Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Large Language Model-Assisted Zoning for Heterogeneous Automatic Voltage Control in Renewable-Rich Regional Power Grids

Jun Zhao

2644665107@qq.com

State Grid Shanxi Electric Power Company Electric Power Research Institute

Xiaoqiang Nan

State Grid Shanxi Electric Power Company (China)

Shifeng Zhang

State Grid Shanxi Electric Power Company Electric Power Research Institute

---

Article

Keywords: renewable energy integration, automatic voltage control, large language models, control-region zoning, multi-agent deep deterministic policy gradient

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11301328/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Large Language Model-Assisted Zoning for Heterogeneous Automatic Voltage Control in Renewable-Rich Regional Power Grids

Jun Zhao[1*], Xiaoqiang Nan[2] & Shifeng Zhang[1]

High penetration of wind and photovoltaic generation complicates reactive power and voltage regulation in regional power grids. A uniform automatic voltage control (AVC) mode may not meet differing regional requirements, while independent control by electrically coupled renewable energy stations can produce poorly coordinated reactive power responses. This paper proposes a heterogeneous AVC framework assisted by a large language model (LLM). First, the LLM uses structured operating features and AVC constraints to group renewable energy stations into control regions, assign control modes, and recommend initial parameters. Then, constant reactive power (CQP) and conventional constant-voltage (CV) strategies are applied in their assigned regions. In strongly coupled regions, an external CV loop determines station-group reactive power requirements, which a multi-agent deep deterministic policy gradient (MADDPG) policy allocates among stations. Finally, power-flow calculations validate the control results, and the validation feedback is used to refine the zoning, mode assignments, and parameters. Case studies on a modified IEEE 39-bus system show improved voltage regulation, more balanced reactive power utilization among stations, and fewer voltage violations than uniform CV control.

Keywords renewable energy integration; automatic voltage control; large language models; control-region zoning; multi-agent deep deterministic policy gradient

Introduction

Renewable-rich power systems are an important part of the low-carbon energy transition[1, 2]. The integration of wind farms and photovoltaic plants changes regional grid operation and increases demands on reactive power and voltage control[3]. Automatic voltage control (AVC) coordinates reactive power resources to regulate grid voltage; its performance affects voltage security and reactive power support when renewable penetration is high[4]. AVC strategies must therefore account for differences in operating conditions across regions.

Regional AVC schemes for renewable energy stations generally include two basic control modes: constant reactive power (CQP) control and constant voltage (CV) control. CQP control operates according to a predefined reactive power reference and lacks direct response to voltage deviations, resulting in limited voltage regulation capability. In contrast, CV control can track voltage references effectively; however, independent regulation among multiple stations may lead to uneven reactive power allocation and reactive power counteraction. Therefore, a single control mode is difficult to satisfy the diversified reactive power and voltage regulation requirements of renewable-rich regional power grids.

Previous work has addressed regional AVC zoning and coordination. Ref. 5 developed an adaptive zone-division AVC system that adjusts control regions to operating conditions. Ref. 6 proposed robust reactive power partitioning under frequent power-flow fluctuations to stabilize region boundaries. Ref. 7 introduced regional-provincial AVC coordination across dispatch levels under renewable energy integration. For multi-station CV control, Ref. 8 adjusted the reactive power regulation step size of AVC substations and imposed output constraints. These changes improved reactive power sharing among wind farms and accelerated voltage regulation. Ref. 9 combined model- and data-driven methods for photovoltaic cluster reactive power and voltage control, balancing system-wide optimization performance with online response capability.

These studies have improved the adaptability of AVC zoning and the coordination of reactive power and voltage control. Most, however, address a particular control level or predefined station groups. Zoning and strategy design are often treated separately, with zoning used mainly to limit the control area rather than to assign different modes. Renewable-rich regional grids therefore need a framework that links regional voltage-control requirements, mode assignment, and performance feedback.

After zoning and mode assignment, CQP or conventional CV control can serve regions with relatively simple operating characteristics. Strongly coupled constant-voltage regions, however, require coordinated allocation of reactive power regulation tasks among stations. Fixed-ratio allocation or independent responses to local voltage deviations can overlook differences in inter-station electrical coupling and response characteristics, leading to uneven reactive power sharing[10, 11].

Deep reinforcement learning (DRL) can learn a policy that maps operating states to control actions through interaction with a power-system environment. Once trained, a policy can reduce the need for repeated online optimization. Ref. 12 applied DRL to autonomous voltage control, and Ref. 13 developed a two-time-scale voltage control method. For control across multiple devices, multi-agent deep reinforcement learning (MADRL) has attracted attention. Ref. 14 formulated active voltage control as a cooperative multi-agent task. Ref. 15 used multi-agent deep deterministic policy gradient (MADDPG) to coordinate photovoltaic inverter reactive power, illustrating its use for coupled devices and continuous actions.

Reward design and training parameters also affect DRL control performance. Ref. 16 used reward shaping to improve learning under complex operating conditions, while Ref. 17 introduced an automatic optimization mechanism during training to reduce reliance on expert tuning. However, studies of regional AVC have given limited attention to coordinated reactive power allocation among stations under CV control. In many methods, reward weights and training parameters remain fixed in advance, limiting their adjustment in response to operating conditions and validation feedback.

Large language models (LLMs) have also been explored for power-system analysis and control. Ref. 18 coupled a GPT agent with DRL to incorporate natural-language operating requirements into real-time optimal power flow. Ref. 19 adapted pretrained LLMs to power-system data and tasks. Ref. 20 used LLM-generated training data to improve the generalization of DRL-based voltage control

[1] Electric Power Research Institute of State Grid Shanxi Electric Power Co., Ltd., Taiyuan 030001, Shanxi, China
[2] State Grid Shanxi Electric Power Co., Ltd., Taiyuan 030021, Shanxi, China
* Correspondence: Jun Zhao, 2644665107@qq.com

across operating conditions. Ref. 21 integrated LLMs into MADRL training and adjusted reward functions using training and testing feedback.

These studies show the potential of LLMs as decision aids in power systems, but their application to regional AVC remains limited. They leave open how an LLM can combine regional operating features, grid information, and AVC rules to support zoning and mode assignment and then refine these decisions using validation feedback. This gap motivates an upper-level LLM module that assists these tasks without generating station-level reactive power commands.

This study proposes an LLM-assisted zoning framework for heterogeneous AVC in renewable-rich regional power grids. First, structured operating features are developed, and the LLM uses them to group stations into control regions, assign modes, and recommend initial parameters. Then, each region applies its assigned strategy. In strongly coupled CV regions, an external CV loop determines station-group reactive power requirements, and CV-MADDPG allocates them among stations. Finally, power-flow calculations validate the control results and provide feedback for revisions to the zoning, modes, and parameters. The LLM also uses training and validation feedback to refine CV-MADDPG reward weights and training parameters.

LLM-Assisted Zoning-Based Heterogeneous AVC Framework

Framework of Zoning-Based Heterogeneous AVC Control

With the high penetration of renewable energy, renewable energy stations, such as wind farms and photovoltaic plants, are commonly connected to regional power grids through collector lines or adjacent grid nodes. The fluctuations in renewable power output can significantly change the reactive power and voltage regulation characteristics of regional power systems. If CQP control is uniformly applied across the entire region, weak voltage nodes may not receive sufficient reactive power support. Conversely, when CV control is adopted uniformly, independent closed-loop regulation among multiple stations may result in overlapping control actions, leading to uneven reactive power allocation, reactive power counteraction at the same nodes, and frequent device operations. Therefore, it is necessary to select heterogeneous AVC control modes according to regional operating characteristics. Figure 1 shows the proposed framework for LLM-assisted zoning and heterogeneous AVC.

