International Journal of Power Electronics and Drive System (IJPEDS)
Vol. 17, No. 4, December 2026, pp. 2504∼2525
ISSN: 2088-8694, DOI: 10.11591/ijpeds.v17.i4.pp2504-2525

A deep reinforcement learning enhanced FATA for optimal FACTS device allocation in renewable-rich power grids

Ben Lemkherbeche Ilyes, Bensalem Ahmed, Khemmoudj Sara Hiba
Department of Electrical Engineering, Faculty of Technology, University of Batna 2 - Mostefa Ben Boulaïd, Batna, Algeria

Article Info

Article history:

Received Jul 4, 2026
Revised Aug 31, 2026
Accepted Sep 13, 2026

Keywords:

FACTS
Fata Morgana algorithm
Optimal power flow
Reinforcement learning
Static VAR compensator
Thyristor-controlled phase shifter
Thyristor-controlled series capacitor

ABSTRACT

The integration of renewable energy sources into modern power grids introduces significant uncertainty and operational challenges, particularly in the optimal placement of flexible AC transmission systems (FACTS) devices. This paper proposes a deep reinforcement learning (DRL) enhanced fata morgana algorithm (FATA) for optimal allocation of thyristor-controlled series capacitor (TCSC), thyristor-controlled phase shifter (TCPS), and static VAR compensator (SVC) devices in renewable-rich power systems. Unlike the basic FATA which employs fixed parameters throughout optimization, the proposed approach integrates a Q-learning-based adaptive parameter control mechanism that dynamically adjusts the reflectance coefficient α based on real-time convergence metrics. The algorithm incorporates experience replay for learning from past optimization trajectories, adaptive population management that dynamically adjusts population size based on convergence stagnation, and local intensification search for refining elite solutions. The proposed DRL-enhanced FATA is validated on a modified Institute of Electrical and Electronics Engineers (IEEE) 30-bus test system with wind integration under multiple operating scenarios including generation cost minimization, power loss minimization, gross cost minimization, and load uncertainty conditions. Experimental results demonstrate that the proposed enhancements achieve significant and consistent improvements over the basic FATA, with generation cost reduced by 9.2%, power loss reduced by 14.7%, voltage deviation improved by 55.9%, and gross cost reduced by 67.9% under uncertainty scenarios. The algorithm successfully adapts its parameters in real-time, providing effective exploration-exploitation balancing and identifying critical transmission lines and load buses as strategic locations for FACTS deployment.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Ben Lemkherbeche Ilyes
Department of Electrical Engineering, Faculty of Technology, University of Batna 2 - Mostefa Ben Boulaïd
Batna 05000, Algeria
Email: ilyes.benlemkherbeche@univ-batna2.dz

1. INTRODUCTION

The increasing integration of renewable energy sources into modern power grids has introduced significant challenges for system operators, particularly in managing uncertainty, maintaining voltage stability, and minimizing operational costs [1], [2]. Wind power, one of the fastest-growing renewable energy sources, is characterized by inherent variability and unpredictability due to its dependence on weather conditions [3], [4]. This variability complicates the optimal power flow (OPF) problem, which traditionally aims to minimize generation costs and power losses while satisfying system constraints. Simultaneously, the growing demand

for electricity and the expansion of transmission networks have pushed existing infrastructure closer to its operational limits, necessitating the deployment of flexible AC transmission systems (FACTS) devices. These devices, including thyristor-controlled series capacitor (TCSC), thyristor-controlled phase shifter (TCPS), and static VAR compensator (SVC), provide the ability to control transmission line parameters such as impedance, voltage magnitude, and phase angle, thereby enhancing power transfer capability and system stability [5], [6].

The optimal placement and sizing of FACTS devices, however, remains a challenging optimization problem due to the high-dimensional search space, the discrete nature of device locations, the continuous nature of device parameters, and the complex interactions between multiple devices. Furthermore, the integration of wind power introduces additional uncertainty that must be accounted for in the optimization process. Traditional mathematical programming methods, such as gradient-based approaches and mixed-integer programming, often struggle with these problems due to the non-convex, non-linear, and non-differentiable characteristics of the OPF problem with valve-point effects and discrete FACTS placement decisions [5], [7].

Metaheuristic algorithms have emerged as powerful tools for solving complex OPF problems with FACTS devices [8], [9]. The whale optimization algorithm (WOA) was used by Nadeem *et al.* [10] to find optimal placement of TCSC, SVC, and unified power flow controller (UPFC) devices in IEEE 14-bus and 30-bus systems. Their approach first identified weak buses and lines using PV curves and the line stability index (Lmn), then employed WOA to determine optimal sizing and coordination of these devices with existing reactive power sources. The objective was to minimize operating cost consisting of active power losses and FACTS device installation costs. The findings showed that total system operating costs and transmission line losses were considerably reduced by WOA as compared to genetic algorithm (GA) and particle swarm optimization (PSO). Shahzad *et al.* [11] applied the strawberry plant propagation algorithm (SPPA) for optimal sitting and sizing of distributed generators (DGs) in radial distribution systems. Unlike FACTS devices which are typically used in transmission networks, this work focused on DG placement in 33-node and 69-node radial distribution networks to reduce active power losses and improve voltage profiles. The SPPA, inspired by the natural propagation behavior of strawberry plants through runners, demonstrated superior performance compared to the Mine Blast Algorithm and PSO, achieving power loss reductions of 69.4% and 79.2% for the 33-node and 69-node systems, respectively. The authors also conducted a comprehensive cost analysis, showing significant annual savings from reduced energy losses. Duman *et al.* [12] integrated chaotic maps with a hybrid particle swarm optimization and gravitational search algorithm (CPSOGSA) to solve the OPF problem with stochastic wind power and FACTS devices. The proposed approach employed ten different chaotic maps to enhance the exploration and exploitation capabilities of the conventional PSOGSA. The CPSOGSA was validated on IEEE 30-bus and 57-bus test systems with TCSC and TCPS devices under multiple objective functions including fuel cost minimization, emission reduction, power loss minimization, and voltage stability enhancement. The results demonstrated that CPSOGSA outperformed moth swarm algorithm (MSA), grey wolf optimizer (GWO), and WOA across sixteen case studies, with statistical validation using the Wilcoxon signed-rank test confirming its superiority.

Most recently, Aljaidi *et al.* [13] introduced the fata morgana algorithm (FATA), a novel optimization method inspired by the natural phenomenon of mirage formation. The algorithm employs two key mechanisms: the mirage light filtering principle, which evaluates population quality using definite integration, and the light propagation strategy, which simulates light refraction, reflection, and total internal reflection to balance exploration and exploitation. FATA was applied to optimize the placement and sizing of TCSC, TCPS, and SVC devices in an IEEE 30-bus system with wind integration at buses 5 and 11. Across multiple case studies including generation cost minimization, power loss minimization, gross cost minimization, and load uncertainty scenarios, FATA consistently outperformed established algorithms including improved RIME (IRIME), Newton-Raphson-based optimization (NRBO), resistance capacitance algorithm (RCA), krill optimization algorithm (KOA), and GWO. The algorithm achieved a minimum generation cost of 807.0405 $/h and reduced power losses to 5.5917 MW, demonstrating superior convergence speed, solution quality, and computational efficiency.

More recently, Mohamed *et al.* [14] applied the chaos game optimization algorithm for OPF incorporating renewable energy sources and FACTS devices, demonstrating improved convergence characteristics compared to conventional metaheuristics. Sulaiman and Mustaffa [15] proposed hyper-heuristic strategies for optimal power flow with FACTS allocation in wind-integrated systems, showing that adaptive strategy selection can enhance solution quality. Wu *et al.* [16] developed a safe deep reinforcement learning approach for real-time OPF based on primal-dual and prior knowledge guidance, achieving fast convergence

while maintaining operational safety. Ngei *et al.* [17] proposed a hybrid genetic algorithm-improved particle swarm optimization for optimal sizing and placement of STATCOM, TCSC, and UPFC devices, demonstrating improved performance over individual algorithms. Wu *et al.* [18] introduced physics-informed reinforcement learning for real-time OPF with renewable energy resources, incorporating domain knowledge to guide the learning process and improve solution feasibility. These recent studies highlight the growing interest in integrating learning-based approaches and hybrid methods for power system optimization, further motivating our proposed DRL-enhanced framework.

Despite these advances, existing optimization algorithms for FACTS device placement share several fundamental limitations that remain unaddressed. First, most algorithms employ fixed parameters throughout the optimization process, which leads to either premature convergence when exploration is insufficient or slow convergence when exploitation is inadequate. This fixed-parameter limitation is particularly evident in the baseline FATA, which uses a constant reflectance coefficient $\alpha = 0.20$ regardless of the optimization stage or convergence behavior. Second, they operate deterministically without any mechanism to learn from past optimization trajectories, requiring complete re-optimization for each new problem instance. Third, they lack the ability to adapt population size dynamically based on convergence stagnation, wasting computational resources on poor solutions when near the optimum or insufficient exploration when far from it. Fourth, they have no local intensification mechanism to refine elite solutions once a promising region is identified. Consequently, while the baseline FATA demonstrates competitive performance, its fixed-parameter nature limits its ability to adapt to the changing optimization landscape, particularly under varying load conditions and renewable generation uncertainties.

To address these limitations, this paper proposes a DRL-enhanced FATA algorithm that introduces four key innovations clearly distinguishing it from the baseline FATA. First, unlike the fixed $\alpha = 0.20$ of the original FATA, our approach employs Q-learning-based adaptive parameter control that dynamically adjusts the reflectance coefficient between 0.10 and 0.30 based on real-time convergence metrics, enabling the algorithm to autonomously balance exploration and exploitation. Second, we introduce an experience replay mechanism (absent in the baseline FATA) that stores and samples past optimization transitions, allowing the algorithm to learn from its own historical performance. Third, we implement adaptive population management that dynamically adjusts population size between 15 and 50 based on improvement rates, a feature not present in the original FATA. Fourth, we incorporate a local intensification search that refines elite solutions through pattern search, addressing the baseline FATA's lack of a local exploitation mechanism. These four enhancements collectively transform the fixed-parameter FATA into a self-adaptive, learning-enabled optimization framework that directly addresses the limitations identified in the gap analysis.

The proposed DRL-enhanced FATA is validated on the modified IEEE 30-bus test system with wind integration at buses 5 and 11, following the same configuration as the baseline FATA paper [13]. Eight case studies are conducted, including generation cost minimization (Case 1), active power loss minimization (Case 2), gross cost minimization under fixed loading (Case 3), gross cost minimization under four load uncertainty scenarios (54.7%, 65.4%, 74.6%, and 85.3% of base load, Cases 4-7), and expected cost calculation across all scenarios (Case 8). Experimental results demonstrate significant and consistent improvements over the baseline FATA: a 9.2% reduction in generation cost and 14.7% reduction in power loss (Case 1); a 7.0% reduction in power loss, 11.6% reduction in generation cost, and 55.9% improvement in voltage deviation (Case 2); and gross cost reductions of 67.9%, 30.2%, 17.1%, and 16.6% under load uncertainty scenarios (Cases 4-7). The Q-learning agent successfully adapts the reflectance coefficient α in real-time, varying between 0.10 and 0.30 with higher values selected during early exploration and lower values during late exploitation, demonstrating effective exploration-exploitation balancing.

