24th International Conference on Renewable Energies and Power Quality (ICREPQ'26)
Santander (Spain), 10th to 12th June 2026
ISBN-13: 978-84-09-82044-3

Impact of Forecast-Aware Observations and Attention Mechanisms in a Multi-Agent Deep Reinforcement Strategy for Voltage Control

Alan Lizarazo[1], G. Osma-Pinto[1] and César Duarte[1]

[1] Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T),
Universidad Industrial de Santander (UIS), Bucaramanga, Colombia.
alan.lizarazo@e3t.uis.edu.co, gealosma@uis.edu.co, cedagua@uis.edu.co

Abstract. High DER penetration, particularly PV, induces sub-hourly voltage excursions and reverse power flows that are difficult to mitigate with slow and discrete devices (OLTCs and switched capacitor banks). We study a multi-agent soft actor–critic (MASAC) voltage controller implemented in PyTorch and quantify the impact of two components: (i) forecast-aware local observations (Fc) using an oracle 1-hour-ahead PV forecast summary and (ii) an attention-based centralized critic (Att) that exploits the electrical neighborhood structure. The environment is simulated with AC power flow in pandapower using static ZIP loads and device capability constraints. Daily trajectories are built as 24-h load/PV pairs $\tau = \{\left(P_{\text{load}}(t), \widehat{P_{\text{PV}}}(t)\right)\}_{t=1}^{T}$ with $T = 288$ at $\Delta t = 5$ min; training uses four trajectories. Four variants (NoFc/NoAtt, Fc/NoAtt, NoFc/Att, Fc/Att) are compared in a 2×2 design on a modified IEEE 33-bus (12.66 kV) feeder with one OLTC, three capacitor banks, and six PV inverters. Inverters update every step, while OLTC/CB actions are gated hourly. Scenario A evaluates the same four pairs, while Scenario B tests generalization on 24 unseen pairs. In Scenario B, Fc/Att attains the lowest average voltage deviation (0.0161 p.u.), the smallest out-of-band fraction (6.5%), and the lowest daily losses (1.145 MWh), resulting in the highest episode return under the proposed reward.

Key words. Voltage control, active distribution networks, multi-agent reinforcement learning, attention, forecasting.

1. Introduction

High penetration of distributed energy resources (DERs), particularly photovoltaics (PV), introduces fast stochastic voltage excursions and reverse power flows that stress legacy assets in medium-voltage (MV) feeders. Classical Volt/VAR coordination based on slow electromechanical devices—on-load tap changers (OLTCs) and switched capacitor banks (CBs)—cannot reliably track sub-hourly variability without excessive wear, and must respect dwell/lockout and minimum on/off constraints [1], [2], [3]. In parallel, smart inverters provide fast Volt/VAR and volt/Watt support [4], [5]. However, feeder-wide coordination becomes challenging under uncertainty, limited sensing/communication, and heterogeneous time scales [6], [7].

Limitations of existing approaches. Surveys of volt/VAR control in active distribution networks (ADN) cover multiple objectives and architectures (local, centralized, decentralized, distributed, and hierarchical). Local droop and rule-based policies are simple and communication-light, but they are intrinsically myopic under high DER variability and may interact poorly across devices and locations [1], [8]. Centralized OPF (optimal power flow)/VVO (Volt/VAR optimization) can deliver near-optimal set-points, yet it depends on accurate models, reliable communications, and computational tractability; with discrete devices and multi-timescale constraints, it often becomes mixed-integer and hard to solve online under uncertainty [9], [10]. Distributed and hierarchical schemes alleviate scalability, but they still require careful tuning across time scales. Additionally, robustness to changing operating conditions remains an open challenge [11], [12].

Deep reinforcement learning (DRL) and, more recently, multi-agent DRL (MADRL) have emerged as promising tools to learn closed-loop voltage controllers directly from interaction data, reducing reliance on explicit optimization models at runtime [13], [14], [15]. In particular, centralized-training and decentralized-execution (CTDE) formulations are a natural fit for heterogeneous device fleets, where deployment policies must be local while training can exploit additional information.