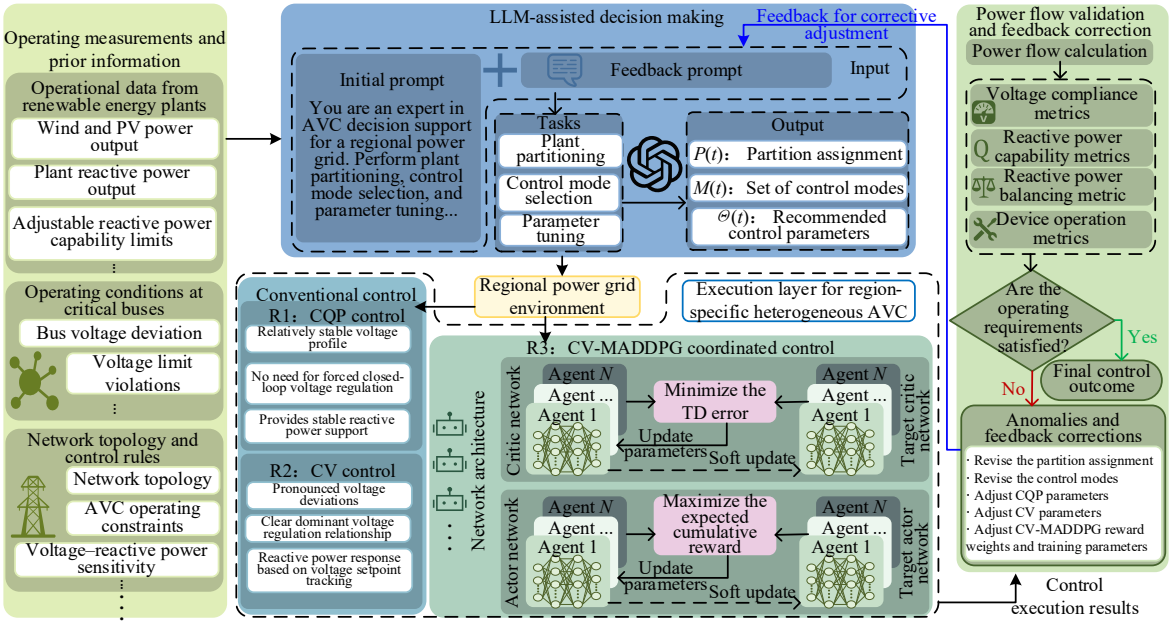


Figure 1. Framework for LLM-assisted zoning and heterogeneous AVC

First, the LLM receives station operating states, key-node voltages, grid topology information, and AVC rules. It recommends station zoning, control modes, and initial parameters. Then, the assigned regions apply CQP, CV, or CV-MADDPG control; CV-MADDPG coordinates reactive power allocation among stations in strongly coupled regions. Finally, power-flow calculations assess voltage limits, reactive power capability, allocation balance, and device operations. If a requirement is not met, the validation results are returned to the LLM for revisions to the zoning, control modes, or parameters.

Let $\mathbf{X}(t)$ denote the regional operating feature matrix, $\mathrm{K}_{\mathrm{grid}}$ represent the grid topology and operational prior knowledge, and $\mathrm{R}_{\mathrm{AVC}}$ indicate the AVC control rules and operating constraints. The LLM-assisted zoning decision is expressed as:

$$[\mathrm{P}(t), \mathrm{M}(t), \boldsymbol{\Theta}(t)] = f_{\mathrm{LLM}}\left(\mathbf{X}(t), \mathrm{K}_{\mathrm{grid}}, \mathrm{R}_{\mathrm{AVC}}\right) \tag{1}$$

where $\mathrm{P}(t)$ represents the station zoning results generated by the LLM, $\mathrm{M}(t)$ denotes the set of control modes assigned to different regions, and $\boldsymbol{\Theta}(t)$ represents the recommended control parameters.

Based on the output of Eq. (1), renewable energy stations are classified into three types of control regions: the CQP control region $\mathrm{G}_{\mathrm{CQP}}$, the CV control region G_{CV}, and the complex coupled control region $\mathrm{G}_{\text{CV-MADDPG}}$. The $\mathrm{G}_{\mathrm{CQP}}$ region generally corresponds to areas with relatively stable voltage profiles and weak closed-loop voltage regulation requirements. In such regions, reactive power support can be provided by maintaining a predefined reactive power reference. The G_{CV} region generally represents areas with noticeable voltage deviations and relatively clear voltage regulation characteristics, where voltage support capability can be improved by tracking voltage references. The $\mathrm{G}_{\text{CV-MADDPG}}$ region corresponds to areas with strong interactions among multiple stations and key nodes. If only CQP control is applied, the voltage variations at multiple nodes may not be responded to in time. Meanwhile, directly applying conventional CV control may cause conflicts among station-level voltage regulation actions, resulting in uneven reactive power allocation and frequent control adjustments. Therefore, coordinated reactive power allocation among multiple stations is further considered in the

complex coupled region. Accordingly, the regional AVC control scheme is transformed from a single control mode into a collaborative process of zoning identification, control-mode matching, and heterogeneous control implementation, providing the basis for subsequent control strategy design.

Regional Operating Constraints and Zoning Feature Modeling

Because regional AVC controls the aggregate reactive power response of renewable energy stations, the reactive power resources within each station are aggregated into an adjustable station-level capability. Let the lower and upper reactive power regulation limits of station i at time t be denoted as $Q_i^{\min}(t)$ and $Q_i^{\max}(t)$, respectively. After zoning-based control, the station reactive power responses and key-node voltages must satisfy the following constraints:

$$\begin{cases} V_j^{\min} \le V_j(t) \le V_j^{\max}, & j \in \mathrm{B} \\ Q_i^{\min}(t) \le Q_i^{\mathrm{c}}(t) \le Q_i^{\max}(t), & i \in \mathrm{G} \\ \left| Q_i^{\mathrm{c}}(t) - Q_i^{\mathrm{c}}(t-1) \right| \le \Delta Q_i^{\max}, & i \in \mathrm{G} \end{cases} \tag{2}$$

where G denotes the set of renewable energy stations, B represents the set of key voltage nodes, $V_j(t)$ denotes the voltage magnitude of the j-th key node, $Q_i^{\mathrm{c}}(t)$ represents the reactive power response of station i after zoning-based control, and $\Delta Q_i^{\max}$ is the maximum allowable change in station reactive power during one control period.

To characterize a station's ability to continue regulating reactive power, define the normalized reactive power margin as:

$$\rho_i(t) = \frac{\min\left(Q_i^{\max}(t) - Q_i(t), Q_i(t) - Q_i^{\min}(t)\right)}{Q_i^{\max}(t) - Q_i^{\min}(t) + \varepsilon} \tag{3}$$

where $Q_i(t)$ denotes the current reactive power output of station i, and ε is a small positive constant introduced to avoid division by zero. A larger $\rho_i(t)$ indicates more bidirectional reactive power headroom at the current operating point.

For LLM-assisted zoning, candidate-region operating states and electrical coupling are organized as structured inputs. Let the set of candidate regions be $\mathrm{R} = \{1, 2, \cdots, N_r\}$. For region r, let its station set and key-node set be $\mathrm{G}_r \subseteq \mathrm{G}$, $\mathrm{B}_r \subseteq \mathrm{B}$, respectively. The regional feature matrix is:

$$\begin{cases} \mathbf{X}(t) = \left[\mathbf{x}_1(t), \mathbf{x}_2(t), \cdots, \mathbf{x}_{N_r}(t) \right]^{\mathrm{T}}, \\ \mathbf{x}_{N_r}(t) = \left[\mathbf{F}_V^r(t), \mathbf{F}_Q^r(t), \mathbf{F}_S^r(t), \mathbf{F}_C^r(t) \right] \end{cases} \tag{4}$$

where $F_V^r(t)$, $F_Q^r(t)$, $F_S^r(t)$, and $F_C^r(t)$ represent the regional voltage state, reactive power capability, electrical coupling, and control action features, respectively. These constraints and features provide the basis for the region-specific AVC strategies developed below.

Heterogeneous AVC Control Strategy

Constant Reactive Power and Conventional Constant Voltage Control Strategies

The voltage regulation characteristics of the CQP control regions and conventional CV control regions are relatively clear, and basic control strategies can be adopted to generate the reactive power responses of renewable energy stations. Specifically, CQP control regions focus on maintaining stable reactive power support, while conventional CV control regions emphasize voltage reference tracking. To avoid frequent control actions caused by small voltage deviations around the acceptable voltage range, a voltage control deadband is introduced, and the effective voltage deviation of key node j is defined as:

$$\tilde{e}_j(t) = \begin{cases} 0, & |V_j^{\mathrm{ref}} - V_j(t)| \le \varepsilon_V \\ V_j^{\mathrm{ref}} - V_j(t), & |V_j^{\mathrm{ref}} - V_j(t)| > \varepsilon_V \end{cases} \tag{5}$$

where V_j^{ref} denotes the reference voltage of key node j, and ε_V represents the voltage control deadband.

