https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

# Reinforcement-Learning–Based Collaborative Optimization of Load Regulation and Opportunistic Maintenance of Wind Turbines

Jie Lin [1], Jing Shi [1], Jianghao Zhu [1], Xubin Li [1], Wei Chen [1],

[1] School of Electrical Engineering and Information Engineering, Lanzhou University of Technology. Lanzhou, 730050. China

*Correspondence to*: Wei Chen (chenlin@lut.edu.cn)

**Abstract.** To address the difficulty in coordinating component-degradation regulation strategies with opportunistic maintenance timings over the full lifecycle of wind turbines, this study proposes a reinforcement-learning–based collaborative optimization method for wind-turbine load regulation and opportunistic maintenance (RL-OppOM).The proposed method establishes a full-lifecycle simulation environment for wind turbines by coupling wind conditions, power-load characteristics, component reliability, and maintenance restoration, and formulates graded power derating, multicomponent maintenance combinations, and operational feasibility constraints within a unified Markov decision process. A lifecycle risk-aware reward function is developed and a factorized dual-clip masked proximal policy optimization algorithm is proposed. By incorporating policy factorization, action masking, and dual clipping, the algorithm reduces the complexity of policy learning and improves training stability. Simulation validation is conducted using SCADA data from a wind farm in northern China. The results show that RL-OppOM achieves a comprehensive cost of 7.320 M CNY, which is 13.80% and 11.65% lower than the costs of CBM and OppM, respectively, while increasing the net profit by 2.36% and 1.93%. The mean and minimum turbine-health indicators are 0.6413 and 0.4491, respectively, and the number of high-risk operating days is zero.

**Keywords:** wind turbine; opportunistic maintenance; load regulation; reinforcement learning

## 1 Introduction

Wind energy, as a clean and renewable energy source, has considerable development potential(McKenna et al., 2022). However, wind turbines are often deployed in remote and harsh environments(McMorland et al., 2022; Tang et al., 2025), including mountainous, desert, and offshore regions, where they are exposed to complex environmental conditions(Østergaard et al., 2023). Operation and maintenance (O&M) expenditures account for approximately 20%–30% of the total lifecycle investment in wind-energy projects(Li et al., 2023). Therefore, intelligent lifecycle O&M strategies that simultaneously consider economic return, equipment reliability, and operational risk are urgently needed.

Current wind-turbine O&M strategies include conventional approaches such as corrective maintenance (CM), preventive maintenance (PM), condition-based maintenance (CBM), and opportunistic maintenance (OppM), and are progressively evolving toward intelligent O&M strategies integrating fault prediction, machine learning, and digital-twin technologies(Chatterjee and Dethlefs, 2021; Cuesta et al., 2025). CM is generally performed after a failure occurs(Ren et al.,

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


2021). Although it can reduce the upfront proactive maintenance expenditure, it can increase unplanned downtime, repair costs, and generation losses(Garcia-Teruel et al., 2022). PM reduces the probability of unexpected failures through scheduled interventions(Aafif et al., 2022); however, heterogeneous component degradation and evolving failure risks often result in predetermined maintenance interventions being either excessive or delayed(Zhao et al., 2023). CBM utilizes monitoring data to track component health and perform early intervention against potential risks(Oh et al., 2024); however, its decision performance is constrained by threshold selection(Chen et al., 2023), multisource information fusion, and state uncertainty(Zhao et al., 2024). OppM exploits existing maintenance opportunities by incorporating other components with potential maintenance demands into the same maintenance opportunity, thereby improving maintenance-resource utilization and reducing O&M costs(McMorland et al., 2023). Nevertheless, the dynamic variation of maintenance opportunities increases the decision complexity(Papadopoulos et al., 2022). Machine-learning–based fault diagnosis and reliability-prediction methods have substantially improved anomaly detection(Badihi et al., 2022), health assessment, and remaining-useful-life prediction for wind turbines(Coraddu et al., 2024; Merainani et al., 2022). Digital-twin–based O&M provides new approaches for turbine-state mapping, real-time information fusion, degradation simulation, and maintenance-plan evaluation(Pacheco-Blazquez et al., 2024; Xia and Zou, 2023). However, its system-level engineering implementation remains constrained by model consistency(VanDerHorn and Mahadevan, 2021), state synchronization, and complex practical deployment(Liu et al., 2021). Overall, the existing methods have improved health-state awareness and maintenance decision-making; however, they still rely primarily on maintenance-side interventions and have not incorporated the effects of operational control on component structural degradation into O&M decisions.

Deep reinforcement learning (DRL) allows decision optimization in dynamic and uncertain environments(Pliego Marugán et al., 2024). Existing studies have applied DRL to offshore maintenance task scheduling(Lee et al., 2025), opportunistic maintenance, and multiunit maintenance decision-making(Najafi and Lee, 2023; Valet et al., 2022), demonstrating its capability for dynamic state-feedback optimization and resource allocation(Najafi and Lee, 2023). Other studies have incorporated uncertain multiobjective optimization and short-term wind–wave forecasting(Ogunfowora and Najjaran, 2023) to improve the adaptability of maintenance strategies to weather variations, risk preferences, and operational windows(Li et al., 2022). More recently, physics-driven RL has begun to embed wind-condition evolution, equipment degradation, and economic and reliability objectives into the training environment, thereby enhancing the physical consistency of long-term maintenance decisions(Zhu et al., 2025). DRL studies based on damage evolution have also demonstrated that degradation states can provide more direct information for maintenance-timing selection and risk intervention(Sun and Yin, 2025). In addition to maintenance interventions, load regulation during turbine operation also affects component-damage accumulation and health evolution(Njiri et al., 2019). As a representative means of operational load regulation, power derating (DR) reduces the turbine power reference(Truong et al., 2022), allowing the turbine to operate below the maximum available power(Sahin and Yavrucuk, 2022). Previous studies have shown that control variables such as blade pitch, generator torque, and rotational speed can be used to regulate turbine dynamic responses, whereas the control strategy, wind conditions, and operating state affect their load-mitigation performance(Capaldo and Mella, 2023; Kipchirchir et al., 2023). In terms of load

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


effects, power DR can reduce aerodynamic thrust and structural dynamic loads, thereby mitigating the ultimate and fatigue-equivalent loads of critical structures such as blades and towers(Kim et al., 2020), and consequently, slowing long-term fatigue-damage accumulation. Overall, although considerable progress has been achieved in intelligent maintenance decision-making and load regulation, existing studies have not yet coordinated the effects of load regulation on component structural degradation with multicomponent opportunistic maintenance decisions, making it difficult to simultaneously optimize degradation mitigation and maintenance timing.

Based on the above considerations, this study proposes an RL-based collaborative optimization method for wind-turbine load regulation and opportunistic maintenance. By integrating graded power DR, opportunistic maintenance, component-reliability evolution, and lifecycle economic accounting into a unified decision-making framework, the proposed method facilitates coordinated optimization of operating control and maintenance scheduling. The main contributions of our study are as follows:

(1) A full-lifecycle collaborative optimization framework integrating power DR and opportunistic maintenance is proposed. Graded power DR and multicomponent maintenance combinations are incorporated into a unified sequential decision-making process, coordinating short-term load regulation and long-term maintenance scheduling.

(2) A state-evolution model coupling wind conditions, loads, component degradation, and maintenance restoration is developed. A Markov decision process (MDP) is employed to jointly characterize DR constraints, multicomponent coordinated maintenance, and field operational constraints, thereby ensuring the physical consistency of state evolution.

(3) A factorized dual-clip masked proximal policy optimization (FAM-DC-PPO) algorithm is proposed for high-dimensional joint operation and maintenance decision-making. The factorized policy structure reduces the learning complexity of the joint action space, state-dependent action masking ensures physical feasibility, and dual-clip mechanism improves the training stability over the long lifecycle horizon.