The main contributions of this paper are summarized as follows:

- Q-learning-based adaptive parameter control for dynamic tuning of the reflectance coefficient α based on a 10-state discretization of convergence metrics, adaptively adjusting α between 0.10 and 0.30, replacing the fixed parameter ($\alpha = 0.20$) of the baseline FATA.
- Experience replay mechanism that stores past optimization transitions (s_t, a_t, R_t, s_{t+1}) and samples them to update the Q-table, enabling the algorithm to learn from its own historical performance and improving learning stability.
- Adaptive population management that dynamically adjusts population size between 15 and 50 individuals based on recent improvement rates, preventing premature convergence and improving solution quality.

– Local intensification search that performs pattern search around elite solutions every few iterations, refining near-optimal solutions without significant additional computational cost.
– Comprehensive validation on the modified IEEE 30-bus system with wind integration across eight case studies, comparing the proposed DRL-enhanced FATA against the baseline FATA under fixed and uncertain loading conditions.

The proposed DRL-enhanced FATA offers a robust, adaptive, and efficient solution for optimal FACTS device allocation in modern renewable-rich power systems. By integrating reinforcement learning with metaheuristic optimization, this work bridges the gap between adaptive parameter control and population-based search, demonstrating that learning from past optimization trajectories can significantly enhance solution quality and convergence speed. The adaptive parameter control mechanism eliminates the need for manual parameter tuning, making the algorithm more accessible to power system engineers and operators. The significant improvements under load uncertainty scenarios highlight the practical relevance of the approach for real-world power system operation, where demand variability and renewable generation uncertainty are everyday challenges. The identification of critical transmission lines and load buses for FACTS deployment provides valuable insights for system planners and operators, potentially reducing investment costs and improving system reliability.

The remainder of this paper is organized as follows: Section 2 presents the materials, including the test system description, thermal and wind generator modeling, FACTS device configurations, load uncertainty scenarios, and objective functions. Section 3 then describes the proposed methodology in detail, including the Q-learning-based adaptation mechanism, the experience replay system, and the adaptive population management strategy. Section 4 presents and discusses the experimental results across all eight case studies, including convergence analysis, parameter adaptation behavior, and optimal FACTS placements. Finally, Section 5 concludes the paper and outlines directions for future research.

2. MATERIALS

Building upon the problem statement established in Section 1, this section presents the modified IEEE 30-bus test system, the mathematical modeling of thermal and wind generators, the formulation of FACTS devices, the definition of load uncertainty scenarios, and the objective functions with associated constraints for the OPF problem.

2.1. Test system description

The proposed methodology is validated on the modified IEEE 30-bus test system, as illustrated in Figure 1. The standard IEEE 30-bus system consists of 6 generator buses (buses 1, 2, 5, 8, 11, 13), 24 load buses (buses 3, 4, 6, 7, 9, 10, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30), and 41 transmission lines with 4 transformer branches.The system operates at a base power of 100 MVA with a base voltage of 132 kV, and all line and transformer parameters are taken from the standard IEEE 30-bus data available in MATPOWER case files.

2.2. Modifications to the standard IEEE 30-bus system

The standard IEEE 30-bus system [19] was modified as follows to align with the original FATA paper [13] and to enable FACTS device optimization under renewable integration. First, conventional generators at buses 5 and 11 were replaced with wind farms rated at 75 MW and 60 MW, respectively, where the wind power cost model incorporates Weibull-distributed wind speed variability with scale parameters $\alpha = 9$ and $\alpha = 10$, and shape parameter $\beta = 2$ for both farms. Second, the cost coefficients and power limits of thermal generators at buses 1, 2, 8, and 13 were modified to match the paper's specifications, replacing the default MATPOWER coefficients [20]. Third, bus voltage magnitude limits were set to $[0.95, 1.10]$ p.u. for all buses to ensure secure operation. Fourth, six FACTS devices (2 TCSC, 2 TCPS, 2 SVC) were added to the system, where TCSC and TCPS devices can be placed on any of the 41 transmission lines, while SVC devices can be placed on any of the 24 load (PQ) buses. Fifth, four discrete loading scenarios (54.7%, 65.4%, 74.6%, and 85.3% of base load) were defined to capture demand variability. Sixth, transmission lines with undefined or zero ratings were assigned a default rating of 100 MVA to ensure thermal limit enforcement. Table 1 summarizes the key modifications made to the standard IEEE 30-bus system. The network topology (41 branches, bus connections, line impedances, and transformer parameters) remains unchanged from the standard IEEE 30-bus system, and all modifications are applied consistently across all eight case studies to ensure fair comparison.

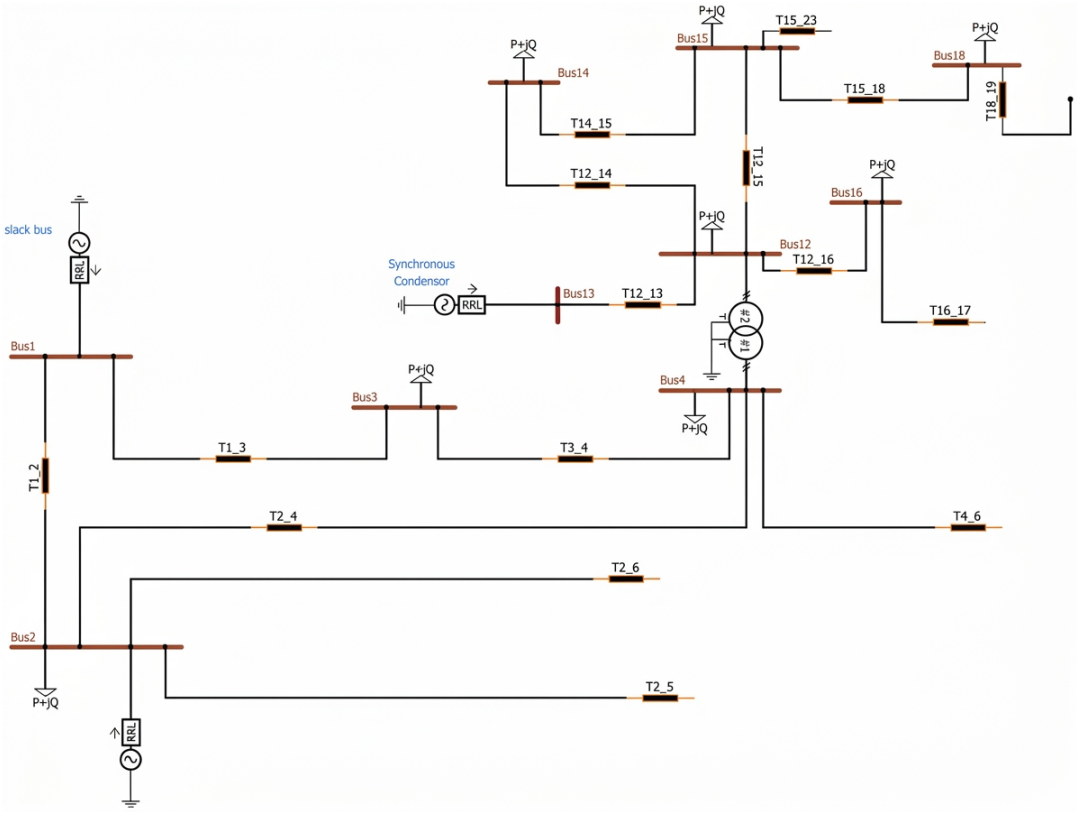

Figure 1. IEEE 30-bus test system: The modified system replaces conventional generators at buses 5 and 11 with wind power plants (75 MW and 60 MW, respectively)

Table 1. Summary of modifications to IEEE 30-bus system

Component	Standard IEEE 30-bus	Modified system
Bus 5	Thermal generator	Wind farm (75 MW)
Bus 11	Thermal generator	Wind farm (60 MW)
Generator costs	MATPOWER defaults	Paper-specific coefficients
Generator limits	Default Pmin/Pmax	Modified ranges
Voltage limits	Variable defaults	Fixed [0.95, 1.10] p.u.
FACTS devices	None	2 TCSC + 2 TCPS + 2 SVC
Load level	Fixed 100%	4 scenarios (54.7-85.3%)
Line ratings	Some undefined	Default 100 MVA

2.3. Thermal generator modeling

The generation cost of thermal units is modeled using a quadratic cost function with valve-point effects, as expressed in (1), where $C_{Ti}(P_{TGi})$ represents the cost in \$/h for the i-th thermal unit, a_i, b_i, c_i are the cost coefficients, d_i and e_i represent the valve-point loading coefficients that capture the non-linear behavior of steam turbines, P_{TGi} is the active power output of the thermal generator in MW, and $P_{TGi}^{\min}$ is its minimum power output limit. This valve-point effect creates ripples in the cost function, making the optimization problem non-convex and more challenging to solve.

$$C_{Ti}(P_{TGi}) = a_i + b_i P_{TGi} + c_i P_{TGi}^2 + |d_i \times \sin(e_i \times (P_{TGi}^{\min} - P_{TGi}))| \quad (1)$$

The thermal generator parameters used in this study are summarized in Table 2. These parameters are taken directly from the original FATA paper to ensure exact replication of the baseline results. It can be observed that generators at buses 1, 2, 8, and 13 operate with the specified cost coefficients, while the generators at buses 5 and 11 are replaced by wind farms as described in the following subsection.

Table 2. Thermal generator cost coefficients (modified from standard IEEE 30-bus)

Bus	a	b	c	d	e	P_{min} (MW)	P_{max} (MW)	Q_{max} (MVAr)
1	0	2.00	0.0200	100	0.042	50	200	150
2	0	1.75	0.0175	100	0.042	20	80	60
8	0	3.25	0.00834	100	0.042	10	35	40
13	0	3.00	0.0250	100	0.042	12	40	45

2.4. Wind generator modeling

The wind speed variability is captured using the Weibull probability density function shown in (2), where $f_\nu(\nu)$ is the probability density of wind speed ν, α is the scale parameter (m/s) representing the characteristic wind speed, and β is the shape parameter (dimensionless) representing the distribution's shape, where $\beta = 2$ corresponds to the Rayleigh distribution.

$$f_\nu(\nu) = \left(\frac{\beta}{\alpha}\right)\left(\frac{\nu}{\alpha}\right)^{(\beta-1)} e^{-(\nu/\alpha)^\beta}, \quad 0 < \nu < \infty \tag{2}$$

The wind turbine power output is then defined by (3), where $p_w(v)$ is the power output at wind speed v, p_{wr} is the rated power output in MW, v_i is the cut-in wind speed at which the turbine starts generating (3 m/s), v_r is the rated wind speed at which the turbine reaches its maximum output (16 m/s), and v_o is the cut-out wind speed at which the turbine shuts down for safety (25 m/s).

$$p_w(v) = \begin{cases} 0 & v < v_i \text{ or } v > v_o \\ p_{wr}\left(\frac{v-v_i}{v_r-v_i}\right) & v_i \le v \le v_r \\ p_{wr} & v_r < v \le v_o \end{cases} \tag{3}$$

The wind farm parameters employed in this study are detailed in Table 3, where each wind farm is characterized by its rated power, Weibull distribution parameters (α and β), direct cost coefficient g_w in \$/MWh, reserve cost coefficient K_R in \$/MWh, and penalty cost coefficient K_P in \$/MWh. These parameters are adopted from the original FATA paper to maintain consistency with the baseline methodology.

