Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-025-10224-8

RESEARCH ARTICLE-ELECTRICAL ENGINEERING

Parallel Computing for Power Quality Disturbances Analysis and Energy Management in Distribution Power Networks

Husham I. Hussein[1] · Ahmed M. Ghadban[1] · Francisco J. Muñoz[2] · Alejandro Rodríguez[2]

Received: 21 November 2024 / Accepted: 14 April 2025

Abstract

Power quality disturbances (PQDs) and energy management parameters (EMPs) pose significant operational challenges in radial distribution systems. Fast and accurate computational solutions are essential for efficient system performance. This study introduces a novel parallel computing approach for clustering-based PQD and EMP analysis, leveraging the density-based spatial clustering of applications with noise (DBSCAN) algorithm. The proposed method reduces processing times by more than 40% compared to sequential implementations, enabling real-time analysis and facilitating operator decision-making. A comparative evaluation with a parallel k-means approach reveals that k-means achieves slightly higher clustering efficiency, while DBSCAN provides automatic cluster selection. The methodology is validated on the IEEE 33-bus radial distribution system, where new correlations between PQD variations and their impact on EMPs are identified, offering insights for energy management optimization. The findings demonstrate that parallel computing enhances clustering performance, significantly improving computational efficiency and integration of PQD and EMP calculations for large-scale distribution networks.

Keywords Power quality · Parallel computing; k-means · DBSCAN

1 Introduction

Power quality has become a growing concern owing to the increasing prevalence of nonlinear loads, power electronics, distributed generation, renewable energy, and electric vehicles. All of these factors collectively challenge the stability and reliability of power systems [1–3]. The term'power quality' refers to sustaining the magnitude and frequency of the near-sinusoidal rated voltage of a power system. Any deviation of these references of rated magnitudes, which may involve various electromagnetic phenomena known as power quality disturbances (PQDs), produces signal deviations that are notable in power system measurements and energy management parameters and, therefore, enormously affect the efficiency and costs of the system [4]. Calculating the severity index for these deviations provides valuable data on disturbance levels in the distribution system [5]. Thus, accurate measurement of power quality is critically important for monitoring power systems. Furthermore, modern energy sources, such as solar and wind, have increased the complexity of electrical systems and the difficulty of energy management. Designing new power networks based on previous ones has increased the energy supply to consumers without incorporating modern technologies for control, measurement and communication [6].

However, modern technologies can facilitate power quality monitoring and energy management, allowing for a reduction in the costs related to purchasing electrical energy. It aims to achieve energy sufficiency within the required limits, while maintaining the system protected at all times. This efficient energy management can be applied in different phases, from the generation, using renewable energy sources and/or

Husham I. Hussein, Ahmed M. Ghadban, Francisco J. Muñoz, and Alejandro Rodríguez have contributed equally to this work.

✉ Alejandro Rodríguez
arodriguezg@uma.es

Husham I. Hussein
husham.idan.78@gmail.com

Ahmed M. Ghadban
ahmadghagh.78@gmail.com

Francisco J. Muñoz
fjmg@uma.es

1 Electrical Power and Machines Department, University of Diyala, Baqubah, Iraq

2 Electrical Engineering Department, University of Málaga, Málaga, Spain

Published online: 13 May 2025

co-generation, through transportation, applying efficient processes and equipment, and using efficient storage and control devices. These techniques have steadily increased, effectively meeting demandside management goals [7]. The shift towards smart grids and the integration of renewable energy emphasize the importance of energy management systems. These systems are designed to oversee, control, optimize, and manage the generation, transmission and distribution systems. They enable utilities to gather, store, and analyze data from hundreds of thousands of points across national or regional networks [7]. These systems support network modeling, power simulation operations, fault detection, outage prevention and participation in energy trading markets and are crucial for modernizing power networks and advancing the development of smart grids and highly automated energy systems in the future [8]. Many researchers have worked on system simulations, signal detection and power system reconfiguration under different conditions [9]. Nevertheless, there is limited literature on comprehensive power quality measurements for studying energy management parameters using new methods. Clustering algorithms, including k-means, are becoming essential tools for studying and modernizing electric power systems; in [10], k-means is used to analyze the energy characteristics of “prosumers”, categorizing them based on two specific indicators, which can help optimize distribution network operations. Density-based spatial clustering of applications with noise (DBSCAN) is the most famous density-based clustering algorithm and is also used in power systems [9, 11]. However, there are many redundant distance computations in the DBSCAN clustering process, which yield high complexity and low efficiency [12].

In power quality, time–frequency transforms are crucial for analyzing power quality issues because they capture both temporal and frequency variations in electrical signals. Unlike traditional Fourier analysis, which only shows the frequency content, time–frequency transforms reveal transient disturbances that impact power systems, aiding in real-time monitoring and diagnosis [13–16]. The use of time–frequency transforms together with clustering algorithms requires high computational requirements. To make operations fast, efficient, and accurate, parallel computing was applied in power systems, introduced in [17], as a tool to improve the time response and solve complexity problems.

Recent developments in DBSCAN methodologies have broadened its applicability and effectiveness across various domains. For example, an extended DBSCAN algorithm has been developed to identify clusters with varying densities by assigning a regional density value to each object and utilizing a dynamic radius for core points, thereby enhancing its ability to detect closely packed clusters [16]. Similarly, EDBSCANH, designed for large satellite image datasets, employs a histogram-based approach to efficiently manage data and introduces additional parameters to better accommodate both dense and sparse regions [18]. Additionally, the HF-DBSCAN algorithm has been refined for analyzing complex trajectory data in smart campus environments, integrating distance measures and grid division methods to enhance clustering accuracy and adaptability for non-convex data structures [19]. DBSCAN clustering has also been incorporated to improve feature selection and execution efficiency [20, 21]. A deep clustering algorithm specifically designed for large datasets has also been proposed [22, 23], although its effectiveness for density-based methods such as DBSCAN remains limited compared to hierarchical approaches such as BIRCH. Taken together, these studies highlight ongoing efforts to refine clustering techniques and optimize computational frameworks for various applications. Moreover, conventional machine-learning approaches for PQD classification are limited by their reliance on sequential processing, leading to longer detection and response times. These limitations decrease the ability to implement real-time monitoring and control strategies in modern smart grids. Parallel computing is crucial in analyzing and mitigating PQD by enabling real-time, high-resolution processing of large-scale power system data.

The present study exploits parallel computing techniques and existing resources, such as parallel clustering algorithms DBSCAN and k-means, to enhance the speed and accuracy of PQD detection. By distributing computational tasks across multiple processing units, the approach significantly reduces execution time, improves scalability, and enables more responsive PQD mitigation strategies. This advancement strengthens the foundation for real-time energy management in smart grids, ensuring greater reliability and stability in both grid-connected and islanded microgrid environments.

The ob jective of this work is to leverage the advantages of parallel computing using DBSCAN to address the challenges posed by PQDs and EMPs in radial distribution systems. These advantages translate into a reduction in the time required to perform the analysis and identification of key power quality metrics such as THD (voltage and current), voltage imbalances and flicker, while determining the associated EMPs. A comparative study with the parallel k-means method is performed to evaluate its effectiveness. The proposed DBSCAN-based approach is validated on the IEEE 33-bus radial distribution system, demonstrating advantages such as time efficiency, high computational performance and robust integration of PQD and EMP calculations.

The contributions of this study are as follows:

1. The first application of parallel computing to clustering algorithms for the combined analysis of PQDs and EMPs in electrical distribution systems. The proposed approach

is validated on the IEEE 33-bus test system, obtaining the results shown in Sect. 4.

2. The novel application of parallel computing to the DBSCAN and k-means clustering algorithms achieves a reduction of more than 40% in processing times and slightly improves efficiency in results, compared to traditional sequential implementations, paving the way for real-time analysis and facilitating decisionmaking in network management.
3. The proposed approach allows the identification of correlations between PQD variations and their impact on EMP, which has important implications for the optimization of energy management.

The paper is divided into five sections to develop the objectives and address the contributions. In addition to this introduction, Sect. 2 describes the DBSCAN and k-means cluster analysis methodologies, and parallel computing. Section 3 introduces the case study, where these methodologies are developed on PQDs and EMPs measurements. Section 4 presents the results obtained from the application of these methodologies using parallel computing and its corresponding discussion. Finally, conclusions are shown.

