number of changes required is calculated from the known initial customers' phases, estimated in the previous step (described in Section 3.3). Given a solution S , the number of phase changes N_S is calculated using Eq. (8), where n is the number of measured customers, and C_i is defined as 0 if the new phase S_i assigned to customer i matches the initial phase x_i , and 1 otherwise. Note that only customers whose phases were estimated in the previous step are considered in the load-balancing process.

$$N_S = \sum_i^n C(S_i), C(S_i) = \begin{cases} 0 & \text{if } S(i) = x_i \\ 1 & \text{if } S(i) \neq x_i \end{cases} \quad (8)$$

The fitness function is defined by Eq. (9), where n is the number of measured customers, B_M is the imbalance corresponding to the solution S , B_T is the target imbalance value, and N_S is the number of phase changes required for load balancing. The parameter $a \in [0, 1]$ enables a trade-off between the two

objectives, typically based on operating conditions. In this work, we use $a = 0.2$ to compensate for the magnitudes of each term and give them equal importance. The B_T parameter is set following operational restrictions (20% in our case). Once this target value is reached, the algorithm prioritizes minimizing the number of reconfiguration changes required by the solution. Consequently, solutions with an imbalance index below B_T are assigned progressively lower fitness values, and further improvement is only possible by reducing the number of switching operations. The selection of $a = 0.2$ is calibrated to balance the objective functions around the system’s practical operating point. In real-world distribution networks, a residual imbalance of approximately 15% represents a highly satisfactory stabilization target, which is typically achieved by keeping physical customer reallocations under 20% to manage logistical and labor costs. Under these target conditions, the weighted components of the fitness function operate within the same order of magnitude (10^1). This alignment prevents scaling bias and ensures proportional selection pressure during the evolutionary search.

$$f(S) = (1 - \alpha)(B_T + |B_M - B_T|) + \alpha \frac{100N_S}{n} \tag{9}$$

Furthermore, this weighting structurally enforces realistic utility planning heuristics. From an operational perspective, a 1% reduction in load imbalance compensates for up to a 4% increase in affected customers. In typical low-voltage distribution sectors—where transformer nodes routinely serve more than 200 customers—the chosen $a = 0.2$ threshold ensures that the penalty incurred by shifting up to 8 users is offset by the technical benefit of the 1% imbalance mitigation. Consequently, the algorithm is constrained to prioritize physical interventions only when they deliver a significant and justifiable return in load imbalance, automatically discarding solutions that demand excessive field labor for marginal technical gains.

The *Hamming* function is used as the distance metric in the solution space required by DC. It allows us to determine whether the phases obtained from two solutions differ, regardless of the numerical difference between the phase values. Regarding the crossover operator, *Uniform Crossover* is used once again. This operator is particularly suitable for the problem, as each gene from the parent solutions has an equal probability of being inherited by the offspring. Note that the DC algorithm in the load-balancing method does not incorporate mutation operators.

Fig. 5 shows each step of the load-balancing procedure applied when the imbalance exceeds a predefined threshold.

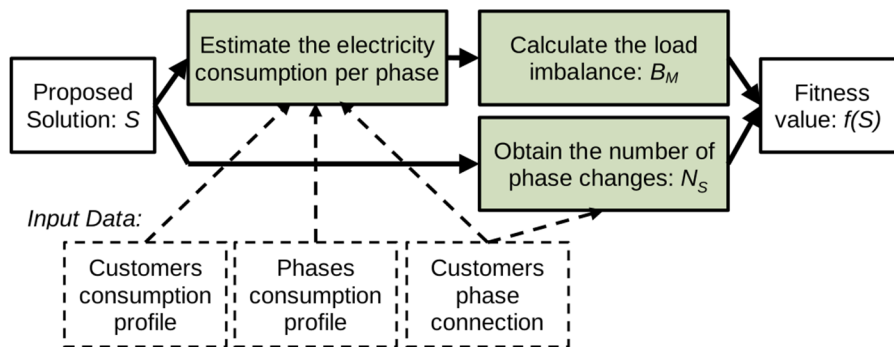


Fig. 5. General scheme of the load balancing process for finding customer phase assignments using Genetic Algorithms, in scenarios where phase reconfiguration is performed manually.