The structure of this paper is shown in **Fig. 1**: Section 2 establishes a physics-driven full-lifecycle coupled simulation environment for wind-turbine operation and maintenance. Section 3 formulates the collaborative optimization problem as an MDP and presents the state space, joint action space, reward function, and proposed FAM-DC-PPO method. Section 4 presents a case study, covering model parameters, comparative strategies, numerical results, and key analyses. Section 5 summarizes the main conclusions and outlines directions for future research.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

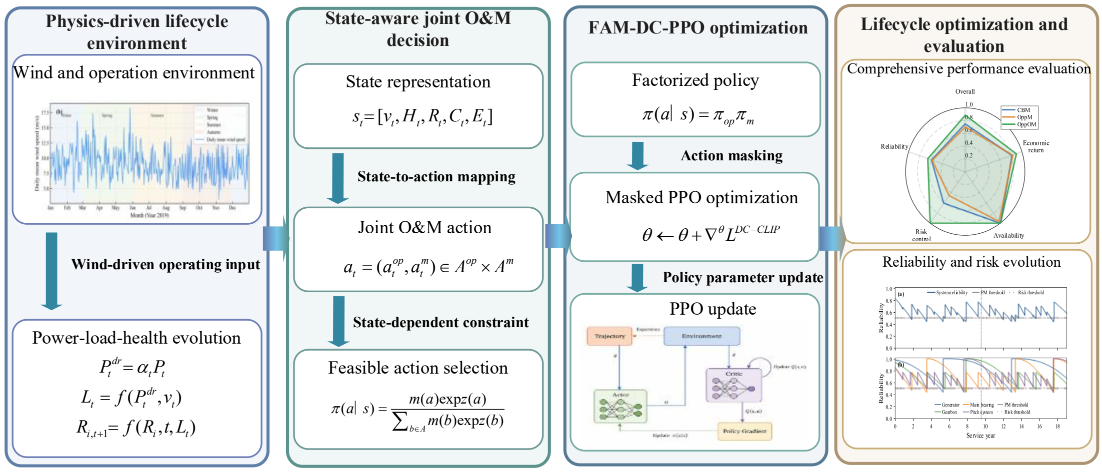


Figure 1: Physics-driven reinforcement-learning framework for full-lifecycle collaborative optimization of wind-turbine power derating and opportunistic maintenance

## 2 Physics-Driven Full-Lifecycle Operation and Maintenance Simulation Environment for Wind Turbines

### 2.1 Coupled Modeling of Continuous Wind Conditions and Power-Load Response

Wind speed is a primary external variable affecting both the power generation and structural loading of wind turbines. To provide consistent environmental inputs for the subsequent reliability evolution and RL decisions, a model mapping continuous wind speed to daily degradation-driving variables was established. Continuous wind-speed sequences were first generated using a Weibull distribution combined with seasonal correction and temporal correlation, after which the theoretical power output was calculated from the turbine power curve. The resulting power sequence was further aggregated into a daily mean load factor and an abnormal-fluctuation driver to characterize long-term fatigue degradation and short-term risk excitation, respectively.

The long-term statistical characteristics of wind speed were described by a Weibull distribution, whose probability density function was given by

$$\boldsymbol{f(v)} = \frac{\boldsymbol{k}}{\boldsymbol{c}}\left(\frac{\boldsymbol{v}}{\boldsymbol{c}}\right)^{\boldsymbol{k-1}}\exp\left[-\left(\frac{\boldsymbol{v}}{\boldsymbol{c}}\right)^{\boldsymbol{k}}\right], \tag{1}$$

where $\boldsymbol{v}$ denotes the wind speed, $\boldsymbol{k}$ is the shape parameter characterizing the distribution profile, and $\boldsymbol{c}$ is the scale parameter reflecting the mean wind-speed level at the site. Considering the temporal continuity and seasonal variability observed in the measured wind speeds, the continuous wind-speed sequence was generated as

$$\boldsymbol{v(t) = \rho v(t-1) + (1-\rho)[v_0 S(t) + \sigma \varepsilon(t)]} \tag{2}$$

where $\boldsymbol{\rho}$is the temporal-correlation coefficient, $\boldsymbol{v_0}$is the baseline wind speed, $\boldsymbol{S(t)}$ is the seasonal-correction factor, $\boldsymbol{\sigma}$ is the amplitude of random fluctuations, and $\boldsymbol{\varepsilon(t)}$ follows a standard normal distribution. The seasonal-correction factor was specified piecewise according to the day of the year:

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026

Open Access
WIND ENERGY SCIENCE DISCUSSIONS

$$
S(t) = \begin{cases} \mathbf{1.10}, \boldsymbol{t} \in \textbf{Spring} \\ \mathbf{0.90}, \boldsymbol{t} \in \textbf{Summer} \\ \mathbf{1.15}, \boldsymbol{t} \in \textbf{Autumn} \\ \mathbf{1.05}, \boldsymbol{t} \in \textbf{Winter} \end{cases} \tag{3}
$$

Here, $\boldsymbol{S(t)}$ took different values according to the seasonal position within the year. These coefficients were calibrated using 30 months of SCADA wind-speed data from a wind farm in northern China to reproduce the site-specific pattern of higher wind speeds in autumn and lower wind speeds in summer. If the model is applied to other regions or turbine capacities, the wind-speed distribution and seasonal-correction parameters must be recalibrated using local SCADA data. **Figure 2** presents the wind-speed sequences generated by the model and their statistical characteristics.

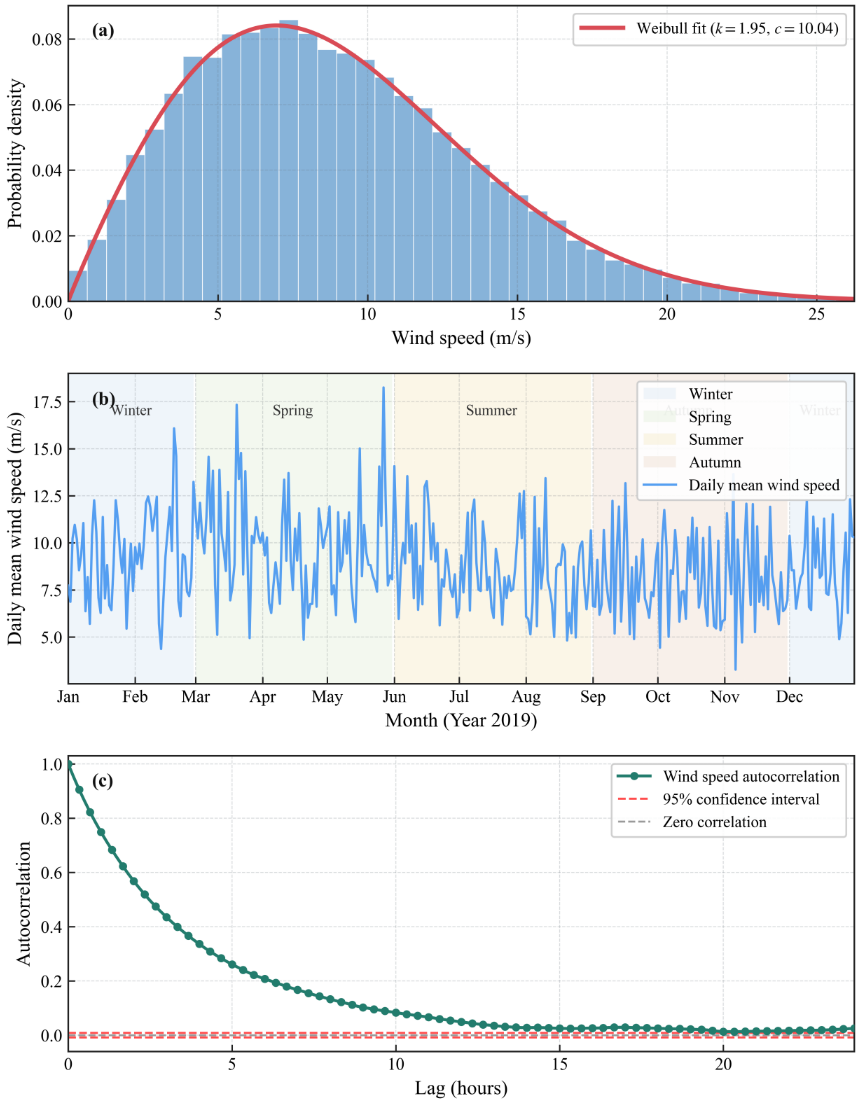


**Figure 2: Characteristics of the wind-speed model: (a) wind-speed probability distribution and Weibull fitting; (b) seasonal evolution of daily mean wind speed; (c) wind-speed autocorrelation.**