2 Methods and Methodology

2.1 DBSCAN and k-Means Algorithms

Clustering is an unsupervised machine-learning technique that organizes data points into groups or clusters based on their similarities. Popular clustering algorithms include DBSCAN [24, 25] and k-means [21]. DBSCAN is a density-based algorithm that groups data points according to their density. Two important parameters are required for DBSCAN. The first is epsilon (ε) which defines the radius of a neighborhood around a point x. The second parameter is the minimum points (MinPts), which refers to the minimum number of neighbors within the ε-radius [26]. It classifies points as core (with many neighbors), border (with few neighbors but near a core point), or outliers (isolated points), as shown in Eqs. (1) and (2). Figure 1a shows the general flowchart of DBSCAN. Core points were connected to form clusters, border points were assigned to these clusters and outliers were excluded. Hence, DBSCAN is effective for identifying clusters of varying shapes and for handling noise [23]. In contrast, k-means is a centroid-based algorithm that divides data into k-clusters by randomly placing centroids, assigning each data point to the nearest centroid and iteratively adjusting the centroids until the clusters stabilize [27]. K-means works well with clearly defined clusters, as shown in Eqs. (3) and (4), and Fig. 1b shows the proccess map for k-means. Both algorithms are valuable tools for clustering, with strengths depending on the nature of the data.

$$\mathrm{Core}(x_i) = \left\{x_i : \left\|x_i - x_j\right\| \le R\right\} \\ \text{and}|\mathrm{Core}(x_i) \ge M \tag{1}$$

$$\varepsilon = \max_{p \in D} d(p, \mathrm{MinPts}) \tag{2}$$

$$x_i \in \mathrm{Core}(x_i) \tag{3}$$

$$v_{ij} = \frac{\sum_{k=1}^{ni} x_{kj}}{n_i} \tag{4}$$

, where V_{ij}: centroid of ith cluster for jth variable; n_i: number of data in i-cluster; X_{kj}: kth data value in the jth cluster for jth; D: the dominant set; MinPts: the minimum samples—th neighbor of p.

$$D_\mathrm{e} = \sqrt{\sum_{i=1}^{p} (n_i - s_i)^2} \tag{5}$$

, where D_e: Euclidean distance; p: variable number; s_i: ith center.

These equations and definitions underpin DBSCAN and k-means clustering algorithms, allowing them to partition data into meaningful groups based on different criteria.

Figure 2 shows a simple example to explain how the DBSCAN and k-mean clustering methods work. DBSCAN clusters points based on two parameters, density (ε), which sets the maximum distance between two points for them to be considered neighbors, and size (MinPts), which sets the minimum number of points needed to form a cluster. It can find clusters of any shape and automatically label as noise those points that are not close to any cluster, making it ideal for detecting outliers and irregular patterns. On the other hand, k-means requires the number of clusters (k) to be specified in advance and works by placing k-centroids, assigning points to the nearest centroid, and iteratively updating the centroids until they stabilize. K-means is fast and better suited to well-defined and uniformly distributed clusters, while DBSCAN excels at handling noise and arbitrary cluster shapes. Both methods are powerful, but serve different purposes depending on the structure of the data.

2.2 Proposed Parallel Computing

Parallel computing involves dividing large and complex scientific procedures into smaller tasks and simultaneously performing them using multiple computer resources. This approach leverages the combined processing power of multicore CPUs and GPUs to perform computations more quickly and handle larger data sets than is possible with a single

Fig. 1. General flow chart for DBSCAN (**a**) and k-means (**b**).

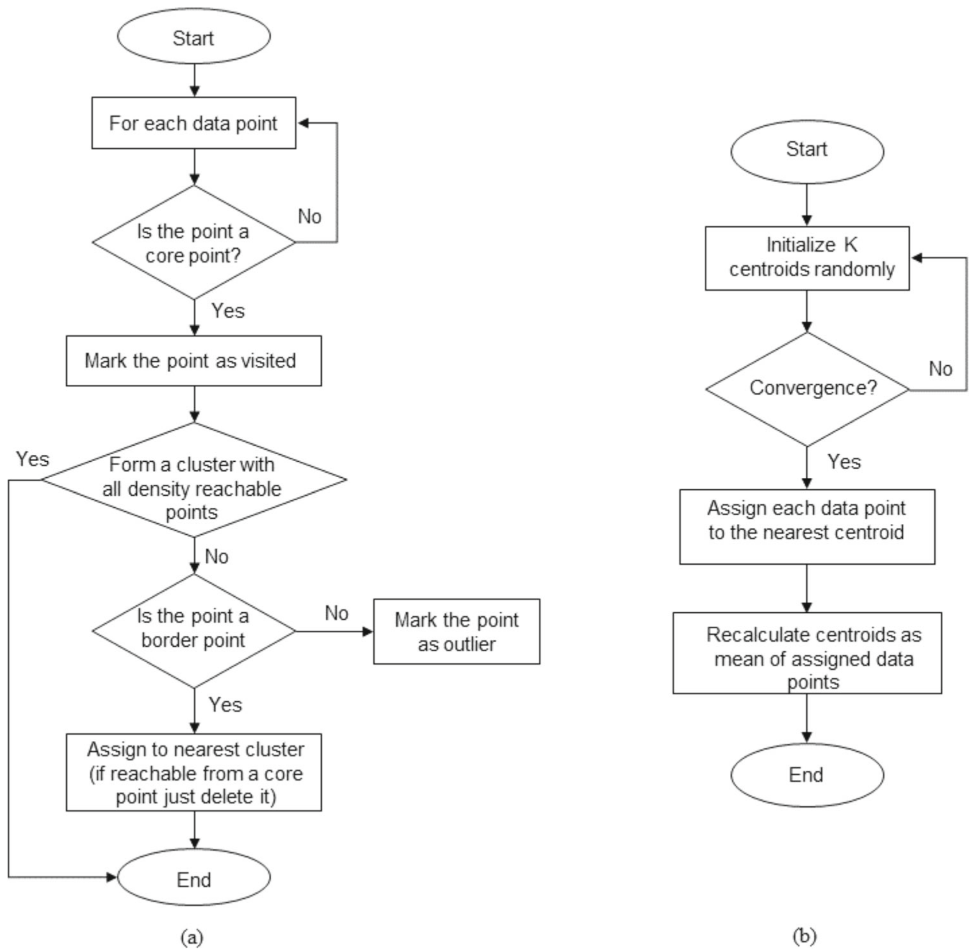
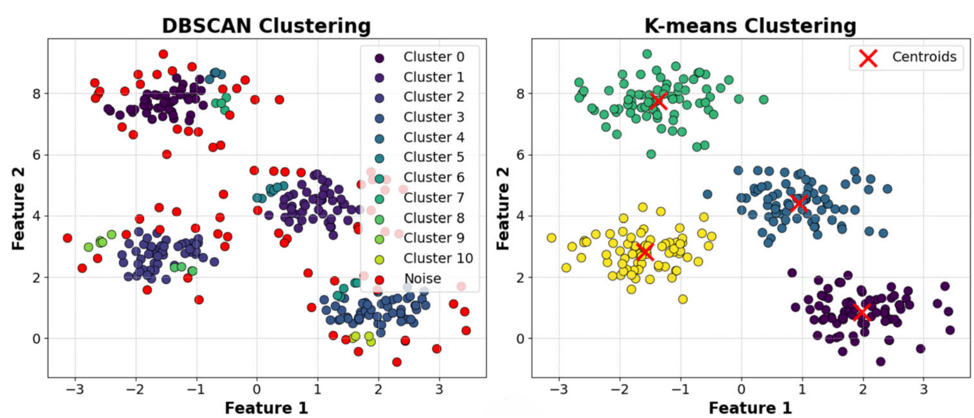


Fig. 2. Graphical example for DBSCAN and k-means.



processing unit. The types of parallel computing include bit-level, instruction level, data level, task, and data parallelism [28]. Data parallelism involves dividing the data into multiple sets, with each set processed by different workers using the same instruction. By contrast, task parallelism assigns multiple activities within a given operation to different workers for simultaneous processing [29]. The present study employed data parallelism to identify PQDs and EMPs.

The power flow for the 33-bus IEEE system was analyzed under each PQD and EMP scenario. The input data for the PQDs and EMPs underwent a preprocessing step to prepare for clustering, normalization, and feature selection. In parallel computing, a set of workers collaborate to perform a given job, communicating with one another during execution. The client session (job scheduler) and the workers operate on the same computer. The evaluation of four indices for PQDs (THDI, THDV, unbalanceandflicker) and four indices for EMPs (production, consumption, active power, and reactive power) were divided into various tasks by the job scheduler. These tasks were then assigned to different workers. Upon completion, the workers submit their results to the job scheduler, which consolidates the evaluations of the PQDs and EMPs. The purpose of analyzing PQDs is to determine their effect on electrical power systems and EMPs. The initial parameter sets for each algorithm were determined: DBSCAN (ε and MinPts) and k-means (number of clusters and initial centroids), where the number of clusters must be specified before running the algorithm. This represents the number of distinct groups that the algorithm wants to identify within the dataset. The initial centroids are the starting points for each cluster, and are frequently selected randomly from the dataset.

Figure 2 illustrates a graphical example for DBSCAN and k-means and Fig. 3 shows the architecture of the proposed parallel computing with DBSCAN and k-means for PQDs and EMPs.

3 Case Study: PQDs and Energy Management in System Model

The assessment of power quality has become increasingly important with the rise of sensitive power electronic devices in daily life. The initial absence of standardized measurement methods caused notable discrepancies in the major parameters calculated using various devices. Consequently, power quality disturbances can result in equipment malfunctions and process interruptions. IEEE Recommended Practice for Monitoring Electric Power Quality [30] and the EN50160 standard [31] address this issue by outlining a specific procedure, mathematical relationships and necessary measurement accuracy for power quality analyzers. For power quality monitoring, a short averaging time can be adequate for evaluating the performance and disturbances related to power quality issues. The primary objective of this study was to analyze power quality issues in power systems and their effects on energy management in the same system. Power quality measurements were obtained from radial power systems, as shown in Fig. 4.