Positioning of this paper. We implement a centralized-training, decentralized-execution multi-agent Soft Actor–Critic (CTDE–MASAC) Volt/VAR controller, i.e., decentralized per-device actors trained with centralized critics that can access joint information during learning, with local measurements every 5 minutes, inverter set-points updated at each step, and hourly gating for OLTC/CB actions. Using a fixed environment and training pipeline (pandapower AC power flow, ZIP load model, and a feasibility layer), we perform an ablation-style assessment of two deployability-relevant components: (i) forecast-aware observations (here modeled as an oracle 1-hour-ahead PV summary) and (ii) attention mechanisms in the centralized critic.
Within that framework, we evaluate two architectural elements:

1) Forecast-aware observations: agents augment local measurements with a compact 1-hour PV forecast summary (4 scalars: mean,standar

deviation, P5, P95 of predicted PV power) to anticipate near-term headroom.

2) Attention-based centralized critics: multi-head attention over electrical neighborhoods exploits topology during training while keeping actors decentralized at runtime.

Contribution of this paper. We quantify the individual and combined impact of forecast information and attention under four training trajectories and two evaluation settings: Scenario A (in-distribution, 4 trajectories used in training) and Scenario B (generalization, 24 additional trajectories disjoint from training). Specifically, we:

1) Define four controller variants in a 2×2 design: NoFc/NoAtt, Fc/NoAtt, NoFc/Att, and Fc/Att.
2) Evaluate the variants on a modified IEEE 33-bus MV (12.66 kV) feeder at 5-minute resolution using voltage-quality, loss, and switching metrics, highlighting trade-offs.

2. Multi-Agent DRL Controller

A. Control setting and Markov game formulation

We consider a radial medium-voltage feeder modeled as a graph $G = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the bus set and $\mathcal{E}$ is the line set. The controllable devices are: (i) one OLTC at the substation, (ii) three switched capacitor banks (CBs), and (iii) six PV inverters located at selected buses $\mathrm{I} \subset \mathcal{N}$. Time is discretized with a base step $\Delta t = 5min$ and episodes span one day ($T = 288$ steps).

We model voltage control as a finite-horizon partially observable Markov game (Dec-POMDP) with $N_a = 10$ agents (six inverter agents, one OLTC agent, and three CB agents). Let $i \in 1, \dots, N_a$ index agents. At time step t, each agent receives a local observation $o_{i,t} \in \mathcal{O}_i$ and outputs an action $a_{i,t} \in \mathcal{A}_i$. The joint action is $a_t = \left(a_{1,t}, \dots, a_{N_a,t}\right) \in \mathcal{A} = \prod_i \mathcal{A}_i$. The environment evolves according to

$$s_{t+1} \sim \mathcal{P}(\cdot \mid s_t, a_t, \xi_t), \tag{1}$$

where $s_t \in \mathcal{S}$ is the global state and ξ_t represents exogenous inputs (load and PV availability trajectories). In our setting, a daily trajectory is a 24-hour pair $\tau = \{\left(P^{\text{load}}(t), \hat{P}^{\text{PV}}(t)\right)\}_{t=1}^{T}$ at $5min$ resolution.

Given (s_t, a_t, ξ_t), we compute the next operating point by applying device constraints (feasibility layer) and solving an AC power flow (pandapower). An episode terminates early only if the AC power flow does not converge; otherwise it runs for the full horizon T. Agents are trained to maximize $\mathbb{E}_\pi[\sum_{t=1}^{T} \gamma^{t-1} r_t]$, where r_t is the one-step reward (Subsec. E) and γ is the discount factor (Table 2).

B. Agents, action spaces, and multi-timescale actuation

All agents produce actions at every base step, but slow discrete devices are gated to update only once per hour (every 12 steps), remaining latched between decision instants, whereas inverter set-points may change every $5min$.

