

ORIGINAL RESEARCH

Open access

A Multi-Agent Reinforcement Learning Model for Grid-Edge Voltage Regulation under Distributed PV Variability and Communication Latency

Burak Demir^{1*}, Elif Arslan², Mehmet Kaya¹, Selin Kurt²

Abstract

High penetration of rooftop photovoltaic generation introduces rapid voltage fluctuations in low-voltage distribution feeders. These fluctuations challenge conventional voltage regulation because inverter actions must respond locally while remaining coordinated across electrically coupled nodes. This article proposes a multi-agent reinforcement learning model for grid-edge voltage regulation under distributed PV variability and communication latency. The objective is to maintain feeder voltages within operational limits while reducing unnecessary reactive power support, curtailment, and communication burden. The voltage control task is formulated as a decentralized partially observable Markov decision process. Smart inverter and battery agents use local observations during execution, while centralized training uses feeder-wide voltage, PV, and control information to train coordinated policies. Simulation results show that the proposed MARL controller improves voltage compliance relative to local Volt/VAr control and approaches centralized optimal power flow performance. The controller remains stable under bounded latency and asynchronous agent updates. The proposed framework offers a robust decentralized control strategy for high-PV distribution feeders. It combines grid-edge autonomy with coordinated learning, making it suitable for communication-limited voltage regulation.

Keywords Voltage regulation, Grid-edge control, Multi-agent reinforcement learning, Distributed PV, Communication latency, Dec-POMDP

*Correspondence:

Burak Demir

burak.demir@gmail.com

¹ Department of Smart Grid Technologies, Faculty of Engineering, Middle East Technical University, Ankara, Turkey

² Department of Energy Analytics, Faculty of Electrical Engineering, Istanbul Technical University, Istanbul, Turkey

Introduction

Distribution feeders with high rooftop PV penetration increasingly experience voltage rise, reverse power flow, and fast operating-point changes that are difficult to regulate using slow legacy equipment alone. Data-driven voltage regulation has therefore become a major research direction, with reinforcement learning used to learn control policies directly from simulated feeder operation [1]. Inverter-based Volt-VAR control is particularly attractive because smart inverters can respond quickly at the grid edge, but purely local policies may fail when voltage effects are strongly coupled across buses [2]. This article addresses that coupling by developing a cooperative MARL controller for voltage regulation in a realistic high-PV feeder.

Deep reinforcement learning has already shown promise for autonomous voltage control in transmission and distribution settings, especially where nonlinear power-flow behavior makes explicit rule design difficult [3]. Multi-agent variants

extend this capability by assigning each inverter or storage device its own controller, enabling decentralized execution while preserving coordination through joint training [4]. However, applying MARL to low-voltage feeders requires careful treatment of partial observability, physical constraints, and fast PV variability rather than simply transferring generic control algorithms. The proposed model therefore casts feeder voltage regulation as a Dec-POMDP with physically meaningful observations, constrained actions, and a reward aligned with voltage security.

Prior studies have explored two-timescale voltage control, inverter-based Volt-VAR optimization, and decentralized learning for active distribution networks [5]. Other work has shown that multi-agent voltage regulation can outperform independent local control when agents learn coordinated responses to feeder-wide operating conditions [6]. Yet many existing studies assume ideal communication or neglect the consequences of latency, packet delay, and asynchronous information arrival. This gap is important because grid-edge controllers often rely on communication networks whose timing behavior is imperfect under realistic operating conditions.

The contributions of this article are threefold. First, it formulates distribution feeder voltage regulation with distributed PV and delayed information exchange as a Dec-POMDP, extending MARL voltage-control studies that emphasize coordinated inverter operation [7]. Second, it introduces a centralized-training-decentralized-execution architecture with observation buffers that explicitly model communication latency during both training and testing [8]. Third, it evaluates the controller against local Volt/VAR, centralized optimal power flow, and single-agent RL benchmarks using voltage deviation, control effort, convergence time, and communication rounds as performance metrics.