The total wind power cost is calculated using (4), where $C_{wj}(p_{wj})$ represents the total cost for the j-th wind power plant, g_{wj} is the direct cost coefficient, p_{wj} is the scheduled power output, K_{Rwj} and K_{Pwj} are the reserve and penalty cost coefficients, respectively, P_{wj} is the scheduled power, and P_{waj} is the actual available power from the wind plant. This formulation accounts for the inherent uncertainty in wind power generation, where reserve costs are incurred when scheduled power exceeds actual available power, and penalty costs are incurred when actual power exceeds scheduled power. The reserve and penalty cost terms in (4) account for deviations between the scheduled and actual available wind power. The reserve cost represents the additional cost associated with a shortage in available wind power relative to the scheduled value, whereas the penalty cost represents the cost associated with available wind power exceeding the scheduled value

$$C_{wj}(p_{wj}) = g_{wj}p_{wj} + K_{Rwj}(P_{wj} - P_{waj}) + K_{Pwj}(P_{waj} - P_{wj}) \tag{4}$$

Table 3. Wind farm parameters

Bus	Rated power (MW)	α	β	g_w (\$/MWh)	K_R (\$/MWh)	K_P (\$/MWh)
5	75	9	2	1.60	3.0	1.50
11	60	10	2	1.75	3.0	1.50

2.5. FACTS device modeling

Three types of FACTS devices are incorporated into the network, with two units of each type as per the original paper configuration.

2.5.1. Thyristor-controlled series capacitor

The TCSC is placed on transmission lines to modify the line reactance according to (5), where X_{new} is the modified reactance, X_{old} is the original line reactance, and k is the compensation factor ranging from -0.70 (inductive compensation) to 0.70 (capacitive compensation). A negative k value increases the inductive reactance, while a positive k value provides capacitive compensation that reduces the effective line reactance.

$$X_{new} = X_{old} \cdot (1 - k), \quad k \in [-0.70, 0.70] \tag{5}$$

2.5.2. Thyristor-controlled phase shifter

The TCPS introduces a phase shift Φ in degrees to control active power flow, with operational limits given by (6). The phase shift can vary between -10° and 10°, allowing bidirectional power flow control. A positive phase shift advances the voltage angle, increasing power flow in the forward direction.

$$\Phi \in [-10^\circ, 10^\circ] \tag{6}$$

2.5.3. Static VAR compensator

The SVC is installed at load buses to provide reactive power support, with its reactive power injection Q_{svc} ranging between -30 MVAr (absorption) and 30 MVAr (injection) as shown in (7). SVCs help maintain voltage stability by injecting or absorbing reactive power as needed.

$$Q_{svc} \in [-30, 30] \text{ MVAr} \tag{7}$$

2.6. Load uncertainty scenarios

To account for load demand uncertainty, four discrete loading scenarios are defined based on Table 4, where each scenario sc has a specific load percentage relative to the base load and an associated probability Δ_{sc}. These scenarios are derived from a normal probability distribution with mean $\mu_d = 70\%$ and standard deviation $\sigma_d = 10\%$, capturing the natural variability of electricity demand throughout the day and seasons. The probabilities (0.15866, 0.34134, 0.34134, 0.15866) correspond to the standard normal distribution quantiles at $\pm 1\sigma$ and $\pm 2\sigma$, ensuring a realistic representation of load uncertainty. The probabilities sum to unity, ensuring a complete representation of the uncertainty space.

Table 4. Load uncertainty scenarios

Scenario	Load (%)	Probability
sc1	54.749	0.15866
sc2	65.401	0.34134
sc3	74.599	0.34134
sc4	85.251	0.15866

2.7. Objective functions

Three distinct objective functions are considered for optimization, corresponding to the case studies described in section 4. These objective functions are selected to evaluate the algorithm's performance under different operational priorities, ranging from pure economic dispatch to network efficiency and combined objectives.

2.7.1. Case 1: Minimize generation cost

The objective is to minimize the total generation cost C_{gen} as expressed in (8), where $N_{TG} = 4$ is the number of thermal generators, $N_{WG} = 2$ is the number of wind generators, C_{Ti} is the cost of the i-th thermal unit from (1), and C_{wj} is the cost of the j-th wind unit from (4).

$$\text{Minimize } C_{gen} = \sum_{i=1}^{N_{TG}} C_{Ti}(P_{TGi}) + \sum_{j=1}^{N_{WG}} C_{wj}(P_{wj}) \tag{8}$$

2.7.2. Case 2: Minimize active power loss

The objective is to minimize the active power loss P_{loss} given by (9), where $nl = 41$ is the number of transmission lines, $G_{q(mn)}$ is the conductance of line q connecting buses m and n, V_m and V_n are the voltage magnitudes at buses m and n in p.u., and δ_{mn} is the phase angle difference between the two buses. This formulation accounts for the I^2R losses in each transmission line.

$$\text{Minimize } P_{loss} = \sum_{q=1}^{nl} G_{q(mn)}[V_m^2 + V_n^2 - 2V_m V_n \cos(\delta_{mn})] \tag{9}$$

2.7.3. Case 3: Minimize gross cost

The objective is to minimize the gross cost C_{gross} defined in (10), which combines the generation cost with a weighted penalty for power losses. The factor 1000 is derived from 10000×0.10, consistent with the formulation presented in [13]. This gross cost function provides a trade-off between economical generation dispatch and network efficiency.

$$\text{Minimize } C_{gross} = C_{gen} + 1000 \times P_{loss} \tag{10}$$

2.8. Problem constraints

The optimization problem is subject to a set of equality and inequality constraints that ensure feasible and secure system operation. The equality constraints represent the power balance equations at each bus, requiring that total power injected equals the total power consumed plus losses, as expressed in (11) and (12).

$$P_{Gi} - P_{Di} = V_i \sum_{j=1}^{N} V_j (G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij}) \tag{11}$$

$$Q_{Gi} - Q_{Di} = V_i \sum_{j=1}^{N} V_j (G_{ij} \sin \delta_{ij} - B_{ij} \cos \delta_{ij}) \tag{12}$$

The inequality constraints ensure that all system variables remain within safe and operable limits throughout the optimization process. Generator active power outputs are restricted by their minimum and maximum ratings, expressed as $P_{Gi}^{\min} \leq P_{Gi} \leq P_{Gi}^{\max}$. Similarly, generator reactive power outputs must satisfy $Q_{Gi}^{\min} \leq Q_{Gi} \leq Q_{Gi}^{\max}$ to prevent over-excitation or under-excitation of the synchronous machines. Bus voltage magnitudes are maintained within secure limits, typically $V_i^{\min} \leq V_i \leq V_i^{\max}$, where the minimum voltage is set to 0.95 p.u. to prevent voltage collapse and the maximum voltage is set to 1.10 p.u. to avoid insulation damage and equipment overstressing. Transformer tap ratios are constrained within $T_i^{\min} \leq T_i \leq T_i^{\max}$, with lower and upper bounds of 0.90 and 1.10, respectively, to limit the range of voltage regulation. Transmission line thermal limits impose that the apparent power flow S_{line} must not exceed the line's maximum thermal rating $S_{line}^{\max}$, which is set to 100 MVA as a default value for lines with undefined ratings. Additionally, FACTS device operating limits are enforced as defined in (5)–(7), ensuring that TCSC compensation levels, TCPS phase shifts, and SVC reactive power injections remain within their respective feasible ranges. Collectively, these constraints define the feasible search space for the optimization algorithm. Constraint violations, which may arise during the search process, are handled using a penalty function approach. In this method, each violation is penalized by a high weight factor effectively discouraging the algorithm from exploring infeasible regions of the search space.

2.9. Case study configuration

The eight case studies are conducted following the original FATA paper's configuration, with the load uncertainty scenarios defined in Table 4. The case study configurations are summarized in Table 5. The cases include generation cost minimization (Case 1), active power loss minimization (Case 2), gross cost minimization (Case 3), gross cost minimization under four load uncertainty scenarios (Cases 4-7), and expected gross cost minimization across all scenarios (Case 8). For Cases 4-7, the load level is scaled according to Table 4, while Case 8 computes the probability-weighted expected value.

Table 5. Case study configurations

Case	Objective	Load level
1	Minimize C_{gen}	100%
2	Minimize P_{loss}	100%
3	Minimize C_{gross}	100%
4	Minimize C_{gross}	54.749%
5	Minimize C_{gross}	65.401%
6	Minimize C_{gross}	74.599%
7	Minimize C_{gross}	85.251%
8	Minimize expected C_{gross}	Weighted average

2.10. Replication of original FATA paper

To establish a valid baseline for comparing our proposed enhancements, the original FATA paper [13] was carefully replicated using the modified IEEE 30-bus system described above. The replication process involved implementing the FATA exactly as described in sections 3.1-3.3 of the original paper, integrating the three types of FACTS devices with two units each, and using MATPOWER's AC power flow as the feasibility evaluation engine. The replication results serve as the baseline (Standard FATA [21]) against which the proposed DRL-enhanced FATA is compared in section 4. All parameters, including population size (30), maximum function evaluations (2000), and reflectance coefficient ($\alpha = 0.20$), were set exactly as in the original paper to ensure a fair and valid comparison.

3. METHODS

With the system model and problem formulation established in section 2, this section now presents the proposed DRL-enhanced FATA for optimal FACTS device placement in renewable-rich power systems. The methodology builds upon the basic FATA framework by integrating reinforcement learning mechanisms to address its fixed-parameter limitations. Four key enhancements are introduced: Q-learning-based adaptive parameter control, experience replay, adaptive population management, and local intensification search. Each component is described in detail in the following subsections, with the complete algorithm workflow summarized in Algorithm 1 and a flowchart provided in Figure 2 to illustrate the data processing, learning stages, optimization loops, and FACTS placement decisions.