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


After the continuous wind-speed sequence was obtained, the output power under the ideal healthy condition was calculated using a typical wind-turbine power curve:

$$\boldsymbol{P^0(v)} = \begin{cases} \mathbf{0} & \boldsymbol{v < v_{\text{in}}} \\ \boldsymbol{P_N \frac{v^3 - v_{\text{in}}^3}{v_r^3 - v_{\text{in}}^3}} & \boldsymbol{v_{\text{in}} \le v < v_r} \\ \boldsymbol{P_N} & \boldsymbol{v_r \le v \le v_{\text{in}}} \\ \mathbf{0} & \boldsymbol{v > v_{\text{out}}} \end{cases}, \tag{4}$$

where $\boldsymbol{P_N}$ is the rated power; and $\boldsymbol{v}_{\text{in}}, \boldsymbol{v}$, and $\boldsymbol{v}_{\text{out}}$ denote the cut-in, rated, and cut-out wind speeds, respectively. Considering the effects of power DR on the energy production and load level, the controlled power output was further expressed as

$$\boldsymbol{P(v, a) = \chi(a) P^0(v)}, \tag{5}$$

where $a$ denotes the operating-control action and $\chi(a)$ is the corresponding power-scaling coefficient. The coefficients for rated operation, mild DR, and deep DR were set to 1.00, 0.90, and 0.80, respectively. Although DR reduced short-term energy production, it decreased the equivalent load input and thereby mitigated the cumulative component degradation. Because both maintenance decisions and reliability updates were performed on a daily basis, the continuous 10-min power sequence was aggregated into a daily load factor:

$$\boldsymbol{L(d) = \frac{1}{N} \sum_{t \in d}^{P_N} \frac{P_{act(d,n)}}{P_N}}, \tag{6}$$

where $\boldsymbol{L(d)}$ is the mean load factor on day $\boldsymbol{d}$ and $\boldsymbol{N = 144}$ is the number of samples per day. This factor characterized the mean intraday power-load level and was subsequently used to modify the long-term component-degradation rate. As this study focused on full-lifecycle maintenance optimization, rather than transient aeroelastic responses, the normalized power was adopted as an approximation of the daily load intensity. This treatment preserved the physically consistent load-degradation trend, while reducing the computational burden of long-term simulation. To further characterize the effects of short-term wind-speed variations and power fluctuations on component risk, an abnormal-fluctuation driver was defined as

$$\boldsymbol{B(d)} = \begin{cases} \boldsymbol{0, F(d) \le F_{\text{th}}} \\ \boldsymbol{\frac{F(d) - F_{\text{th}}}{1 - F_{\text{th}}}, F(d) > F_{\text{th}}} \end{cases}, \tag{7}$$

where $F(d)$ represents the wind-speed or power-fluctuation intensity on day $d$, $F_{th}$ is the activation threshold for abnormal fluctuations, and $B(d)$ quantifies the additional excitation associated with pronounced fluctuations. Under this formulation, normal power loading primarily contributed to long-term fatigue degradation, whereas only strong fluctuations entered the risk-correction term, thereby preventing routine stochastic variations from causing persistent risk accumulation. In addition, the feasibility of maintenance actions was constrained by daily wind conditions. Maintenance actions were included in the feasible action set when the daily mean wind speed and gust level satisfied the safety requirements; under high-wind restrictions or extreme wind conditions, maintenance actions were prohibited.

### 2.2 Component Reliability and Turbine-Level Health-State Model

The overall operational risk of a wind turbine is affected by multiple critical components. A single turbine-level health indicator cannot adequately characterize the differences in component degradation, whereas component-level analysis alone cannot reflect the system-level shutdown risk. Therefore, a component reliability and turbine-level health-state model was established to characterize the heterogeneous degradation among multiple components and its impact on the overall operating conditions of the turbine. The generator, gearbox, main bearing, and pitch system were considered, and the Weibull distribution was adopted to characterize their nonlinear reliability degradation. The reliability of component $i$ on day $d$ was expressed as

155 $$\boldsymbol{R_i(d) = \exp\left[-\left(\frac{\tau_i(d)}{\eta_i}\right)^{\beta_i}\right]}, \quad (8)$$

where $R_i(d)$ denotes the reliability of component $i$, ranging from $[0, 1]$; $\tau_i(d)$ is the equivalent service time; $\eta_i$ is the scale parameter characterizing the baseline lifetime of the component; and $\beta_i$ is the shape parameter describing the variation in failure rate with service time. The Weibull parameters of the critical wind-turbine components were adopted from the lifetime parameters reported in Ref.(Sarker and Faiz, 2016), as listed in **Table 1**.

160 **Table 1. Weibull reliability parameters of key wind-turbine components**

| Component | Scale parameter, $\eta_i$(d) | Shape parameter, $\beta_i$ |
|---|---|---|
| Generator | 3750 | 3.0 |
| Gearbox | 3300 | 2.4 |
| Main bearing | 2400 | 2.2 |
| Pitch system | 1858 | 3.0 |

The equivalent service time of each component was updated on a daily basis as $\boldsymbol{\tau_i(d+1) = \tau_i(d) + \Delta\tau_i(d)}$, (9)
where the daily equivalent aging increment was given by $\boldsymbol{\Delta\tau_i(d) = \kappa_i[1 + \alpha_i L(d) + \mu_i B(d)]}$, (10)
where $\kappa_i$ is the baseline aging coefficient of component $i$, $\alpha_i$ and $\mu_i$ are the mean-load acceleration coefficient and abnormal-fluctuation acceleration coefficient, respectively; and $L(d)$ and $B(d)$ denote the daily load factor and abnormal-fluctuation driver on day $d$, respectively. High-load operation and pronounced fluctuations increased the equivalent aging rate of the components, whereas the power DR reduced the load factor $F(d)$, thereby indirectly mitigating the cumulative component degradation.

Considering the critical components as a series system, the overall turbine reliability was calculated as

$$\boldsymbol{R_{\mathrm{sys}}(d) = \prod_i R_i(d)}\,, \quad (11)$$

170 where $R_{\mathrm{sys}}(d)$ denotes the overall turbine reliability on day $d$ and characterizes the system-level operational risk arising from the combined effects of multiple critical components. The bottleneck-component reliability was defined as

$$\boldsymbol{R_{\min}(d) = \min_i R_i(d)}, \quad (12)$$

where $R_{\min}(d)$ represents the reliability level of the currently weakest component and is adopted as the turbine health indicator to characterize the bottleneck health condition limiting safe turbine operation. The overall turbine reliability $R_{\mathrm{sys}}(d)$ reflects the aggregate system risk associated with multiple critical components, whereas the turbine health indicator $R_{\min}(d)$ represents the health margin of the weakest critical component. Both variables were incorporated into the subsequent RL state representation and risk assessment.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


To classify the turbine health state, three reliability thresholds were introduced, including the minor-maintenance warning threshold $R_{\text{minor}}$, high-risk threshold $R_{\text{risk}}$, and failure threshold $R_{\text{failure}}$; their values are listed in **Table 2**. The turbine health state was defined based on $R_{\min}(d)$ as

$$\boldsymbol{H_d} = \begin{cases} \textbf{Normal; } \boldsymbol{R}_{\min}(\boldsymbol{d}) > \boldsymbol{R}_{\text{minor}} \\ \textbf{Warning; } \boldsymbol{R_{risk}} < \boldsymbol{R}_{\min}(\boldsymbol{d}) \le \boldsymbol{R}_{\text{minor}} \\ \textbf{High-risk; } \boldsymbol{R_{failure}} < \boldsymbol{R}_{\min}(\boldsymbol{d}) \le \boldsymbol{R}_{\text{risk}} \\ \textbf{Failure; } \boldsymbol{R}_{\min}(\boldsymbol{d}) \le \boldsymbol{R}_{\text{failure}} \end{cases}. \tag{13}$$

These thresholds were used only for health-state identification and risk statistics, and not as fixed maintenance triggers for the RL policy. In contrast, the conventional condition-based maintenance strategy directly used these thresholds to generate maintenance decisions. By integrating Weibull reliability evolution, wind-load–induced aging acceleration, series-system reliability, and health-risk classification, the proposed model characterized the heterogeneous degradation of critical components as well as the lifecycle evolution of turbine health and system risk, providing a modeling basis for the subsequent power DR, maintenance-induced reliability restoration, and RL-based collaborative optimization.