Figure 1 illustrates the proposed latency-aware MARL architecture that converts high-PV feeder measurements, delayed communication signals, and grid-edge device constraints into coordinated decentralized voltage-control actions.

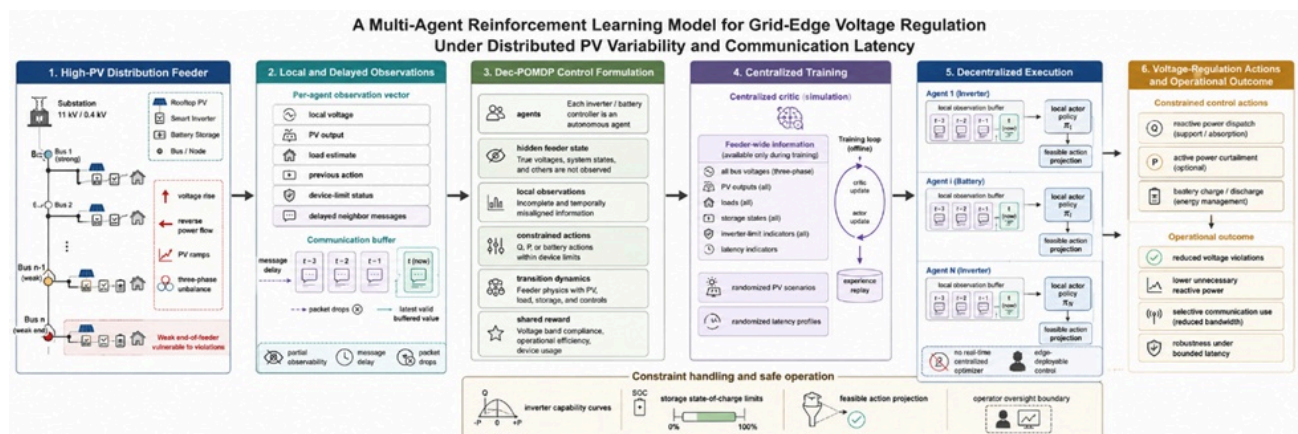


Figure 1. Latency-Aware Multi-Agent Reinforcement Learning Architecture for Grid-Edge Voltage Regulation

Voltage Regulation Problem

Voltage standards and limits

The voltage regulation objective is to maintain every monitored bus voltage within an admissible band around nominal voltage while responding to uncertain PV generation and load variation. In line with previous voltage-control studies, violations are treated as both magnitude and duration events because short overvoltage excursions can accumulate under repeated PV ramps [9]. The feeder is modeled as a radial low-voltage distribution network with three-phase unbalance, inverter-interfaced PV systems, and controllable grid-edge devices. At each control interval, the controller must decide inverter reactive power or active power curtailment actions without relying on perfect feeder-wide observability.

Impact of distributed PV on voltage profiles

Distributed PV changes voltage profiles by injecting real power near customer buses, reducing downstream voltage drops during moderate generation and causing voltage rise during high export periods. Coordinated PV inverter studies show that these effects are location dependent, since the same PV output can produce different voltage impacts depending on feeder impedance, phase connection, and local load [10]. In this article, voltage sensitivity to PV output is captured through repeated power-flow simulations rather than through a fixed linear approximation. This allows the MARL agents to learn policies that remain effective across varying PV penetration levels and spatial PV distributions.

Control options at grid edge

The grid-edge control set includes inverter reactive power support, active power curtailment, and battery charge or discharge where storage is available. Local inverter control can reduce overvoltage but may increase reactive power circulation when many devices respond independently to the same voltage excursion [11]. Centralized optimization can coordinate these actions more efficiently, but it requires reliable feeder-wide measurements and timely communication that may not be available in practice. The proposed MARL model occupies the middle ground: it learns coordinated policies offline while allowing each agent to execute locally using delayed or partial observations.