One part of the data will be used to evaluate the solutions with the fitness function, while another subset of the data is reserved to validate the final solutions and determine their durability. A solution will only preserve load balance over a limited time horizon if consumption patterns change rapidly, for example, in

fast-growing urban areas or regions with extensive use of air-conditioning systems, among other factors. Consequently, an adequate observation period depends on the dynamics and historical variability of electricity consumption. The amount of data used in the optimization process must therefore be sufficient to ensure that the resulting phase configuration remains stable and maintains three-phase balance throughout the validation period. One week for the search process and three weeks to validate the final solution should be adequate, because the variations in this problem are slow, requiring the installation of new customers, construction, or similar alterations to consumption patterns.

4. Results and Discussion

The verification of the analysis used to determine the number of customers required for load balancing (presented in Section 3.2) was based on simulations using real customer measurements from Tucumán, Argentina. The data were provided by the local electricity distribution company. In particular, data from a typical residential neighborhood was used, with 159 single-phase customers supplied by the same transformer substation that supplies them with electricity. Smart meters are installed at each customer's residence, providing electricity consumption measurements at 15-minute intervals.

The tests were performed by creating scenarios with varying numbers of smart meters. Additionally, each scenario was tested at different levels of imbalance, assuming that the imbalance is mainly due to an uneven distribution of customers across phases. To simulate this imbalance, the phase for each customer is assigned to generate a load with varying imbalance levels at the transformer substation. Once the phases are established, the total consumption is simulated, thus obtaining the necessary measurements to apply the proposed methodology. So, the different scenarios were generated with the following criteria:

- A percentage of customers with available measurements is selected, from 10% to 100%. Then there will be customers whose electricity consumption is unknown and cannot be used by the algorithms.
- More customers will be connected in one phase than in the other two. The number of customers connected to the most heavily loaded phase ranges from 40% to 60%.
- The customer phases are selected at random, in accordance with the proportion mentioned above.
- Each scenario with a given percentage of smart meters and a given percentage of customers in each phase is repeated 20 times to ensure that the results are not biased by the distribution of randomly selected phases.

Simulated scenarios with varying smart meter penetration rates are used to identify the minimum number of smart meters required for the application of the selected load balancing method.

4.1. Phase Identification Accuracy

The step before the load balancing is the phase identification of each customer with the method detailed in Section 3.3. The results reported in Fig. 4 show that the phase identification method requires 20 days of data or less when the grid has 200 customers in total (assuming at least 30% of customers have available measurements). In this paper, datasets containing up to 30 days of measurements are employed to evaluate the algorithm using the new dataset. Fig. 6 presents the accuracy results obtained for the considered test scenarios.

A population size of 500 individuals and a mutation rate of 10% were used. As shown, the accuracy achieved is consistent with that reported in [9], exceeding 95% for 30% of customers. Accuracy increases as more measured customers become available. On the other hand, the accuracy is more stable as more customers are measured since the deviation of the obtained values decreases as the smart meter rate increases.



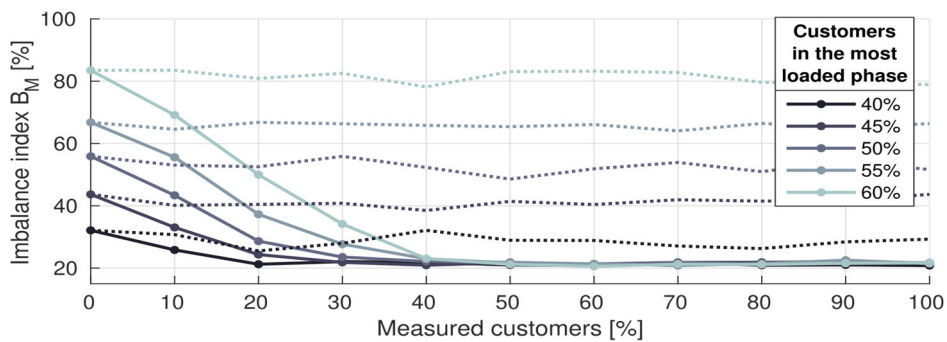
Fig. 6. Performance accuracy of the phase identification algorithm under different numbers of available measured customers.

4.2. Phase Reassignment Efficiency

Once the customer phase has been identified, the algorithm described in Section 3.4 is used to determine the new customer phase to mitigate load imbalance at the transformer substation. The same dataset used to identify the customer phase in the previous step is used for load balancing, but the correct phase information will be used here to test only the balance method. That is, it is assumed that the phase estimation was performed with 100% accuracy for these tests.