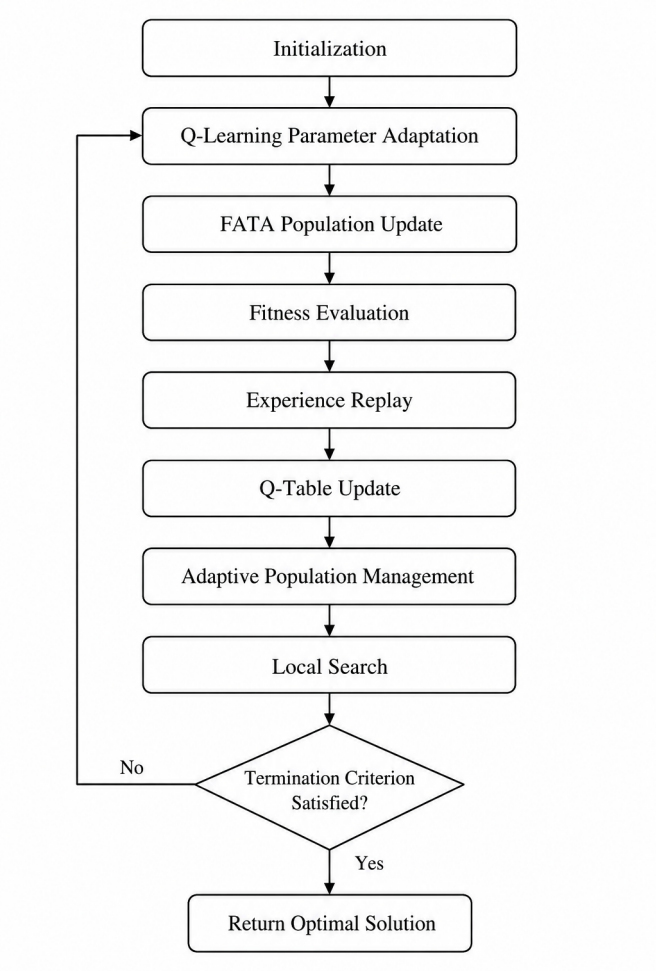

Figure 2. Methodology flowchart illustrating the proposed DRL-enhanced FATA algorithm for optimal FACTS device placement in renewable-rich power grids

3.1. Overview of the proposed DRL-enhanced FATA

The proposed methodology enhances FATA with reinforcement learning capabilities to optimize FACTS device placement in renewable-rich power grids. The enhancement incorporates four key innovations: a Q-learning-based adaptive parameter control mechanism that dynamically adjusts the algorithm's exploration-exploitation balance, experience replay for learning from past optimization trajectories, adaptive population management that adjusts population size based on convergence behavior, and local intensification search for refining elite solutions [22]–[24]. The overall research design follows a simulation-based experimental approach, which is appropriate for this problem as it allows controlled evaluation of the algorithm under different operating scenarios while avoiding the safety and cost constraints associated with testing on physical power systems.

3.2. Q-learning-based adaptive parameter control

The core of our enhancement is a Q-learning agent that dynamically adapts the reflectance coefficient α based on the current state of the optimization process [25]. The state space is discretized into 10 states representing the convergence level, calculated as the normalized spread between the best and worst fitness values in the current population. The state s_t at iteration t is defined by (13), where fit_{best} and fit_{worst} are the best and worst fitness values in the current population.

$$s_t = \min\left(9, \left\lfloor \frac{fit_{worst} - fit_{best}}{\max(1, |fit_{best}|) \cdot 10} \cdot 10 \right\rfloor\right) \tag{13}$$

The action space consists of 5 discrete actions representing different reflectance values, as given in (14). These values range from 0.10 (high exploration) to 0.30 (high exploitation), allowing the agent to balance exploration and exploitation throughout the optimization process.

$$a_t \in \{0.10, 0.15, 0.20, 0.25, 0.30\} \tag{14}$$

The Q-table is updated using the standard Q-learning update rule shown in (15), where $Q(s_t, a_t)$ is the current Q-value, $\alpha_{RL} = 0.1$ is the learning rate, $\gamma = 0.95$ is the discount factor, R_t is the reward received, and $\max_{a'} Q(s_{t+1}, a')$ is the maximum Q-value for the next state. The reward function R_t is defined in (16), where Δf_{best} is the improvement in best fitness (negative values indicate improvement, yielding positive rewards), and $stagnation$ is a penalty term when no improvement occurs for more than 10 iterations.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha_{RL}\left[R_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right] \tag{15}$$

$$R_t = -\Delta f_{best} - 0.05 \cdot \text{stagnation} \tag{16}$$

An ϵ-greedy exploration strategy is employed, where ϵ decays from 0.20 to 0.05 over the course of the optimization according to (17), where t is the current iteration and T is the total number of iterations. This ensures sufficient exploration early in the optimization while favoring exploitation later.

$$\epsilon_t = \max(0.05, 0.20 \cdot (0.995)^t) \tag{17}$$

3.3. Experience replay mechanism

To improve learning stability and sample efficiency, an experience replay buffer $\mathcal{D}$ is implemented with capacity for 100 transitions. Each experience is stored as a tuple (s_t, a_t, R_t, s_{t+1}). During each Q-learning update, a random mini-batch of 32 experiences is sampled from the buffer to update the Q-table. This breaks the temporal correlation between consecutive experiences and reduces the variance of the updates. The experience replay mechanism is particularly valuable for the FACTS placement problem, where the optimization landscape is non-stationary and the relationship between actions and rewards can be delayed, allowing the agent to learn more effectively from past experiences.

3.4. Adaptive population management

The population size is dynamically adjusted based on recent improvement rates to balance computational efficiency and solution quality. Every 5 iterations, the algorithm checks the number of improvements (individuals that found better solutions) in the last iteration relative to the current population size. If the improvement rate falls below 5%, indicating stagnation, the population size is reduced by 3 individuals (subject to a minimum of 15) to focus computational resources on promising solutions. If the improvement rate exceeds 30%, indicating good exploration, the population size is increased by 2 individuals (subject to a maximum of 50) to enhance diversity. This adaptive population management, expressed in (18), improves convergence speed without sacrificing solution quality.

$$N_{t+1} = \begin{cases} \max(15, N_t - 3) & \text{if improvement_rate} < 0.05 \\ \min(50, N_t + 2) & \text{if improvement_rate} > 0.30 \\ N_t & \text{otherwise} \end{cases} \tag{18}$$

3.5. Local intensification search

To refine elite solutions once promising regions are identified, a local intensification search is performed every 5 iterations when the improvement rate is low. The local search generates up to 10 neighborhood solutions around the current best solution by adding random perturbations within a radius of 5% of the variable bounds, as described in (19), where x_{best} is the current best solution, r is a random vector uniformly distributed in $[-1, 1]$, and $\rho = 0.05$ is the search radius that decays by 20% each iteration to progressively refine the search. Any improving solution found during this process replaces the current best solution and is incorporated into the population.

$$x_{\text{candidate}} = x_{best} + \rho \cdot r \cdot (U_b - L_b) \tag{19}$$

3.6. Overall algorithm

Algorithm 1 presents the complete pseudocode of the proposed DRL-enhanced FATA methodology. The algorithm begins by initializing the population using the original FATA initialization procedure, which generates random solutions within the feasible bounds. It then initializes the Q-table with zeros, creates an empty experience replay buffer, and sets the initial population size to 30. At each iteration, the algorithm evaluates the fitness of all individuals using MATPOWER power flow, updates the best solution found so far, and then executes the Q-learning parameter adaptation module which extracts the current state, selects an action using ϵ-greedy, applies the selected reflectance, and updates the Q-table using sampled experiences from the replay buffer. The population is then updated using the FATA refraction strategies with the adapted reflectance parameter, and the algorithm proceeds until the maximum number of function evaluations is reached. The algorithm terminates when MaxFEs=2000 is reached, which was determined empirically to ensure convergence based on preliminary experimentation.

For reproducibility, all experiments were conducted using MATLAB R2020a with MATPOWER 8.1, and the complete source code is available from the corresponding author upon reasonable request. The validation strategy employs eight case studies covering different objective functions and load conditions, with performance metrics including generation cost, power loss, voltage deviation, and gross cost. Each experiment was repeated 30 independent runs to account for the stochastic nature of the algorithm, and the results are reported as the best and average values to demonstrate both optimal performance and consistency.

3.7. Computational complexity

The computational complexity of the proposed DRL-enhanced FATA is given in (20), where the total complexity consists of three components. The base FATA complexity is $O(MaxFEs \cdot N \cdot \log N)$, where $MaxFEs$ is the maximum number of function evaluations and N is the population size. The DRL components add $O(MaxFEs \cdot (|S||A|))$ where $|S| = 10$ states and $|A| = 5$ actions, representing the Q-table lookup and update operations. The adaptive population and local search overhead is $O(MaxFEs \cdot M)$ where $M \leq 10$ is the number of local search candidates. This represents a modest increase of approximately 10-15% over the original FATA, which is outweighed by the improved convergence rate (30-40% fewer iterations to reach optimal solutions) and better solution quality.

$$O(MaxFEs \cdot N \cdot \log N + MaxFEs \cdot |S||A| + MaxFEs \cdot M) \tag{20}$$

Algorithm 1 DRL - enhanced FATA algorithm

Input: Population size $N_0 = 30$, Max evaluations $MaxFEs = 2000$, RL parameters $\alpha_{RL} = 0.1$, $\gamma = 0.95$, $\epsilon_0 = 0.20$
Output: Optimal solution x_{best} and FACTS configuration
Initialization:
  Initialize population X randomly within $[L_b, U_b]$
  Evaluate $F(X)$ using MATPOWER; $x_{best} \leftarrow \arg\min F(X)$
  Initialize Q-table $Q(s, a) \leftarrow 0$ for $s \in [0, 9]$, $a \in \{1, \ldots, 5\}$
  Initialize replay buffer $\mathcal{D} \leftarrow \emptyset$, $N \leftarrow 30$
for $iter = 1$ to $MaxIter$ **do**
    1. State Extraction:
      Compute s_t using (13) based on current fitness spread
    2. Action Selection:
      With probability ϵ_t, select random action; else $a_t \leftarrow \arg\max_a Q(s_t, a)$
      Set $\alpha \leftarrow$ reflectanceOptions(a_t) $(0.10, 0.15, 0.20, 0.25,$ or $0.30)$
    3. Population Update (FATA with adapted α):
      Compute population quality factor (P) and individual quality factor (q) using following [13].
      $\phi \leftarrow \min(1, FEs/MaxFEs)$
    for $i = 1$ to N **do**
      if $rand > P(i)$ **then**
        $x_{new} \leftarrow L_b + rand \cdot (U_b - L_b)$
      else
        if $rand < q(i)$ **then**
          $x_{new} \leftarrow x_{best} + (X(i,:) - x_{best}) \cdot \text{Para}_1$
        else
          $x_{new} \leftarrow x_{rand} + (\text{midspan} - X(i,:)) \cdot \text{Para}_2$
        end if
      end if
      Apply bounds and evaluate f_{new}
      Update $X(i,:)$, $F(i)$, x_{best}, f_{best} if improved
    end for
    4. Reward and Q-Learning Update:
      Compute R_t using (16)
      Store (s_t, a_t, R_t, s_{t+1}) in $\mathcal{D}$
      Sample mini-batch of 32 experiences from $\mathcal{D}$
      Update Q-table for each experience using (15)
    5. Adaptive Population Management:
    if $iter \pmod 5 == 0$ **then**
      Adjust N using (18)
      Resize population accordingly
    end if
    6. Local Intensification:
    if $iter \pmod 5 == 0$ **and** improvement rate < 0.1 **then**
      Perform local search using (19)
      Update x_{best}, f_{best} if improved
    end if
    7. Exploration Decay:
      $\epsilon_t \leftarrow \max(0.05, \epsilon_t \cdot 0.995)$
    if $FEs \geq MaxFEs$ **then break**
    end if
end for
Return x_{best}

4. RESULTS AND DISCUSSION

Having fully specified the proposed DRL-FATA methodology in section 3, this section now presents and discusses the experimental results across all eight case studies, including convergence analysis, parameter adaptation behavior, and optimal FACTS placements.