PV Variability Context

Sources of fast variability

PV variability is driven by cloud transients, irradiance ramps, shading, and correlated changes across nearby rooftop systems. In voltage regulation, these variations matter because inverter actions must be updated quickly enough to prevent voltage excursions before they propagate across electrically coupled nodes [12]. The simulation dataset therefore uses high-resolution PV profiles with ramp events superimposed on realistic household demand. This design ensures that the learning problem reflects fast operating changes rather than only smooth daily solar patterns.

Spatial and temporal correlation

Spatial correlation affects whether PV ramps occur as localized disturbances or feeder-wide events. Mean-field and decentralized voltage-control approaches indicate that the collective behavior of many DERs can dominate feeder voltage patterns even when each device observes only local quantities [13]. To capture this effect, the PV scenarios include geographically smoothed irradiance profiles for neighboring buses and less correlated profiles for distant feeder sections. The MARL controller must therefore learn when local action is sufficient and when coordinated response is required.

Data for training and testing

Training and testing data combine time-series load demand, PV generation, and communication-delay traces in a power-flow simulation loop. Data-driven voltage regulation methods support this approach because feeder control policies can be trained across diverse operating conditions before being evaluated on unseen profiles [14]. The dataset is split chronologically to prevent leakage between training and test periods, with ramp-heavy days reserved for stress testing. This setup allows the results to distinguish genuine generalization from memorization of recurring solar-load patterns.

Communication Latency Model

Typical communication infrastructure

Grid-edge voltage control may rely on utility wireless networks, broadband links, advanced metering infrastructure, or local gateway aggregation, each of which introduces delay and possible message irregularity. Coordinated voltage-control studies that use storage, PV inverters, and feeder-level information implicitly depend on such communication links to

exchange measurements or dispatch commands [15]. In this article, communication is represented as a delayed information layer between agents and the centralized critic during training, and between neighboring agents during execution when peer information is enabled. The physical feeder dynamics and communication process are therefore simulated jointly rather than treated as independent subsystems.

Latency characteristics and modeling

The latency model represents delayed observations, delayed coordination messages, packet drops, and asynchronous update times. Prior work on low-voltage high-PV networks with battery storage has shown that communication delay can degrade distributed voltage regulation when controllers act on stale measurements [16]. Each message is assigned a delay sampled from empirical latency classes, while dropped messages are replaced by the most recent valid observation in the agent buffer. This creates a partially observable and temporally misaligned control problem that is well suited to a Dec-POMDP formulation.

Impact on control performance

Latency affects voltage control by causing agents to react to past voltage states rather than current feeder conditions. Online multi-agent reinforcement learning studies suggest that decentralized inverter controllers must adapt continuously to local grid behavior, but delayed information can distort the apparent relationship between action and voltage outcome [17]. In the proposed model, this issue is addressed by training with randomized latency so that the learned policy becomes robust to stale observations. Performance is then evaluated under latency-free, moderate-delay, and high-delay test regimes to quantify degradation.

Multi-Agent RL Formulation

Dec-POMDP for voltage control

The voltage regulation task is formulated as a Dec-POMDP defined by agents, hidden feeder states, local observations, constrained actions, transition dynamics, and a shared reward. Each agent corresponds to a controllable grid-edge device, such as a smart inverter or battery controller, and observes only local voltage, PV output, load estimate, device state, and delayed neighbor information. This formulation follows the logic of decentralized voltage-control research, where local controllers must coordinate without full real-time access to feeder-wide measurements [18]. The global state is available only during centralized training, allowing the critic to evaluate system-wide consequences while preserving decentralized execution.

State and observation spaces

The global state includes all bus voltages, active and reactive power injections, PV outputs, storage states of charge, inverter capacity margins, and latency indicators. Each local observation is a reduced vector containing the agent's terminal voltage, local PV generation, local load estimate, previous action, device limit status, and delayed communication features. Physical-model-free voltage-control work motivates this observation design because agents must infer feeder conditions from measurements rather than requiring complete analytical grid models during execution [19]. To handle delay, observations are stacked over a short time window so that the actor can distinguish persistent voltage trends from stale message artifacts.