### 2.3 Maintenance Actions, Reliability Restoration, and Opportunistic-Maintenance Window Model

Maintenance actions directly affect component health, turbine downtime, and lifecycle costs. To characterize state restoration under different maintenance intensities, four maintenance actions were considered: no maintenance, minor maintenance, major maintenance, and corrective maintenance. Under no maintenance, components degraded naturally according to the model described in Section 2.2. Minor maintenance provided partial health restoration, major maintenance provided substantial restoration, and corrective maintenance was performed when a component failed.

For any component requiring maintenance, its reliability before and after maintenance was denoted by $R$ and $R^+$, respectively. Minor maintenance was modeled as an imperfect restoration process:

$$\boldsymbol{R^+ = R + \eta(1 - R)}, \tag{14}$$

where $\eta \in (0,1)$ denotes the minor-maintenance restoration coefficient characterizing the effect of imperfect maintenance. Major maintenance restored the component to a near-initial health condition:

$$\boldsymbol{R^+ = R_{\text{new}}}, \tag{15}$$

where $R_{\text{new}}$ denotes the target reliability after major maintenance. If the component reliability reached the failure threshold, corrective maintenance was performed to repair or replace the component and restore it to the post-replacement health state. Accordingly, the equivalent service time was recalculated from the restored reliability as

$$\boldsymbol{\tau^+ = \lambda[-\ln(R^+)]^{1/\beta}}, \tag{16}$$

where $\tau^+$ denotes the equivalent service time after maintenance, and $\lambda$ and $\beta$ are the Weibull scale and shape parameters, respectively, of component $i$. Maintenance actions could only be executed when safe operating conditions were satisfied. A daily maintenance-feasibility indicator was therefore defined as

$$\boldsymbol{W(d)} = \begin{cases} \boldsymbol{1};\ \boldsymbol{\overline{v}(d) \le v_{\text{safe}}}\,\text{且}\,\boldsymbol{v_{\max}(d) \le v_{\text{gust}}} \\ \boldsymbol{0};\ \textbf{else} \end{cases}, \tag{17}$$

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

where $W(d) = 1$ indicated that maintenance is feasible on day $d$; $\overline{v}(d)$ and $v_{\max}(d)$ denote the daily mean wind speed and maximum gust speed, respectively; and $v_{\text{safe}}$ and $v_{\text{gust}}$, lim are the corresponding safety thresholds. When $W(d) = 0$, maintenance is temporarily infeasible, and the turbine continues to evolve under its current operating state.

To reduce repeated downtime and access costs caused by asynchronous degradation among multiple components, opportunistic maintenance was introduced. When a component developed a clear maintenance demand and the daily operating conditions permitted maintenance, an exploitable maintenance opportunity became available, and other components whose health states were close to the maintenance region were included in the opportunistic-maintenance candidate set:

$$\boldsymbol{O(d) = \{i \mid R_{\text{low}} < R_i(d) \leq R_{\text{opp}}\}}, \tag{18}$$

where $O(d)$ denotes the set of opportunistic-maintenance candidates, $R_{\text{opp}}$ denotes the opportunistic-maintenance boundary, and $R_{\text{low}}$ denotes the lower bound of the candidate interval. The availability of a maintenance opportunity did not directly trigger maintenance; whether the opportunity was utilized and which components were maintained were determined by the subsequent decision policy. Multiple components maintained within the same opportunity could share the fixed turbine-access and dispatch cost. The total maintenance cost was expressed as

$$\boldsymbol{C(d) = C_{\text{fix}} + \sum_{i \in M(d)} C_i}, \tag{19}$$

where $C_{\text{fix}}$ denotes the fixed access and dispatch cost associated with a maintenance opportunity, $M(d)$ is the set of components actually maintained, and $C_i$ denotes the direct maintenance cost of component $i$. The parameter settings are listed in **Table 2**.

**Table 2. Maintenance, operational constraint, and economic parameters**

| Category | Parameter | Value |
|---|---|---|
| Health boundary | Minor-maintenance candidate | 0.60 |
| | Major-maintenance candidate | 0.45 |
| | High-risk threshold | 0.35 |
| | Failure threshold | 0.01 |
| Opportunistic maintenance | Opportunistic-maintenance upper bound | 0.66 |
| Maintenance recovery | Target reliability after minor maintenance | 0.78 |
| | Minimum reliability improvement | 0.05 |
| | Reliability after corrective maintenance | 1.00 |
| Weather constraint | Maximum allowable wind speed for minor/major maintenance | 16/12 |
| Downtime | Minor/major/corrective maintenance | 1/7/15 d |
| Maintenance cost | Minor/major/corrective maintenance | 0.210/1.600/2.400 M CNY |
| Window cost | Minor/major/corrective maintenance | 0.040/0.100/0.150 M CNY |
| Electricity price | Base electricity price | 0.50 CNY/kWh |

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


| | | |
|---|---|---|
| Revenue model | Revenue scaling coefficient | 1.51 |
| Operating control | Mode-switching cost | 0.005 M CNY |

## 3 RL-based Collaborative Optimization of Adaptive Load Regulation and Opportunistic Maintenance

### 3.1 Lifecycle Operation–Maintenance Co-optimization Framework

A collaborative optimization framework for wind-turbine operation and maintenance was established, in which adaptive power DR and opportunistic maintenance were formulated as a unified sequential decision-making process, as illustrated in **Fig. 3**. Based on wind conditions, component reliability, turbine health state, and maintenance feasibility, the agent jointly selected the operating mode and maintenance action. The simulation environment subsequently updated the power output, degradation evolution, reliability restoration, and lifecycle returns. With the objective of maximizing the expected long-term cumulative reward over the entire lifecycle, the optimal policy was formulated as:

$$\boldsymbol{\pi}^* = \boldsymbol{arg}\,\max_{\boldsymbol{\pi}} \mathbb{E}\left[\sum_{d=0}^{T=1} \boldsymbol{\gamma}^{d} \boldsymbol{r}(\boldsymbol{d})\right], \qquad (20)$$

where $\pi$ denotes the decision policy, $T$ is the lifecycle horizon, $\gamma$ is the discount factor, and $r(d)$ is the composite reward at decision period $d$, jointly accounting for power-generation revenue, maintenance expenditure, and reliability risk.