For CQP control regions, the regional reactive power support requirement is obtained by weighting the voltage deviations of key nodes and is allocated according to the available reactive power capacity and voltage regulation contribution of each station. The equivalent reactive power response of station i is expressed as:

$$\begin{cases} Q_i^{\mathrm{ctrl}}(t) = Q_i(t) + \alpha_i(t) K_r \sum_{j \in \mathrm{B}_r} \omega_j \tilde{e}_j(t), \\ \alpha_i(t) = \dfrac{\rho_i(t) S_i \bar{S}_{i,r}(t)}{\sum_{k \in \mathrm{G}_r} \rho_k(t) S_k \bar{S}_{k,r}(t) + \varepsilon}, \quad i \in \mathrm{G}_{\mathrm{CQP}} \end{cases} \tag{6}$$

where B_r and G_r represent the sets of key voltage nodes and renewable energy stations in region r, respectively; K_r denotes the regional reactive power regulation coefficient; ω_j represents the voltage weight of key node j; $\alpha_i(t)$ is the reactive power allocation coefficient of station i; S_i denotes the rated capacity of station i; and $\bar{S}_{i,r}(t)$ represents the average voltage regulation contribution of station i to the key nodes in region r.

Conventional CV regions primarily track voltage references. Their dominant station-to-node voltage relationships are relatively clear, allowing each station's reactive power response to be determined from voltage deviations and voltage–reactive power sensitivities. Let $\lambda_{ji}(t)$ denote the voltage deviation weight of station i with respect to key node j, and define the weighted voltage deviation for station i as:

$$\begin{cases} \bar{e}_i(t) = \sum_{j \in \mathrm{B}_r} \lambda_{ji}(t) \tilde{e}_j(t), \\ \lambda_{ji}(t) = \dfrac{\left| \partial V_j / \partial Q_i \right|}{\sum_{k \in \mathrm{B}_r} \left| \partial V_k / \partial Q_i \right| + \varepsilon} \end{cases} \tag{7}$$

The equivalent reactive power response of station *i* in the conventional CV control region is expressed as:

$$Q_i^{\mathrm{ctrl}}(t) = Q_i(t) + K_i \frac{\bar{e}_i(t)}{\bar{S}_{i,r}(t) + \varepsilon}, \quad i \in \mathrm{G}_{\mathrm{CV}} \tag{8}$$

where K_i denotes the voltage control response coefficient of station *i*.

Coordinated Control Strategy for Complex Constant-Voltage Regions

For $\mathrm{G}_{\text{CV-MADDPG}}$, several stations significantly affect the voltages of multiple key nodes, creating strong voltage–reactive power coupling. Independent conventional CV control may cause stations to respond repeatedly to similar voltage deviations, leading to counteracting reactive power adjustments, uneven outputs, and frequent device operations. The proposed control scheme therefore separates the regional reactive power requirement from its allocation among stations. An external CV loop determines the requirement from key-node voltage deviations, and CV-MADDPG allocates it among stations in the region. This coordinated allocation problem is formulated as a decentralized partially observable Markov decision process (Dec-POMDP) and addressed using CV-MADDPG.

Dec-POMDP Formulation for Complex Constant-Voltage Regions

Let the key-node set in a complex constant-voltage region be $B = \{b_1, b_2, \cdots, b_m\}$, and the renewable energy station set be $G = \{g_1, g_2, \cdots, g_n\}$. Each key node is associated with a station group, which is treated as an agent. The agent set and its station groups are:

$$N = \{1, 2, \cdots, m\}, \quad G = \bigcup_{a=1}^{m} G_a \tag{9}$$

where G_a denotes the set of renewable energy stations assigned to agent *a*.

At time *t*, the external CV loop determines the reactive power requirement for agent *a* from key-node voltage deviations and the voltage–reactive power sensitivities of its associated stations.

$$\begin{cases} \Delta V_a(t) = V_a^{\mathrm{ref}} - V_a(t) \\ S_{BG}(t) = \left[S_{ji}(t) \right]_{m \times n}, \qquad S_{ji}(t) \approx \dfrac{V_j^{(i)}(t) - V_j^0(t)}{\Delta Q_i} \\ \Delta Q_a^{\mathrm{req}}(t) = f_{\mathrm{CV}}\left[\Delta V_a(t), S_{BG_a}(t) \right] \end{cases} \tag{10}$$

where V_a^{ref} and $V_a(t)$ denote the voltage reference and actual voltage of the key node corresponding to agent *a*, respectively; $S_{BG_a}(t)$ represents the voltage-reactive power sensitivity between the station group G_a and the corresponding key-node set; $V_j^0(t)$ denotes the reference voltage of key node *j* under the current operating condition; $V_j^{(i)}(t)$ represents the node voltage obtained from power-flow calculation; and $f_{\mathrm{CV}}(\cdot)$ represents the external CV control regulation logic of the region.

The external CV loop's reactive power requirements are clipped according to the stations' current outputs and adjustable limits so that they remain within the region's available regulation range. The resulting commands are denoted by $\Delta Q_a^{\mathrm{cmd}}(t)$. The allocation problem is then formulated as a Dec-POMDP:

$$M = \langle N, S, O, A, P, R, \gamma \rangle \tag{11}$$

where *N* denotes the set of agents; *S* represents the state space; *O* denotes the observation space; *A* represents the action space; *P* is the state transition function; *R* denotes the reward function; and γ represents the discount factor.

1) State space: At time t, the state is:

$$s(t) = \left[\Delta V(t), Q_G(t), Q_G^{\min}(t), Q_G^{\max}(t), \ S_{BG}(t), \ \Delta Q^{\mathrm{cmd}}(t) \right] \tag{12}$$

where $\Delta V(t)$ denotes the voltage deviation vector of key nodes; $Q_G(t)$ represents the reactive power output vector of renewable energy stations; and $Q_G^{\min}(t)$ and $Q_G^{\max}(t)$ denote the lower and upper reactive power output limits of renewable energy stations, respectively.

2) Observation space: During decentralized execution, agent *a* uses the voltages of its corresponding key nodes, the operating states of its station group, and the reactive power requirement from the external CV loop. Its observation is:

$$o_a(t) = \left[\Delta V(t), Q_{G_a}(t), Q_{G_a}^{\min}(t), Q_{G_a}^{\max}(t), S_{BG_a}(t), \ \Delta Q_a^{\mathrm{cmd}}(t) \right] \tag{13}$$

where $Q_{G_a}(t)$, $Q_{G_a}^{\min}(t)$, and $Q_{G_a}^{\max}(t)$ represent the reactive power outputs and the lower and upper reactive power limits of stations within the station group G_a, respectively; $S_{BG_a}(t)$ denotes the voltage-reactive power sensitivity between the station group G_a and the key-node set B.

3) Action space: A CV-MADDPG action specifies how a station group's reactive power requirement is allocated among its stations. The action of agent *a* is defined as:

$$u_a(t) = \left[\alpha_i(t) \right]_{g_i \in G_a}, \alpha_i(t) \geq 0, \sum_{g_i \in G_a} \alpha_i(t) = 1 \tag{14}$$

where $\alpha_i(t)$ denotes the allocation coefficient of station *i* for the station group's reactive power requirement. The resulting reactive power reference for each station is:

$$Q_i^{\mathrm{ref}}(t) = Q_i(t) + \alpha_i(t)\Delta Q_a^{\mathrm{cmd}}(t), \qquad g_i \in G_a \tag{15}$$

CV-MADDPG allocates the reactive power requirement among stations. If a station's resulting reference exceeds its adjustable limit, that reference is clipped and the remaining requirement is redistributed among stations with available reactive power margins. The joint action is:

$$u(t) = \left[u_1(t), u_2(t), \cdots, u_m(t)\right] \tag{16}$$

4) State transition function: Applying the agents' actions changes the station reactive power references and hence the calculated power flow. Key-node voltages and station operating states are then updated. The transition is:

$$s(t+1) \sim P\left(s(t+1) \mid s(t), u(t)\right) \tag{17}$$

During training, the power-flow calculation environment implements this transition.