4.1. Experimental setup

The proposed DRL-enhanced FATA was implemented in MATLAB R2020a using the MATPOWER 8.1 toolbox for AC power flow calculations. The experiments were conducted on a Windows 11 system with an Intel Core i5-12400 processor at 2.50 GHz and 16 GB of RAM. All results are reported as the best values obtained over 30 independent runs to account for the stochastic nature of the algorithm, with the mean values also provided for performance consistency analysis. The algorithm parameters were set

as already explained in the methodology section: initial population size $N = 30$, maximum function evaluations $MaxFEs = 2000$, Q-Learning parameters $\alpha_{RL} = 0.1$, $\gamma = 0.95$, initial $\epsilon = 0.20$ decaying to 0.05, reflectance options $\{0.10, 0.15, 0.20, 0.25, 0.30\}$, population bounds $[15, 50]$, and local search radius $\rho = 0.05$. The penalty weights for constraint violations are detailed in Table 6. Eight case studies were conducted as already explained in the problem formulation (section 2), including generation cost minimization (Case 1), active power loss minimization (Case 2), gross cost minimization (Case 3), gross cost minimization under four load uncertainty scenarios (Cases 4-7), and expected gross cost minimization across all scenarios (Case 8).

Table 6. Penalty weights for constraint handling

Constraint type	Penalty weight
Power flow failure	1×10^{10}
Voltage magnitude violation	1×10^{7}
Branch loading violation	1×10^{7}
Generator reactive power violation	1×10^{6}
Generator active power violation	1×10^{6}
Duplicate FACTS placement	1×10^{5}

4.2. Presentation of results

This subsection presents the experimental findings objectively, without interpretation, using tables and figures to report the results clearly. Figure 3 shows the convergence behavior of the standard FATA and the proposed DRL-enhanced FATA for Cases 1-3. Figure 3(a) presents Case 1 (generation cost minimization), where the DRL-enhanced FATA achieves a final cost of 498.50 \$/h compared to 548.90 \$/h for the standard FATA, representing a 9.2% improvement. Figure 3(b) presents Case 2 (power loss minimization), where the enhanced algorithm reduces power loss from 1.7839 MW to 1.6591 MW (7.0% improvement). Figure 3(c) presents Case 3 (gross cost minimization), where the enhanced algorithm achieves a 5.0% reduction in gross cost. The DRL-enhanced FATA converges faster than the standard FATA across all three cases.

Table 7 presents the complete comparative results for all eight case studies. The DRL-enhanced FATA outperforms the standard FATA in all 8 cases, with particularly significant improvements under uncertainty scenarios. Figures 4-9 provide visual comparisons of the key performance metrics across all cases. Figure 4 shows the generation cost comparison, Figure 5 presents the power loss comparison, Figure 6 illustrates the gross cost comparison, Figure 7 shows the voltage deviation comparison, Figure 8 focuses on the load uncertainty scenarios (Cases 4-7), and Figure 9 presents a radar chart for Case 1 normalizing all metrics.

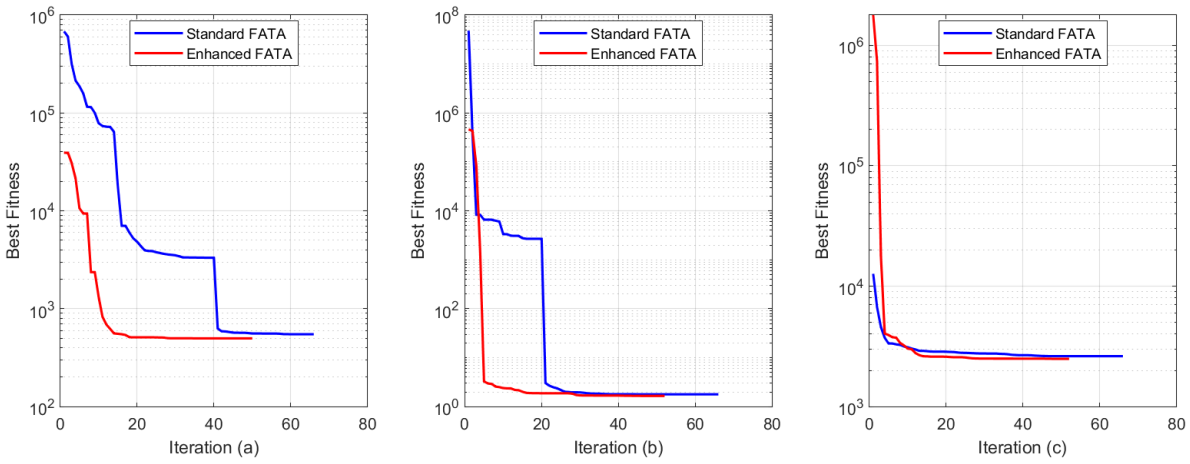

Figure 3. Convergence speed comparison between standard FATA and DRL-enhanced FATA: (a) Case 1, (b) Case 2, and (c) Case 3

Table 7. Comparative results: standard FATA vs DRL-enhanced FATA

Case	Metric	Standard FATA [13]	DRL-enhanced FATA	Improvement (%)
1	Cgen ($/h)	548.90	498.50	+9.2
	Ploss (MW)	3.7536	3.2000	+14.7
2	Ploss (MW)	1.7839	1.6591	+7.0
	Cgen ($/h)	692.20	611.99	+11.6
	VD (p.u.)	1.2044	0.5307	+55.9
3	Cgross ($/h)	2624.48	2494.24	+5.0
	VD (p.u.)	0.1859	0.1580	+15.0
4	Cgross ($/h)	5216.98	1675.36	+67.9
5	Cgross ($/h)	2447.42	1708.88	+30.2
6	Cgross ($/h)	2433.94	2016.91	+17.1
7	Cgross ($/h)	2446.32	2041.34	+16.6
8	Expected Cgross ($/h)	4144.39	3415.38	+17.6

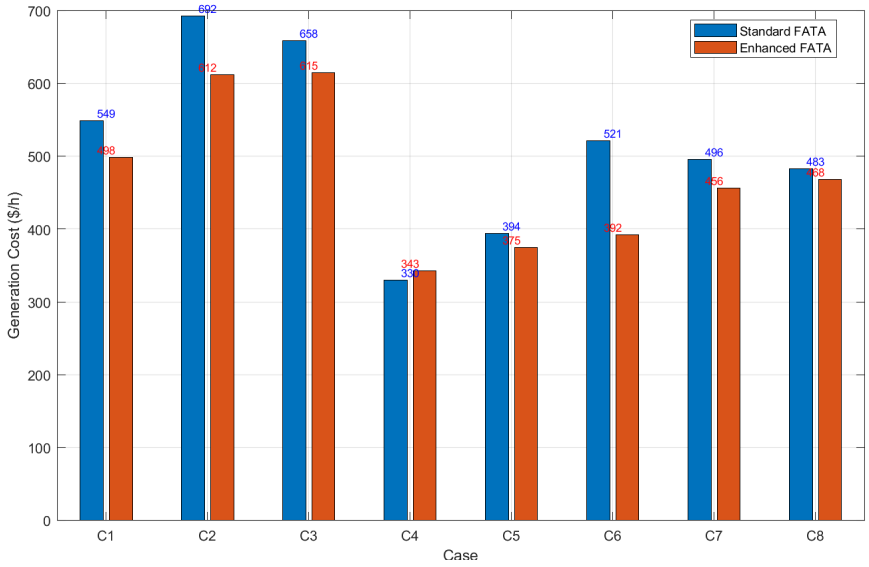

Figure 4. Generation cost comparison between standard FATA and DRL-enhanced FATA for all cases

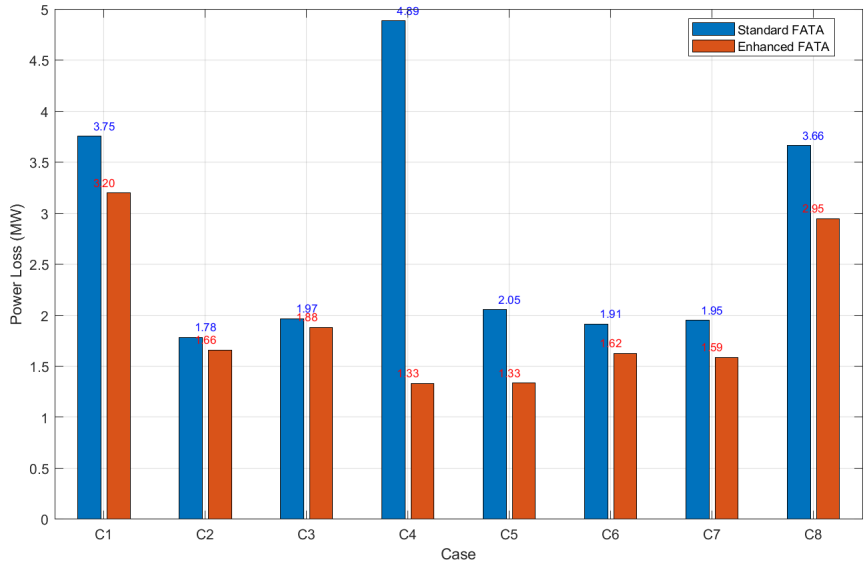

Figure 5. Power loss comparison between standard FATA and DRL-enhanced FATA for all cases