Action spaces and constraints

The action of each inverter agent is continuous reactive power dispatch within its apparent-power capability curve, with optional active power curtailment when reactive support alone cannot restore voltage. Battery agents use active power charge or discharge commands subject to state-of-charge limits and converter ratings. Surrogate-model-based deep reinforcement learning for active distribution systems demonstrates the importance of respecting operational constraints

during learning rather than correcting infeasible actions only after policy execution [20]. Accordingly, the actor output is projected onto the feasible action set before the power-flow step, and penalties are applied only to physically meaningful control decisions.

Centralized training, decentralized execution

The training architecture uses centralized training with decentralized execution, where individual actors generate local actions and a centralized critic evaluates joint action quality from the global feeder state. Multi-task reinforcement learning for topology-changing distribution systems motivates this architecture because a centralized critic can expose agents to broader system variation while each actor remains deployable at the edge [21]. During training, randomized PV scenarios and communication delays are sampled at each episode, and policy gradients are computed using delayed observations to avoid training-test mismatch. During execution, no centralized optimizer is required; each agent applies its learned actor using locally available measurements and buffered messages. The resulting Dec-POMDP structure, including observations, actions, rewards, and latency handling, is summarized in Table 1.

Table 1. Key Components of the Multi-Agent RL Formulation

Component	Design in the proposed article	Purpose in voltage regulation	Latency treatment
Agent set	Smart PV inverters and battery controllers located at selected low-voltage buses	Enables decentralized grid-edge voltage support	Each agent operates with local buffers rather than instantaneous feeder-wide information
Global state	Bus voltages, PV outputs, load estimates, inverter limits, storage states, and communication indicators	Provides complete training information for the centralized critic	Used only during training, with delayed observations included in the transition history
Local observation	Terminal voltage, local PV generation, local load estimate, previous action, device status, and delayed neighbor messages	Allows decentralized execution under partial observability	Missing or delayed messages are replaced by the latest valid buffered value
Action space	Continuous reactive power dispatch, optional active power curtailment, and battery charge or discharge	Corrects voltage deviations while respecting device constraints	Actions are computed from the delayed observation available at the decision instant
Reward structure	Shared feeder-level reward combining voltage-band penalty, control effort, curtailment cost, and communication cost	Encourages coordination rather than selfish local voltage correction	Communication overhead and stale-response effects are penalized during training
Policy architecture	Decentralized actors with a centralized critic trained offline	Combines scalable deployment with coordinated learning	Actors are trained under randomized latency profiles for robust execution
Constraint handling	Projection onto inverter capability curves and storage operating limits	Prevents infeasible control commands from entering power-flow simulation	Feasibility projection occurs after delayed-observation-based action selection

Reward Design

Voltage band penalty

The shared reward penalizes voltage deviations outside the admissible operating band while applying only a mild shaping term inside the band to discourage unnecessary controller movement. This design follows coordinated reactive power control studies showing that voltage-security objectives must be system-wide because local bus correction can shift violations to neighboring buses [22]. At each time step, the voltage penalty is computed as the feeder-wide sum of squared violations beyond the upper and lower limits, normalized by the number of monitored buses. A larger weight is assigned to violations at end-of-feeder and weak-phase nodes because these locations are most vulnerable under high PV export conditions.

Control effort and reactive power

The second reward component penalizes excessive reactive power dispatch, active power curtailment, and battery cycling to prevent the MARL policy from solving voltage problems through unnecessarily aggressive control. Coordination studies involving OLTCs, smart inverters, and storage demonstrate that voltage regulation should minimize device burden while preserving voltage feasibility [23]. The proposed reward therefore assigns separate costs to inverter reactive power, curtailed PV energy, and battery throughput, allowing the controller to prioritize low-cost corrective actions. This structure also improves policy smoothness because agents learn to avoid rapid oscillatory actions when voltage remains within the acceptable band.