5) Reward function: CV-MADDPG coordinates station reactive power allocation while pursuing the regional voltage references. Its reward contains voltage tracking, reactive power balancing, and action smoothing terms:

$$\begin{cases} R(t) = -\left[w_v R_v(t) + w_b R_b(t) + w_s R_s(t)\right] \\ R_v(t) = \dfrac{1}{|B|}\displaystyle\sum_{a=1}^{m}\left(V_a(t) - V_a^{\mathrm{ref}}(t)\right)^2 \\ R_b(t) = \dfrac{1}{|N|}\displaystyle\sum_{a=1}^{m}\dfrac{1}{|G_a|}\sum_{g_i \in G_a}\left(\chi_i(t) - \bar{\chi}_a(t)\right)^2 \\ R_s(t) = \dfrac{1}{|G|}\displaystyle\sum_{g_i \in G}\left(\dfrac{Q_i^{\mathrm{ref}}(t) - Q_i^{\mathrm{ref}}(t-1)}{Q_i^{\max}(t) - Q_i^{\min}(t) + \varepsilon}\right)^2 \end{cases} \tag{18}$$

where $R_v(t)$ denotes the voltage control term, $R_b(t)$ represents the reactive power balancing term, and $R_s(t)$ denotes the action smoothing term. w_v, w_b, and w_s are the corresponding weighting coefficients, respectively. $\chi_i(t)$ and $\bar{\chi}_a(t)$ denote the normalized reactive power utilization ratio of station i and the mean ratio within station group a, respectively.

$$\begin{cases} \chi_i(t) = \dfrac{Q_i^{\mathrm{ref}}(t) - Q_i^{\min}(t)}{Q_i^{\max}(t) - Q_i^{\min}(t) + \varepsilon}, \\ \bar{\chi}_a(t) = \dfrac{1}{|G_a|}\displaystyle\sum_{g_i \in G_a}\chi_i(t) \end{cases} \tag{19}$$

CV-MADDPG Solution Process

The Dec-POMDP is addressed with CV-MADDPG using centralized training and decentralized execution (CTDE). During training, each Critic evaluates joint actions using the global state. During execution, the external CV loop determines each station group's reactive power requirement from key-node voltage deviations. Each agent then uses its observation to produce allocation coefficients for its stations. Combining these coefficients with the group requirement yields station-level reactive power references. Figure 2 shows this process.

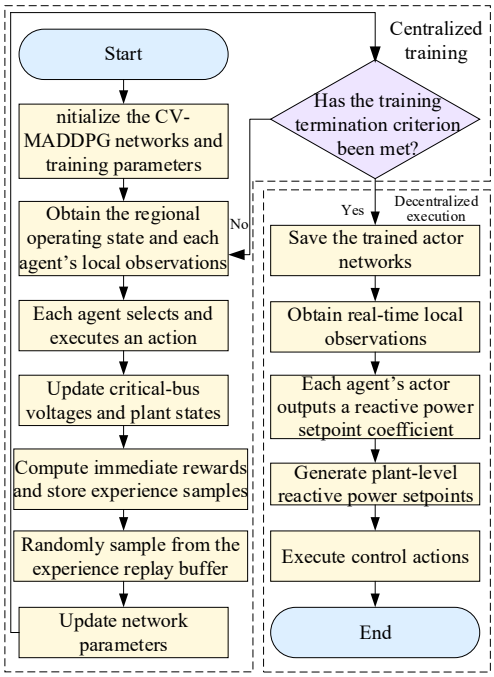


Figure 2. Coordinated control procedure for CV-MADDPG

Let the Actor network and Critic network of agent a be denoted as:

$$u_a(t) = \mu_a(o_a(t); \theta_a), \quad Q_a = Q_a(s(t), u(t); \phi_a) \tag{20}$$

where $\mu_a(\cdot)$ denotes the policy function of agent a, $Q_a(\cdot)$ represents its action-value function, and θ_a and ϕ_a denote the parameters of the Actor network and Critic network, respectively.

Centralized training proceeds as follows.

1) Network initialization. Initialize each agent's Actor, Critic, and corresponding target networks. Configure the replay buffer, discount factor γ, learning rates, soft update coefficients, and exploration noise parameters.

2) Action generation and execution. At time *t*, the external CV loop calculates the reactive power requirement for each station group $\Delta Q_a^{\mathrm{cmd}}(t)$ from key-node voltage deviations. The requirement, voltage deviations, sensitivity information, and station-group operating states form the agent observations. Each agent uses its observation $o_a(t)$ to generate allocation coefficients. During training, exploration noise is added to the Actor output:

$$u_a(t) = \mu_a(o_a(t);\theta_a) + \mathrm{N}_a(t) \tag{21}$$

where $\mathrm{N}_a(t)$ denotes the action exploration noise of agent *a*.

3) Reactive power reference mapping and environment interaction. The allocation coefficients and group requirements from the external CV loop determine station-level reactive power references. These references are applied to the power-flow calculation environment, which updates key-node voltages, station reactive power outputs, and adjustable limits. The updated states are used to calculate the voltage tracking, reactive power balancing, and action smoothing terms.

4) Experience storage. Store each interaction in replay buffer D. A transition sample is:

$$e(t) = \left(s(t), o(t), u(t), R(t), s(t+1), o(t+1)\right) \tag{22}$$

5) Network parameter update. Sample mini-batches from replay buffer D. Each Critic uses the global state and joint actions to estimate the action value and minimizes the temporal-difference error. Each Actor updates its policy using the corresponding Critic's gradient to favor actions with higher estimated long-term returns. The target for agent *a* is:

$$y_a(t) = R(t) + \gamma Q_a^{-}\left(s(t+1), u^{-}(t+1); \phi_a^{-}\right) \tag{23}$$

where Q_a^{-} denotes the target Critic network, ϕ_a^{-} represents the parameters of the target Critic network, and $u^{-}(t+1)$ denotes the next-step joint action generated by the target Actor networks. The loss function of the Critic network is defined as:

$$L(\phi_a) = \frac{1}{|\mathrm{B}|}\sum_{\mathrm{B}}\left[Q_a(s(t), u(t); \phi_a) - y_a(t)\right]^2 \tag{24}$$

where B denotes the mini-batch training sample set.

6) Target network update. Soft updates of the target Actor and Critic networks reduce training oscillations:

$$\theta_a^{-} \leftarrow \tau\theta_a + (1-\tau)\theta_a^{-}, \quad \phi_a^{-} \leftarrow \tau\phi_a + (1-\tau)\phi_a^{-} \tag{25}$$

where τ denotes the soft update coefficient.

7) Training termination. Training ends when a specified episode limit, reward convergence criterion, or voltage tracking criterion is met. Otherwise, interaction and network updates continue from Step 2).

During decentralized execution, only the trained Actor networks are used. Agent *a* selects an action from its current observation $o_a(t)$:

$$u_a(t) = \mu_a(o_a(t);\theta_a) \tag{26}$$

The action mapping converts each agent's output into reactive power references for its stations. At execution, the trained Actors generate these references from their observations without running the Critic networks or a centralized optimization solver.

Power-Flow Validation and LLM Feedback Correction Mechanism

Once the heterogeneous AVC controller generates station-level reactive power references, a power-flow calculation checks the candidate control result. As operating conditions change, a single zoning scheme may cease to meet the operational requirements. Validation indices and detected violations are returned to the LLM to recommend revisions to the zoning, control modes, or parameters.

At time *t*, let the zoning controller generate the station reactive power reference vector $\mathbf{Q}^{\mathrm{ref}}(t)$. Applying these references to the power-flow model yields calculated node voltages, station reactive power outputs, and modeled equipment states:

$$\begin{cases} Q^{\mathrm{ref}}(t) = \left[Q_1^{\mathrm{ref}}(t), Q_2^{\mathrm{ref}}(t), \cdots, Q_n^{\mathrm{ref}}(t)\right]^{\mathrm{T}}, \\ \left[V(t), Q(t), A(t)\right] = F_{\mathrm{pf}}\left(P(t), Q^{\mathrm{ref}}(t), D(t), Y\right) \end{cases} \tag{27}$$

The power-flow results are summarized by a validation index vector:

$$I(t) = \left[I_V(t), I_Q(t), I_B(t), I_A(t)\right]^{\mathrm{T}} \tag{28}$$

where $I_V(t)$, $I_Q(t)$, $I_B(t)$, and $I_A(t)$ represent the voltage validation index, reactive power capability validation index, reactive power balancing index, and equipment operation index, respectively. The definitions of these indices are given as follows:

$$\begin{cases} I_V(t) = \dfrac{1}{|B|}\displaystyle\sum_{b_j \in B}\left[\left(V_j(t) - V_j^{\mathrm{ref}}(t)\right)^2 + \kappa_V \begin{pmatrix} \max(0, V_j(t) - V_j^{\max})^2 \\ +\max(0, V_j^{\min} - V_j(t))^2 \end{pmatrix}\right] \\ I_Q(t) = \displaystyle\sum_{g_i \in G}\left[\max(0, Q_i(t) - Q_i^{\max}(t)) + \max(0, Q_i^{\min}(t) - Q_i(t))\right] \\ \chi_i(t) = \dfrac{Q_i(t) - Q_i^{\min}(t)}{Q_i^{\max}(t) - Q_i^{\min}(t) + \varepsilon}, \quad \bar{\chi}_a(t) = \dfrac{1}{|G_a|}\displaystyle\sum_{g_i \in G_a}\chi_i(t) \\ I_B(t) = \sqrt{\dfrac{1}{|N|}\displaystyle\sum_{a=1}^{m}\dfrac{1}{|G_a|}\sum_{g_i \in G_a}\left(\chi_i(t) - \bar{\chi}_a(t)\right)^2} \\ I_A(t) = \dfrac{1}{|G|}\displaystyle\sum_{g_i \in G}\left|Q_i^{\mathrm{ref}}(t) - Q_i^{\mathrm{ref}}(t-1)\right| \end{cases} \tag{29}$$

where $\chi_i(t)$ denotes the normalized reactive power utilization ratio of station g_i, and $\bar{\chi}_a(t)$ represents the average reactive power utilization ratio within the station group G_a.