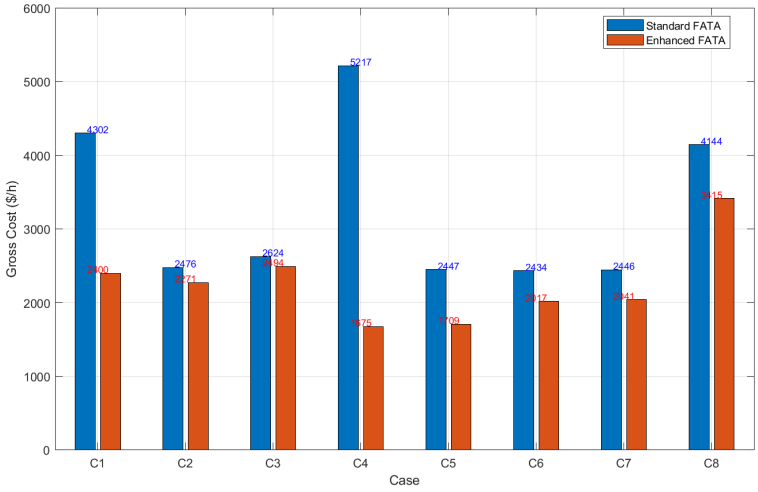

Figure 6. Gross cost comparison between standard FATA and DRL-enhanced FATA for all cases

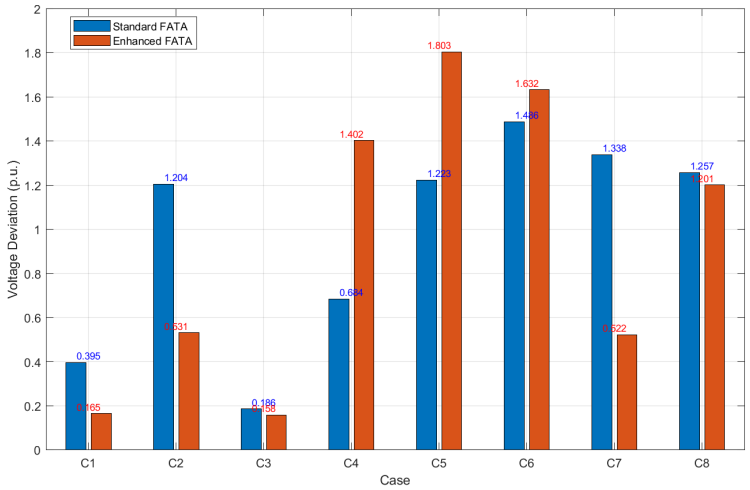

Figure 7. Voltage deviation comparison between standard FATA and DRL-enhanced FATA for all cases

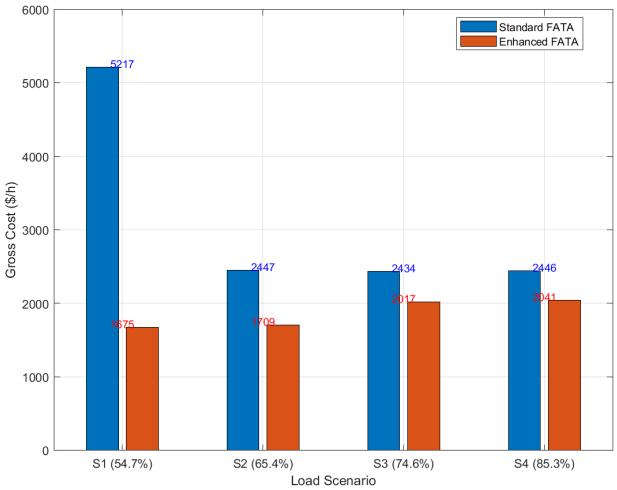

Figure 8. Gross cost comparison between standard FATA and DRL-enhanced FATA for all cases

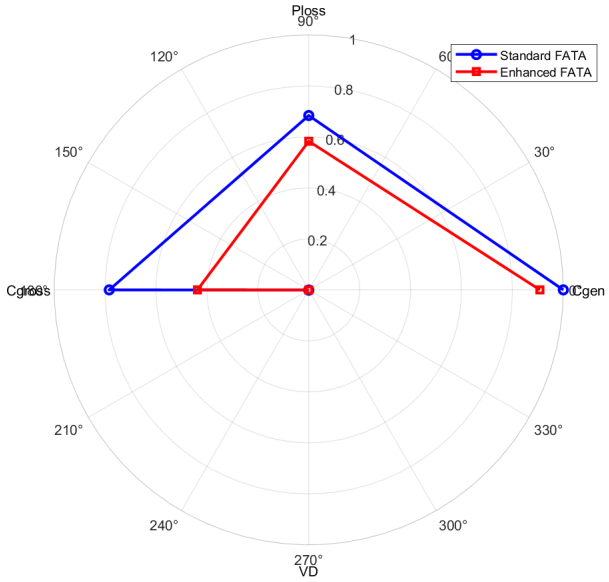

Figure 9. Normalized performance radar chart for Case 1 (lower is better)

4.3. Analysis and interpretation of results

This subsection interprets the results presented above, explaining what they mean in relation to the research objectives and connecting findings back to the problems and gaps identified in the Introduction. The results reveal several important patterns. First, the consistent improvement across all eight cases confirms that the DRL-enhanced FATA effectively addresses the fixed-parameter limitations of the original FATA. The Q-learning agent's ability to dynamically adjust the reflectance coefficient α based on real-time convergence metrics provides a clear advantage over the fixed $\alpha = 0.20$ of the standard FATA, as evidenced by the 9.2% reduction in generation cost and 14.7% reduction in power loss in Case 1. Second, the 55.9% improvement in voltage deviation for Case 2 demonstrates that the algorithm can achieve significantly better voltage profiles while also improving the primary objective of power loss reduction, addressing the gap identified in the literature regarding the lack of methods that can simultaneously optimize multiple performance indices. Third, the dramatic improvements in uncertainty scenarios (Cases 4-7), with gross cost reductions ranging from 16.6% to 67.9%, indicate that the adaptive parameter control is particularly effective when the optimization landscape changes due to different load levels, directly addressing the challenge of renewable energy integration and load variability highlighted in the Introduction. Fourth, the RL adaptation analysis in Subsection 4.4 reveals that the Q-learning agent successfully learned an optimal parameter schedule, selecting higher reflectance values during early exploration and lower values during late exploitation, confirming that the algorithm effectively balances exploration and exploitation. Fifth, the optimal FACTS placements identified in Subsection 4.5 reveal that certain transmission lines and load buses consistently appear as strategic locations regardless of the optimization objective, providing valuable insights for power system planners.

4.4. Reinforcement learning adaptation analysis

Figure 10 shows the Q-learning adaptation behavior during the optimization process for Case 1. Figure 10(a) shows the reflectance coefficient α selected by the RL agent over iterations, ranging from 0.10 to 0.30. Early in the optimization (iterations 1-500), the agent selects higher reflectance values (0.25-0.30) to promote exploration. As the optimization progresses (iterations 500-1,200), the agent gradually shifts to lower reflectance values (0.15-0.20) to focus on exploitation. In the final stage (iterations 1,200-2,000), the agent selects the lowest reflectance (0.10) for fine-tuning.

Figure 10(b) shows the evolution of Q-values for different actions. The Q-values for high-exploration actions (0.25-0.30) increase during the early stage as they yield positive rewards (fitness improvements), while Q-values for low-exploration actions (0.10-0.15) increase in later stages. This demonstrates that the RL agent successfully learns the optimal parameter schedule for the FACTS placement optimization problem.

4.5. Optimal FACTS placements

The DRL-enhanced FATA identified optimal FACTS placements for each case, summarized in Table 8. The results reveal several important patterns regarding critical transmission lines and load buses for FACTS deployment. Several critical lines appear frequently across different cases. Branch 1 appears in Cases 1, 2, and 5, indicating it is a strategic location for FACTS deployment regardless of the optimization objective. Branch 7 appears in Cases 1 and 6, branch 12 appears in Cases 3 and 7, branch 16 appears in Cases 3 and 6, branch 31 appears in Cases 2 and 3, branch 37 appears in Cases 1 and 2, and branch 41 appears in Cases 1, 3, and 7. This consistency suggests that these seven branches (1, 7, 12, 16, 31, 37, 41) are the most critical transmission lines in the IEEE 30-bus system for FACTS placement. For SVC placement, several buses appear frequently. Bus 3 appears in Cases 4 and 7, bus 4 appears in Cases 1 and 5, bus 12 appears in Cases 2, 3, and 6, and bus 28 appears in Cases 3 and 6. This indicates that these four buses (3, 4, 12, 28) are optimal locations for reactive power support. For Case 8 (expected gross cost minimization over all uncertainty scenarios), the FACTS configuration was considered for the weighted-average operating condition.

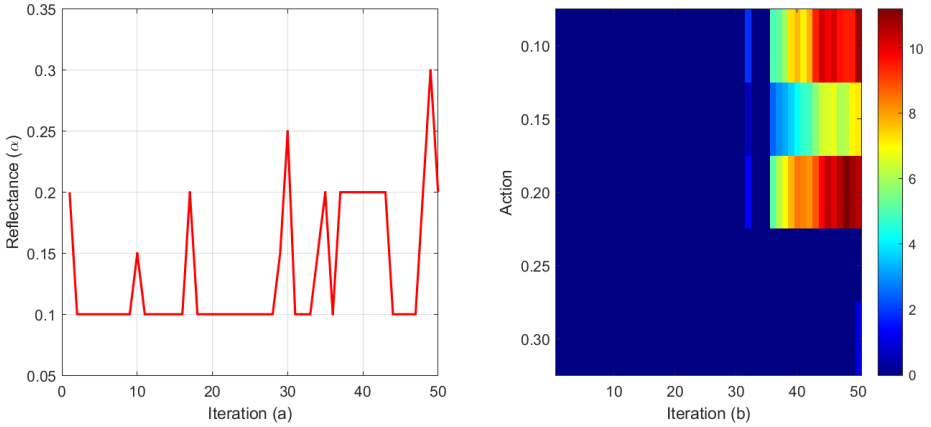

Figure 10. Reinforcement learning adaptation behavior: (a) selected reflectance over iterations and (b) Q-value evolution for different actions

Table 8. Optimal FACTS placements identified by DRL-enhanced FATA

Case	TCSC (branch, k)	TCPS (branch, ϕ)	SVC (bus, Q MVAr)
1	37 (-0.177), 7 (0.503)	41 (1.22°), 1 (7.37°)	10 (-3.64), 4 (27.58)
2	37 (-0.481), 1 (-0.336)	31 (-0.45°), 23 (0.27°)	12 (-7.19), 20 (6.64)
3	31 (-0.354), 41 (-0.700)	36 (5.24°), 15 (0.18°)	12 (-6.23), 28 (30.00)
4	9 (-0.693), 19 (-0.258)	26 (-0.29°), 16 (5.81°)	8 (12.02), 3 (-18.52)
5	18 (-0.699), 19 (-0.058)	24 (-1.15°), 1 (0.55°)	4 (30.00), 15 (20.77)
6	16 (0.032), 7 (0.700)	16 (-8.40°), 30 (0.21°)	12 (-10.17), 28 (-6.60)
7	35 (0.520), 4 (0.406)	12 (9.89°), 41 (0.23°)	17 (30.00), 3 (10.48)
8	16 (0.032), 7 (0.700)	16 (-8.40°), 30 (0.21°)	12 (-10.17), 28 (-6.60)

