A PROPOSED MODEL FOR DETERMING BIDDING ZONES IN THE ELECTRICITY MARKET USING LOCATION MARGINAL PRICES AND CLUSTERING ALGORITHMS

Le Hong Lam*, Tran Nguyen Thu Trang, Le Trong Minh Duc, Nguyen Hai Duc

The University of Danang - University of Science and Technology

*Corresponding author: lhlam@dut.udn.vn

(Received: May 05, 2025; Revised: June 18, 2025; Accepted: June 22, 2025)

DOI: 10.31130/ud-jst.2025.23(9C).520E

Abstract - Currently, electricity demand in Vietnam is rising rapidly due to industrialization, modernization, and population growth, creating risks of shortages, especially in the dry season when hydropower declines. To ensure supply, coal and oil generation is often dispatched, increasing costs and pollution. Renewable energy is expanding but faces issues of price volatility and supply reliability. A zonal electricity market design is therefore seen as a promising solution to optimize pricing, attract investment, and promote clean energy development, contributing to carbon reduction. This paper proposes a two-step zonal market model: first, node indices are computed under different scenarios using DC Optimal Power Flow (DC-OPF); second, nodes are grouped into bidding zones via Spectral Clustering. The model is tested on a 118-bus system with solar and rainy-day datasets to identify efficient and practical zonal configurations.

Key words - clustering, bidding zone, DC-OPF, locational marginal price, bus connection

1. Introduction

Vietnam's electricity market is currently transitioning from a monopoly model to a competitive market, aiming toward a fully competitive generation and wholesale system. However, the current electricity pricing mechanism does not adequately reflect regional differences in generation and transmission costs. This results in uniform pricing, which fails to incentivize investment and does not accurately reflect grid infrastructure constraints. Several regions of Vietnam's power system frequently face transmission congestion, especially along lines from the Central to the North and from the South to the Central regions. The absence of a zonal pricing mechanism prevents generators in surplus regions from operating at full capacity, while deficit regions still rely on distant sources, leading to higher transmission losses and increased risk of supply-demand imbalance. In this context, market zoning becomes essential for improving operational efficiency, transparency, and investment signals.

Countries like the U.S., Australia, and Brazil have implemented clustering models based on Locational Marginal Pricing (LMP) [1] combined with grid topology to form independent price zones [2]. Notably, Italy-whose elongated geography is similar to Vietnam-has adopted a zonal electricity market with 21 price zones [3]. This approach better reflects transmission congestion, encourages local generation development, and ensures greater transparency and pricing accuracy. This study employs the DC Optimal Power Flow (DC-OPF) model to determine LMPs-key indicators reflecting the marginal cost of delivering electricity to each node. While DC-OPF is a standard tool that models optimal power dispatch under linear constraints [4], it is not a clustering algorithm; it does not segment the grid into coherent zones as it overlooks network connectivity and nodal similarity.

An alternative is K-means clustering [5], favored for its simplicity and speed. Yet, K-means relies solely on Euclidean distance and fails to incorporate physical network structure, congestion, or grid topology, leading to technically inaccurate zones in complex power systems like Vietnam.

This paper proposes a market zoning method for Vietnam's electricity market based on the Spectral Clustering algorithm, with a key contribution of simultaneously combining economic data (locational marginal price – LMP) and technical data (physical linkage matrix and congestion matrix) to identify price zones that accurately reflect the actual operation of the system. This method addresses the limitations of traditional approaches such as pure OPF or K-means, which have not fully considered the physical connections between nodes and congestion phenomena in the grid. The model has been tested on the IEEE 118-bus system under four typical operating scenarios (High Sun, Low Sun, High Rain, Low Rain), thereby constructing a unified zoning scheme that ensures feasibility and accuracy when applied to Vietnam's power system. The research results not only provide a scientific basis for designing zonal pricing mechanisms but also contribute to policy orientation, infrastructure investment incentives, reduction of transmission losses, and improvement of market transparency.

2. Optimal Power Flow (OPF)

The Optimal Power Flow (OPF) algorithm for calculating market indices has been developed. This model is fully described in [6], [7]. OPF captures key aspects of actual transmission network operations, such as explicit N-1 security criteria, the capability to incorporate special contingencies like double circuit line outages, and the consideration of both preventive and corrective remedial actions.