The new phase configuration proposed by the Genetic Algorithm is evaluated by simulating the consumption from each substation’s phase after the changes are applied. For this purpose, the first half of the data (corresponding to the first 15 days) is used to determine the new customer phases, while the other half is used to verify whether the solution maintains the balance of the phase loads. As for the Genetic Algorithm parameters, 200 individuals and 500 generations were used, with termination allowed after 50 consecutive generations without finding a better solution. Fig. 7 shows the variation of the imbalance index and the number of required phase-switching operations as a function of the number of customers, considering customer distributions among the phases.

It shows the average values obtained for each case reported with each repetition. A significant reduction in the imbalance index is observed after reconfiguring the customer phases, as shown by comparing the dotted lines (before balancing) with the solid lines (after balancing). The final imbalance value is 20% from 30–40% of B_M customers measured (as a function of the initial imbalance condition). However, when fewer meters are available, it is impossible to fully balance the loads. This behavior relates to the probability of finding suitable metered customers connected to the most heavily loaded phase that can be reassigned to mitigate the imbalance. This observation is consistent with the analysis presented in Section 3.2 regarding the number of metered customers required.



(a)

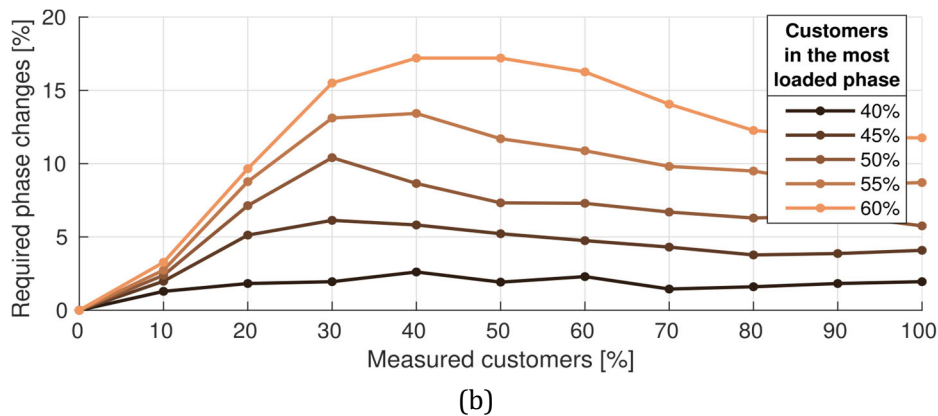


Fig. 7. Results obtained from the load balancing simulations considering different numbers of metered customers available and different imbalance values: (a) Average load imbalance index. (b) Average phase changes.

To explicitly address the limits of this behavior under highly adverse conditions, it is valuable to analyze whether a 30% smart meter penetration rate remains effective under the most severe scenario tested, where a 60% extreme initial imbalance is concentrated on a single phase. Under these conditions, the optimization algorithm's performance depends heavily on the probability that the randomly selected smart meters happen to be installed within that heavily overloaded phase. If a representative portion of the 30% metered subpopulation resides on the overloaded phase, the Genetic Algorithm retains sufficient degrees of freedom to execute targeted phase-swapping operations, achieving a significant reduction in the imbalance index. However, if the unmonitored customers are the primary drivers of this 60% concentration, a 30% penetration rate hits a strict structural ceiling, as the algorithm cannot alter unmetered loads. As shown in Fig. 7(a), while a 30% smart meter penetration rate significantly mitigates moderate imbalances, resolving a larger imbalance requires increasing the number of metered customers. However, in this scenario, the algorithm can still significantly reduce the imbalance (from 80% to 35%) even if it does not reach the desired value below 20%.

5. Conclusions

This paper presented a strategy for acquiring smart meters and applying a load-balancing technique by switching single-phase customers to a single phase in an uncovered area. It suggests installing meters for a small percentage of customers, analyzing data to identify the phases to which they are connected, and then identifying new phases for a subset of them to balance the load in the transformer substation.

The techniques used for phase identification and optimal phase connection scheme selection are adequate for solving the problem when not all customers have measurements available. Based on the proposed analysis and tests, it is concluded that for relatively low imbalance levels (up to 60%), load balancing can be achieved with only 30% of the metered customers selected at random. Only 10% of customers, or fewer, should be switched to achieve this performance. However, if 40% of customers are metered, load balancing can be ensured for more severe cases while maintaining the number of customers to be changed.