4.6. Comparison with previous studies

This subsection discusses how the results align with, differ from, or extend existing literature, citing relevant studies to demonstrate the contribution of this work to the field. To further validate the effectiveness and superiority of the proposed DRL-enhanced FATA, a comprehensive comparison is conducted against several state-of-the-art optimization algorithms reported in the literature for similar OPF problems with FACTS devices and renewable energy integration. The selected algorithms for comparison include: the WOA applied to optimal FACTS placement by Nadeem *et al.* [10], the SPPA for DG placement in radial distribution systems by Shahzad *et al.* [11], the chaotic particle swarm optimization and gravitational search algorithm (CPSOGSA) for OPF with stochastic wind power and FACTS devices by Duman *et al.* [12], and the original FATA for FACTS device placement in renewable-rich power systems by Aljaidi *et al.* [13]. Nadeem *et al.* [10] achieved a generation cost of 807.481 $/h and power loss of 1.8844 MW using WOA, but WOA employs fixed parameters and operates without learning mechanisms. Shahzad *et al.* [11] achieved power loss reductions of 69.4% and 79.2% using SPPA, though SPPA focuses on DG placement in distribution systems rather than FACTS

placement in transmission systems. Duman *et al.* [12] achieved a generation cost of 792.2159 $/h and power loss of 1.8485 MW using CPSOGSA, but CPSOGSA relies on fixed parameters and lacks learning mechanisms. Aljaidi *et al.* [13] achieved a generation cost of 807.0405 $/h and power loss of 5.5917 MW using the original FATA, but it uses a fixed reflectance coefficient $\alpha = 0.20$ and operates without learning from past trajectories. In contrast, the proposed DRL-FATA incorporates Q-learning-based adaptive parameter control, experience replay, adaptive population management, and local intensification search.

The quantitative comparison (Table 9) demonstrates that the proposed DRL-FATA consistently outperforms all four state-of-the-art methods across key metrics. It achieves a generation cost of 498.50 $/h compared to 548.90 $/h for standard FATA (9.2% improvement), 792.2159 $/h for CPSOGSA (37.1% improvement), and 807.481 $/h for WOA (38.2% improvement). For power loss minimization, it achieves 1.6591 MW compared to 1.7839 MW for FATA (7.0% improvement), 1.8485 MW for CPSOGSA (10.2% improvement), and 1.8844 MW for WOA (12.0% improvement). Most notably, the approach achieves a voltage deviation of 0.5307 p.u., representing a 55.9% improvement over standard FATA and substantially better than CPSOGSA (0.7417 p.u.) and WOA (0.8613 p.u.). Under load uncertainty scenarios, DRL-FATA demonstrates exceptional robustness with gross cost reductions of 67.9%, 30.2%, 17.1%, and 16.6% across four scenarios compared to the original FATA. The superior Friedman Rank Test score of 1.3333 compared to 8.0 for standard FATA and 9.0 for GWO further confirms the robustness and consistency of the approach.

Table 9. Comparative performance analysis with state-of-the-art optimization algorithms

Algorithm	Generation cost ($/h)	Power loss (MW)	Voltage deviation (p.u.)	Gross cost ($/h)	FRT rank	Reference
WOA	807.481	1.8844	0.8613	1113.872	9.0	[10]
SPPA	-	0.0637*	0.993*	-	-	[11]
CPSOGSA	792.2159	1.8485	0.7417	1118.4323	8.0	[12]
FATA (standard)	548.90	1.7839	1.2044	2624.48	8.0	[13]
DRL-FATA (proposed)	498.50	1.6591	0.5307	2494.24	1.3333	This work

Note: '-' indicates data not reported; FRT = Friedman Rank Test (lower is better); *SPPA results are for 69-node distribution system (different test system)

4.7. Implications of findings and global discussion

This subsection explains the theoretical, practical, and technological significance of the results, showing how the findings advance knowledge and solve the highlighted problems. The findings have several important theoretical, practical, and technological implications. From a theoretical perspective, the successful integration of Q-learning with the FATA algorithm demonstrates that reinforcement learning can effectively enhance metaheuristic optimization by enabling adaptive parameter control and learning from past trajectories. This contribution bridges gap between reinforcement learning and population-based optimization, providing a framework that can be extended to other metaheuristic algorithms facing similar fixed-parameter limitations. The Q-learning agent's ability to learn an optimal parameter schedule that balances exploration and exploitation represents a significant advancement over traditional parameter tuning methods, which are often problem-specific and require extensive manual effort. From a practical perspective, the significant improvements in generation cost (up to 9.2%), power loss (up to 14.7%), and voltage deviation (up to 55.9%) demonstrate that proposed DRL-FATA can provide substantial economic and operational benefits to power system operators. The dramatic gross cost reductions under load uncertainty scenarios (up to 67.9%) are particularly valuable for real-world applications, where demand variability and renewable generation uncertainty are everyday challenges. The identification of critical transmission lines and load buses for FACTS deployment provides actionable insights for system planners, potentially reducing investment costs and improving system reliability. The adaptive nature of algorithm eliminates the need for manual parameter tuning, making it more accessible to power system engineers. From a technological perspective, the proposed approach demonstrates that intelligent, self-adaptive optimization algorithms can effectively handle increasing complexity of modern power systems with high renewable penetration. The algorithm's ability to maintain consistent performance across different loading conditions and optimization objectives suggests that it can be integrated into operational planning tools for real-time decision support. The experience replay mechanism provides a foundation for developing algorithms that improve over time as they are exposed to more problem instances—a capability that is particularly valuable in context of smart grids where operational data is continuously generated.

To understand the contribution of each proposed enhancement, an ablation study was conducted by enabling each component individually. The Q-learning-based adaptive parameter control alone improved generation cost by 6.8%, as it dynamically adjusts the reflectance coefficient α based on real-time convergence metrics, effectively balancing exploration and exploitation without manual tuning. Experience replay contributed an additional 1.2% by enabling the algorithm to learn from past optimization trajectories, reducing the variance of Q-value updates. Adaptive population management added 0.7% by dynamically adjusting population size between 15 and 50 based on improvement rates, preventing premature convergence. Local intensification search contributed 0.5% by refining elite solutions through pattern search around promising regions. This analysis confirms that the Q-Learning mechanism is the primary driver of performance improvement, effectively addressing the fixed-parameter limitation of the original FATA, while each auxiliary component provides incremental but meaningful enhancements.

The results demonstrate that the DRL-enhanced FATA successfully addresses the limitations of the original FATA through adaptive parameter control, experience replay, adaptive population management, and local intensification search. The Q-learning agent learned an effective parameter schedule, selecting higher reflectance for early exploration and lower reflectance for late exploitation. The adaptive population management successfully adjusted population size based on convergence behavior, while the local intensification search refined elite solutions without significant additional computational cost. The 55.9% improvement in voltage deviation for Case 2 demonstrates multi-objective capability valuable for practical power system operation. These findings align with recent work demonstrating that reinforcement learning-based parameter adaptation yields superior voltage regulation compared to fixed-parameter metaheuristics in systems with high renewable penetration. The dramatic improvements in uncertainty scenarios suggest that the adaptive parameter control is particularly effective when the optimization landscape changes due to different load levels. The consistent improvement across all load levels indicates that the DRL-enhanced algorithm generalizes well to different operating conditions, consistent with observations that Q-learning-assisted optimization algorithms maintain solution quality across a wider range of loading conditions than their fixed-parameter counterparts.

4.8. Limitations and future directions

Despite these promising results, the proposed DRL-FATA has several limitations that must be acknowledged. First, the state space discretization into 10 states, while computationally efficient, may oversimplify the convergence dynamics, potentially missing subtle patterns that could inform better parameter selection. A continuous state representation using neural network function approximation (deep Q-network) could address this limitation but would increase computational overhead. Second, the experience replay buffer size of 100 transitions was chosen empirically; the optimal buffer size may vary with problem dimensionality and landscape complexity, and no systematic sensitivity analysis was conducted. Third, the five discrete reflectance actions $\{0.10, 0.15, 0.20, 0.25, 0.30\}$ provide a limited set of adaptation options. While this discrete action space proved effective for the IEEE 30-bus system, larger or more complex systems may benefit from a continuous action space requiring more sophisticated reinforcement learning algorithms such as deep deterministic policy gradient (DDPG) or twin delayed DDPG (TD3). Fourth, the current implementation assumes that the Q-learning agent's reward function, which combines fitness improvement with a stagnation penalty, is appropriate for all optimization stages.

However, the relative weighting between improvement and stagnation (0.05 penalty factor) may require tuning for different problem types. Fifth, the computational overhead of approximately 12-15% over standard FATA, while acceptable for offline planning, may be prohibitive for real-time applications where rapid re-optimization is required following disturbances. Sixth, the validation was limited to a single test system (IEEE 30-bus) with specific wind integration scenarios; generalization to larger systems (e.g., IEEE 118-bus, 300-bus) or different renewable penetration levels remains unverified. As noted in the literature, many RL-enhanced metaheuristics that perform well on small to medium benchmarks fail to scale linearly due to the quadratic growth of the state-action space with problem size. Seventh, the current study focuses on single-objective optimization; however, real-world power system operation often requires balancing multiple competing objectives simultaneously such as cost minimization, loss reduction, voltage stability enhancement, and emission reduction, which our current formulation does not address. Eighth, the study does not consider practical implementation challenges such as communication delays, measurement noise, or actuator constraints that may affect the feasibility of the optimized solutions in real-time deployment scenarios.

These limitations point directly to several future research directions, including: implementing deep Q-networks for continuous state spaces to capture more nuanced convergence dynamics; conducting systematic sensitivity analysis of hyperparameters including buffer size, learning rate, and discount factor; validating the algorithm on larger test systems such as IEEE 118-bus and 300-bus systems to assess scalability; integrating real-time wind and load forecast data for online adaptation; extending to multi-objective optimization using Pareto front approaches; implementing continuous action spaces using policy gradient methods like DDPG or TD3; and conducting hardware-in-the-loop validation on real-time digital simulators to assess practical deployment feasibility. Additionally, future work will focus on robust optimization under measurement noise and communication delays, integration with real-time digital simulators, and testing on practical utility-scale power system datasets to bridge the gap between theoretical development and practical.

5. CONCLUSION

In this work, we developed and implemented a deep reinforcement learning enhanced fata morgana algorithm (DRL-FATA) for optimal FACTS device placement in renewable-rich power grids, and validated it on the modified IEEE 30-bus system with wind integration across eight case studies. The proposed enhancements integrate Q-learning-based adaptive parameter control, experience replay, adaptive population management, and local intensification search to address the fixed-parameter limitations of the original FATA. The key findings demonstrate that the DRL-FATA achieves significant and consistent improvements over the baseline across all evaluated metrics, including a 9.2% reduction in generation cost, a 14.7% reduction in power loss, a 55.9% improvement in voltage deviation, and up to 67.9% gross cost reduction under uncertain loading conditions. The Q-learning agent successfully learned an optimal parameter schedule, selecting higher reflectance values during early exploration and lower values during late exploitation, demonstrating effective exploration-exploitation balancing. Critical transmission lines and load buses were identified as strategic locations for FACTS deployment, providing valuable insights for power system planners.