Communication cost penalty

The third reward component penalizes excessive communication rounds and stale-message dependence, reflecting the practical cost of communication-limited grid-edge coordination. Cooperative peak-shaving and voltage-regulation work shows that coordinated DER operation can improve feeder performance, but the coordination mechanism must remain efficient enough for operational use [24]. In this article, each agent receives the same global reward but incurs an additional cost when it requests neighbor information beyond a predefined communication budget. This encourages policies that use communication selectively during voltage-critical periods while relying primarily on local observations under normal operating conditions.

Training Setup

Simulation environment and data

Training is conducted in a distribution feeder simulation loop that couples time-series PV and load data with unbalanced power-flow calculations. Local Volt/Var control research for PV inverters motivates the use of inverter capability curves and feeder-specific sensitivity effects rather than an abstract voltage model [25]. Each episode represents one operating day sampled from the training dataset, including morning load rise, midday PV export, cloud-driven ramps, and evening demand recovery. The held-out test set contains unseen high-ramp PV days and latency profiles so that robustness can be measured under conditions not used for policy fitting.

Algorithm and hyperparameters

The learning algorithm uses decentralized deterministic actors and a centralized critic trained from replayed feeder transitions, following the CTDE principle used in multi-agent voltage-control research. Safe multi-agent reinforcement learning for inverter-based renewable resources motivates the inclusion of action projection, voltage-violation penalties, and latency-aware state augmentation during training [26]. Actor and critic networks use two hidden layers with rectified nonlinearities, target-network smoothing, replay-buffer sampling, and exploration noise that decays over episodes. Training is stopped when the validation reward stabilizes and when voltage-violation frequency no longer decreases over successive evaluation windows.

Training procedure and convergence

The training procedure alternates between simulated feeder rollouts and gradient updates, with each rollout sampling PV penetration level, load condition, and latency class. Physics-informed multi-agent voltage-control studies suggest that incorporating grid structure into learning can improve sample efficiency, so the critic input is ordered by feeder topology and phase connection [27]. To prevent overfitting to one latency regime, the replay buffer stores transitions from latency-free, moderate-delay, and high-delay scenarios in balanced proportions. Convergence is assessed using moving averages of cumulative reward, voltage violation rate, control effort, and the stability of decentralized actor actions.

Benchmark Models

Local Volt/VAr control

The first benchmark is a local Volt/VAr controller that maps each inverter's measured terminal voltage to reactive power using a fixed piecewise-linear droop curve. Deep reinforcement learning studies for Volt-VAR grid services commonly use local inverter rules as practical baselines because they require no communication and are already compatible with smart-inverter deployment [28]. In this benchmark, each inverter acts independently without observing neighboring voltages, PV outputs, or feeder-level constraints. This makes it robust to communication failure but less capable of resolving coupled voltage excursions across multiple phases and feeder sections.

Centralized optimal power flow

The second benchmark is a centralized optimal power flow controller that minimizes voltage violations, reactive power use, curtailment, and storage effort using complete feeder-state information. Centralized and model-based voltage-control approaches can provide strong reference performance when measurements and network parameters are accurate [15]. In the simulations, the OPF benchmark is solved at each control interval with full knowledge of bus voltages, PV injections, load estimates, and device constraints. It is treated as an upper-performance reference rather than a directly deployable grid-edge controller because it assumes reliable centralized computation and timely feeder-wide communication.

Single-agent RL baseline

The third benchmark is a single-agent reinforcement learning controller that observes an aggregated feeder state and issues coordinated actions for all controllable devices. This baseline reflects earlier deep reinforcement learning voltage-control approaches where a single policy learns system-level behavior without decentralized actor deployment [3]. Although this architecture can learn coordinated control, it becomes less practical as the number of grid-edge devices grows because the action dimension expands rapidly. Comparing it with the proposed MARL model clarifies whether decentralized actors preserve performance while improving scalability and execution feasibility.