DC OPF Model [8].

Equations (1) - (4) present the mathematical formulation of the DC-OPF model. The objective function

(1) aims to minimize the total generation cost:

$$\text{MinOPF}\sum_{g1,g2} a_g.(P_g)^2 + b_g.P_g + c_g \quad (1)$$

Eqs (2)-(4) present the network constraints. Here, (2) presents power flow of the line ij, (3) guaranties the balance power at each node, and (4) presents the limitation of transfer capacity of the line ij. Finally, the angle of slack bus (i) is fixed to zero (5).

$$\text{St:} \quad P_{ij} = \frac{\delta_i - \delta_j}{x_{ij}^2} \ \forall l \in L \quad (2)$$

$$P_g^i - P_d^i - P_{ij} + P_{ji} = 0 \ \forall i \quad (3)$$

$$-P_{ij}^{max} \leq P_{ij} \leq P_{ij}^{max} \ \forall l \in L \quad (4)$$

$$\partial^{slack} = 0 \quad (5)$$

where, $g \in G$ is generator and $d \in D$ is demand, and i, j are nodes between line $l \in L$. a_g, b_g, c_g are parameters in cost function of generator.

Because of using DC-OPF model, some assumptions are adopted: (i) Only consider active power flow; (ii) the branch impedance is equal to the reactance only; (iii) all voltage magnitudes are equal to 1 p.u; (iv) voltage angles are close to each other, so sin (δi - δj) = δi - δ.

Locational Marginal Prices (LMPs) represent the marginal cost of electricity at each bus and are typically calculated using the Lagrange multipliers associated with the constraints in the Optimal Power Flow (OPF) problem [7], [9]. In this study, LMPs are derived from the proposed DC Optimal Power Flow (DC-OPF) model, where system security constraints are explicitly modeled. This approach ensures that the calculated LMPs accurately reflect the applied assumptions, thereby simultaneously capturing both the technical requirements of power system operation and the economic efficiency of the electricity market, along with the relevant market operation rules.

3. Spectral Clustering Algorithm

Spectral Clustering [10] operates by representing the dataset as an undirected weighted graph G = (V, E), where:

- V is the set of vertices, with each vertex representing a data point.

- E is the set of edges, representing the relationships or similarities between data points.

- The weight $w_{ij} \geq 0$ indicates the degree of similarity between data points x_i and x_j.

From this graph, a weight matrix $W \in R^{n \times n}$ is defined, where $W_{ij} = w_{ij}$. Additionally, a degree matrix $D \in R^{n \times n}$ is constructed as a diagonal matrix with entries:

$$D_{ii} = \sum_j W_{ij}$$

This matrix captures the total connection weight of each vertex in the graph.

Spectral Decomposition: In Spectral Clustering, the number of selected eigenvectors and the number of clusters k critically influence clustering quality. Eigenvectors derived from the Laplacian matrix project the data into a new subspace where the inherent cluster structure becomes clearer. Selecting too few or too many eigenvectors can result in misgrouping or over-fragmentation. Similarly, choosing k must balance technical and economic characteristics of the power system (e.g., LMP distribution, congestion) with operational feasibility. This study evaluates multiple operating scenarios and applies criteria like the eigenvalue gap and the Elbow method to optimize these parameters, ensuring that market zoning is both accurate and practically applicable.

Advantages:

• **Flexible with nonlinear data:** The algorithm can handle clusters with complex shapes, even when clusters are not spherically distributed.

• **Effective for graphs and social networks:** Spectral Clustering performs well on graph-structured data, making it useful for community detection and clustering within networks.

• **Insensitive to initialization:** Unlike K-means, this algorithm is not sensitive to the initial placement of cluster centers.

Disadvantages:

• **High computational cost:** Matrix decomposition involves eigenvalue computations, which can be resource-intensive for large datasets.

• **Dependence on the affinity matrix:** The clustering quality heavily depends on how the affinity (similarity) matrix between data points is constructed.

• **Difficult to determine the number of clusters:** The algorithm requires the number of clusters k to be specified in advance, which is not always straightforward.