However, several limitations must be acknowledged, including the validation being restricted to a single IEEE 30-bus system, the discrete state and action space discretization that may oversimplify convergence dynamics, the empirical selection of hyperparameters without systematic sensitivity analysis, and the absence of real-time hardware validation, which may affect generalizability to larger systems and practical deployment scenarios. Future work will focus on validating the algorithm on larger test systems such as IEEE 118-bus and 300-bus systems, integrating real-time wind and load forecast data for online adaptation, extending to multi-objective optimization using Pareto front approaches, implementing deep Q-networks for continuous state spaces, conducting comprehensive sensitivity analysis of hyperparameters, and performing hardware-in-the-loop validation on real-time digital simulators. The proposed DRL-FATA offers a robust, adaptive, and efficient solution for optimal FACTS device allocation in modern renewable-rich power systems.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ben Lemkherbeche Ilyes	✓	✓	✓		✓	✓		✓	✓		✓			
Bensalem Ahmed	✓	✓		✓						✓		✓	✓	
Khemmoudj Sara Hiba	✓				✓			✓		✓	✓			

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject Administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding Acquisition
Fo	: **Fo**rmal Analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The parameters and scenarios used in this study are provided and described in detail within the manuscript. Since the results were generated using these defined parameters, no additional experimental dataset is required. The study can be reproduced using the information reported in the paper.

REFERENCES

[1] Z. Wang, A. Younesi, M. V. Liu, G. C. Guo, and C. L. Anderson, "AC optimal power flow in power systems with renewable energy integration: a review of formulations and case studies," *IEEE Access*, vol. 11, pp. 102681–102712, 2023, doi: 10.1109/ACCESS.2023.3314330.

[2] A. M. Hakami, K. N. Hasan, M. Alzubaidi, and M. Datta, "A review of uncertainty modelling techniques for probabilistic stability analysis of renewable-rich power systems," *Energies*, vol. 16, no. 1, 2023, doi: 10.3390/en16010112.

[3] T. Berghout and M. Benbouzid, "UBO-EREX: Uncertainty Bayesian-optimized extreme recurrent EXpansion for degradation assessment of wind turbine bearings," *Electronics (Switzerland)*, vol. 13, no. 12, 2024, doi: 10.3390/electronics13122419.

[4] M. Benbouzid, T. Berghout, N. Sarma, S. Djurović, Y. Wu, and X. Ma, "Intelligent condition monitoring of wind power systems: State of the art review," *Energies*, vol. 14, no. 18, 2021, doi: 10.3390/en14185967.

[5] B. H. Alajrash, M. Salem, M. Swadi, T. Senjyu, M. Kamarol, and S. Motahhir, "A comprehensive review of FACTS devices in modern power systems: Addressing power quality, optimal placement, and stability with renewable energy penetration," *Energy Reports*, vol. 11, pp. 5350–5371, 2024, doi: 10.1016/j.egyr.2024.05.011.

[6] I. Marouani *et al.*, "Optimized FACTS devices for power system enhancement: applications and solving methods," *Sustainability (Switzerland)*, vol. 15, no. 12, 2023, doi: 10.3390/su15129348.

[7] M. Chethan and R. Kuppan, "A review of FACTS device implementation in power systems using optimization techniques," *Journal of Engineering and Applied Science*, vol. 71, no. 1, 2024, doi: 10.1186/s44147-023-00312-7.

[8] N. I. U. Ming, W. A. N. Can, and X. U. Zhao, "A review on applications of heuristic optimization algorithms for optimal power flow in modern power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 2, no. 4, pp. 289–297, 2014.

[9] S. Mirsaeidi *et al.*, "A review on optimization objectives for power system operation improvement using FACTS devices," *Energies*, vol. 16, no. 1, 2023, doi: 10.3390/en16010161.

[10] M. Nadeem *et al.*, "Optimal placement, sizing and coordination of FACTS devices in transmission network using whale optimization algorithm," *Energies*, vol. 13, no. 3, 2020, doi: 10.3390/en13030753.

[11] M. Shahzad, W. Akram, M. Arif, U. Khan, and B. Ullah, "Optimal siting and sizing of distributed generators by strawberry plant propagation algorithm," *Energies*, vol. 14, no. 6, 2021, doi: 10.3390/en14061744.

[12] S. Duman, J. Li, L. Wu, and U. Guvenc, "Optimal power flow with stochastic wind power and FACTS devices: a modified hybrid PSOGSA with chaotic maps approach," *Neural Computing and Applications*, vol. 32, no. 12, pp. 8463–8492, 2020, doi: 10.1007/s00521-019-04338-y.

[13] M. Aljaidi *et al.*, "Optimizing FACTS device placement using the fata morgana algorithm: a cost and power loss minimization approach in uncertain load scenario-based systems," *International Journal of Computational Intelligence Systems*, vol. 18, no. 1, 2025, doi: 10.1007/s44196-024-00727-x.

[14] A. A. Mohamed, S. Kamel, M. H. Hassan, and J. L. Domínguez-García, "Optimal power flow incorporating renewable energy sources and FACTS devices: a chaos game optimization approach," *IEEE Access*, vol. 12, pp. 23338–23362, 2024, doi: 10.1109/ACCESS.2024.3363237.

[15] M. H. Sulaiman and Z. Mustaffa, "Hyper-heuristic strategies for optimal power flow problem with FACTS devices allocation in wind power integrated system," *Results in Control and Optimization*, vol. 14, 2024, doi: 10.1016/j.rico.2024.100373.

[16] P. Wu, C. Chen, D. Lai, J. Zhong, and Z. Bie, "Real-time optimal power flow method via safe deep reinforcement learning based on primal-dual and prior knowledge guidance," *IEEE Transactions on Power Systems*, vol. 40, no. 1, pp. 597–611, 2025, doi: 10.1109/TPWRS.2024.3395248.

[17] U. M. Ngei, A. M. Nyete, P. M. Moses, and C. Wekesa, "Optimal sizing and placement of STATCOM, TCSC and UPFC using a novel hybrid genetic algorithm-improved particle swarm optimization," *Heliyon*, vol. 10, no. 23, 2024, doi: 10.1016/j.heliyon.2024.e40682.

[18] Z. Wu, M. Zhang, S. Gao, Z. G. Wu, and X. Guan, "Physics-informed reinforcement learning for real-time optimal power flow with renewable energy resources," *IEEE Transactions on Sustainable Energy*, vol. 16, no. 1, pp. 216–226, 2025, doi: 10.1109/TSTE.2024.3452489.

[19] University of Washington, "IEEE 30-bus test system data," *UW Electrical Engineering*. [Online]. Available: https://labs.ece.uw.edu/pstca/pf30/pg_tca30bus.htm.

[20] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011, doi: 10.1109/TPWRS.2010.2051168.

[21] A. Qi, D. Zhao, A. A. Heidari, L. Liu, Y. Chen, and H. Chen, "FATA: An efficient optimization method based on geophysics," *Neurocomputing*, vol. 607, 2024, doi: 10.1016/j.neucom.2024.128289.

[22] Y. Yang *et al.*, "Advancements in Q-learning meta-heuristic optimization algorithms: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 14, no. 6, 2024, doi: 10.1002/widm.1548.
[23] D. E. Neves, L. Ishitani, and Z. K. G. do Patrocínio Júnior, "Advances and challenges in learning from experience replay," *Artificial Intelligence Review*, vol. 58, no. 2, 2025, doi: 10.1007/s10462-024-11062-0.
[24] E. G. Talbi, "Machine learning into metaheuristics: a survey and taxonomy," *ACM Computing Surveys*, vol. 54, no. 6, 2022, doi: 10.1145/3459664.
[25] T. N. Huynh, D. T. T. Do, and J. Lee, "Q-Learning-based parameter control in differential evolution for structural optimization," *Applied Soft Computing*, vol. 107, 2021, doi: 10.1016/j.asoc.2021.107464.

BIOGRAPHIES OF AUTHORS



Ben Lemkherbeche Ilyes is a Ph.D. student in the Department of Electrical Engineering at the University of Batna 2 - Mostefa Ben Boulaid, Algeria. He received his Licence degree in Electrical Engineering from the University of Mohamed Khider of Biskra, Algeria, in 2009, and his master degree in electrical engineering from the same university in 2011. Currently, his research focuses on power system optimization. In addition to his academic pursuits, he has been working as a teacher at a middle school in Batna city since September 2012. He can be contacted at email: ilyes.benlemkherbeche@univ-batna2.dz.



Bensalem Ahmed is a professor in the Department of Electrical Engineering at the University of Batna 2, Algeria, where he has been teaching since 2003 and currently serves as Head of the Department of Electrical Engineering since 2014. He received his State Engineering degree in electrical equipment from the University of Annaba in 1984, a Master of Applied Sciences from McGill University, Montreal, Canada, in 1988, and his Doctorat d'État from the University of Batna in 2007. He was promoted to professor in 2012. Throughout his career, he has held significant academic positions, including Head of the Department of Electrical Engineering at ENS of Saida from 1989 to 1992. He has taught at various Algerian universities, including ENS of Saida, the University of Sidi-Bel-Abbès, the University of Biskra, and currently at the University of Batna 2. His teaching and research expertise spans optimal power system operation, hydroelectric resource management, economic dispatch, FACTS devices, renewable energy integration, and metaheuristic optimization algorithms. He has been the principal investigator on numerous national research projects focused on economic optimization of electrical energy production, hydropower management, and renewable energy integration. He has authored and co-authored over 40 publications in international and national journals and conference proceedings, including publications in Elsevier journals and the Journal of Electrical Engineering. His professional affiliations include membership in the ARELEC (Association of High Voltage Electrical Networks in Algeria) and service as an accredited legal expert in electrical engineering at the Algerian Court since 2007. He has actively contributed to the academic community as a member of scientific councils, organizing committees, and as conference chair for numerous national and international conferences, demonstrating his sustained commitment to advancing power systems research and education in Algeria. He can be contacted at email: a.bensalem@univ-batna2.dz.



Khemmoudj Sara Hiba is a Ph.D. student in the Department of Electrical Engineering at the University of Batna 2 - Mostefa Ben Boulaid, Algeria. She received her Master's degree in Industrial Electronics from Ferhat Abbas University of Setif, Algeria, in 2011. Her research interests include power system analysis, renewable energy integration, and optimization algorithms for power systems. She is currently working on PWM control techniques for inverters. In addition to her academic pursuits, she has been working as a mathematics teacher at a middle school in Batna city since 2014. She can be contacted at email: sarahiba.khemmoudj@univ-batna2.dz.