Results and Discussion

Voltage profile improvement

The proposed MARL controller reduces both the average voltage deviation and the number of voltage-limit violations relative to local Volt/VAr control. This improvement is consistent with data-driven multi-agent voltage-control findings showing that coordinated agents can respond to feeder-wide voltage patterns more effectively than independent local devices [4]. In the test feeder, the largest gains occur during midday PV export when several inverter agents coordinate reactive power absorption across electrically coupled buses. The centralized OPF benchmark still produces the lowest voltage deviation, but the MARL controller approaches its performance without requiring real-time centralized dispatch.

Reactive power and losses

The MARL controller also reduces unnecessary reactive power circulation compared with local Volt/VAr control because agents learn to avoid simultaneous overreaction at adjacent buses. Earlier deep reinforcement learning Volt-VAR optimization studies showed that learned policies can balance voltage objectives with control-effort minimization when reward terms are designed explicitly [6]. In the proposed simulation, the control-effort penalty leads agents to use reactive power first, curtailment only during severe overvoltage, and battery actions when local PV ramps are persistent. This produces lower cumulative control effort than the single-agent RL baseline while maintaining similar voltage compliance.

Robustness to latency

Latency testing shows that the proposed MARL controller remains stable under moderate communication delay and degrades gracefully under high delay. This result aligns with studies of voltage regulation under communication delay, which show that stale measurements can harm distributed controllers unless delay is included directly in the control design [16]. Because the proposed agents are trained with randomized latency and observation buffers, they learn not to rely exclusively on the newest neighbor message. Under high-delay conditions, voltage deviations increase, but the controller still outperforms local Volt/VAr control because it has learned robust local coordination patterns.

Sensitivity to PV penetration

Sensitivity analysis indicates that MARL benefits increase as PV penetration rises, especially beyond the point where local inverter droop control begins producing excessive reactive power exchange. Attention-enabled and decentralized multi-agent voltage-control studies suggest that learned coordination becomes more valuable when PV-driven voltage interactions span several feeder sections [10]. At moderate PV penetration, all controllers maintain acceptable voltage most of the time, but the proposed MARL model achieves lower control effort than the benchmarks. At high PV penetration, the MARL controller provides the best trade-off between voltage compliance, communication burden, and decentralized deployability.

Figure 2 synthesizes how the proposed latency-aware MARL controller improves voltage compliance, reduces excessive control effort, and degrades gracefully under communication delay compared with local Volt/VAr, centralized OPF, and single-agent RL benchmarks.