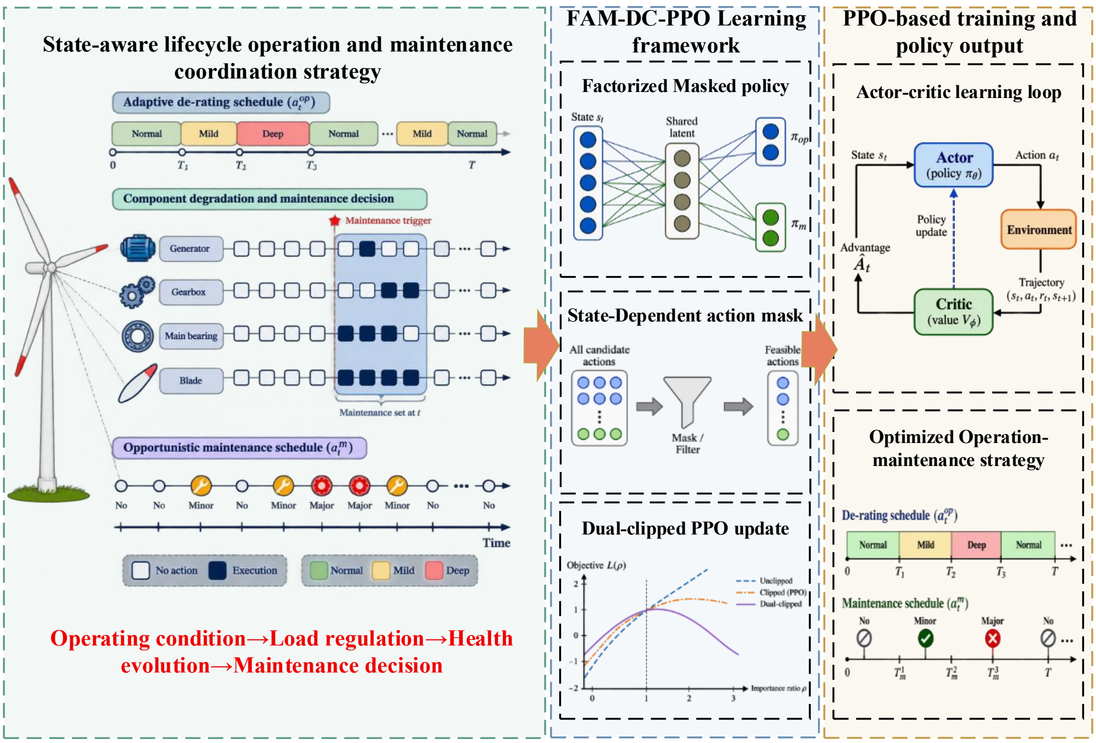

**Figure 3: Lifecycle operation–maintenance co-optimization framework**

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


### 3.2 MDP Formulation: States, Actions, and State Transitions

The full-lifecycle operation and maintenance optimization of the wind turbine was formulated as a fully observable MDP. The simulation environment explicitly updated the component reliability, turbine health state, wind and load conditions, operating mode, and maintenance state, all of which were observable to the agent. The decision process was represented by the following five-tuple:

$$\boldsymbol{\mathcal{M}} = (\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{R}}, \boldsymbol{\gamma}), \tag{21}$$

where $\mathcal{S}$ denotes the state space, $\mathcal{A}$ denotes the action space, $\mathcal{P}$ represents the state-transition dynamics, $\mathscr{R}$ represents the immediate reward function, and $\gamma$ is the discount factor. On decision day $d$, the system state was represented as

$$\boldsymbol{s}(\boldsymbol{d}) = [\boldsymbol{R}(\boldsymbol{d}), \boldsymbol{R}_{\mathbf{sys}}(\boldsymbol{d}), \boldsymbol{R}_{\mathbf{min}}(\boldsymbol{d}), \boldsymbol{E}(\boldsymbol{d}), \boldsymbol{Q}(\boldsymbol{d}), \boldsymbol{H}(\boldsymbol{d})], \tag{22}$$

where $R(d)$ denotes the reliability vector of the four critical components; $R_{\mathrm{sys}}(d)$ and $R_{\mathrm{min}}(d)$ denote the overall turbine reliability and bottleneck-component reliability, respectively; $E(d)$ contains wind speed, load factor, and abnormal-fluctuation information; $Q(d)$ describes the seasonal position and lifecycle progress; and $H(d)$ represents the current operating mode, remaining downtime, and maintenance-opportunity status. All continuous variables were normalized to improve training stability. On each decision day, the agent simultaneously selected the operating-control and maintenance actions and the joint action was expressed as

$$\boldsymbol{a}(\boldsymbol{d}) = (\boldsymbol{a}^{\mathrm{op}}(\boldsymbol{d}), \boldsymbol{a}^{\boldsymbol{m}}(\boldsymbol{d})). \tag{23}$$

The joint action incorporated both operating-control and maintenance decisions. The agent selected from among the rated operation, mild DR, and deep DR, while determining whether minor or major maintenance was to be performed on a combination of critical components, according to their current states. This joint decision structure enabled collaborative optimization of power-DR control and multicomponent opportunistic maintenance. The system state was recursively updated according to the physics-driven simulation model:

$$\boldsymbol{s}(\boldsymbol{d}+\mathbf{1}) = \boldsymbol{F}(\boldsymbol{s}(\boldsymbol{d}), \boldsymbol{a}(\boldsymbol{d}), \boldsymbol{\xi}(\boldsymbol{d})), \tag{24}$$

where $F(\cdot)$denotes the physics-driven state-transition function, which jointly describes wind-speed evolution, power output, load variation, component degradation, DR-induced degradation mitigation, maintenance-induced reliability restoration, and downtime processes; and $\xi(d)$ represents stochastic disturbances such as wind-condition fluctuations. Maintenance actions were further subject to execution constraints associated with weather conditions, downtime status, and maintenance opportunities. Therefore, the action actually selected was restricted to a feasible action set corresponding to the current state:

$$\boldsymbol{a}(\boldsymbol{d}) \in \boldsymbol{A}_{\mathbf{feas}}(\boldsymbol{s}(\boldsymbol{d})) \tag{25}$$

In summary, the MDP formulation allowed the agent to continuously perceive the system state under physical constraints and learn a collaborative strategy for operating control and maintenance scheduling through long-term interaction.

### 3.3 State-Aware Coordination Mechanism for Opportunistic Maintenance and Power DR

A state-aware coordination mechanism for opportunistic maintenance and cross-timescale power DR was developed. Component reliability was used to identify the primary maintenance demands and opportunistic-maintenance candidates,

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026




whereas the policy based on wind conditions, system health state, and long-term return autonomously determined the operating mode, maintenance intensity, and component combination. Accordingly, the primary maintenance-demand set and opportunistic-maintenance candidate set were defined as

$$\boldsymbol{M(d) = \{i \mid R_i(d) \leq R_m\}}, \quad (26)$$

$$\boldsymbol{O(d) = \{i \mid R_m < R_i(d) \leq R_{\mathrm{opp}}\}}, \quad (27)$$

where $M(d)$ denotes the set of components whose health states have entered the primary maintenance region; and $O(d)$ denotes the set of components whose health states have not yet entered the primary maintenance region, but have potential for early coordinated maintenance; $R_m$ and $R_{\mathrm{opp}}$ denote the maintenance reference boundary and opportunistic-maintenance candidate boundary, respectively, with $R_{\mathrm{opp}} > R_m$. The two sets jointly characterized the current multicomponent maintenance demand but neither directly triggered maintenance nor requires all candidate components to be maintained simultaneously. The actual maintenance components were determined by the maintenance decision in the joint action defined in Section 3.2:

$$\boldsymbol{J(d) = g(a^m(d))}, \quad (28)$$

where $J(d)$ denotes the actual combination of components selected for maintenance and $g(\cdot)$represents the mapping from the maintenance action to the corresponding component combination. When the maintenance conditions defined in Section 3.2 were satisfied, the corresponding maintenance combination became feasible. Multiple components maintained within the same maintenance opportunity shared the fixed access cost, highlighting the economic advantage of opportunistic maintenance. The reliability boundaries were used only for state identification and opportunity-value characterization, whereas the final maintenance action was autonomously determined by the policy network.

Power DR further provides cross-timescale risk buffering before maintenance. When the load level increased or system health deteriorated while maintenance was temporarily infeasible, or when immediate maintenance was less favorable in terms of long-term returns, the policy could select mild or deep DR to reduce the daily load factor and thereby slow the equivalent aging of the components. Once suitable maintenance conditions became available, the maintenance combination was selected based on the updated health states. This dynamic relationship could be summarized as

$$\boldsymbol{a^{\mathrm{op}}(d) \rightarrow L(d) \rightarrow R(d+1) \rightarrow a^m(d+1)}, \quad (29)$$

where the operating-control action first modified the current load level, which subsequently affected the component reliability in the next period and resulting maintenance demand. If a maintenance action was executed, the downtime state took precedence over operating control; therefore, the power DR had no additional effect on the power output or degradation process that day.