PV inverter agents (6). Each inverter uses a Gaussian policy producing a normalized continuous action $u_{i,t} \in [-1,1]$, which is mapped to a reactive-power command $Q_{i,t}$ and then projected onto the instantaneous P–Q capability curve. Each PV inverter is rated at $P_{\max} = 0.63MW$ and $S_{\max} = 0.7MVA$, and must satisfy

$$0 \le P_{i,t} \le \hat{P}_i^{\text{PV}}(t), \quad P_{i,t}^2 + Q_{i,t}^2 \le S_{\max}^2, \tag{2}$$

where $\hat{P}_i^{\text{PV}}(t)$ is the available PV active power from the daily trajectory at the inverter location.

OLTC agent (1). The OLTC uses a categorical policy that selects a tap position from a discrete set. Tap positions lie in $[-16,15]$ with a per-tap step of 0.625%. The selected tap is applied only at hourly decision instants.

Capacitor-bank agents (3). Each CB uses a categorical policy that selects one of three discrete reactive states {0, 0.35, 0.7} MVAR, applied only at hourly decision instants.

This multi-timescale actuation reflects the practical separation between fast inverter control and slower electromechanical devices, while keeping a single base-step simulation.

C. CTDE–MASAC with optional attention

We adopt a CTDE–MASAC scheme. During training, centralized critics access global or topological information, while at execution each actor uses only its local observation.

1) **Actors**: per-device neural networks with layer normalization mapping $o_{i,t}$ to $a_{i,t}$. Temperature parameters are adapted automatically per agent.
2) **Critics**: twin centralized critics $Q_{\phi,i}$ estimating soft state–action values for each agent i.

In the attention variants, the critics use a shared encoder to map the joint input (s_t, a_t) to embeddings, followed by a multi-head attention block that aggregates information over electrical neighbourhoods. The attended representations feed per-agent *multi-layer perceptron* (MLP) heads to output $Q_{\phi,i}$. The attention mask is derived from an electrical adjacency rule and restricts each agent to attend only to neighbouring buses, improving interpretability and controlling complexity. Target networks are updated via Polyak averaging, and prioritized replay is used to improve sample efficiency.

D. Forecast-aware local observations

For inverter agents, the local observation includes (i) the local voltage magnitude at the point of connection and (ii) the local active and reactive powers of the PV unit and co-located load (when present).

In the *forecast-aware* variants, this vector is augmented with a compact summary of a 1-hour-ahead PV active-power forecast at the same location (computed over the next 12 five-minute samples),

$$\mu_{\mathrm{PV}}, \quad \sigma_{\mathrm{PV}}, \quad q_{5\%}, \quad q_{95\%}, \tag{3}$$

where μ_{PV} and σ_{PV} are the mean and standard deviation, and $q_{5\%}/q_{95\%}$ are the 5th/95th percentiles. In this paper we use an oracle (perfect) 1-hour-ahead forecast derived from the realized PV trajectory to quantify an upper bound on the potential value of forecast-aware observations. In the *no-forecast* variants, these features are removed and the policy relies only on current/past measurements.

E. Reward and feasibility layer

The one-step reward is a weighted combination of voltage quality, line-loss efficiency, and discrete switching penalties:

$$r_t = w_V r_V(t) + w_L r_L(t) + w_S r_S(t), \tag{4}$$

with default weights $w_V = 0.5$, $w_L = 0.1$, and $w_S = 0.4$.

Voltage term. Let $d_b(t) = |V_b(t) - 1.0|$ be the per-bus deviation in p.u., and $d_{\max}(t) = \max_{b\in\mathcal{N}} d_b(t)$. We enforce a voltage-first safety rule with an admissible band $d_{\max}(t) \le v_{\mathrm{band}}$ (default $v_{\mathrm{band}} = 0.05$, *i.e.*, $V \in [0.95, 1.05]$ p.u.). If the band is violated at any bus, the step reward is forced to the minimum value ($r_t = -1$). This voltage-band violation does *not* terminate the episode; the rollout continues so the policy can recover. Only AC power-flow non-convergence triggers termination (with $r_t = -1$). Otherwise, $r_V(t)$ is shaped as a smooth quadratic function of the mean absolute deviation.