An abnormality set is formed from the index thresholds:

$$\mathrm{E}(t)=\left\{k|\ I_k(t)>\bar{I}_k, k\in\{V,Q,B,A\}\right\} \tag{30}$$

where $\bar{I}_k$ denotes the allowable threshold of the k-th index. If $\mathrm{E}(t)=\varnothing$, the result passes the specified power-flow checks. Otherwise, the abnormality set, validation indices, current zoning, control modes, and parameters are returned to the LLM for correction suggestions.

The LLM may recommend changes to zoning, control modes, or their parameters. Parameter revisions may involve the reactive power allocation parameters in CQP regions, voltage control parameters in conventional CV regions, or reward weights and training parameters for CV-MADDPG in complex constant-voltage regions. The revised scheme is evaluated again by power flow. This process stops when the operational requirements are met or the maximum number of correction iterations is reached.

LLM-Assisted Prompt Design

In the proposed heterogeneous AVC framework, the LLM acts as an upper-level decision-support module. Using structured operating features, AVC rules, and power-flow validation results, it recommends station zoning, control modes, initial parameters, and feedback revisions. The LLM does not generate station-level reactive power setpoints; these are calculated by the CQP, conventional CV, and CV-MADDPG control modules, and the resulting control scheme is checked by power-flow calculations.

The initial prompt provides regional voltage conditions, station reactive power capability, electrical coupling information, and operating constraints. It guides the LLM to assign stations to control regions, select the corresponding control modes, and recommend control parameters. For complex constant-voltage regions using CV-MADDPG, the recommendations also include initial reward weights and training parameters. The prompt requires structured outputs with brief rationales and constrains recommendations to respect reactive power capability limits and avoid unjustified frequent changes in control mode.

When power-flow validation detects an abnormality, the feedback prompt supplies the current zoning, control modes, parameters, validation indices, and abnormality types to the LLM. Parameter revisions are preferred when they may resolve the issue; zoning or mode assignments are reconsidered if the abnormality persists and appears related to an unsuitable region or mode. Each suggestion specifies the target, magnitude, and rationale for the change. The revised control scheme is checked again by power flow until the operational requirements are met or the maximum number of correction iterations is reached.

Case Study

Case Study Setup

The proposed LLM-assisted zoning and heterogeneous AVC method is evaluated on a modified IEEE 39-bus system with eight wind farms and eight photovoltaic plants connected to the corresponding key nodes. Figure 3 shows the modified system. The simulation platform uses Python 3.9, pandapower for time-series power-flow calculations, and PyTorch for CV-MADDPG. Simulations are run on an Intel Core i7-13700K CPU with an NVIDIA GeForce RTX 4070 Ti GPU (12 GB) and 64 GB RAM. GPT-5.6 is used for the LLM-assisted decision module.

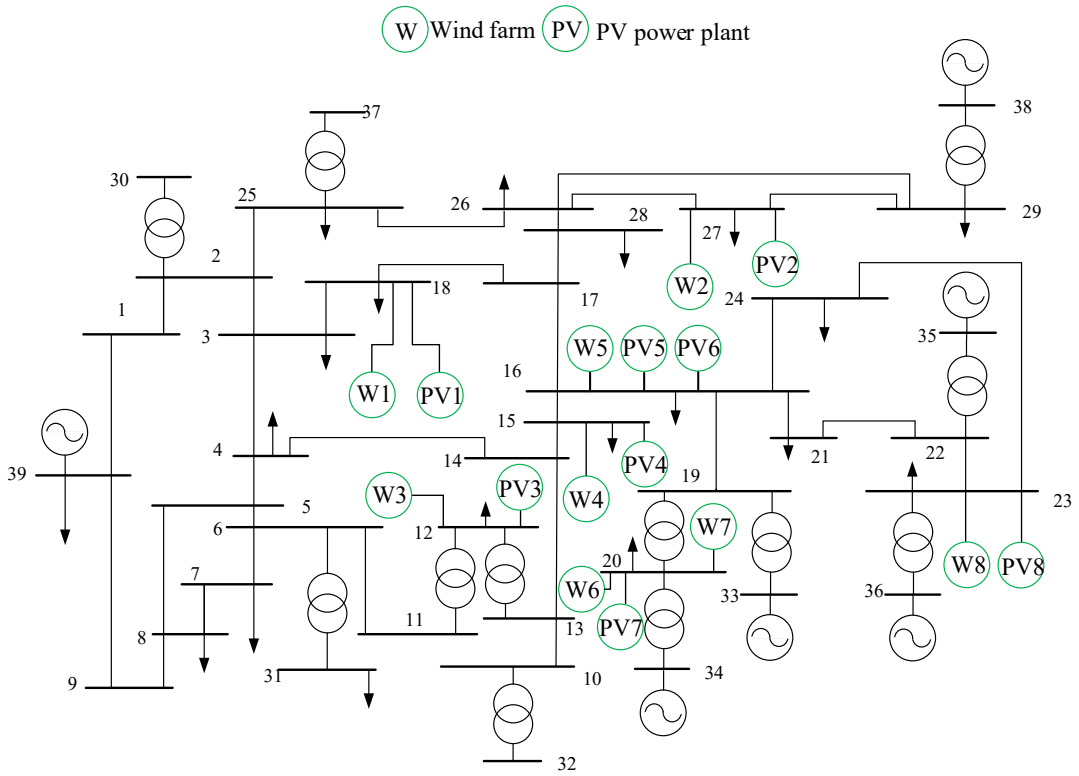


Figure 3. Configuration of the modified IEEE 39-bus system

Load and renewable generation measurements were collected at 5 min intervals over one year from a municipal power grid in Shanxi Province, China. Days are randomly selected using seasonal stratification to form training, validation, and test sets of 200, 30,

and 60 days, respectively. Each CV-MADDPG training episode uses one day of operating data, and final performance is evaluated over the full test set. The key-node voltage reference is 1.0 p.u., the allowable voltage range is 0.95-1.05 p.u., and the deadband of CV control is set to ±0.002 p.u. The regional AVC decision interval is 5 min, and station-level AVC controllers respond to commands within 2 min. Table 1 lists the baseline CV-MADDPG training parameters.

Parameter	Value
Actor learning rate	0.0001
Discount factor	0.99
Mini-batch size	128
Critic learning rate	0.0003
Soft-update coefficient	0.005
Initial exploration noise	0.30
Experience replay buffer capacity	100000

Table 1. Baseline training parameters for CV-MADDPG

Validation of LLM-Assisted Zoning and Heterogeneous AVC Control Performance

The operating features of station groups associated with different key nodes are first examined to assess the LLM-generated zoning and control-mode assignments. These features and the voltage–reactive power sensitivity matrix characterize regional operating differences and electrical coupling. The resulting heterogeneous control scheme is then compared with alternative AVC strategies.

Figure 4 shows the normalized voltage deviation, voltage violation rate, reactive power margin, electrical coupling strength, and response heterogeneity of the station groups associated with each key node. Both bubble size and color intensity indicate the normalized feature values.