Voltage levels generated by power station centers typically increase to 70–500 kV through transformers on transmission lines. This high voltage was then reduced to a medium or primary distribution voltage, typically 20 kV. At the distribution substation, the primary distribution voltage was further lowered to a low voltage of 380/220 V before being distributed to consumers. To perform the initial step of this work (load flow), the necessary data include resistance (R), reactance (X), active load (P) and reactive load (Q). These parameters, along with point data and line lengths, define the IEEE 33-bus test system [32]. The 33-bus radial distribution

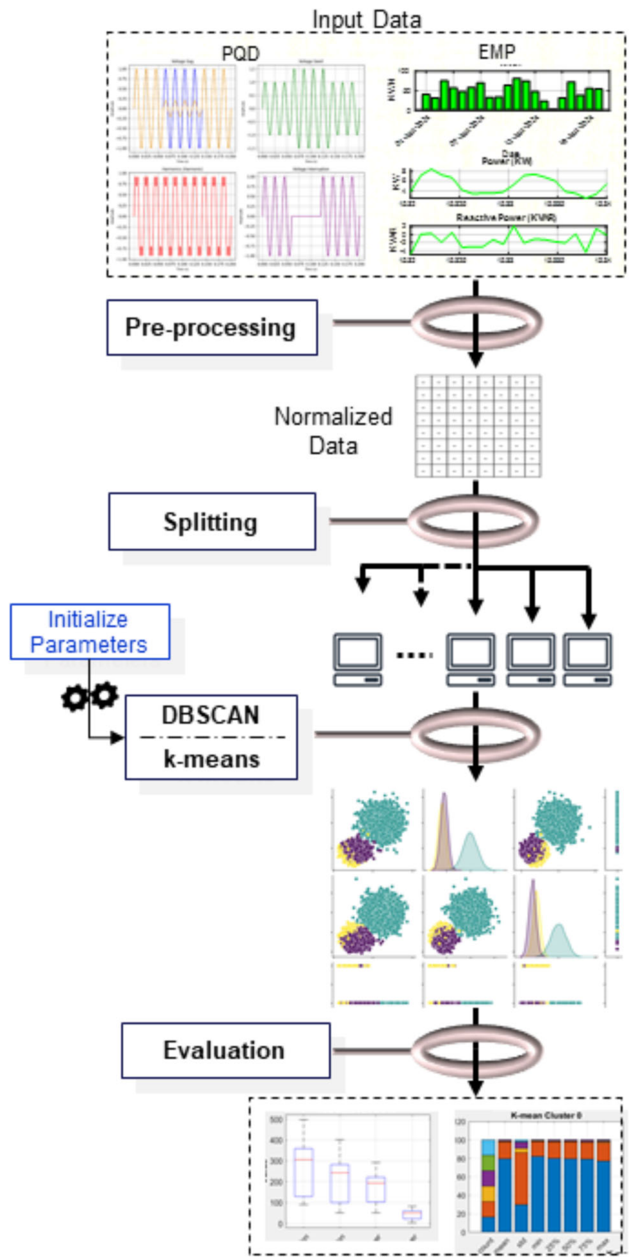


Fig. 3 The architecture of proposed parallel computing with DBSCAN and k-means for PQDs and EMPs

system comprises an unbalanced three-phase system and a balanced power system. This study considers the unbalanced three-phase distribution system to be associated with the distributed and renewable energy resource and demand response mechanisms, along with appropriate reactive power compensation units [33]. The voltage system is a 12.66 kV one-feeder substation that encompasses 33 buses, 32 branches and four distributed generation units (installed on buses 18, 22, 25 and 33) and two reactive power compensation systems connected to buses 18 and 33 with 0.4 MV Ar and 0.6 MV Ar, respectively. The total active and reactive demands are 3.715 MW

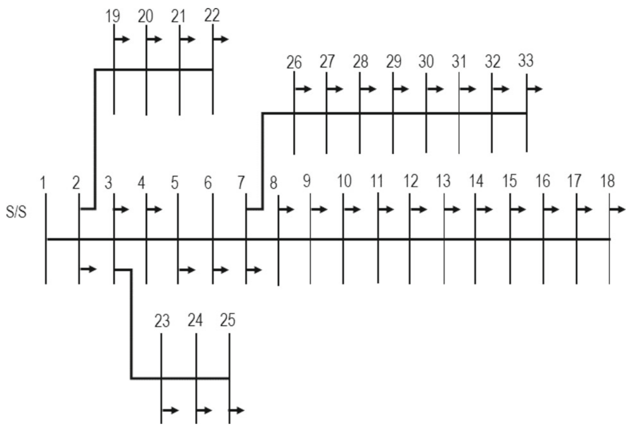


Fig. 4. Single line diagram of the 33-bus IEEE test system [32].

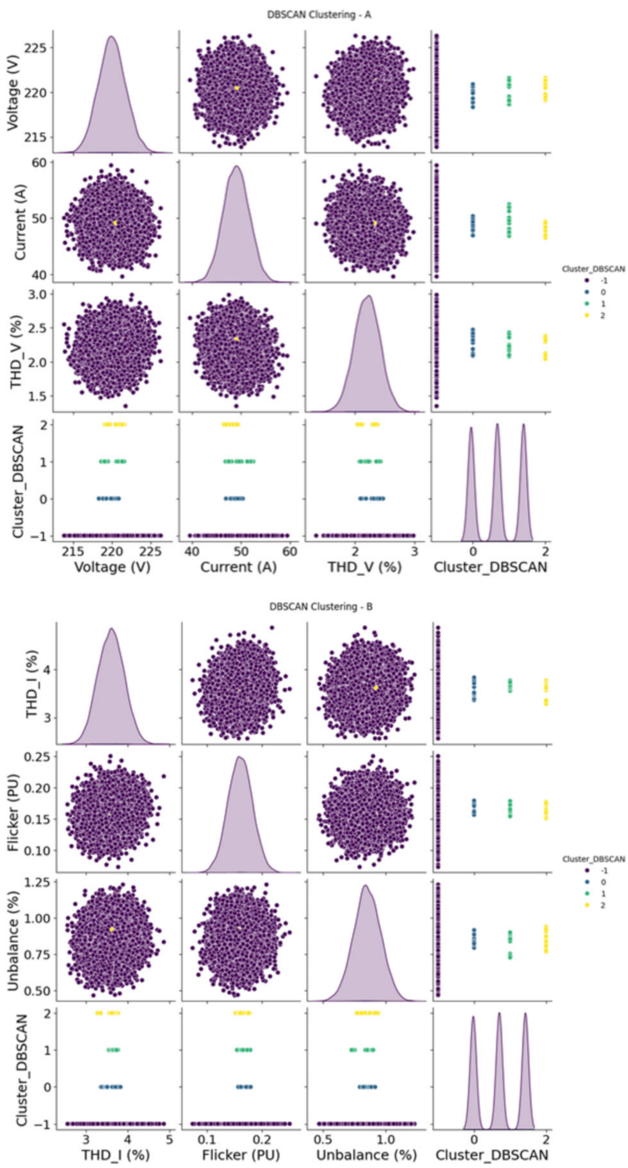


Fig. 5. Results of DBSCAN clustering for PQDs using parallel computing.

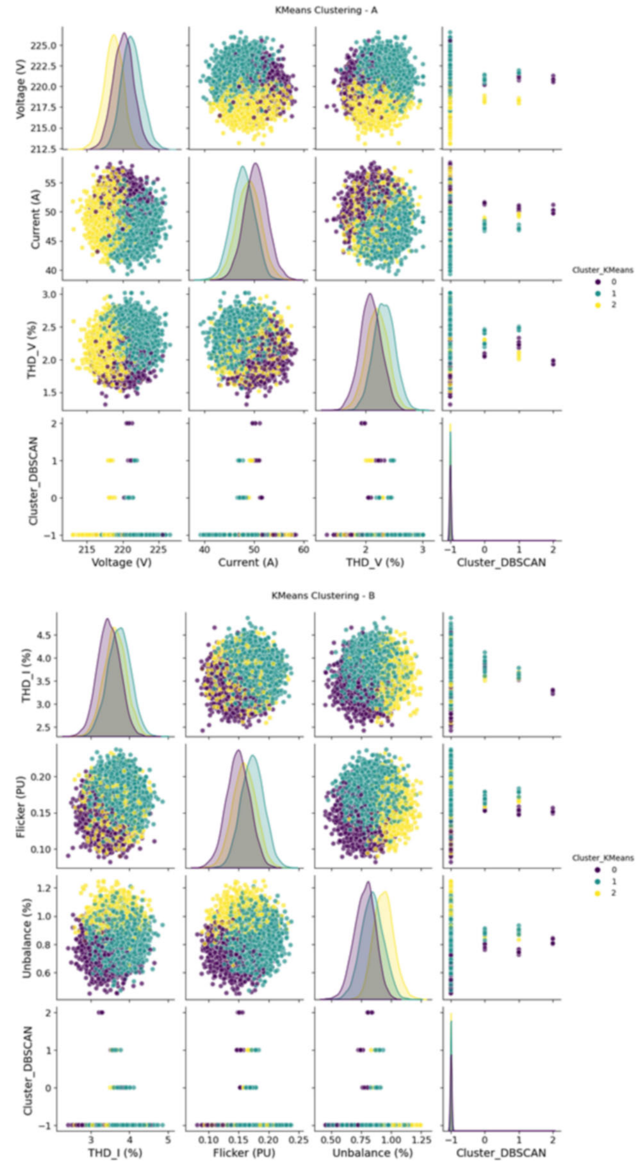


Fig. 6. Results of k-means clustering for PQDs using parallel computing.