Loss term. Line losses are computed from the AC power-flow solution and normalized by a reference threshold (default $P_{\mathrm{loss,ref}} = 0.03\, P_{\mathrm{load,tot}}$), interpreted as an admissible loss margin: $r_L(t) = 1$ if $P_{\mathrm{loss}}(t) \le P_{\mathrm{loss,ref}}$, and it decreases quadratically when losses exceed this margin (default scale $k_L = 2.0$).

Switching term. The switching penalty $r_S(t) \in [-1, 0]$ depends on discrete changes at the current step. For the OLTC, the penalty depends on the tap step magnitude (0, 1, or ≥ 2 steps). For CBs, it depends on the number of switching transitions across all banks in the step. The final switching penalty is the minimum of both contributions.

Feasibility layer and safety handling. Before solving the power flow, inverter actions are projected onto the instantaneous P–Q capability curve, and discrete actions are clipped to device limits and hourly gating constraints. AC power-flow convergence is checked at each step. If the solver does not converge, the environment returns $r_t = -1$ and terminates the episode; failures are logged for post-processing.

3. Controller Variants

To isolate the effect of forecast information and attention, we define four MASAC variants:

1) **V1 – NoFc/NoAtt**: baseline MASAC without PV forecast features in the observations and with a standard MLP-based centralized critic (no attention).
2) **V2 – Fc/NoAtt**: same as V1 but with forecast-aware inverter observations; critics are still MLP-based without attention.
3) **V3 – NoFc/Att**: same as V1 but replacing the critic with an attention-based architecture over electrical neighbourhoods; inverter observations exclude forecasts.
4) **V4 – Fc/Att**: full architecture with both forecast-aware inverter observations and attention-based critics (reference design).

Table 1 summarizes the configuration of each variant.

Table 1: MASAC variants considered in this study.

Variant	Forecast features	Attention critic
V1 – NoFc/NoAtt	No	No
V2 – Fc/NoAtt	Yes	No
V3 – NoFc/Att	No	Yes
V4 – Fc/Att	Yes	Yes

This 2×2 design enables separating the main effect of forecast-aware observations from the effect of attention, and assessing whether their combination is synergistic or brittle under out-of-distribution solar irradiance.

All variants share the same network model, ZIP loads, device limits, reward weights, and training hyperparameters. This allows fair comparison focused on the architectural ingredients of interest.

4. Experimental Setup

A. Test Feeder and Load Model

Experiments are conducted on a modified IEEE 33-bus (12.66 kV) distribution feeder with canonical line impedances, bas e voltage and power. The controllable devices (OLTC, CBs, and PV inverters) are placed at representative locations, dividing the feeder into four zones (near substation, central trunk, far end, and lateral branch). Loads are represented with a ZIP model (40 % constant-power, 30 % constant-current, 30% constant-impedance components).

Training uses four representative operating-day trajectories. Two of them keep the solar irradiance trace fixed while varying the load profile, and two keep the load profile fixed while varying solar irradiance, so that learning is exposed to both sources of variability. During training, each episode samples one of these four trajectories. In forecast-aware variants, the 1-hour-ahead PV summary is an oracle computed from the realized PV trajectory (upper-bound setting). For reporting, we define two evaluation settings: Scenario A evaluates in-distribution performance on the same four training trajectories used during training, while Scenario B evaluates generalization on 24 additional daily load/PV trajectories that are disjoint from the training set.

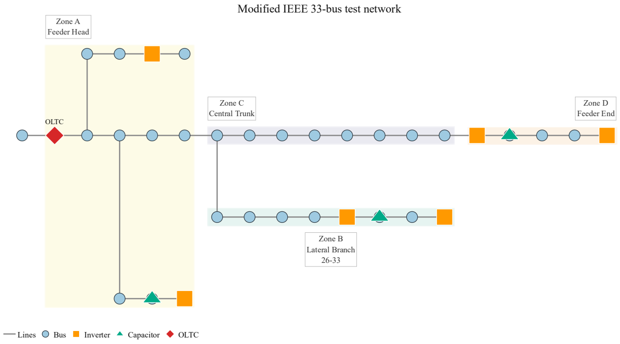

Figure 1: Modified IEEE 33-bus feeder (12.66 kV) and controllable device placement. PV inverters are connected at buses 14, 18, 21, 25, 30, and 33; CBs at buses 15, 24, and 31; and the OLTC is located at the substation transformer (bus 34–1).