Compared to K-means, Spectral Clustering offers significant advantages when applied to electricity market zoning. K-means relies solely on Euclidean distances between data points, which makes it effective for identifying spherical, uniformly distributed, and relatively simple clusters. However, in power systems, buses are interconnected through a complex network with various technical characteristics such as transmission congestion, voltage and power variations, locational marginal prices (LMPs), and regions without clearly defined geometric boundaries. K-means is unable to accurately capture these zones because it does not consider the underlying grid topology or physical connectivity between nodes.

In contrast, Spectral Clustering leverages the adjacency matrix and the graph Laplacian to reflect the true structure of the network. It is not constrained by cluster shape and can identify irregular, non-uniform zones, making it particularly suitable for systems with complex physical interconnections like power grids. By integrating both LMP information and grid connectivity, Spectral Clustering provides more accurate zonal partitions, enhancing the effectiveness of price zone identification, system dispatch, and congestion management.

So, we chose to apply Spectral Clustering because its advantages align well with the structure of the data input. The power outputs from loads are nonlinear, resulting in LMPs that also exhibit nonlinear characteristics. Furthermore, the system includes various models and connectivity states between buses, making Spectral Clustering a more suitable and optimal choice compared to the K-means algorithm.

Theoretical basis and algorithm implementation process:

Step 1: Construct the adjacency matrix W

Based on the Gaussian Kernel or other similarity measures:

$$W_{ij} = e^{\frac{-(LMP_i - LMP_j)^2}{2\sigma^2}} \tag{6}$$

Step 2: Build the Laplacian matrix

- Compute the degree matrix D (a diagonal matrix where each entry is the sum of the connections of each point):

$$D_{ii} = \sum_j W_{ij} \tag{7}$$

- Then compute the Laplacian matrix:

Unnormalized Laplacian: L = D – W (8.1)

Or Normalized Laplacian:

$$L_{sym} = D^{\frac{-1}{2}}\, L\, D^{\frac{-1}{2}} \tag{8.2}$$

Step 3: Select k eigenvectors

Calculate the eigenvectors of L.

Setp 4: Map the Data into a New Space

Select the k smallest eigenvectors (excluding the first eigenvector if all its elements are zeros) to form the new space, denoted as vecs_k.

- Each column in vecs_k represents an eigenvector, allowing the data to be represented in the newly constructed feature space.

- Each row in vecs_k corresponds to a data point projected into the new space.

- Selecting the first five eigenvectors helps to reduce the data dimensionality while still preserving the clustering information.

Note: Only the real part of the eigenvectors instead follows the actual connectivity structure of the buses.

Step 5. Execute K-means Clustering in the New Space

- Use the K-mean algorithm cluster the data point based on their new coordinates.

- In this transformed space, clustering no longer relies solely on the coordinates (i.e., Euclidean distances) but steps:

Step 6.1: Initialize Cluster Centroids

Randomly select **k** points from the dataset to serve as the initial cluster centroids.

Step 6.2: Assign Each Data Point to the Nearest Cluster

Each data point originally located at x_i in the original space is transformed to a new coordinate y_i in the eigenvector space.

The distance between two points y_i and y_j is computed using the Euclidean distance formula:

$$d_{(y_i \cdot y_j)} = \sqrt{\sum_{m=1}^{k} (y_{i,m} - y_{j,m})^2} \tag{9}$$

Where: $y_{i,m}$ is the m-th component of point y_i in the eigenvector space. k is the number of smallest eigenvectors selected.

Step 6.3: Update Cluster Centroids

Suppose cluster C_j contains N_j data points $y_1, y_2, \dots y_j$ in the eigenvector space (each point has k coordinates corresponding to the k smallest eigenvectors).

The centroid μ_j of cluster C_j is computed as the mean of all points within the cluster:

$$\mu_j = \frac{1}{N_j} \sum_{i \epsilon\, C_j} y_i \tag{10}$$

Where: μ_j *is the new centroid vector of cluster j,* y_i *is the coordinate vector of data point i in the eigenvector space,* N_j *is the number of data points in cluster* C_j.

4. Proposed Algorithm

1. Start with input data: including Locational Marginal Prices (LMPs) at buses, the connectivity (adjacency) matrix, and the congestion matrix.