and 2.3 MV Ar, respectively. The power quality parameters considered in this work are THDI, THDV, unbalance, flicker, accompanied by voltage and current measurements. The selection of these variables is based on their relevance in steady-state operation and long-term power quality analysis. Meanwhile, other issues, such as sag and swell, are critical but they are typically monitored in specific scenarios, wherein transient events are the primary concern. Flicker is included because of its effects on lighting and human comfort, making it a critical parameter in certain environments. EMPs (production [kW], consumption [kW], active power [kW] and reactive power [kV Ar]) are studied to understand the effects of PQDs on a power system [27]. The measurement unit for production and consumption performed with

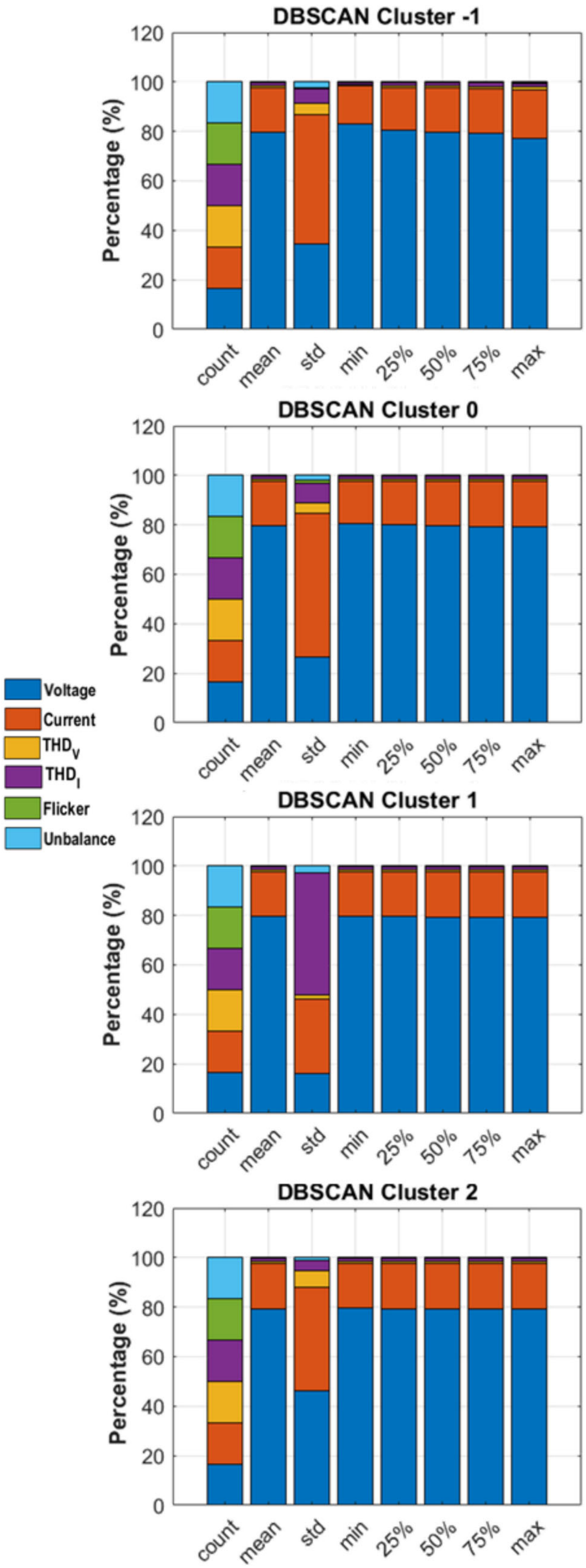


Fig. 7. Statistical analysis of DBSCAN clustering for PQDs using parallel computing.

power measurements is kW. The matrix for PQDs is [13105 $\times$ 6] and that for EMPs is [13105 $\times$ 4], indicating that the total samples are 78,630 and 52,420, respectively. For each dataset with 128 samples per cycle, sampling frequency is 6400 Hz and system frequency is 50 Hz.

4 Results and Discussion

In this section, the performance of the DBSCAN and k-means algorithms, both implemented in parallel computing, is evaluated using two data sets, PQD and EMP, obtained from the analysis of the IEEE bus-33 network.

4.1 Results for PQDs

The inputs for analyzing PQD patterns are the measurements of THDI, THDV, unbalance, and flicker, accompanied by the voltage and intensity measurements. These inputs preserve temporal relationships and patterns in the data, which are important for time series analysis. To perform the analysis using DBSCAN it is first necessary to set the parameters ε and MinPts that control, respectively, the density and the cluster size. The values of these parameters were determined by trial and error. A comparative analysis of the results according to the values of these parameters is shown in Table 1. The values chosen were $\varepsilon = 0.5$ and MinPts $= 5$. The result of the clustering using parallel DBSCAN, configured with the indicated parameter values, is shown in Fig. 5.

Four parallel clusters, labeled with − 1, 0, 1 and 2, were automatically established and chosen under the supervision of DBSCAN parallel computing technique. Cluster (− 1) contains the majority of data points and is considered noise by the parallel DBSCAN. High standard deviations indicate significant variability within this cluster. Clusters (0) and (1) were relatively small with low variability, indicating closely related points within each cluster, and finally, cluster (2) was the smallest cluster and thus had low variability and high cohesion among its points. On the other hand, for the k-means clustering algorithm, the parameter required to be set prior to its execution is the number of clusters (n cluster). A random-state parameter was used to ensure reproducibility. It sets the seed for the random-number generator used in the k-means algorithm. By fixing the seed, the initial cluster centroids were selected in the same manner every time the algorithm was run, resulting in consistent and repeatable clustering results. This condition is particularly useful for comparing results across different runs or when sharing results with others. Again, using trial and error, it was set to three, as this number was found to better match the underlying structure of the data and provide more easily interpretable results, see Table 1. The results were three Clusters (0), (1),

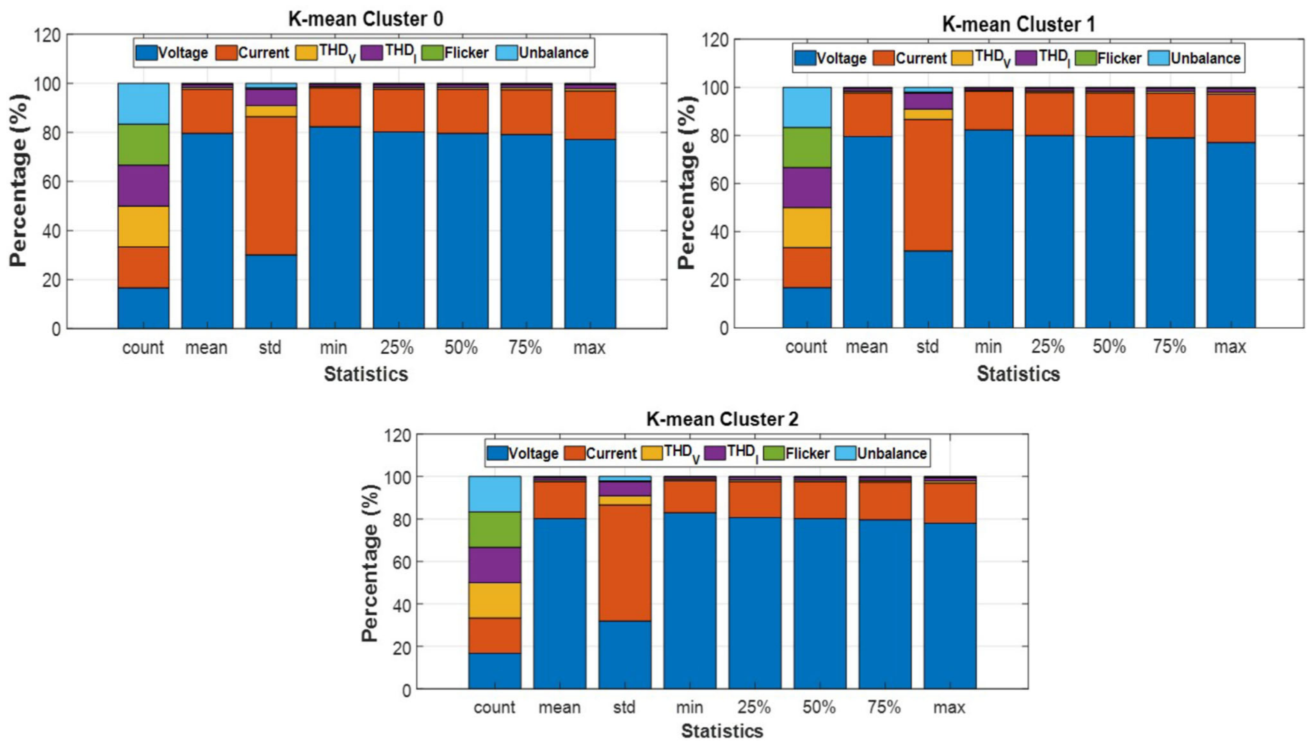


Fig. 8. Statistical analysis of k-means clustering for PQDs using parallel computing.

and (2) that enable a high analysis level of the PQD data, as shown in Fig. 6.