B. Daily load and PV trajectories

Daily trajectories are built as 24-h profiles sampled every 5 min ($T = 288$). Load profiles follow representative residential and commercial daily shapes reported in [16], while PV active-power availability is derived from solar irradiance measurements taken on the rooftop terrace of the E3T building at Universidad Industrial de Santander (UIS), Bucaramanga, Colombia, and mapped to $\hat{P}^{\mathrm{PV}}(t)$ through a normalized PV profile. Training uses four trajectories: two vary the load profile under a fixed solar irradiance trace, and two vary solar irradiance under a fixed load profile. Scenario B evaluates 24 additional load/PV trajectories not used during training.

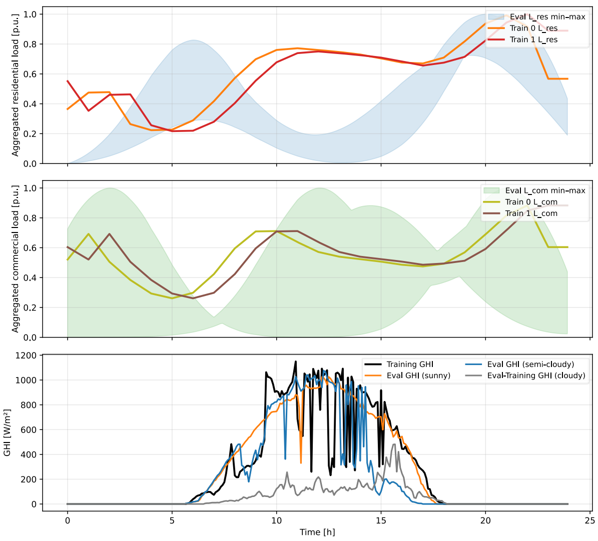

Figure 2: Aggregated daily load and solar irradiance profiles used to build 24-h trajectories ($\Delta t = 5$ min, $T = 288$). Top: aggregated residential load profile applied to zones A and C; middle: aggregated commercial load profile applied to zones B and D. Thick lines denote the two training load days (weekday/weekend), while the shaded bands indicate the min-max range spanned by the six evaluation load profiles. Bottom: measured Global Horizontal Irradiance (GHI) trace used for training and the additional irradiance days used to form combined load+PV evaluation trajectories.

C. Training hyperparameters

Table 2 reports the most relevant training settings used across all variants (values from the training configuration used in the evaluation study).

Table 2: Training hyperparameters.

Parameter	**Value**
Discount γ	0.99
Actor/Critic learning rate	1×10^{-4} / 3×10^{-4}
Critic LR cap	up to 1×10^{-3}
Batch size	256
Replay buffer	10,000 (prioritized; $\alpha \approx 0.61$, $\beta \to 1$)
Polyak ρ	1×10^{-4}
Attention heads	4
Warm-up steps	288
Entropy target (continuous)	-1.0
Discrete entropy coefficient	0.7

5. Results and Discussion

A. Training behaviour of MASAC variants

Fig. 3 shows the MA-10 episodic return over 100 training episodes. All variants start with negative returns due to voltage-violation penalties during exploration and improve as policies stabilize. Variants with forecast features (V2, V4) learn faster and reach higher steady returns than no-forecast variants (V1, V3). The full Fc/Att design (V4) exhibits the most stable training dynamics in our runs.

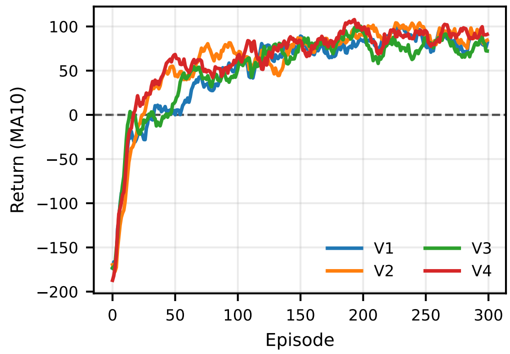

Figure 3: Training curves of the four MASAC variants. The plot shows the 10-episode moving average of the episodic return (MA-10) over 100 training episodes.