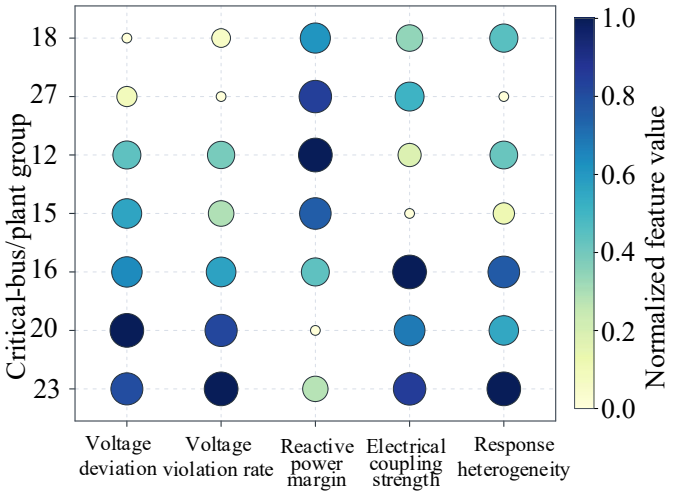


Figure 4. Operating features of station groups associated with key nodes

Figure 4 shows clear differences in voltage conditions, reactive power capability, and coupling characteristics among station groups. The groups at nodes 18 and 27 have relatively small voltage deviations and low violation rates, together with sufficient reactive power margins. These features indicate relatively stable voltages and a limited need for continuous closed-loop voltage regulation. The groups at nodes 12 and 15 require voltage regulation but retain sufficient reactive power margins. Their lower electrical coupling strength and response heterogeneity suggest relatively clear dominant voltage regulation relationships between stations and key nodes.

The groups at nodes 16, 20, and 23 have larger voltage deviations and higher violation rates. Some also have limited reactive power margins, stronger electrical coupling, and greater response heterogeneity. These features suggest strong coupling between individual stations and multiple key nodes. Independent control based only on local voltage deviations may therefore distribute reactive power duties unevenly and create interactions among station controllers.

Figure 5 shows the voltage–reactive power sensitivity matrix used to examine the electrical basis of the station grouping. Key nodes and stations are ordered by the candidate groups generated by the LLM. Darker colors indicate larger absolute sensitivities of node voltage to station reactive power changes.

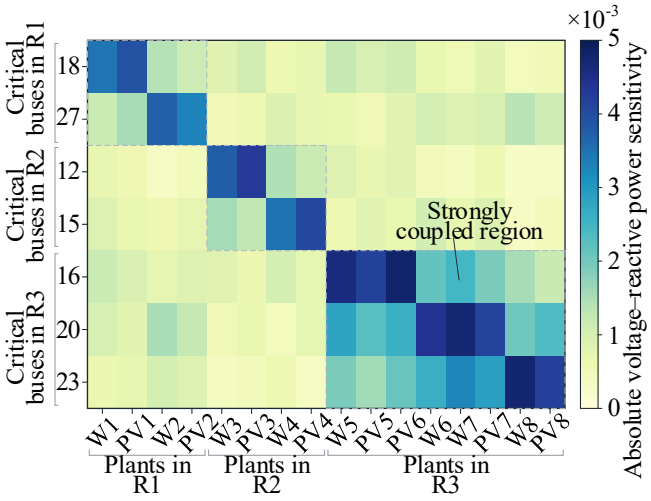


Figure 5. Voltage–reactive power sensitivities between key nodes and renewable energy stations

The sensitivity matrix in Figure 5 has a clear block structure. Node 18 is mainly influenced by W1 and PV1, and node 27 by W2 and PV2. Node 12 has relatively high sensitivities to W3 and PV3, while node 15 is mainly influenced by W4 and PV4. These groups have identifiable dominant station-to-node voltage relationships, consistent with the lower electrical coupling strength and response heterogeneity shown in Figure 4.

The rows for nodes 16, 20, and 23 contain more high-sensitivity entries. W5, PV5, and PV6 strongly influence node 16; W6, W7, and PV7 are closely associated with node 20; and W8 and PV8 mainly affect node 23. Several stations also have high sensitivities to multiple key nodes, indicating strong coupling within these regions. Under independent constant-voltage control, stations with faster responses or higher sensitivities may take on a larger share of reactive power regulation, resulting in uneven allocation among stations [8].

Nonzero sensitivities across regions show that the control regions are not electrically isolated, although the dominant station-to-node relationships remain within each region. Together, Figures 4 and 5 show how the LLM-generated zoning reflects differences in voltage regulation needs, reactive power capability, and electrical coupling. Three regions are formed: R1 comprises the groups at nodes 18 and 27 and uses CQP control; R2 comprises the groups at nodes 12 and 15 and uses conventional CV control; and R3 comprises the groups at nodes 16, 20, and 23 and uses CV-MADDPG to coordinate station reactive power allocation under constant-voltage control objectives.

Four schemes are compared in Table 2: the base operating condition, uniform CQP control, uniform CV control, and the proposed heterogeneous AVC scheme. The base condition has no additional AVC regulation. The two uniform schemes apply CQP or CV control, respectively, to all renewable energy stations. The proposed scheme uses CQP, CV, and CV-MADDPG in R1, R2, and R3, following the LLM-generated zoning. All schemes use the same operating data, equipment parameters, and power-flow constraints.

Control strategy	Mean voltage deviation (p.u.)	Maximum voltage rise (p.u.)	Maximum voltage drop (p.u.)	Voltage violation rate (%)	Reactive power utilization imbalance	Intra-group reverse reactive power adjustment rate (%)	Mean magnitude of reactive power adjustment (MVAr)
Base operating condition	0.0287	0.074	0.092	9.84	—	—	—
Uniform CQP control	0.0225	0.063	0.079	6.12	0.16	0	2.78
Uniform CV control	0.0126	0.052	0.059	2.07	0.34	12.86	5.61
Proposed method	0.0113	0.053	0.054	1.16	0.11	0	3.96

Table 2. Performance comparison of AVC control strategies

Table 2 shows that the base operating condition has a mean voltage deviation of 0.0287 p.u. and a voltage violation rate of 9.84%, indicating inadequate voltage regulation under the evaluated renewable generation fluctuations. The uniform CQP control reduces the average voltage deviation and voltage violation rate to 0.0225 p.u. and 6.12%, respectively, and no reverse regulation among stations within the same group occurs. However, since the reactive power references of stations cannot be dynamically adjusted according to voltage states, its voltage regulation performance remains limited. Uniform CV control reduces the mean voltage deviation and violation rate to 0.0126 p.u. and 2.07%, respectively. Each station responds independently to local voltage deviations, and differences in coupling and response characteristics can produce uneven reactive power allocation. The reactive power utilization imbalance is 0.34, the intra-group reverse reactive power adjustment rate is 12.86%, and the mean magnitude of reactive power adjustment is 5.61 MVAr. Thus, improved voltage regulation is accompanied by less balanced allocation and larger adjustments.

The proposed scheme achieves a mean voltage deviation of 0.0113 p.u. and a voltage violation rate of 1.16%, reductions of 60.63% and 88.21% from the base operating condition. Relative to uniform CV control, the reductions are 10.32% and 43.96%, respectively. The maximum voltage drop decreases from 0.059 p.u. to 0.054 p.u., while the maximum voltage rise remains nearly unchanged. These results indicate less severe undervoltage without a substantial increase in overvoltage in the evaluated cases. Reactive power utilization imbalance decreases to 0.11, a reduction of 67.65% from uniform CV control. The intra-group reverse reactive power adjustment rate decreases to 0, and the mean magnitude of reactive power adjustment decreases from 5.61 MVAr to 3.96 MVAr.

These results are consistent with the division of control duties among regions. R1 and R2 use CQP and CV control suited to their operating characteristics, avoiding unnecessary coordinated allocation where the dominant control relationships are clear. In the strongly coupled R3 region, the external CV loop determines each station group's reactive power requirement, and CV-MADDPG allocates that requirement within the group. This coordination reduces opposing reactive power adjustments and distributes regulation duties more evenly.

Validation of LLM-Assisted CV-MADDPG Optimization Performance

After R3 is identified, the LLM assists in refining the CV-MADDPG reward weights and training parameters. It uses training and validation feedback on voltage regulation, reactive power allocation, and action smoothness to recommend parameter revisions. Its role is limited to parameter refinement; station-level reactive power references are obtained from the trained CV-MADDPG policy outputs.

Four training schemes are compared: the baseline setting, reward-weight optimization, training-parameter optimization, and LLM-based joint optimization. The baseline uses the initial reward weights and training parameters. Reward-weight optimization adjusts only the weights of the voltage tracking, reactive power balancing, and action smoothing terms. Training-parameter optimization adjusts selected CV-MADDPG training parameters while retaining the initial reward weights. LLM-based joint optimization refines both parameter groups. The schemes are trained independently using the same data, network architecture, and remaining training conditions.