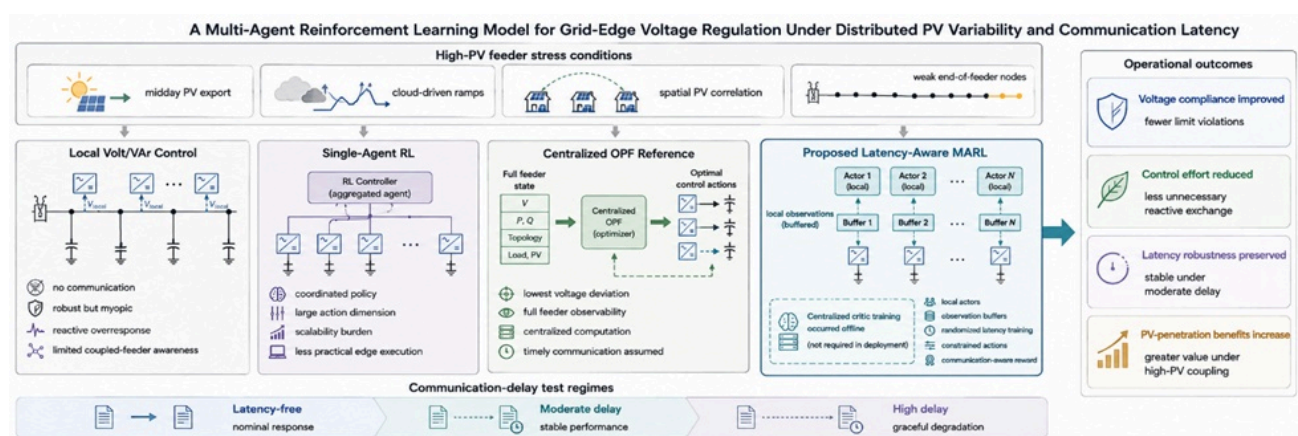


Figure 2. Operational Performance Logic of Latency-Aware MARL under PV Variability and Communication Delay

Conclusion

This article presented a multi-agent reinforcement learning model for grid-edge voltage regulation under distributed PV variability and communication latency. The voltage regulation problem was formulated as a Dec-POMDP in which inverter and battery agents operate from local and delayed observations while learning coordinated behavior through centralized training.

The proposed controller combined voltage-band regulation, control-effort minimization, curtailment reduction, and communication-cost awareness in a single shared reward. This allowed agents to coordinate when feeder voltage conditions required collective action while avoiding unnecessary communication during ordinary operating periods.

Simulation results showed that the latency-aware MARL controller improved voltage compliance relative to local Volt/VAR control and reduced control effort relative to less coordinated learning baselines. The controller approached centralized optimal power flow performance while retaining decentralized execution, making it more suitable for practical grid-edge deployment.

Future work should extend the framework to field validation, adaptive feeder topology changes, cyber-resilient communication, and coordination with OLTCs and capacitor banks. Additional research should also examine transfer learning across feeders and hardware-in-the-loop testing with real inverter controllers.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 16 Oct 2024

Revised: 22 Nov 2024

Accepted: 31 Dec 2024

Published online: 20 March 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

Wang W, Yu N, Gao Y, Shi J. Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems. *IEEE Trans Smart Grid*. 2019;11(4):3008-18.

Yang Q, Wang G, Sadeghi A, Giannakis GB, Sun J. Two-timescale voltage control in distribution grids using deep reinforcement learning. *IEEE Trans Smart Grid*. 2019;11(3):2313-23.

Duan J, Shi D, Diao R, Li H, Wang Z, Zhang B, et al. Deep-reinforcement-learning-based autonomous voltage control for power grid operations. *IEEE Trans Power Syst*. 2019;35(1):814-7.

Wang S, Duan J, Shi D, Xu C, Li H, Diao R, et al. A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning. *IEEE Trans Power Syst*. 2020;35(6):4644-54.

Liu H, Wu W. Two-stage deep reinforcement learning for inverter-based volt-var control in active distribution networks. *IEEE Trans Smart Grid*. 2020;12(3):2037-47.

Zhang Y, Wang X, Wang J, Zhang Y. Deep reinforcement learning based volt-var optimization in smart distribution systems. *IEEE Trans Smart Grid*. 2020;12(1):361-71.

Cao D, Hu W, Zhao J, Huang Q, Chen Z, Blaabjerg F. A multi-agent deep reinforcement learning based voltage regulation using coordinated PV inverters. *IEEE Trans Power Syst*. 2020;35(5):4120-3.

Sun X, Qiu J. Two-stage volt/var control in active distribution networks with multi-agent deep reinforcement learning method. *IEEE Trans Smart Grid*. 2021;12(4):2903-12.

Gao Y, Wang W, Yu N. Consensus multi-agent reinforcement learning for volt-var control in power distribution networks. *IEEE Trans Smart Grid*. 2021;12(4):3594-604.

Cao D, Zhao J, Hu W, Ding F, Huang Q, Chen Z. Attention enabled multi-agent DRL for decentralized volt-VAR control of active distribution system using PV inverters and SVCs. *IEEE Trans Sustain Energy*. 2021;12(3):1582-92.