Cluster (0) is characterized by a slightly lower voltage and higher imbalance compared with the other clusters, indicating that this group may represent an unbalanced voltage disturbance or a set of equipment with these characteristics. Cluster (1) had a higher voltage and current with lower T HDV and T HDI values, representing a more stable operational condition with relatively lower harmonic distortions and flicker values. Cluster (2) features the highest voltage and flicker values, with higher THDV and THDI, indicating that points in this cluster may experience more voltage variations and harmonic distortions. The comparison between parallel DBSCAN and parallel k-means clustering results revealed notable differences in the statistical characteristics of the clusters across various electrical parameters. In the parallel DBSCAN clustering results, clusters exhibited wider ranges and more variability in most metrics than the parallel k-means clusters. For instance, in the Voltage parameter, parallel DBSCAN clusters range from a minimum of 213.88 V to a maximum of 227.57 V, whereas parallel k-means clusters span the same range. However, the mean values for Voltage show slight variations, with parallel DBSCAN clusters ranging from 218.64 to 221.30 V and parallel k-means clusters ranging from 218.64 to 220.01 V. Similarly, in terms of Current, THDV, THDI, Unbalance and Flicker, parallel DBSCAN clusters generally exhibit wider ranges and slightly higher variability in mean values than parallel k-means clusters, as shown in Figs. 7 and 8.

These results suggest that parallel DBSCAN may identify clusters with more diverse and spread-out data points, whereas k-means clusters tend to be more compact and less varied in their statistical profiles. These differences underscore the distinct clustering approaches of DBSCAN, which identifies clusters based on density, and parallel k-means, which partitions data into clusters based on centroid proximity. These approaches influence the definition of clusters and their statistical distributions across the analyzed electrical parameters. Figures 5 and 7 show the minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum values for each parallel cluster in both DBSCAN and k-means results across the variables voltage, current, THDV, T HDI, unbalanced and flicker.

4.2 Results for EMPs

According to the EMPs, the DBSCAN and k-means clustering results present different perspectives on the data, reflecting the unique methodologies of each clustering algorithm. For EMPs, DBSCAN established three clusters (− 1, 0 and 1) as shown in Fig. 9. In contrast, the k-means clustering results delineate the data into three distinct clusters based on their centroids, showing a different distribution of variables, as shown in Fig. 10.

Table 1 Comparison of results for different parameters of the parallel DBSCAN and k-means methods

Method	DBSCAN	DBSCAN	k-means
Parameter	ε 0.5 0.3, 0.7, 1.0	MinPts	n cluster
Chosen value		5	3
Alternatives values		3, 10	2, 4, 5
Results with chosen values	Balanced clusters numbers and noise points	Reasonable cluster sizes without excessive noise	Aligns with the synthetic data
Results with alternative values	$\varepsilon = 0.3$: Too many noise points and small clusters. $\varepsilon = 0.7/1.0$: Fewer clusters, may over-simplify	MinPts = 3: Smaller less meaningful clusters MinPts = 10: Fewer clusters, more noise	n clusters = 2: Oversimplifies, merges distinct sources. n clusters = 4/5: Overfits, creates less meaningful clusters
Justification for chosen values	$\varepsilon = 0.5$, a good balance between identifying meaningful clusters and filtering out noise	MinPts = 5 ensures clusters are dense and meaningful without restrictive	n clusters = 3: matches the underlying structure of the data and provides interpretable results

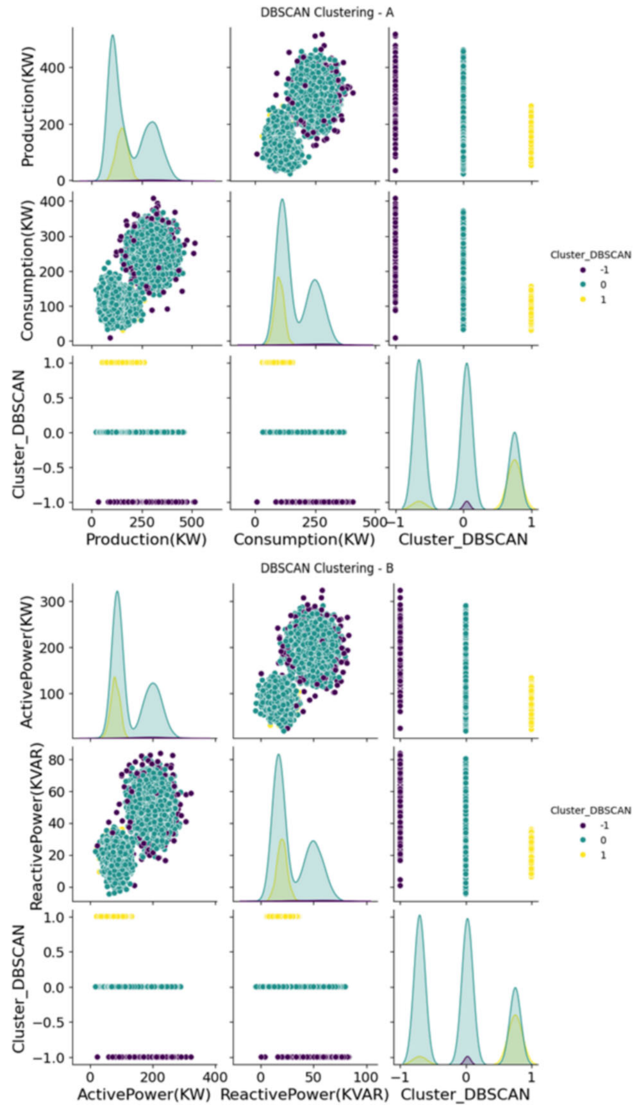


Fig. 9. Results of DBSCAN clustering for EMPs using parallel computing.

For DBSCAN, cluster (− 1), representing noise points which exhibit significantly higher average values in all variables (Production, Consumption, Active Power and Reactive Power). Specifically, the mean Production for Cluster (1) is 305.89 kW, Consumption is 242.70 kW, Active Power is 193.14 kW and Reactive Power is 49.78 kV AR, which are considerably higher than the means of DBSCAN Clusters (0) and (1). DBSCAN Cluster (0), which includes the majority of the data points, shows lower mean values: Production at 188.34 kW, Consumption at 166.46 kW, Active Power at 130.75 kw and reactive power at 29.81 kV AR. DBSCAN Cluster (1), with fewer data points than Cluster (0) but more than the noise cluster, has the lowest mean values: Production at 150.98 kW, Consumption at 100.26 kW, Active Power at 79.92 kW and Reactive Power at 20.17 kV AR, as shown in Fig. 11. These differences highlight how DBSCAN effectively segregates noise from structured clusters, emphasizing the significant variances within the data.

The k-means Cluster (0) groups the data points with the lowest mean values: Production at 100.26 kW, Consumption at 120.12 kW, Active Power at 89.47 kW and Reactive Power at 14.84 kV AR. This cluster represents data points with generally lower values. k-means Cluster (1), on the contrary, includes data points with significantly higher mean values: Production at 299.82 kW, Consumption at 249.60 kW, Active Power at 199.21 kW and Reactive Power at 50.39 kV AR. Cluster (2) falls between these extremes, with mean values higher than Cluster 0 but lower than Cluster 1: Production at 155.00 kW, Consumption at 97.58 kW, Active Power at 79.70 kW and Reactive Power at 20.56 V AR. The standard deviations within k-means Cluster (1) are notably higher, indicating greater variability in data points compared with Clusters (0) and (2) Fig. 12. This clear separation by k-means reflects its method of minimizing variance within clusters and maximizing variance between clusters, offering a complementary perspective to DBSCAN's density-based approach.

Table 2 Parallel DBSCAN summary during PQD events

Parameter	Value	Cluster − 1	Cluster 0	Cluster 1
Active power	Mean	122.70	91.80	57.99
(kW)	Std. dev.	57.99	12.20	12.20
Consumption	Mean	157.06	107.87	73.10
(kW)	Std. dev.	73.10	1.33	1.33
Production	Mean	183.79	100.74	91.75
(kW)	Std. dev.	91.75	14.39	14.39
Reactive power	Mean	28.45	14.23	17.35
(kV Ar)	Std. dev.	17.35	1.83	1.83

Table 3 Parallel k-mean summary during PQD events.