2. Initialize parameters: set the sensitivity parameter σ (for the adjacency matrix) and iteration index i=1.

3. Compute the adjacency matrix W using a Gaussian similarity function based on LMP differences, reflecting similarity between buses.

$$W_{ij} = e^{\frac{-(LMP_i - LMP_j)^2}{2\sigma^2}}$$

4. Construct the Laplacian matrix L from W, capturing the network's structural connectivity.

$$L = I - D^{-1} W D^{-1}$$

5. Select eigenvalues and eigenvectors from L to construct a new spectral feature space.

6. Apply K-means clustering in the spectral space:

Initialize cluster centroids.

Assign data points to the nearest centroid.

Update centroids.

Iterate until convergence.

7. Evaluate clustering results:

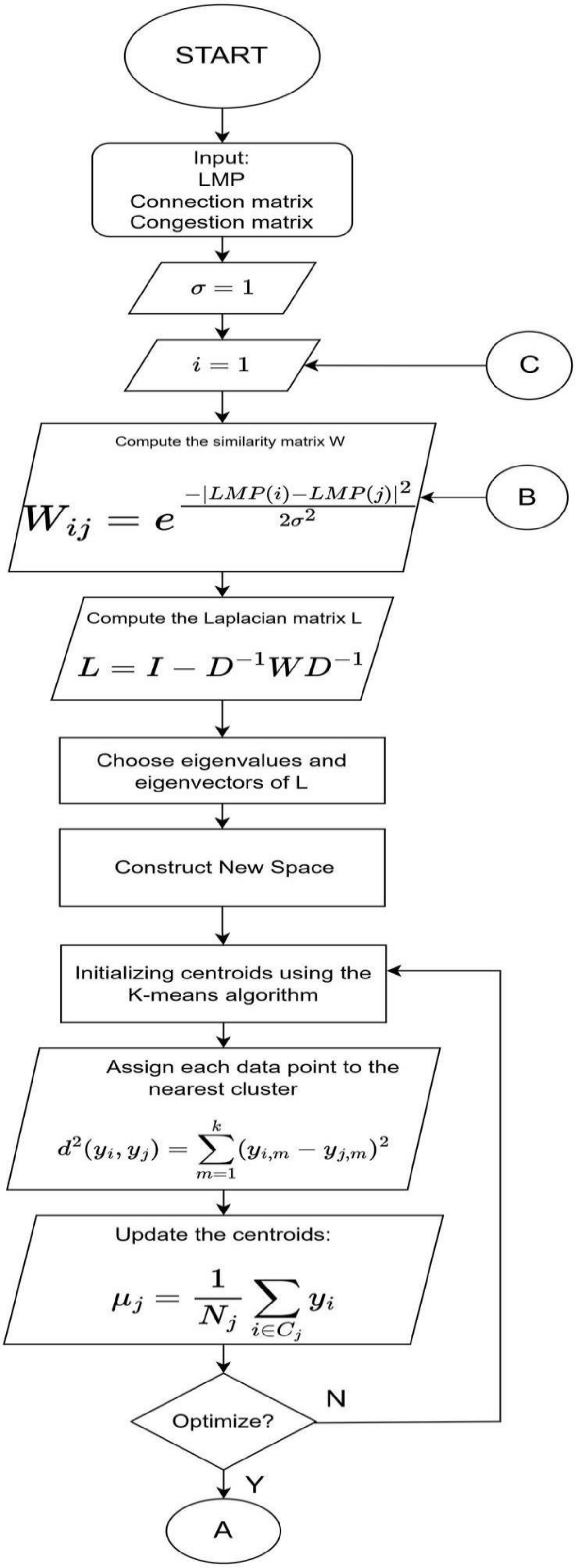
Check physical connectivity among buses in each cluster.

Ensure minimal internal congestion within clusters.

8. If not satisfactory, adjust σ\sigmaσ or continue iteration i until convergence criteria are met.

9. Identify the optimal clustering result and finalize the zonal part

Four hypothetical scenarios are developed and tested using a 118-node IEEE model electricity market to apply the Spectral clustering algorithm under peak and off-peak load conditions during the dry and rainy seasons. Clustering is based on local marginal prices (LMPs) to group nodes with similar LMP characteristics, ensuring node connectivity and accounting for transmission congestion, which reflects cost and performance correlations in the electricity market. The test data includes hourly LMPs for four days representing different operating scenarios:



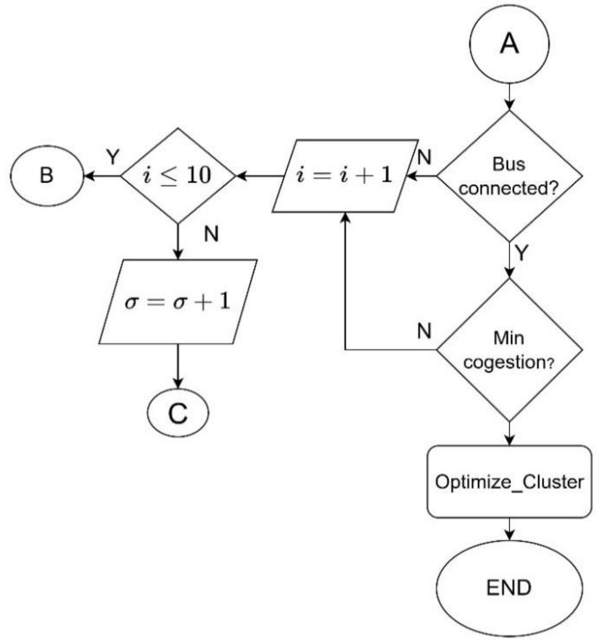


Figure 1. Clustering Algorithm flowchart

Peak sunny day scenario: Under sunny conditions, solar power production increases, leading to a decrease in LMP in areas with abundant solar resources due to low marginal costs. However, sudden increases in supply can cause congestion on transmission lines if the infrastructure is not capable of handling the large amount of power, leading to voltage fluctuations and even system instability.

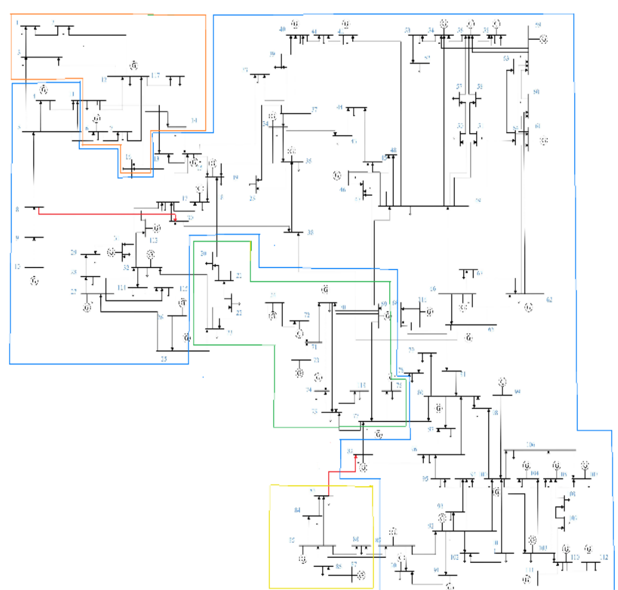


Figure 2. Peak Sun Day Zoning

Low sunny day scenario: When sunlight is low, solar power production decreases, causing LMP to increase due to the need to mobilize more expensive power sources. Although transmission congestion may be reduced due to lower transmission demand, the reliance on other power sources can put pressure on other parts of the system, affecting voltage stability and system reliability.

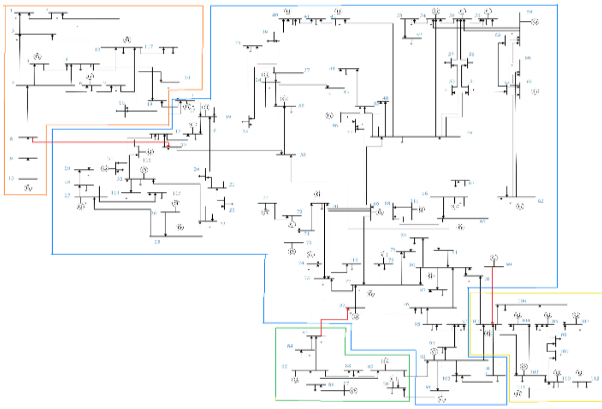


***Figure 3.** Low sun day zoning*

Peak rainy day scenario: Under heavy rain conditions, if the system has many hydropower plants, hydropower output increases, leading to a decrease in LMP in these areas. However, similar to the sunny scenario, the increase in local supply can cause transmission congestion if not well managed, affecting voltage and system stability.