B. Voltage quality and losses across MASAC variants

Beyond the overall fraction of time steps with voltage violations and the daily energy losses, we also quantify how far the voltages are from the nominal value 1.0 p.u. and how intensively the available discrete devices are operated. For each episode (one operating day), we compute the average absolute voltage deviation

$$\bar{\Delta V} = \frac{1}{TN_b}\sum_{t=1}^{T}\sum_{b=1}^{N_b}|V_b(t) - 1.0|, \tag{5}$$

where $T = 288$ is the number of 5-minute time steps in a day and N_b is the number of buses. We also count the number of tap changes of the OLTC transformer, N_{tap}, and the number of switching operations for each CB, N_{cap0}, N_{cap1}, and N_{cap2}, by counting transitions in the corresponding discrete position signals. We additionally report the daily voltage extrema $V_{\min}/V_{\max}$ (minimum/maximum across all buses and time steps), *FracStepsOut* (fraction of time steps in which at least one bus voltage is outside $[0.95,1.05]$ p.u.), and *Loss* [MWh], computed as

$$\text{Loss [MWh]} = \sum_{t=1}^{T} P_{\text{loss}}(t)\,\Delta t, \Delta t = \frac{5}{60}\text{ h}, \tag{6}$$

where $P_{\text{loss}}(t)$ is the total line-loss power obtained from the AC power-flow solution at time step t.

Table 3 summarizes the average daily performance metrics for two evaluation settings: Scenario A evaluates in-distribution performance on the same four trajectories used for training (pairs 0–3), while Scenario B evaluates generalization on 24 additional daily load/PV trajectories that are disjoint from the training set (pairs 4–27).

Scenario A (in-distribution). All variants satisfy the voltage band on the four training curves (FracStepsOut =0), as expected for in-distribution conditions. Differences are mainly expressed in efficiency and in the use of discrete devices. The full Fc/Att variant (V4) yields the lowest daily losses (1.088 MWh) and the lowest average voltage deviation ($\Delta\bar{V} = 0.0111$ p.u.), but it relies more on OLTC and capacitor switching ($N_{\text{tap}} \approx 10.3$). In contrast, the baseline (V1) avoids OLTC actions but exhibits high capacitor switching, especially on cap1 ($N_{\text{cap1}} \approx 17.8$), reflecting a more reactive control policy.

Scenario B (generalization). Under the 24 evaluation curves with simultaneous demand and irradiance variations, the full Fc/Att agent (V4) achieves the lowest out-of-band fraction (6.5%), the smallest deviation ($\Delta\bar{V} = 0.0161$ p.u.) and the lowest daily losses (1.145 MWh) among the tested variants. Forecast features alone (V2) remain competitive in voltage quality (7.5% out-of-band), but at higher losses (1.243 MWh). The attention-only variant (V3) shows higher $\Delta\bar{V}$ and FracStepsOut than V1/V2 in Scenario B, indicating that attention provides benefits in this setup only when combined with forecast-aware observations.

Overall, Scenario A confirms constraint satisfaction on seen trajectories, while Scenario B exposes the trade-off between voltage quality and discrete-device wear: the controller with the best Scenario B voltage-quality metrics (V4) is also the most aggressive in OLTC and capacitor operations. This motivates incorporating explicit wear costs and tuning switching penalties when transferring these controllers to field deployments.

Data-driven selection and value of the 2×2 ablation. The four-variant design is necessary to attribute performance changes to specific components: comparing V1 vs. V2 isolates the effect of forecast-aware observations, comparing V1 vs. V3 isolates the effect of attention, and comparing V2 vs. V4 tests whether attention provides additional benefit once forecasts are available. Based on Table 3, V4 is preferred when voltage quality and losses are prioritized (best $\Delta\bar{V}$, FracStepsOut, and Loss in Scenario B), while V2 can be preferred when reducing discrete-device operations is a primary objective, since it remains competitive in voltage quality with fewer OLTC/CB actions.