Parameter	Value	Cluster − 1	Cluster 0	Cluster 1
Active power	Mean	199.11	84.97	84.51
(kW)	Std. dev.	29.42	39.63	39.63
Consumption	Mean	251.36	110.10	110.27
(kW)	Std. dev.	39.63	39.63	39.63
Production	Mean	300.29	124.71	126.99
(kW)	Std. dev.	49.04	49.04	49.04
Reactive power	Mean	50.69	17.36	17.42
(kV Ar)	Std. dev.	10.03	10.03	10.03

4.3 Effects of PQDs on EMPs

Parallel DBSCAN clustering identifies three clusters with varying characteristics during PQD events. In Table 2 shows the data to explain the effects of PQDs on EMPs. Clusters characterized by lower variability in consumption and reduced energy usage tend to experience fewer power PQDs. In contrast, Cluster (− 1), identified as noise or outliers, exhibits the highest mean values for active power and production. Additionally, this cluster demonstrates elevated THDV (2.2%), THDI (3.6%), and flicker (0.16 pu), signifying notable power quality deviations. The findings further indicate that an increase in THDI (3.6%) correlates with higher reactive power consumption, underscoring the necessity for enhanced compensation strategies to mitigate harmonic distortions and optimize energy efficiency in variable operating conditions. Cluster (0) exhibits moderate mean values for all parameters with low variability in consumption, suggesting that it captures a more stable subset of data. In contrast, Cluster (1) is characterized by the lowest mean values for active power and production, with low variability, indicating that it represents a distinct pattern of low energy use during PQD events.

Meanwhile, parallel k-means clustering defines three distinct clusters with different energy profiles. Cluster (0) exhibits the highest mean values for all the parameters, with considerable standard deviations, reflecting a pattern of high energy use and variability. Cluster (1) presents moderate mean values and standard deviations, suggesting a balance between high and low energy states.

Finally, Cluster (2) is characterized by the lowest mean values and standard deviations across all the parameters, indicating low energy use and stability during PQD events, as shown in Table 3.

4.4 Performance and Accuracy

The performance of the DBSCAN and k-means clustering algorithms, along with their parallel implementations in the IEEE 33-bus case study, was evaluated, The results are shown in Table 4. The standard DBSCAN algorithm required 2.2351 s to complete the task, whereas the parallel version significantly improved the execution time to 1.00183 s. Conversely, the k-means algorithm demonstrated better overall performance, with the standard version completing in 2.121 s and the parallel version further reducing the time to only 1.0009 s. When examining efficiency, standard DBSCAN achieved 78%, which is slightly lower than 80% of its parallel version's 80%. In comparison, k-means exhibited higher efficiency, with the standard implementation at 80.45% and the parallel implementation achieved the highest efficiency at 83.4%. This analysis underscores the superior performance of the parallel k-means algorithm in terms of both speed and

Table 4 Performance of the proposed parallel algorithms for IEEE 33 case study

	DBSCAN	Parallel DBSCAN	k-means	Parallel k-means
Time (s)	2.235	1.00183	2.121	1.0009
Efficiency	78%	80%	80.45%	83.4%

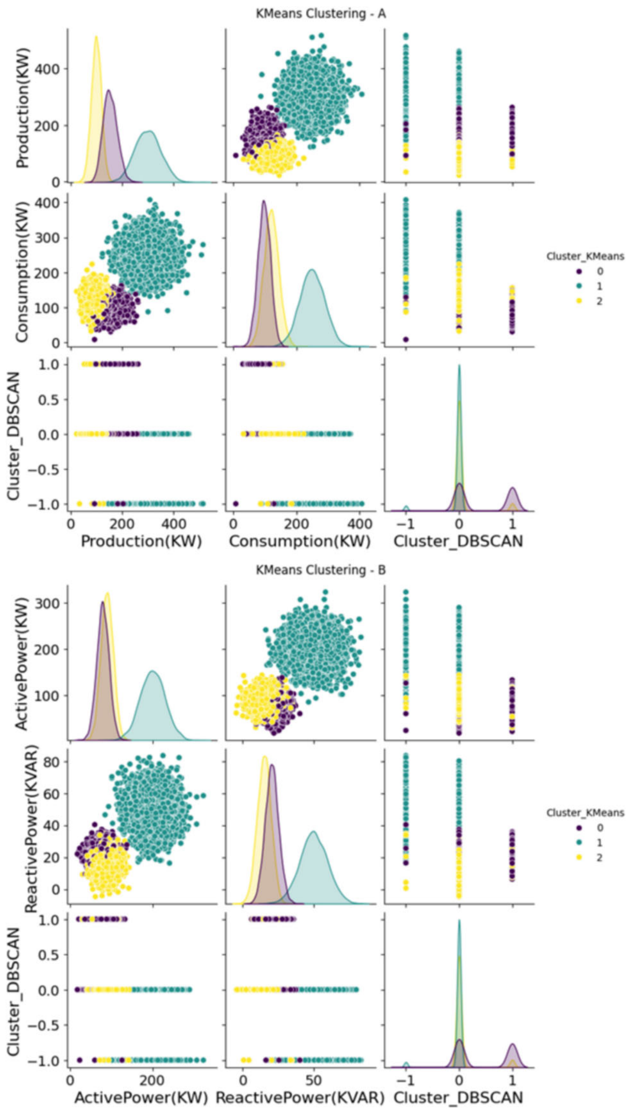


Fig. 10. Results of k-means clustering for EMPs using parallel computing.

efficiency, making it the most effective among the evaluated methods.

Table 5 collects the data regarding accuracy and noise robustness. The results highlight that parallel DBSCAN achieves higher accuracy (83%) and noise robustness (100%) compared to parallel k-means (accuracy 56%, robustness 64%). This is because DBSCAN inherently handles noise and irregular cluster shapes better. In terms of scalability in the application of the methods, the data obtained are shown

Table 5 Accuracy and robustness to noise comparison

Method	Accuracy (%)	Robustness to noise (%)
Parallel DBSCAN	83	100
Parallel k-means	56	64

in Fig. 13. DBSCAN is outperformed by k-means in execution speed. For DBSCAN the execution time increases as the size of the data set increases. This makes k-means more suitable for large-scale data despite its limitations in handling noise and complex structures. Taken together, these results demonstrate the trade-offs between accuracy, robustness and scalability in clustering algorithms.

5 Conclusions

This study proposed a parallel computing technique using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to accelerate data query times for power distribution system analysis, aiming to improve the efficiency in identifying anomalies and energy management parameters. The results obtained confirm the hypothesis that the use of parallel computing with DBSCAN allows for a significant reduction in query times, enhancing the efficiency in identifying anomalies and energy management parameters. Experiments on a radial distribution system demonstrated that parallel computing significantly enhanced processing speed by leveraging multiple CPU cores. DBSCAN effectively identified clusters of varying shapes and detected outliers, although parameter tuning was necessary to minimize noise misclassification. In contrast, k-means produced well-defined clusters with clear boundaries but required a predefined number of clusters. Statistical analysis based on the EN 50160 standard confirmed that both clustering techniques provided meaningful insights: DBSCAN excelled in anomaly detection, while k-means delivered consistent categorizations of PQDs and EMPs. The reduction in processing times achieved with parallel computing not only streamlines the analysis but also enables real-time analysis, which is fundamental for decision-making in grid management and fault prevention. DBSCAN's ability to accurately detect anomalies facilitates the early identification of problems in the grid, allowing for faster and more effective interventions. While the study focused on a specific radial