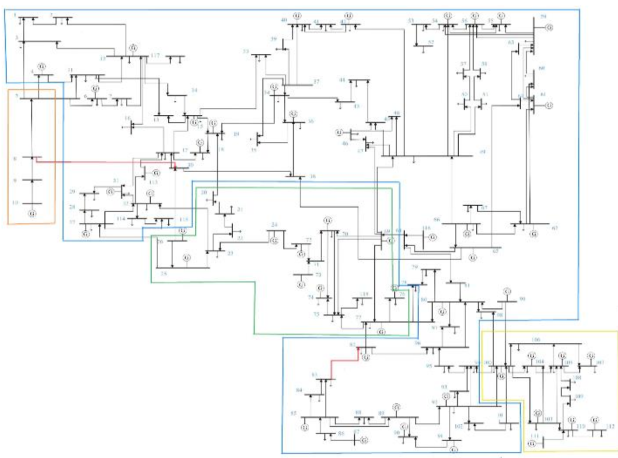


***Figure 4.** Peak Rainy Day Zoning*

Low rain day scenario: When rainfall is low, hydropower generation decreases, leading to increased LMP due to the need to mobilize other, more expensive sources of power. This can reduce transmission congestion in areas that have traditionally relied on hydropower, but can also put pressure on other sources of power and affect voltage stability.

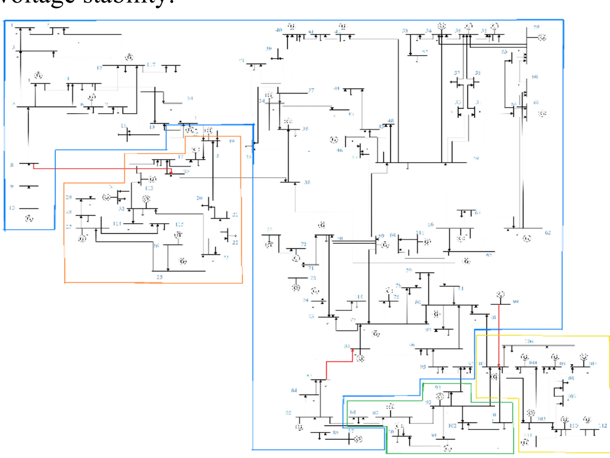


***Figure 5.** Low Rainy Day Zoning*

To cope with these fluctuations, it is necessary to define a unified partition for all four scenarios.

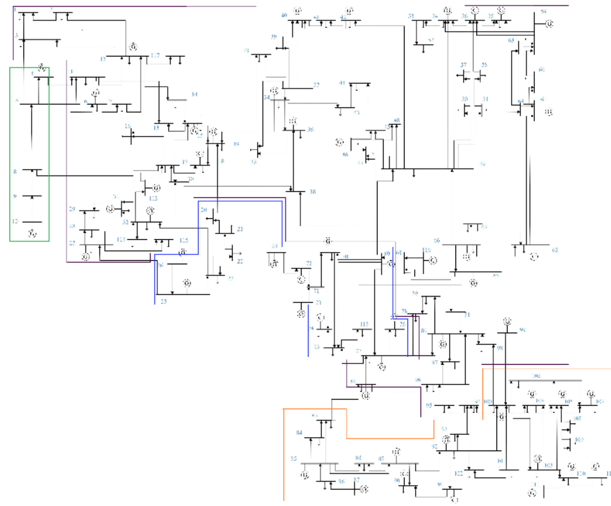


***Figure 6.** Algorithm Partitioning Results*

The power system zoning results in 96 different schemes due to dynamic data over 24 hours from four hypothetical scenarios. However, considering all schemes is impractical, so we focus on the peak load case-when the system faces the most severe conditions, prone to congestion and instability. Selecting zoning based on this scenario ensures the optimal scheme performs well under extreme conditions while maintaining efficiency during lower load periods. Analysis of the zoning schemes at peak load reveals that most results converge to one main scheme, indicating it is the most suitable according to the proposed criteria. Specifically:

1. VoltageVoltage Stabilizer: Ensures voltage remains within allowable limits, promoting stable system operation and preventing voltage-related incidents.

2. Optimizing LMP in Clusters: Buses within the same cluster have similar LMPs, reducing operating costs and ensuring effective zoning.