The main advantage of our approach is that, by not needing to meter all customers, the investment in smart meters decreases. The penetration of smart meters can be done gradually, allowing a larger area to be covered in less time. From another perspective, it would allow balancing a larger number of substations with the same investment in meters.

Future work will include additional aspects of the search for a new customer phasing scheme, including

reducing voltage drop at the end points of the distribution feeder.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Victor A. Jimenez; methodology, Victor A. Jimenez; software, Victor A. Jimenez; validation, Victor A. Jimenez; formal analysis, Adrian L. E. Will; investigation, Victor A. Jimenez; resources, Adrian L. E. Will; data curation, Victor A. Jimenez; writing—original draft preparation, Victor A. Jimenez; writing—review and editing, Adrian L. E. Will; visualization, Victor A. Jimenez; supervision, Adrian L. E. Will; project administration, Adrian L. E. Will; funding acquisition, Adrian L. E. Will. All authors had approved the final version.

Acknowledgment

We would like to thank the local energy distribution company, Empresa de Distribución Eléctrica de Tucumán (EDET S.A.), for its collaboration and the data supplied.

References

- [1] Agarwal, A., Kumar, S., & Ali, S. (2012). A research review of power quality problems in electrical power system. *MIT International Journal of Electrical and Instrumentation Engineering*, 2(2), 88–93.
- [2] El-hawary, M. E. (2014). The smart grid—State-of-the-art and future trends. *Electric Power Components and Systems*, 42(3–4), 239–250. <https://doi.org/10.1080/15325008.2013.868558>
- [3] U. S. Government (2011). *Nist Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0*. General Books. <https://doi.org/10.6028/NIST.SP.1108r4>
- [4] Tirupathamma, G. L., Kumar, I. V., & Scholar, P. (2017). Dynamic switching of residential customers among three phases by low voltage feeders. *International Journal of Engineering Science*, 14586.
- [5] Fidalgo, J. N., Moreira, C., & Cavalheiro, R. (2019). Impact of load unbalance on low voltage network losses. *Proceedings of 2019 IEEE Milan PowerTech* (pp. 1–5). <https://doi.org/10.1109/PTC.2019.8810710>
- [6] Byun, H. J., Zheng, Y. P., Choi, S. J., & Shon, S. G. (2018). New identification method for power transformer and phase in distribution systems. *Applied Mechanics and Materials*, 878, 291–295. <https://doi.org/10.4028/www.scientific.net/AMM.878.291>
- [7] Tripaldi, J. C., & Moreno, D. A. (2015). AEA 92559—1: Towards an Argentinean smart grid vision. *Proceedings of 2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)* (pp. 768–773). <https://doi.org/10.1109/ISGT-LA.2015.7381252>
- [8] Donato, P. G., Carugati, I., Strack, J., Maestri, S., Orallo, C., Hadad, M., & Funes, M. (2018). Overview of the status of smart metering systems in Argentina and future perspectives. *Proceedings of 2018 IEEE Biennial Congress of Argentina (ARGENCON)* (pp. 1–7). <https://doi.org/10.1109/ARGENCON.2018.8646080>
- [9] Jimenez, V. A., & Will, A. (2021). A new data-driven method based on Niching Genetic Algorithms for phase and substation identification. *Electric Power Systems Research*, 199, 107434. <https://doi.org/https://doi.org/10.1016/j.epsr.2021.107434>
- [10] Pappu, S. J., Bhatt, N., Pasumarthy, R., & Rajeswaran, A. (2018). Identifying topology of low voltage distribution networks based on smart meter data. *IEEE Transactions on Smart Grid*, 9(5), 5113–5122. <https://doi.org/10.1109/TSG.2017.2680542>