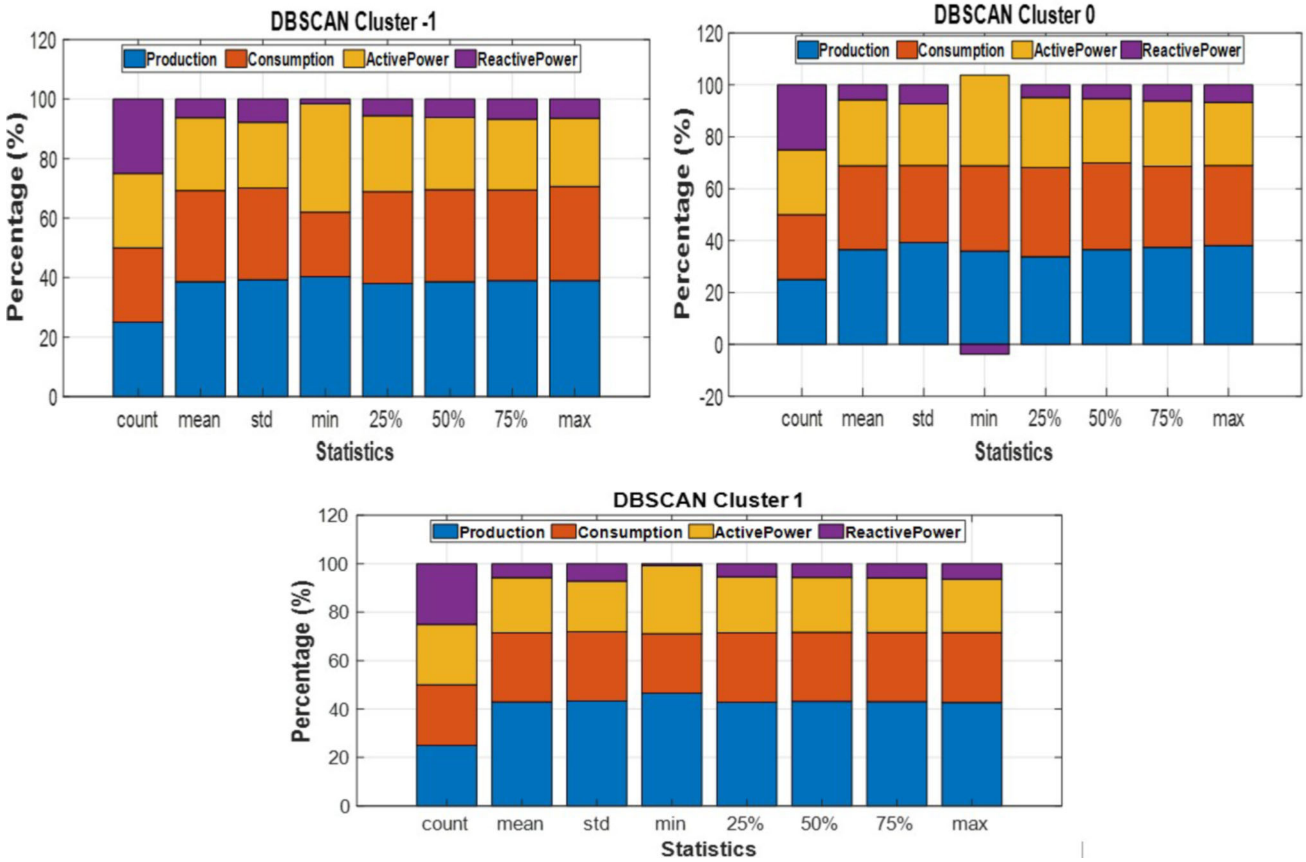


Fig. 11. Statistical analysis of DBSCAN clustering for EMPs using parallel computing.

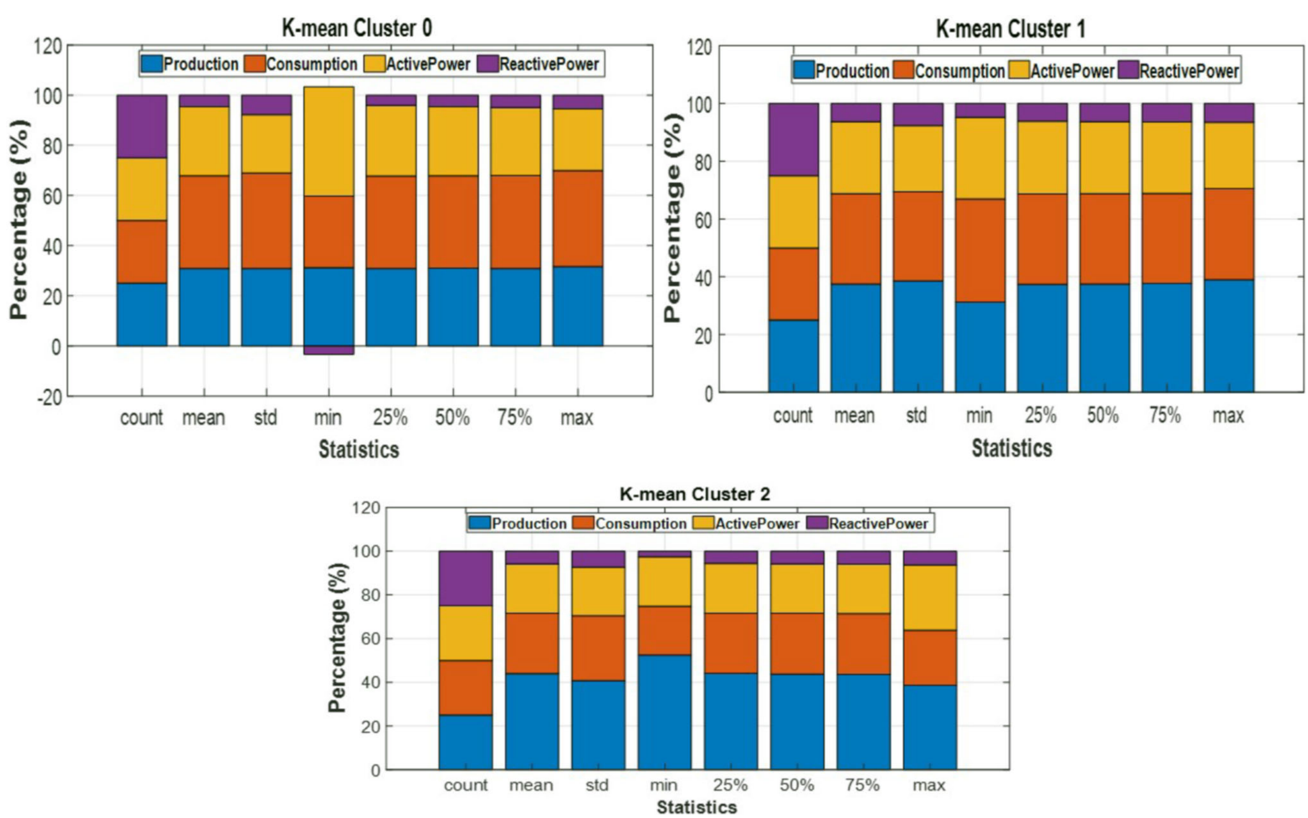


Fig. 12 Statistical analysis of k-means clustering for EMPs using parallel computing

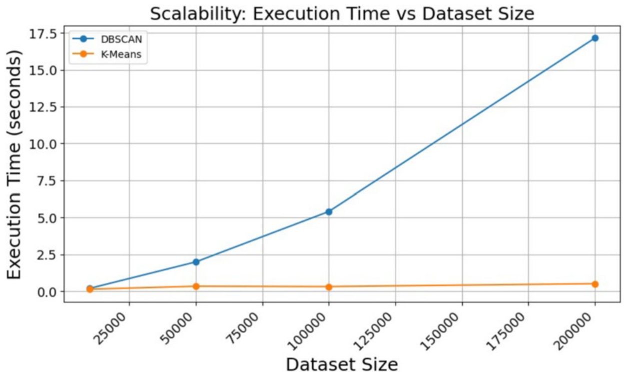


Fig. 13 Scalability: execution time versus dataset size

distribution system, the proposed methodology is adaptable to other network topologies. Future work could also extend the validation of the proposed methodology to other standard power system models, such as the IEEE 123-bus or the IEEE 8500-node test feeders, to assess its performance and generalization across different network configurations. Moreover, exploring the hybridization of DBSCAN and k-means with optimization algorithms, such as genetic algorithms or particle swarm optimization, could be a promising direction to automatically determine the optimal hyperparameter values for each clustering method, thereby improving the accuracy and robustness of the analysis. Future research could explore its application in meshed networks and with different load characteristics, along with the development of automatic methods for the optimal selection of DBSCAN parameters. Taken together, these findings demonstrate the significant potential of parallel clustering techniques to transform the analysis of electrical systems, opening new avenues for intelligent energy management and improving the reliability of the electrical grid.

Funding Funding for open access publishing: Universidad de Málaga/CBUA. Financing for open access position: Universidad de Málaga/CBU