### 3.4 Lifecycle Risk-Aware Reward Function

Lifecycle decision-making for wind turbines requires balancing the current power-generation revenue, maintenance expenditure, and long-term failure risk. Maximizing power-generation revenue alone may cause the turbine to operate at low reliability for prolonged periods, whereas excessive emphasis on health preservation may lead to frequent maintenance and

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026



unnecessary downtime. Therefore, a lifecycle risk-aware reward function was constructed to jointly incorporate the economic returns, maintenance cost, and reliability risk into the policy feedback:

$$\boldsymbol{r(d) = G(d) - w_C C(d) - w_R P(d) - w_F F(d)}, \tag{30}$$

where $G(d)$ denotes the actual power-generation revenue on day $d$, $C(d)$ is the comprehensive operation and maintenance cost, $P(d)$ is the reliability-risk penalty, and $F(d)$ is the expected failure-risk cost. $w_C$, $w_R$ and $w_F$ are the corresponding weighting coefficients, with $w_C = 1.0$, $w_R = 1.55$, and $w_F = 2.0$. The actual power-generation revenue was calculated from the power output after applying the operating-control action. Therefore, the revenue variations caused by power DR and downtime were accounted for. Accordingly, the associated generation losses were not deducted repeatedly. The comprehensive operation and maintenance cost was expressed as:

$$\boldsymbol{C(d) = C_m(d) + C_{\mathrm{cm}}(d) + C_{\mathrm{op}}(d)}, \tag{31}$$

where $C_m(d)$ denotes the planned-maintenance and fixed access costs, $C_{\mathrm{cm}}(d)$ is the corrective-maintenance cost, and $C_{\mathrm{op}}(d)$ represents the operating-mode switching cost and other additional operating costs. When multiple components were maintained within the same maintenance opportunity, they shared the fixed access cost, thereby allowing the reward function to capture the economic advantage of coordinated maintenance.

To mitigate delayed risk feedback caused by sparse failure events, a region-dependent nonlinear risk penalty was constructed based on the bottleneck-component reliability:

$$\boldsymbol{P(d) = \beta_1 \left[\frac{R_{\mathrm{ref}} - R(d)}{R_{\mathrm{ref}} - R_{\mathrm{risk}}}\right]_{\div}^{2} + \beta_2 \left[\frac{R_{\mathrm{ref}} - R(d)}{R_{\mathrm{ref}} - R_{\mathrm{failure}}}\right]_{\div}^{2} + \beta_3 \left[\frac{R_{\mathrm{near}} - R(d)}{R_{\mathrm{near}} - R_{\mathrm{failure}}}\right]_{\div}^{2}}, \tag{32}$$

where $R_{\mathrm{min}}(d)$ denotes the current bottleneck-component reliability; $R_{\mathrm{ref}}$, $R_{\mathrm{risk}}$, and $R_{\mathrm{failure}}$ denote the warning, high-risk, and failure thresholds, respectively; $R_{\mathrm{near}}$ is the near-failure threshold; $\beta_1$, $\beta_1$, and $\beta_3$ are the corresponding risk-penalty coefficients; and $[x]_+ = \max(0, x)$ denotes the positive-part operator. The risk penalty increased progressively as the bottleneck-component reliability deteriorated and became more pronounced as the reliability approached the failure boundary, thereby providing continuous risk feedback to the agent. Furthermore, $F(d)$ was calculated from the current failure probability of each component and corresponding failure consequence to characterize potential failures that had not yet occurred but already involved significant economic risk. This term alleviated the reward sparsity caused by the low occurrence frequency of actual failure events and allowed the policy to balance the cost of early maintenance against the risk of delayed intervention over the turbine lifecycle.

### 3.5 FAM-DC-PPO for Policy Optimization

Full-lifecycle wind-turbine optimization is characterized by a high-dimensional state space, long decision horizon, and sparse risk feedback, making standard policy-gradient methods susceptible to extreme maintenance costs and failure events. To improve the training stability, a FAM-DC-PPO strategy was developed based on PPO, integrating joint-action modeling, physical-feasibility constraints, and long-horizon policy updates within a unified actor–critic framework. The policy network employed a shared encoder to extract features related to wind conditions, component reliability, lifecycle position, and operating state, and used separate output branches for operating-control and maintenance decisions. To eliminate infeasible

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

actions, a state-dependent action mask was further introduced to filter actions that violated weather, downtime, or maintenance constraints. The masked policy distribution was expressed as

$$\boldsymbol{\pi(a \mid s)} = \frac{\boldsymbol{m(a)}\mathbf{exp}\boldsymbol{z(a)}}{\sum_{b\in A} \boldsymbol{m(b)}\mathbf{exp}\boldsymbol{z(b)}}, \tag{33}$$

where $z(a)$ denotes the score of action a and $m(b)$ indicates its feasibility. Infeasible actions were assigned zero probability, such that policy exploration and updating were restricted to the current feasible action space. PPO improved training stability by constraining the variation between the updated and previous policies. The policy probability ratio was defined as

$$\boldsymbol{\rho} = \frac{\boldsymbol{\pi_\theta(a|s)}}{\boldsymbol{\pi}_{\mathbf{old}}\boldsymbol{(a|s)}}. \tag{34}$$

Considering that corrective maintenance and near-failure events could produce large negative advantages, a dual-clip objective was adopted:

$$\mathrm{LDC} = \boldsymbol{E}\begin{cases}\mathbf{min}[\boldsymbol{\rho A},\mathbf{clip}(\boldsymbol{\rho}, \mathbf{1}-\boldsymbol{\varepsilon}, \mathbf{1}+\boldsymbol{\varepsilon})\boldsymbol{A}]; & \boldsymbol{A} \geq \mathbf{0} \\ \mathbf{max}\{\mathbf{min}[\boldsymbol{\rho A},\mathbf{clip}(\boldsymbol{\rho}, \mathbf{1}-\boldsymbol{\varepsilon}, \mathbf{1}+\boldsymbol{\varepsilon})\boldsymbol{A}], \boldsymbol{cA}\}; & \boldsymbol{A} < \mathbf{0}\end{cases}, \tag{35}$$

where $A$ denotes the advantage estimate, $\varepsilon$ is the standard PPO clipping coefficient, and $c > 1$ is the additional clipping coefficient for negative advantages. This mechanism limited excessive policy updates induced by extreme negative returns, thereby improving training stability over the long lifecycle horizon. The advantage function was estimated using generalized advantage estimation (GAE):

$$\boldsymbol{\delta(d)} = \boldsymbol{r(d)} + \boldsymbol{\gamma V(s(d+1))} - \boldsymbol{V(s(d))}, \tag{36}$$

$$\boldsymbol{A^{(d)}} = \sum_{\ell=0}^{\infty} \boldsymbol{(\gamma\lambda)^{\ell}\delta(d+\ell)}, \tag{37}$$

where $V(\cdot)$denotes the state-value function and $\lambda$ is the GAE smoothing coefficient. To reduce the influence of large maintenance costs and failure losses on value-function learning, the critic employs a clipped value-function update with Huber loss, while entropy regularization was introduced to maintain effective exploration during training. The overall optimization objective was formulated as:

$$\boldsymbol{L} = -\boldsymbol{L}_{\mathrm{DC}} + \boldsymbol{\omega_V L}_{\mathrm{Huber}} - \boldsymbol{\omega_H H(\pi)}, \tag{38}$$

where $\boldsymbol{L}_{\mathrm{Huber}}$ denotes the Huber value loss, $\boldsymbol{H(\pi)}$ denotes the policy entropy, and $\boldsymbol{\omega_V}$ and $\boldsymbol{\omega_H}$ are the corresponding weighting coefficients. During training, the policy network continuously interacted with the physics-driven simulation environment established in Section 2. Based on state feedback, the operating-control and maintenance decisions were iteratively updated. The main hyperparameters are listed in **Table 3**. Through iterative policy optimization, an adaptive collaborative strategy for full-lifecycle return optimization was ultimately obtained.