- [11] Jimenez, V. A., Will, A. L. E., & Lizondo, D. F. (2021). Phase reassignment for load balance in low-voltage distribution networks. *International Journal of Electrical Power & Energy Systems*, 137, 107691. <https://doi.org/10.1016/j.ijepes.2021.107691>
- [12] Korovkin, N. V., Vu, Q. S., & Yazenin, R. A. (2016). A method for minimization of unbalanced mode in three-phase power systems. *Proceedings of 2016 IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference (EIconRusNW)* (pp. 611–614). <https://doi.org/10.1109/EIconRusNW.2016.7448257>
- [13] Shahnian, F., Wolfs, P. J., & Ghosh, A. (2014). Voltage unbalance reduction in low voltage feeders by dynamic switching of residential customers among three phases. *IEEE Transactions on Smart Grid*, 5(3), 1318–1327. <https://doi.org/10.1109/TSG.2014.2305752>
- [14] Bao, G., & Ke, S. (2019). Load transfer device for solving a three-phase unbalance problem under a low-voltage distribution network. *Energies*, 12(15), 2842. <https://doi.org/10.3390/en12152842>
- [15] Haq, S. U., Arif, B., Khan, A., & Ahmed, J. (2018). Automatic three phase load balancing system by using fast switching relay in three phase distribution system. *Proceedings of 2018 1st International Conference on Power, Energy and Smart Grid (ICPESG)* (pp. 1–6). <https://doi.org/10.1109/ICPESG.2018.8384514>
- [16] Mansani, S., & Udaykumay, R. Y. (2016). An optimal phase balancing technique for unbalanced three-phase secondary distribution systems. *Proceedings of 2016 IEEE 7th Power India International Conference (PIICON)* (pp. 1–6). <https://doi.org/10.1109/POWERI.2016.8077172>
- [17] Grigoras, G., Gavrilas, M., Neagu, B.-C., Ivanov, O., Tristiu, I., & Bulac, C. (2019). An efficient method to optimal phase load balancing in low voltage distribution network. *Proceedings of 2019 International Conference on Energy and Environment (CIEM)* (pp. 323–327). <https://doi.org/10.1109/CIEM46456.2019.8937577>
- [18] Chitra, R., & Neelaveni, R. (2011). A realistic approach for reduction of energy losses in low voltage distribution network. *International Journal of Electrical Power & Energy Systems*, 33(3), 377–384. <https://doi.org/10.1016/j.ijepes.2010.08.033>
- [19] Ma, K., Fang, L., & Kong, W. (2020). Review of distribution network phase unbalance: Scale, causes, consequences, solutions, and future research direction. TechRxiv preprint, <https://doi.org/10.36227/techrxiv.11401056.v2>
- [20] Huang, M.-Y., Chen, C.-S., Lin, C.-H., Kang, M.-S., Chuang, H.-J., & Huang, C.-W. (2008). Three-phase balancing of distribution feeders using immune algorithm. *IET Generation, Transmission & Distribution*, 2(3), 383–392. <https://doi.org/10.1049/iet-gtd:20070206>
- [21] Toma, N., Ivanov, O., Neagu, B., & Gavrila, M. (2018). A PSO algorithm for phase load balancing in low voltage distribution networks. *Proceedings of 2018 International Conference and Exposition on Electrical and Power Engineering (EPE)* (pp. 0857–0862). <https://doi.org/10.1109/ICEPE.2018.8559805>
- [22] Schweickardt, G. A., & Wiman, G. (2018). Fepso multi-objective metaheuristic. An application for load balance analysis in low-voltage distribution networks. *Energética*, 41, 33–48. <https://doi.org/10.15446/energetica>
- [23] Wang, K., Skiena, S., & Robertazzi, T. G. (2013). Phase balancing algorithms. *Electric Power Systems Research*, 96, 218–224. <https://doi.org/10.1016/j.epsr.2012.10.012>
- [24] Abril, I. P. (2016). Multi-objective optimization of the balancing of phases in primary distribution circuits. *International Journal of Electrical Power & Energy Systems*, 82, 420–428. <https://doi.org/10.1016/j.ijepes.2016.03.047>
- [25] Homaei, O., Najafi, A., Dehghanian, M., Attar, M., & Falaghi, H. (2019). A practical approach for