References

1. Rao, C.S.P.; Pandian, A.; Reddy, C.R.; Gulzar, M.M.; Khalid, M.: A novel hybrid RERNN-SCSO technique-based unified power quality conditioner of microgrid in an EV charging station. Arab. J. Sci. Eng. **49**(5), 7277–7306 (2024). https://doi.org/10.1007/s13369-024-08765-5
2. Razmi, D.; Lu, T.; Papari, B.; Akbari, E.; Fathi, G.; Ghadamyari, M.: An overview on power quality issues and control strategies for distribution networks with the presence of distributed generation resources. IEEE Access **11**, 10308–10325 (2023). https://doi.org/10.1109/ACCESS.2023.3238685
3. Yahdou, A.; Belhadj Djilali, A.; Bounadja, E.; Boudjema, Z.: Power quality improvement through backstepping super-twisting control of a DFIG-based dual rotor wind turbine system under grid voltage drop. Arab. J. Sci. Eng. **49**, 1–18 (2024). https://doi.org/10.1007/s13369-023-08699-4
4. Hossain, E.; Tür, M.R.; Padmanaban, S.; Ay, S.; Khan, I.: Analysis and mitigation of power quality issues in distributed generation systems using custom power devices. IEEE Access **6**, 16816–16833 (2018). https://doi.org/10.1109/ACCESS.2018.2814981
5. Polat, O.; Gul, O.: Development of a probabilistic short-term voltage quality assessment method with the weak point detection capability through the dynamic analyses. Appl. Energy **326**, 120003 (2022). https://doi.org/10.1016/j.apenergy.2022.120003
6. De Carne, G.; Maroufi, S.M.; Beiranvand, H.; De Angelis, V.; D'Arco, S.; Gevorgian, V., et al.: The role of energy storage systems for a secure energy supply: a comprehensive review of system needs and technology solutions. Electr. Power Syst. Res. **236**, 110963 (2024). https://doi.org/10.1016/j.epsr.2024.110963
7. Lee, D.; Lin, C.: Universal artificial intelligence workflow for factory energy saving: ten case studies. J. Clean. Prod. **468**, 143049 (2024). https://doi.org/10.1016/j.jclepro.2024.143049
8. Khalid, M.: Smart grids and renewable energy systems: perspectives and grid integration challenges. Energy Strategy Rev. **51**, 101299 (2024). https://doi.org/10.1016/j.esr.2024.101299
9. Al-Falahi, M.D.; Tarasiuk, T.; Jayasinghe, S.G.; Jin, Z.; Enshaei, H.; Guerrero, J.M.: AC ship microgrids: control and power management optimization. Energies **11**(6), 1458 (2018). https://doi.org/10.3390/en11061458
10. Kabeyi, M.J.B.; Olanrewaju, O.A.: Smart grid technologies and application in the sustainable energy transition: a review. Int. J. Sustain. Energy **42**(1), 685–758 (2023). https://doi.org/10.1080/14786451.2023.2222298
11. Miraftabzadeh, S.M.; Colombo, C.G.; Longo, M.; Foiadelli, F.: K-means and alternative clustering methods in modern power systems. IEEE Access **11**, 119596–119633 (2023). https://doi.org/10.1109/ACCESS.2023.3327640
12. Zichen, Z., Liangxiao, L., Zhenlin, L., Miaomiao, C., Wenchu, T., Yang, W.: Harmonic pollution zoning method based on improved DBSCAN clustering. In: 2022 China International Conference on Electricity Distribution (CICED), pp. 1099–1103. IEEE (2022)
13. Babichev, S., Lytvynenko, V., Osypenko, V.: Implementation of the objective clustering inductive technology based on DBSCAN clustering algorithm. In: 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), vol. 1, pp. 479–484. IEEE (2017)
14. Li, S.S.: An improved DBSCAN algorithm based on the neighbor similarity and fast nearest neighbor query. IEEE Access **8**, 47468–47476 (2020). https://doi.org/10.1109/ACCESS.2020.2972034
15. Tang, Q.; Qiu, W.; Zhou, Y.: Classification of complex power quality disturbances using optimized S-transform and kernel SVM. IEEE Trans. Ind. Electron. **67**(11), 9715–9723 (2019). https://doi.org/10.1109/TIE.2019.2952823

16. Arranz, R.; Paredes, A.; Rodríguez, A.; Muñoz, F.: Fault location in transmission system based on transient recovery voltage using Stockwell transform and artificial neural networks. Electr. Power Syst. Res. **201**, 107569 (2021). https://doi.org/10.1016/j.epsr.2021.107569
17. Vishwanath, Y.U.; Esakkirajan, S.; Keerthiveena, B.; Pachori, R.B.: A Generalized classification framework for power quality disturbances based on synchrosqueezed wavelet transform and convolutional neural networks. IEEE Trans. Instrum. Meas. (2023). https://doi.org/10.1109/TIM.2023.3308235
18. Kumar, R.; Saxena, A.; Kumar, R.; Marwaha, S.; Singh, J.; Singh, G.K.: A comprehensive overview on modified versions of stockwell transform for power quality monitoring. IEEE Access **10**, 91963–91975 (2022). https://doi.org/10.1109/ACCESS.2022.3202309
19. Pomarico, A.; Berizzi, A.; Giannuzzi, G.M.; Pisani, C.: Topology-based clustering techniques for graph partitioning applied to the Italian transmission network. IEEE Access (2024). https://doi.org/10.1109/ACCESS.2024.3414494
20. Paramita, A.S.; Hariguna, T.: Comparison of K-means and DBSCAN algorithms for customer segmentation in e-commerce. J. Digit. Mark. Digit. Curr. **1**(1), 43–62 (2024). https://doi.org/10.47738/jdmdc.v1i1.3
21. Hamdipour, A.; Basiri, A.; Zaare, M.; Mirjalili, S.: Artificial rabbits optimization algorithm with automatically DBSCAN clustering algorithm to similarity agent update for features selection problems. J. Supercomput. **81**(1), 150 (2025). https://doi.org/10.1007/s11227-024-06606-8
22. Al-Batah, M.S.; Al-Kwaldeh, E.R.; Wahed, M.A.; Alzyoud, M.; Al-Shanableh, N.: Enhancement over DBSCAN satellite spatial data clustering. J. Electr. Comput. Eng. **2024**(1), 2330624 (2024). https://doi.org/10.1155/2024/2330624
23. Sardar, T.H.: Reflecting on a decade of evolution: mapreduce-based advances in partitioning-based, hierarchical-based, and density-based clustering (2013–2023). Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **14**(6), e1566 (2024). https://doi.org/10.1002/widm.1566
24. Li, L.; Yu, K.: Application of improved HF-DBSCAN algorithm in analyzing complex trajectory data of Wi-Fi users in smart campus. Appl. Artif. Intell. **38**(1), 2403260 (2024). https://doi.org/10.1080/08839514.2024.2403260
25. Wan, L.; Zheng, W.; Yuan, X.: Efficient interdevice task scheduling schemes for multidevice co-processing of data-parallel kernels on heterogeneous systems. IEEE Access **9**, 59968–59978 (2021). https://doi.org/10.1109/ACCESS.2021.3073955
26. Zhou, S.; Xu, H.; Zheng, Z.; Chen, J.; Li, Z.; Bu, J., et al.: A comprehensive survey on deep clustering: taxonomy, challenges, and future directions. ACM Comput. Surv. **57**(3), 1–38 (2024). https://doi.org/10.1145/3689036
27. Guan, J.; Li, S.; He, X.; Chen, J.: Clustering by fast detection of main density peaks within a peak digraph. Inf. Sci. **628**, 504–521 (2023). https://doi.org/10.1016/j.ins.2023.01.144
28. Paul, K.; Sinha, P.; Mobayen, S.; El-Sousy, F.F.; Fekih, A.: A novel improved crow search algorithm to alleviate congestion in power system transmission lines. Energy Rep. **8**, 11456–11465 (2022). https://doi.org/10.1016/j.egyr.2022.08.267
29. Bilal, M.; Habib, H.A.; Mehmood, Z.; Saba, T.; Rashid, M.: Single and multiple copy–move forgery detection and localization in digital images based on the sparsely encoded distinctive features and DBSCAN clustering. Arab. J. Sci. Eng. **45**(4), 2975–2992 (2020). https://doi.org/10.1007/s13369-019-04238-2
30. Bhandari, K.; Kumar, K.; Sangal, A.L.: Alleviating class imbalance issue in software fault prediction using DBSCAN-based induced graph under-sampling method. Arab. J. Sci. Eng. **49**, 1–39 (2024). https://doi.org/10.1007/s13369-024-08740-0
31. Li, Q.; Yue, S.; Ding, M.: Volume and surface area-based cluster validity index. IEEE Access **8**, 24170–24181 (2020). https://doi.org/10.1109/ACCESS.2020.2968938
32. Scitovski, R.; Sabo, K.: DBSCAN-like clustering method for various data densities. Pattern Anal. Appl. **23**(2), 541–554 (2020). https://doi.org/10.1007/s10044-019-00809-z
33. Zuo, Y.; Hu, Z.; Yuan, S.; Zheng, J.; Yin, X.; Li, B.: Identification of convective and stratiform clouds based on the improved DBSCAN clustering algorithm. Adv. Atmos. Sci. **39**(12), 2203–2212 (2022). https://doi.org/10.1007/s00376-021-1223-7
34. Alweshah, M.; Almiani, M.; Alkhalaileh, S.; Kassaymeh, S.; Hezzam, E.A.; Alomoush, W.: Parallel metaheuristic algorithms for solving imbalanced data classification problems. IEEE Access (2023). https://doi.org/10.1109/ACCESS.2023.3324061
35. Standard. IEEE recommended practice for monitoring electric power quality. IEEE Standard (1995). https://doi.org/10.1109/IEEESTD.1995.79050
36. Standard: Voltage Characteristics of Electricity Supplied by Public Electricity Networks. German Institute for Standardisation (2010)
37. Lin, C.H.; Chen, C.S.; Chuang, H.J.; Huang, M.Y.; Huang, C.W.: An expert system for three-phase balancing of distribution feeders. IEEE Trans. Power Syst. **23**(3), 1488–1496 (2008). https://doi.org/10.1109/TPWRS.2008.926472
38. Baran, M.E.; Wu, F.F.: Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE Trans. Power Deliv. **4**(2), 1401–1407 (1989). https://doi.org/10.1109/61.25627
39. Ghorbanian, M.; Dolatabadi, S.H.; Siano, P.: Game theory-based energy-management method considering autonomous demand response and distributed generation interactions in smart distribution systems. IEEE Syst. J. **15**(1), 905–914 (2020). https://doi.org/10.1109/JSYST.2020.2984730