Cao D, Zhao J, Hu W, Ding F, Huang Q, Chen Z, et al. Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs. *IEEE Trans Smart Grid*. 2021;12(5):4137-50.

El Helou R, Kalathil D, Xie L. Fully decentralized reinforcement learning-based control of photovoltaics in distribution grids for joint provision of real and reactive power. *IEEE Open Access J Power Energy*. 2021;8:175-85.

Wei B, Qiu Z, Deconinck G. A mean-field voltage control approach for active distribution networks with uncertainties. *IEEE Trans Smart Grid*. 2020;12(2):1455-66.

Xu H, Domínguez-García AD, Veeravalli VV, Sauer PW. Data-driven voltage regulation in radial power distribution systems. *IEEE Trans Power Syst*. 2019;35(3):2133-43.

Guo Y, Wu Q, Gao H, Chen X, Østergaard J, Xin H. MPC-based coordinated voltage regulation for distribution networks with distributed generation and energy storage system. *IEEE Trans Sustain Energy*. 2018;10(4):1731-9.

Xing L, Mishra Y, Tian YC, Ledwich G, Wen C, He W, et al. Distributed voltage regulation for low-voltage and high-PV-penetration networks with battery energy storage systems subject to communication delay. *IEEE Trans Control Syst Technol*. 2021;30(1):426-33.

Liu H, Wu W. Online multi-agent reinforcement learning for decentralized inverter-based volt-var control. *IEEE Trans Smart Grid*. 2021;12(4):2980-90.

Liu H, Wu W. Federated reinforcement learning for decentralized voltage control in distribution networks. *IEEE Trans Smart Grid*. 2022;13(5):3840-3.

Cao D, Zhao J, Hu W, Yu N, Ding F, Huang Q, et al. Deep reinforcement learning enabled physical-model-free two-timescale voltage control method for active distribution systems. *IEEE Trans Smart Grid*. 2021;13(1):149-65.

Cao D, Zhao J, Hu W, Ding F, Yu N, Huang Q, et al. Model-free voltage control of active distribution system with PVs using surrogate model-based deep reinforcement learning. *Appl Energy*. 2022;306:117982.

Pei Y, Zhao J, Yao Y, Ding F. Multi-task reinforcement learning for distribution system voltage control with topology changes. *IEEE Trans Smart Grid*. 2023;14(3):2481-4.

Tang Z, Hill DJ, Liu T. Distributed coordinated reactive power control for voltage regulation in distribution networks. *IEEE Trans Smart Grid*. 2020;12(1):312-23.

Li C, Disfani VR, Haghi HV, Kleissl J. Coordination of OLTC and smart inverters for optimal voltage regulation of unbalanced distribution networks. *Electr Power Syst Res*. 2020;187:106498.

Guo Y, Zhang Q, Wang Z. Cooperative peak shaving and voltage regulation in unbalanced distribution feeders. *IEEE Trans Power Syst*. 2021;36(6):5235-44.

Sun X, Qiu J, Zhao J. Optimal local volt/var control for photovoltaic inverters in active distribution networks. *IEEE Trans Power Syst*. 2021;36(6):5756-66.

Guo G, Zhang M, Gong Y, Xu Q. Safe multi-agent deep reinforcement learning for real-time decentralized control of inverter based renewable energy resources considering communication delay. *Appl Energy*. 2023;349:121648.

Zhang B, Cao D, Hu W, Ghias AM, Chen Z. Physics-Informed Multi-Agent deep reinforcement learning enabled distributed voltage control for active distribution network using PV inverters. *Int J Electr Power Energy Syst*. 2024;155:109641.

Hossain R, Gautam M, Thapa J, Livani H, Benidris M. Deep reinforcement learning assisted co-optimization of Volt-VAR grid service in distribution networks. *Sustain Energy Grids Netw*. 2023;35:101086.