6. Conclusions and Limitations

This paper evaluated a CTDE–MASAC multi-agent Volt/VAR controller on a modified IEEE 33-bus feeder, focusing on a 2×2 ablation of (i) forecast-aware local observations and (ii) attention mechanisms in the centralized critic. Key findings from Scenario A (in-distribution, four training trajectories, pairs 0–3) and Scenario B (generalization, 24 unseen trajectories, pairs 4–27) are:

1) On the training trajectories (Scenario A), all variants keep voltages within $[0.95,1.05]$ p.u. (FracStepsOut= 0). Differences appear mainly in technical losses and the usage of discrete devices.
2) Under generalization (Scenario B), combining forecast-aware observations with attention (V4) yields the lowest $\bar{\Delta V}$, the lowest FracStepsOut, and the lowest losses among the tested variants in Scenario B (Table 3), at the cost of higher OLTC/CB operations.
3) Forecast-aware observations alone (V2) remain competitive in voltage quality in Scenario B, while attention alone (V3) does not improve Scenario B voltage-quality metrics in this study.

Because the forecast-aware variants use an oracle (perfect) 1-hour-ahead PV forecast, the reported gains should be interpreted as an upper bound; assessing robustness under realistic forecast errors is required before deployment-oriented claims. Moreover, improvements in voltage quality and losses come with increased discrete operations, so the switching penalty (or explicit wear constraints) should be calibrated to meet operational limits.

Limitations and threats to validity. (i) Forecast-aware features are modeled as an *oracle* 1-hour-ahead PV summary (perfect information); future work must evaluate sensitivity to forecast error and potential distribution shift in PV and load. (ii) Training used only four representative daily trajectories; broader training sets with higher variability in load and irradiance are required before making stronger general claims about deployability. (iii) Discrete-device constraints are represented via hourly gating and switching penalties; additional field constraints (*e.g.*, explicit dwell/lockout logic) should be included for deployment-grade validation.

Table 3: Summary of average daily performance metrics for Scenario A (in-distribution evaluation on the four training trajectories) and Scenario B (generalization evaluation on 24 disjoint load/PV trajectories), including daily voltage extrema (V_{min}, V_{max}).

Scenario	Variant	$V_{\min}$ [p.u.]	$V_{\max}$ [p.u.]	$\bar{\Delta}V$	FracStepsOut	Loss [MWh]	N_{tap}	N_{cap0}	N_{cap1}	N_{cap2}
A	V1	0.9530	1.0299	0.0117	0.000	1.161	0.0	0.0	17.8	2.0
A	V2	0.9628	1.0350	0.0123	0.000	1.183	10.0	3.0	8.2	9.8
A	V3	0.9556	1.0362	0.0117	0.000	1.212	1.0	0.0	7.5	21.8
A	V4	0.9625	1.0350	0.0111	0.000	1.088	10.2	8.5	10.0	4.8
B	V1	0.9662	1.0673	0.0179	0.081	1.235	0.0	4.8	8.8	3.1
B	V2	0.9638	1.0675	0.0179	0.075	1.243	9.5	2.7	4.7	7.5
B	V3	0.9702	1.0719	0.0198	0.124	1.184	1.0	0.0	7.3	6.9
B	V4	0.9678	1.0698	0.0161	0.065	1.145	10.3	8.1	6.4	4.5

Acknowledgement

The authors wish to thank the Department of Electrical, Electronics and Telecommunications Engineering (Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones), and the Vice-Rectorate for Research and Extension (Vicerrectoría de Investigación y Extensión) from the Universidad Industrial de Santander.

References

[1] P. Srivastava, R. Haider, V. Nair, V. Venkataramanan, A. M. Annaswamy, and A. K. Srivastava, "Voltage regulation in distribution grids: A survey," *Annual Reviews in Control*, vol. 55, pp. 165–181, 2023, doi: 10.1016/j.arcontrol.2023.03.008.