3. Reduce Congestion: This solution significantly reduces grid congestion, improving power transmission without overloading lines.

4. Maintain Connectivity: Ensures all nodes in a cluster remain connected, avoiding isolation and enhancing system flexibility.

Zoning based on criteria like operational synchronization, price sensitivity, and technical constraints optimizes power system operation, ensuring safe, stable, and efficient supply. This zoning scheme is selected as the unified solution, representing the optimal structure for varying conditions. Limiting congestion addresses transmission issues by reducing current overloads, preventing voltage instability that impacts power quality. While other schemes may arise due to data fluctuations, they often fail to balance criteria, either increasing LMP differences or losing connectivity. This scheme was chosen as the reference model because it best meets the requirements for voltage stability, LMP, congestion management, and interconnectivity, ensuring stable and economical system operation.

5. Conclusion

In the field of power system analysis, traditional clustering methods like K-Means rely on geometric distance and often overlook physical connections and congestion in power systems, leading to clusters that may not reflect the system's true structure. To address this, the Spectral Clustering algorithm is applied, considering both physical connections and congestion. By using eigenvalues from the connection matrix, it identifies clusters with stronger physical ties. Spectral Clustering offers several advantages in power system analysis, such as accurately identifying bidding zones, improving transmission congestion management, and enhancing price transparency. It also helps pinpoint system bottlenecks, supports congestion management, and encourages renewable energy participation. Despite challenges with large datasets, this method shows promise for improving the efficiency of Vietnam's electricity market.

REFERENCES

[1] G. Chicco *et al.*, "Overview of the Clustering Algorithms for the Formation of the Bidding Zones". *2019 54th International Universities Power Engineering Conference (UPEC). IEEE*, 2019.

[2] F. C. Schweppe, M. C. Caramanis, R. D. Tabors, R. E. Bohn, *Spot Pricing of Electricity*, Kluwer Academic Publishers, 1988.

[3] L. Michi *et al.*, "Optimal Bidding Zone Configuration: Investigation on Model-based Algorithms and their Application to the Italian Power System", *Proceedings of the 111th Annual AEIT International Conference, Florence, Italy,* 18-20 Sept. 2019

[4] C. W. Tan, D. W. H. Cai, and Xin Lou, "DC optimal power flow: Uniqueness and algorithms", *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm), Tainan,* 2012, pp. 641-646, doi: 10.1109/SmartGridComm.2012.6486058.

[5] B. Chong, "K-means clustering algorithm: a brief review", *Data Science and Big Data Technology, Shanxi University of Finance and economics, Taiyuan, Shanxi,* ISSN 2616-5775 Vol. 4, Issue 5: 37-40, DOI: 10.25236/AJCIS.2021.040506

[6] C. Bovo, V. Ilea, E. M. Carlini, M. Caprabianca, F. Quaglia, L. Luzi, and G.Nuzzo, "Optimal computation of Network indicators for Electricity Market Bidding Zones configuration", *Proceedings of the 2020 55th International Universities Power Engineering Conference (UPEC), Torino, Italy,* 1–9 September 2020, pp. 1–6.

[7] C. Bovo, V. Ilea, E. M. Carlini, M. Caprabianca, F. Quaglia, L. Luzi, and G. Nuzzo, "Optimal Computation of Network Indicators for Electricity Market Bidding Zones Configuration Considering Explicit N-1 Security Constraints", *Energies,* vol. 14, art. 4267, 2021.

[8] L. H. Lam and V. D. T. An. "A Strategy to Identify Congestions of Transmission Networks in N-1 Contingency: Khanh Hoa Case Study". *The University of Danang - Journal of Science and Technology*, vol. 21, no. 7, pp. 1-6, 2023.

[9] Y. M. Al-Abdullah and M. Sahraei-Ardakani, "Differences in locational marginal prices: Deterministic vs. stochastic market formulations", *2018 5th International Conference on Renewable Energy: Generation and Applications (ICREGA), Al Ain, United Arab Emirates*, 2018, pp. 268-272, doi: 10.1109/ICREGA.2018.8337598.

[10] U. V. Luxburg, U. A tutorial on spectral clustering. *Stat Comput*, vol. 17, pp. 395–416, 2007. https://doi.org/10.1007/s11222-007-9033-z