**Table 3. Main hyperparameter settings of the FAM-DC-PPO algorithm**

| Hyperparameter | Value | Description |
|---|---|---|
| Random seed | 42 | Random seed for reproducible training |
| Number of parallel environments | 4 | Number of environments used for parallel rollout collection |

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

| | | |
|---|---|---|
| Rollout steps | 256 | Number of interaction steps collected per rollout |
| Total updates | 1200 | Maximum number of policy-update iterations |
| Minibatch size | 256 | Number of samples used in each gradient update |
| PPO epochs | 4 | Number of optimization epochs per rollout |
| Learning rate | 0.0001 | Learning rate of the Adam optimizer |
| Discount factor, $\gamma$ | 0.995 | Discount factor used for reinforcement-learning returns |
| GAE coefficient, $\lambda$ | 0.95 | Smoothing coefficient for generalized advantage estimation |
| PPO clipping coefficient | 0.15 | Clipping range of the policy probability ratio |
| Value clipping coefficient | 0.15 | Clipping range used for value-function updates |
| Dual-clip coefficient | 3.0 | Additional clipping coefficient for large negative advantages |
| Entropy coefficient | 0.022–0.004 | Linearly decayed exploration regularization coefficient |
| Value-loss coefficient | 0.50 | Weight of the Critic loss |
| Maximum gradient norm | 0.50 | Threshold used for gradient clipping |
| Target KL divergence | 0.015 | Threshold for controlling excessive policy updates |
| Huber-loss threshold | 10.0 | Threshold of the Huber value loss |
| Hidden-layer dimensions | 256/128 | Dimensions of the shared feature-extraction layers |
| Evaluation interval | 20 updates | Interval between periodic policy evaluations |
| Evaluation episodes | 5 | Number of lifecycle episodes used for each evaluation |

### 3.6 Multi-Attribute Lifecycle-Evaluation Framework

To comprehensively evaluate the performances of different operation and maintenance strategies over the entire service life, a lifecycle-evaluation framework was established based on five dimensions—economic return, operational availability, system reliability, actual energy production, and risk-control capability—and an overall performance score was calculated.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026




This evaluation framework was not involved in RL training and was used only for consistent performance comparison among CBM, OppM, and the proposed strategy. Lifecycle economic return was used to quantify the long-term economic performance of each strategy and was defined as

$$\boldsymbol{\Pi = pE - C}, \tag{39}$$

where $\Pi$ denotes the lifecycle economic return, $p$ is the electricity price, $E$ is the cumulative actual energy produced, and $C$ is the cumulative operation and maintenance cost, including that of minor maintenance, major maintenance, corrective maintenance, and fixed access costs. As the actual energy production already accounted for the effects of power DR and maintenance-induced downtime, the corresponding energy losses were not deducted again. The system availability characterized the ability of the turbine to remain operational over its lifecycle and was defined as

$$\boldsymbol{A = 1 - \frac{D}{T}}, \tag{40}$$

where $A$ denotes the system availability, $D$ is the cumulative downtime, and $T$ is the total lifecycle duration. As the turbine remained capable of generating power under DR operation, DR periods are not counted as downtime. The mean system reliability was used to evaluate the ability of a strategy to maintain long-term system reliability and was defined as

$$\boldsymbol{\overline{R_{sys}} = \frac{1}{T}\sum_{d=1}^{T} R_{sys}(d)}, \tag{41}$$

where $\overline{R_{sys}}$ denotes the lifecycle mean system reliability and $R_{\text{sys}}(d)$ denotes the overall turbine reliability on day $d$, calculated using the series-system reliability model established in Section 2.2. This indicator characterized the long-term system reliability resulting from the combined effects of multiple critical components.

The actual energy production $E$ was used to quantify the lifecycle energy output under the combined effects of stochastic wind conditions, power DR, component health states, and maintenance-induced downtime. In addition, the O&M cost per unit of energy production, $K$, was introduced as a supplementary economic indicator to quantify the maintenance expenditure per unit of energy output and was defined as

$$\boldsymbol{K = \frac{C}{E}}, \tag{42}$$

where $K$ denotes the O&M cost per unit of energy produced, expressed in CNY/kWh. Unlike the actual energy produced $E$, $K$ was used only as a supplementary indicator to evaluate the economic expenditure of different strategies and was not included in the calculation of the overall performance score.

To evaluate the capability of different strategies to control long-term operational risk, the number of high-risk exposure days was used to characterize the lifecycle risk exposure. When the turbine health indicator $R_{min}(d)$ entered the high-risk region, day $d$ was counted as a high-risk operating day. The cumulative number of high-risk exposure days was defined as

$$\boldsymbol{N_{risk} = \sum_{d=1}^{T} 1[R_{min}(d) \le R_{risk}]}, \tag{43}$$

where $N_{risk}$ denotes the number of high-risk exposure days over the lifecycle, $R_{min}(d)$ is the turbine health indicator on day $d$, $R_{risk}$ denotes the high-risk threshold, and$1[\cdot]$is the indicator function. A smaller $N_{risk}$ indicated better risk-control capability.

To eliminate the differences in dimensionality and numerical ranges among the evaluation indicators, all basic indicators were normalized before comprehensive evaluation. The lifecycle economic return, system availability, mean system reliability, and actual energy production are treated as benefit-type indicators, for which larger values indicated better performances. High-risk exposure days were treated as a cost-type indicators and were inversely normalized to obtain the risk-control score, such that a larger value represents better risk-control performance. On this basis, the overall performance score was defined as

$$\boldsymbol{\Omega = w_{\Pi} Z_{\Pi} + w_A Z_A + w_R Z_R + w_E Z_E + w_Q Z_Q}, \tag{44}$$

where $\Omega$ denotes the overall performance score; $Z_{\Pi}$, $Z_A$, $Z_R$, $Z_E$, and $Z_Q$ denote the dimensionless evaluation values of the lifecycle economic return, system availability, mean system reliability, actual energy production, and risk-control capability, respectively; and $w_{\Pi}$, $w_A$, $w_R$, $w_E$, and $w_Q$ are the corresponding weighting coefficients satisfying $w_{\Pi} + w_A + w_R + w_E + w_Q = 1$.The final lifecycle evaluation vector was expressed as $Z = [\Omega, Z_{\Pi}, Z_A, Z_R, Z_E, Z_Q]$. This evaluation framework facilitated consistent comparison of different operation and maintenance strategies across five dimensions— economic performance, operational availability, system reliability, energy production, and risk control.

## 4 Case Study

### 4.1 Case Setup

#### 4.1.1 Data Source and Study Object

A 2 MW onshore wind turbine located at a wind farm in northern China was selected as the study object. The site exhibited pronounced seasonal variations in wind resources, providing a representative application scenario for investigating the long-term wind-turbine operation-control and maintenance decision-making under stochastic wind conditions. The SCADA dataset covered the period from January 2019 to July 2021 and mainly included wind speed, power output, ambient temperature, and turbine operating status.

The measured data were primarily used to characterize the statistical distribution, seasonal variation, and temporal correlation of site-specific wind resources, as well as to analyze the wind-speed probability distribution, wind-speed autocorrelation, and wind-speed–power relationship. Based on these analyses, the continuous wind-speed generation model, power-load response model, component-reliability evolution model, and maintenance restoration model established in Section 2 were integrated to construct a physics-driven simulation environment for full-lifecycle operation and maintenance optimization. The case-study horizon was set to 6,935 days, corresponding to approximately 19 years of turbine service life. Continuous wind conditions were first generated at relatively high temporal resolutions and then aggregated to obtain the daily load and abnormal-fluctuation drivers. Operating-mode selection, reliability updating, maintenance decision-making, and economic accounting were all performed with a 1-d time step, thereby balancing the long-term simulation efficiency and temporal resolution required for operation and maintenance decisions.

#### 4.1.2 Comparative Strategies and Experimental Setup

To validate the effectiveness of the proposed RL-OppOM strategy, CBM and OppM were selected as benchmark strategies. CBM determined the maintenance timing according to component reliability and predefined health thresholds, and performed minor maintenance, major maintenance, or corrective maintenance accordingly. Building on CBM, OppM exploited the available maintenance opportunities by incorporating other components whose health states were close to the maintenance region into the coordinated maintenance. In contrast, RL-OppOM autonomously optimized the operating mode, maintenance intensity, and maintenance-component combination based on wind conditions, component reliability, turbine health state, and maintenance feasibility, thereby facilitating joint decision-making for power DR and opportunistic maintenance.