Figure 6 illustrates a representative sequence of parameter revisions based on training and validation feedback. Red dashed boxes identify changes to reward weights, and blue dashed boxes identify changes to training parameters.

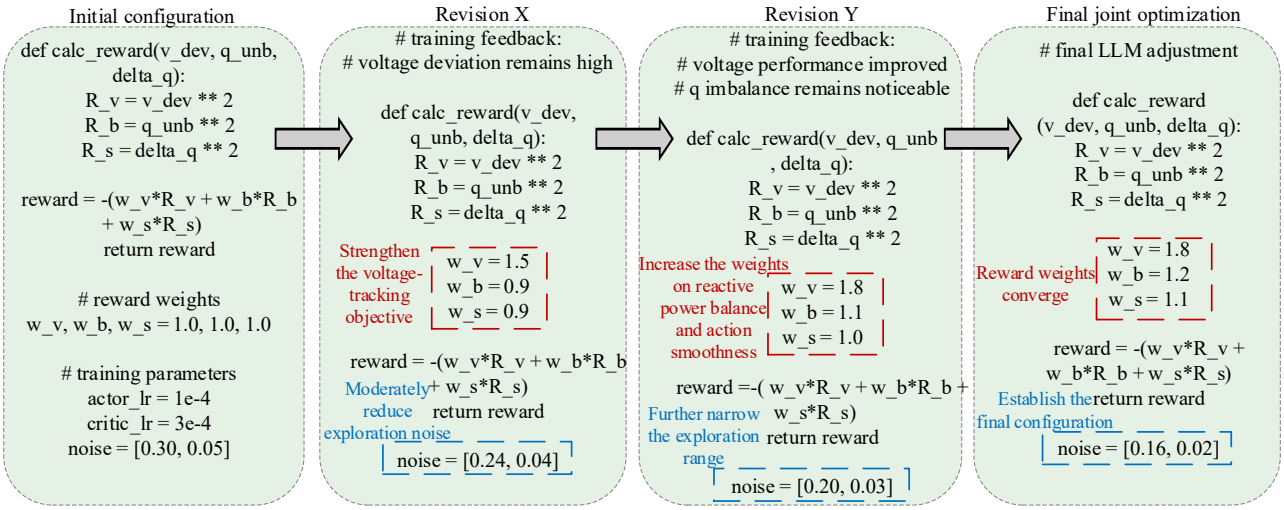
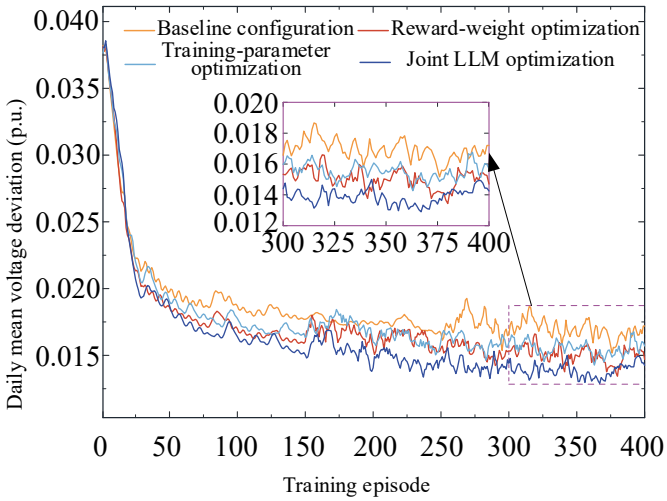


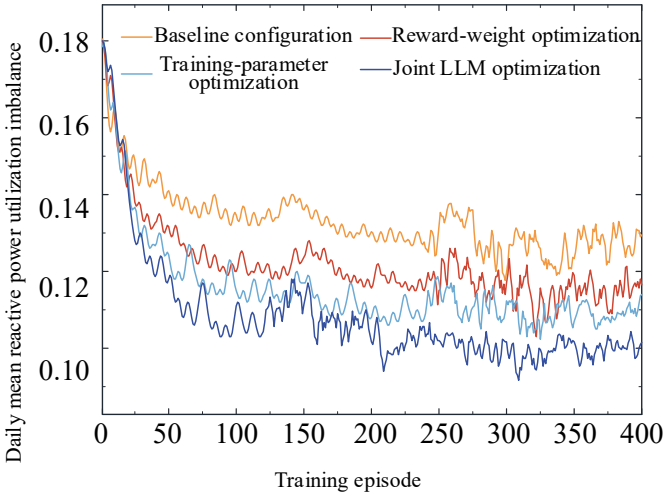
Figure 6. Example of LLM-assisted parameter refinement for CV-MADDPG

In Figure 6, the early revisions place greater emphasis on voltage tracking. Later revisions adjust the balance among voltage tracking, reactive power balancing, and action smoothness, together with the training parameters, to obtain the final joint configuration. The LLM recommends these parameter revisions, while the trained CV-MADDPG policy supplies the outputs used to determine station reactive power references.

In R3, CV-MADDPG replaces independent station responses with coordinated reactive power allocation under a common voltage control objective. This control structure eliminates the mechanism that leads to reverse reactive power regulation among stations within the same group. Training performance is therefore assessed using the daily mean key-node voltage deviation and daily mean station reactive power utilization imbalance, as shown in Figure 7. Each episode covers one training day with 288 intervals of 5 min.



(a) Daily mean voltage deviation at key nodes in R3



(b) Daily mean reactive power utilization imbalance among stations in R3

Figure 7. CV-MADDPG training performance under different parameter settings

In Figure 7, both indicators decrease rapidly early in training and then fluctuate around relatively stable levels. This trend is consistent with CV-MADDPG learning to allocate station reactive power in response to key-node voltage regulation requirements. The remaining fluctuations and occasional curve crossings may reflect differences among training days and changes in available reactive power margins. Both forms of individual parameter refinement improve the indicators relative to the baseline. Reward-weight optimization yields lower voltage deviations later in training, whereas training-parameter optimization yields lower reactive power utilization imbalance.

LLM-based joint optimization adjusts both reward weights and training parameters. Later in training, it yields the lowest voltage deviation, while reactive power utilization imbalance stabilizes at approximately 0.10, below the other three schemes. The joint configuration therefore provides a better balance between voltage tracking and station reactive power allocation in these training runs.

The trained policies are evaluated on the same test set to compare performance on data not used for training. Conventional CV control generates independent station responses from the corresponding key-node voltage deviations. The other schemes use CV-MADDPG for coordinated station reactive power allocation. Table 3 reports the results.

Control scheme	Mean voltage deviation at key nodes in R3 (p.u.)	Voltage violation rate at key nodes in R3 (%)	Reactive power utilization imbalance among stations in R3	Mean change in reactive power setpoints between consecutive intervals (MVAr)
Conventional CV control	0.0151	3.26	0.36	6.18
Baseline configuration	0.0144	2.81	0.145	4.83
Reward-weight optimization	0.0131	2.04	0.132	4.39
Training-parameter optimization	0.0138	2.35	0.122	3.26
Joint LLM optimization	0.0123	1.42	0.105	3.08

Table 3. Test-set performance of conventional CV control and CV-MADDPG with different parameter settings

Relative to conventional CV control, the baseline CV-MADDPG setting reduces mean voltage deviation and the voltage violation rate by 4.64% and 13.80%, respectively (Table 3). Reactive power utilization imbalance and the mean change in reactive power setpoints between consecutive intervals decrease by 59.72% and 21.84%, respectively. These results show that coordinated allocation improves voltage regulation, allocation balance, and setpoint smoothness in the test cases even with the initial reward weights and training parameters.

Reward-weight optimization reduces mean voltage deviation and the voltage violation rate to 0.0131 p.u. and 2.04%, respectively, reductions of 9.03% and 27.40% from the baseline. Training-parameter optimization reduces reactive power utilization imbalance and the mean change in reactive power setpoints between consecutive intervals to 0.122 and 3.26 MVAr, respectively, reductions of 15.86% and 32.51%. Thus, among the two schemes that refine one parameter group, reward-weight optimization produces better voltage regulation, while training-parameter optimization produces more balanced allocation and smoother setpoints.

LLM-based joint optimization performs best on all four indicators in Table 3. Relative to conventional CV control, mean voltage deviation, the voltage violation rate, reactive power utilization imbalance, and the mean change in reactive power setpoints between consecutive intervals decrease by 18.54%, 56.44%, 70.83%, and 50.16%, respectively. Relative to baseline CV-MADDPG, the corresponding reductions are 14.58%, 49.47%, 27.59%, and 36.23%. These results support joint refinement of reward weights and training parameters to improve voltage tracking, allocation balance, and setpoint smoothness in R3 under the evaluated conditions.