- distribution network load balancing by optimal re-phasing of single phase customers using discrete genetic algorithm. *International Transactions on Electrical Energy Systems*, 29(5), e2834. <https://doi.org/10.1002/2050-7038.2834>
- [26] Ivanov, O., Neagu, B.-C., Gavrilas, M., Grigoras, G., & Sfintes, C.-V. (2019). Phase load balancing in low voltage distribution networks using metaheuristic algorithms. *Proceedings of 2019 International Conference on Electromechanical and Energy Systems (SIELMEN)* (pp. 1–6). <https://doi.org/10.1109/SIELMEN.2019.8905900>
- [27] Lisowski, M., Masnicki, R., & Mindykowski, J. (2019). PLC-enabled low voltage distribution network topology monitoring. *IEEE Transactions on Smart Grid*, 10(6), 6436–6448. <https://doi.org/10.1109/TSG.2019.2904681>
- [28] Chen, C.-S., Ku, T.-T., & Lin, C.-H. (2012). Design of phase identification system to support three-phase loading balance of distribution feeders. *IEEE Transactions on Industry Applications*, 48(1), 191–198. <https://doi.org/10.1109/ICPS.2011.5890874>
- [29] Dusabimana, E., & Yoon, S.-G. (2020). A survey on the micro-phasor measurement unit in distribution networks. *Electronics*, 9(2). <https://doi.org/10.3390/electronics9020305>
- [30] Xu, M., Li, R., & Li, F. (2016). Phase identification with incomplete data. *IEEE Transactions on Smart Grid*, 9(4), 2777–2785. <https://doi.org/10.1109/TSG.2016.2619264>
- [31] Zhang, M., Luan, W., Guo, S., & Wang, P. (2018). Topology identification method of distribution network based on smart meter measurements. *Proceedings of 2018 China International Conference on Electricity Distribution (CICED)* (pp. 372–376). <https://doi.org/10.1109/CICED.2018.8592228>
- [32] Watson, J. D., Welch, J., & Watson, N. R. (2016). Use of smart-meter data to determine distribution system topology. *The Journal of Engineering*, 2016(5), 94–101. <https://doi.org/10.1049/joe.2016.0033>
- [33] Olivier, F., Sutera, A., Geurts, P., Fonteneau, R., & Ernst, D. (2018). Phase identification of smart meters by clustering voltage measurements. *Proceedings of 2018 Power Systems Computation Conference (PSCC)* (pp. 1–8). <https://doi.org/10.23919/PSCC.2018.8442853>
- [34] Wang, W., & Yu, N. (2017). Advanced metering infrastructure data driven phase identification in smart grid. *Proceedings of The Second International Conference on Green Communications, Computing and Technologies*.
- [35] Ni, F., Liu, J. Q., Wei, F., Zhu, C. D., & Xie, S. X. (2017). Phase identification in distribution systems by data mining methods. *Proceedings of 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)* (pp. 1–6). <https://doi.org/10.1109/EI2.2017.8245748>
- [36] Vycital, V., Ptacek, M., Toman, P., Topolánek, D., Drápela, J., & Zamphiropolos, J. (2019). Phase Identification in smart metering pilot project Komorany. *Proceedings of 25th International Conference on Electricity Distribution*. <https://doi.org/http://dx.doi.org/10.34890/944>
- [37] Foggo, B., & Yu, N. (2018). A Comprehensive Evaluation of Supervised Machine Learning for the Phase Identification Problem. *International Scholarly and Scientific Research & Innovation*, 12(6), 419–427.
- [38] Arya, V., Seetharam, D. P., Kalyanaraman, S., Dontas, K., Pavlovski, C. J., Hoy, S., & Kalagnanam, J. (2011). Phase identification in smart grids. *Proceedings of 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)* (pp. 25–30). <https://doi.org/10.1109/SmartGridComm.2011.6102329>
- [39] Jimenez, V. A., Will, A., & Rodriguez, S. (2020). Phase identification and substation detection using data analysis on limited electricity consumption measurements. *Electric Power Systems Research*, 187, 106450. <https://doi.org/https://doi.org/10.1016/j.epsr.2020.106450>
- [40] Arghavani, H., & Peyravi, M. (2017). Unbalanced current-based tariff. *CIGRE-Open Access Proceedings Journal*, 2017(1), 883–887. <https://doi.org/10.1049/oap-cired.2017.0129>

- [41] Casaravilla, G., & Echinope, V. (2005). Imbalances: Study of alternatives for their estimation. Meeting on Power, instrumentation and measurement, *EPIM 05*. Montevideo, Uruguay. <https://hdl.handle.net/20.500.12008/21167>
- [42] National Electrical Manufacturers Association Standard (NEMA). (1998). Motors and generators, MG-1.
- [43] Mahfoud, S. W. (1995). Niching methods for genetic algorithms. *Urbana*, 51(95001), 62–94. <http://hdl.handle.net/2142/22047>
- [44] Will, A., Bustos, J., Bocco, M., Gotay, J., & Lamelas, C. (2013). On the use of niching genetic algorithms for variable selection in solar radiation estimation. *Renewable Energy*, 50, 168–176. <https://doi.org/10.1016/j.renene.2012.06.039>

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).