[2] A. Savasci, A. Inaolaji, and S. Paudyal, "Optimal tuning of local voltage control rule of load tap changers for dynamic operation of unbalanced distribution networks," *IEEE Transactions on Industry Applications*, vol. 60, no. 1, pp. 1322–1331, 2024, doi: 10.1109/TIA.2023.3313128.

[3] Y. Liu, J. Bebic, B. Kroposki, J. de Bedout, and W. Ren, "Distribution system voltage performance analysis for high-penetration PV," in *2008 IEEE energy 2030 conference*, 2008, pp. 1–8. doi: 10.1109/ENERGY.2008.4781069.

[4] Y. Gui *et al.*, "Voltage support with PV inverters in low-voltage distribution networks: An overview," *IEEE journal of emerging and selected topics in power electronics*, vol. 12, no. 2, pp. 1503–1522, 2024, doi: 10.1109/JESTPE.2023.3280926.

[5] C. W. G. C6.31, "Smart inverters for distribution systems," *CIGRE Technical Brochure*, vol. No. 837, 2020.

[6] F. Matsushima, M. Aoki, Y. Nakamura, S. C. Verma, K. Ueda, and Y. Imanishi, "Multi-timescale voltage control method using limited measurable information with explainable deep reinforcement learning," *Energies*, vol. 18, no. 3, p. 653, 2025, doi: 10.3390/en18030653.

[7] H. Wang, Y. Yao, J. Zhao, and F. Ding, "Data-driven mean-corrected recursive estimation-based optimal DER dispatch for distribution system voltage control," *IEEE Transactions on Sustainable Energy*, vol. 16, no. 4, pp. 2861–2873, 2025, doi: 10.1109/TSTE.2025.3570218.

[8] X. Zhang, Z. Wu, Q. Sun, W. Gu, S. Zheng, and J. Zhao, "Application and progress of artificial intelligence technology in the field of distribution network voltage control: A review," *Renewable & Sustainable Energy Reviews*, vol. 192, p. 114282, 2024.

[9] Y. Zheng, H. Zhang, Z. Long, S. Gao, Q. Yang, and H. Ji, "Data-driven multi-mode adaptive control for distribution networks with multi-region coordination," *Processes*, vol. 13, no. 10, 2025, doi: 10.3390/pr13103046.

[10] N. R. Chowdhury, A. K. Rao, Y. Bichpuriya, V. Sarangan, and S. A. Soman, "Provably robust online voltage control for distribution networks with line parameter estimation," *European Journal of Control*, vol. 85, p. 101252, 2025, doi: 10.1016/j.ejcon.2025.101252.

[11] W. Jiao *et al.*, "Analytical target cascading based real-time distributed voltage control for MV and LV active distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 159, p. 110024, 2024, doi: 10.1016/j.ijepes.2024.110024.

[12] Z. Zhang, Z. Y. Dong, and D. Yue, "Multiple time-scale voltage regulation for active distribution networks via three-level coordinated control," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 4429–4439, 2024, doi: 10.1109/TII.2023.3304340.

[13] X. Sun and J. Qiu, "Two-stage volt/VAR control in active distribution networks with multi-agent deep reinforcement learning method," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 2903–2912, 2021, doi: 10.1109/TSG.2021.3052998.

[14] D. Cao *et al.*, "Deep reinforcement learning enabled physical-model-free two-timescale voltage control method for active distribution systems," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 149–165, 2022, doi: 10.1109/TSG.2021.3121792.

[15] H. Peng *et al.*, "Deep reinforcement learning based multi-timescale volt/VAR control in distribution networks considering network reconfiguration," *IEEE Transactions on Sustainable Energy*, vol. 16, no. 4, pp. 2948–2958, 2025, doi: 10.1109/TSTE.2025.3574806.

[16] S. Toledo-Cortés, J. S. Lara, A. Zambrano, F. A. G. Osorio, and J. R. Garcia, "Characterization of electricity demand based on energy consumption data from colombia," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 5, pp. 4798–4809, 2023, doi: 10.11591/ijece.v13i5.pp4798-4809.