The LLM analyzes daily operating features and control performance in the test set to select representative days for comparing conventional CV control with the jointly refined CV-MADDPG scheme. Figure 8 shows the variation in reactive power utilization imbalance among R3 stations on one selected day.

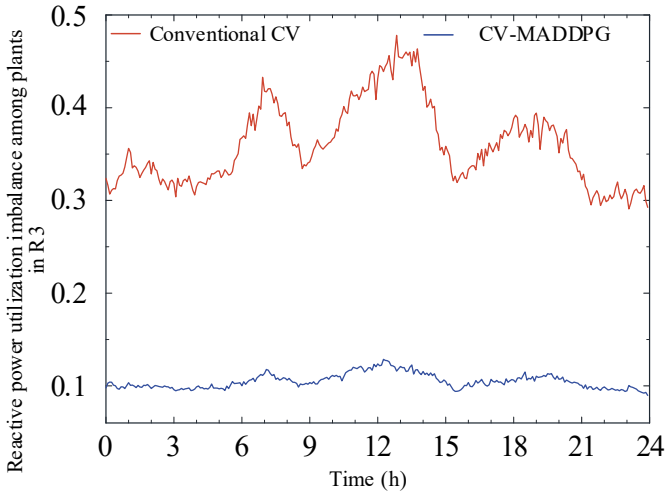


Figure 8. Reactive power utilization imbalance among R3 stations on a representative day

On the day shown in Figure 8, conventional CV control produces relatively high reactive power utilization imbalance, with peaks approaching 0.48. The imbalance varies with operating conditions. Under independent station control, differences in response characteristics, reactive power margins, and voltage–reactive power sensitivities can concentrate regulation duties at particular stations.

With CV-MADDPG, reactive power utilization imbalance remains within 0.09–0.13 for most of the day and fluctuates less. The controller allocates a common reactive power requirement across stations while accounting for their available capability and effects on key-node voltages. This coordinated allocation yields more balanced reactive power utilization as generation and load vary over the illustrated day.

Conclusion

This paper presents an LLM-assisted approach to zoning and heterogeneous AVC for renewable-rich regional power grids. The LLM uses structured regional operating features to recommend station zoning, control modes, and initial parameters. CQP, conventional CV, and CV-MADDPG are assigned according to regional operating characteristics. In complex constant-voltage regions, the external CV loop determines reactive power requirements, and CV-MADDPG coordinates their allocation among stations. Power-flow validation provides feedback for LLM-assisted revisions to control parameters, reward weights, and training parameters. In simulations on the modified IEEE 39-bus system, the method improves key-node voltage regulation and reactive power allocation balance while reducing opposing reactive power adjustments within station groups and changes in reactive power references.

Future work will investigate online zoning updates under changing operating conditions, safety validation for LLM-assisted decisions, and coordination between renewable energy stations and conventional reactive power control devices across multiple timescales.

Data availability

The data that support the findings of this study were obtained from a municipal power grid in Shanxi Province, China. Access to these data is subject to restrictions imposed by the data provider. The data may be made available upon reasonable request and with permission from the data provider.

References

1. National Development and Reform Commission, National Energy Administration. Guiding Opinions on Promoting the Consumption and Regulation of New Energy: NDRC Energy [2025] No. 1360[Z] (2025). (in Chinese).
2. National Development and Reform Commission, National Energy Administration. Guiding Opinions on Promoting High-Quality Development of the Power Grid: NDRC Energy [2025] No. 1710[Z] (2025). (in Chinese).
3. Sun, H., Guo, Q., Qi, J. et al. Review of challenges and research opportunities for voltage control in smart grids. IEEE Trans. Power Syst. 34(4), 2790–2801 (2019).
4. Guo, Q., Sun, H., Wang, B. et al. Hierarchical automatic voltage control for integration of large-scale wind power: Design and implementation. Electr. Power Syst. Res. 120, 234–241 (2015).
5. Sun, H. B., Guo, Q. L., Zhang, B. M. et al. An adaptive zone-division-based automatic voltage control system with applications in China. IEEE Trans. Power Syst. 28(2), 1816–1828 (2013).
6. Li, P., Wang, J. H., Xu, X. et al. Robust reactive power partitioning method for frequent power flow fluctuation in new power system. Autom. Electr. Power Syst. 46(11), 102–110 (2022). (in Chinese).
7. Qu, G. & Yuan, H. Automatic voltage control scheme of zonal-provincial coordination with renewable energy access. In 2022 9th International Forum on Electrical Engineering and Automation, 858–862 (2022).
8. Zhu, X., Liu, W., Ye, S. et al. Wind farm group voltage control strategy considering both regulation speed and reactive power equalization. Trans. China Electrotech. Soc. 40(13), 4216–4228 (2025). (in Chinese).
9. Lu, X., Wu, Z., Li, P. et al. Combined model- and data-driven Volt/Var control for distribution network with photovoltaic clusters. Autom. Electr. Power Syst. 48(9), 97–106 (2024). (in Chinese).
10. Guo, J., Qiao, Y., Lu, Z. et al. Reactive power unbalance/circulation suppression for wind farm group. Power Syst. Technol. 47(1), 109–121 (2023). (in Chinese).
11. Li, S., Zhang, Y., Yu, X. et al. Study on reactive power sharing of wind farms with improved droop control and bi-level reactive power optimization. Power Syst. Prot. Control 47(13), 1–7 (2019). (in Chinese).
12. Duan, J., Shi, D., Diao, R. et al. Deep-reinforcement-learning-based autonomous voltage control for power grid operations. IEEE Trans. Power Syst. 35(1), 814–817 (2020).

13. Cao, D., Zhao, J., Hu, W. et al. Deep reinforcement learning enabled physical-model-free two-timescale voltage control method for active distribution systems. IEEE Trans. Smart Grid 13(1), 149–165 (2022).
14. Wang, J., Xu, W., Gu, Y. et al. Multi-agent reinforcement learning for active voltage control on power distribution networks. In Advances in Neural Information Processing Systems 34, 3271–3284 (2021).
15. Cao, D., Hu, W., Zhao, J. et al. A multi-agent deep reinforcement learning based voltage regulation using coordinated PV inverters. IEEE Trans. Power Syst. 35(5), 4120–4123 (2020).
16. Cheng, H., Luo, H., Liu, Z. et al. Reinforcement learning based robust Volt/Var control in active distribution networks with imprecisely known delay. IEEE Trans. Comput. Soc. Syst. (2025). https://doi.org/10.1109/TCSS.2025.3564500.
17. Liu, X., Zhang, C., Li, S. & Zhu, J. A tuning friendly deep reinforcement learning method for inverter-based Volt/Var control. IEEE Trans. Smart Grid 16(6), 5669–5672 (2025).
18. Yan, Z. & Xu, Y. Real-time optimal power flow with linguistic stipulations: Integrating GPT-agent and deep reinforcement learning. IEEE Trans. Power Syst. 39(2), 4747–4750 (2024). https://doi.org/10.1109/TPWRS.2023.3338961.
19. Gao, M., Zhou, S., Gu, W. et al. Adaptation approach for pre-trained LLMs towards power system tasks: Representation learning and supervised fine-tuning with multi-source heterogeneous data. Proc. CSEE 46(10), 3967–3980 (2026). https://doi.org/10.13334/j.0258-8013.pcsee.252646. (in Chinese).
20. Yan, L. & Cheng, C. Voltage control for distribution networks based on large language model-assisted deep reinforcement learning. IEEE Access 13, 76072–76084 (2025). https://doi.org/10.1109/ACCESS.2025.3565280.
21. Yu, Z., Li, J., Wang, T. et al. Volt-var control strategy for distribution networks based on large language model-assisted multi-agent deep reinforcement learning. Proc. CSEE (2026). https://doi.org/10.13334/j.0258-8013.pcsee.260314. (in Chinese).

Author contributions

J.Z. contributed to conceptualization, methodology, supervision, and manuscript review and editing. X.N. contributed to methodology, software development, simulation experiments, and formal analysis, and drafted the manuscript. S.Z. contributed to data curation, validation, and manuscript review and editing. All authors reviewed and approved the final manuscript.

Funding

This work was supported by the Science and Technology Project of State Grid Corporation of China, entitled “Research on Coordinated Control Technologies for Multiple Reactive Power Sources in Regional Power Grids with High Renewable Energy Penetration” (Project No. 52053025001L).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.Z.