To ensure fair comparison, the three strategies adopted identical lifecycle horizons, initial health states, wind-condition sequences, component-degradation parameters, maintenance-restoration models, maintenance costs, downtime durations, and weather-related operational constraints, with differences retained only in their decision-making mechanisms. After training, the network parameters of the RL strategy are fixed, and independent evaluations were conducted under identical simulation conditions. The strategies were compared primarily in terms of the lifecycle net profit, O&M cost, system availability, mean system reliability, number of maintenance actions, and Number of high-risk operating days. All performance indicators were aggregated from daily simulation results using a consistent calculation procedure.

### 4.2 RL Training and Convergence Analysis

**Figure 4** illustrates the instantaneous reward, reward trend, and cumulative evaluation performance of the FAM-DC-PPO strategy during training. The instantaneous reward reflected the stepwise decision feedback received by the agent at different training stages, while the smoothed trend characterized the overall evolution of policy performance. The cumulative evaluation performance was used to track the progressive improvement of candidate policies throughout training. The dashed line indicated the approximate point at which the training entered a relatively stable stage.

At the early stage of training, the policy network had not yet learned an effective operation and maintenance decision pattern, and the agent therefore performed extensive exploration within the joint action space. Consequently, the instantaneous reward exhibited pronounced fluctuations, while the cumulative evaluation performance remained relatively low. As the number of training updates increased, the overall reward trend gradually improves and cumulative evaluation performance increased markedly, indicating that the agent began to learn the dynamic relationships among wind-condition variations, power DR, component-health evolution, and maintenance interventions. The local fluctuations observed during the intermediate training stage mainly resulted from continuous policy adjustments among different DR levels, maintenance timings, and component combinations, rather than a persistent deterioration in policy performance.

After approximately 700 updates, the variation in the reward trend gradually decreased and the cumulative evaluation performance entered a relatively stable range, indicating that policy updates had approached convergence under the current simulation environment and parameter configuration. During training, candidate policies were independently evaluated over

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


the full lifecycle at fixed intervals, and the final policy snapshot was selected according to economic, reliability, and safety metrics. The network parameters of the selected policy were then fixed for the subsequent full-lifecycle performance comparison and decision-mechanism analysis.

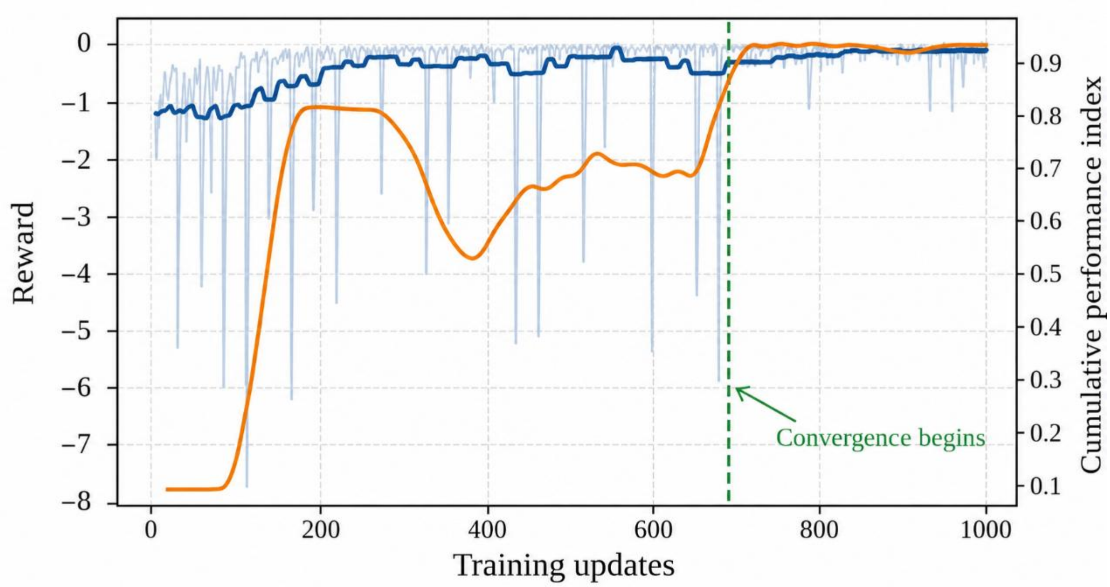


**Figure 4: Training process and convergence performance of the RL-OppOM strategy**

### 4.3 Full-Lifecycle Comprehensive Performance Comparison

**Table 4** summarizes the economic performance, energy production, operational availability, reliability, and health-risk indicators of CBM, OppM, and RL-OppOM over the simulation horizon. Figure 5 compares five normalized performance indicators: Overall, Economic return, Availability, Risk control, and Reliability. The Overall score is obtained by weighting the economic return, operational availability, system reliability, actual energy production, and risk-control capability, where the actual energy production is included in the Overall score, but is not displayed separately in **Fig. 5**. Under the specified multiobjective preference weights, RL-OppOM achieved an Overall score of 0.885, exceeding that of CBM (0.746) and OppM (0.702), indicating a better overall balance among lifecycle economic performance, availability, reliability, energy production, and risk control.

In terms of economic performance, RL-OppOM achieved a full-lifecycle net profit of 50.89 M CNY, representing improvements of 2.36% and 1.93% over that of CBM and OppM, respectively. Its actual energy production reached 75.27 GWh, which was 0.23% higher than that of CBM, but 0.09% lower than that of OppM. The full-lifecycle comprehensive cost of RL-OppOM was 7.32 M CNY, lower than the values of 8.49 M CNY of CBM and 8.29 M CNY of OppM, corresponding to reductions of 13.80% and 11.65%, respectively. These results indicated that the proposed strategy applied moderate power DR during high-load or health-degradation periods and coordinated maintenance timing, thereby reducing the long-term health-related energy losses and maintenance expenditure. Consequently, a higher lifecycle net profit was achieved despite the short-term energy loss associated with active DR.

In terms of turbine health and operational risk, the mean and minimum turbine health indicators of RL-OppOM were 0.6413 and 0.4491, respectively, both higher than those of the two benchmark strategies. Moreover, the number of high-risk exposure days decreased from 376 d for CBM and 521 d for OppM to zero. The operational availability of RL-OppOM was 0.9877, identical to that of CBM and slightly higher than the value of 0.9864 achieved by OppM. These results demonstrated that RL-OppOM maintained high operational availability while suppressing component degradation through power DR and coordinating subsequent maintenance interventions, thereby achieving a comprehensive lifecycle trade-off among turbine health, operational risk, economic return, and energy production.

**Table 4. Full-lifecycle performance comparison of different O&M strategies**

| Performance indicator | Unit | CBM | OppM | RL-OppOM |
|---|---|---|---|---|
| Potential energy production | GWh | 77.104 | 77.104 | 77.104 |
| Actual energy production | | 75.097 | 75.337 | 75.269 |
| Total energy loss | | 2.007 | 1.767 | 1.835 |
| Downtime energy loss | | 1.125 | 0.663 | 1.119 |
| Derating energy loss | | 0 | 0 | 0.346 |
| Health-related energy loss | | 0.882 | 1.104 | 0.371 |
| Maintenance cost | M CNY | 6.977 | 6.951 | 5.934 |
| Generation-loss cost | | 1.515 | 1.334 | 1.385 |
| Lifecycle comprehensive cost | | 8.492 | 8.285 | 7.320 |
| Lifecycle net profit | | 49.722 | 49.928 | 50.894 |
| Operational availability | – | 0.9877 | 0.9864 | 0.9877 |
| Mean turbine health indicator | | 0.5975 | 0.6016 | 0.6413 |
| Minimum turbine health indicator | | 0.1399 | 0.0985 | 0.4491 |

| | | | | |
|---|---|---|---|---|
| High-risk exposure days | day | 376 | 521 | 0 |
| Minor maintenance count | | 43 | 33 | 29 |
| Major maintenance count | – | 6 | 6 | 8 |
| Corrective maintenance count | | 0 | 0 | 0 |

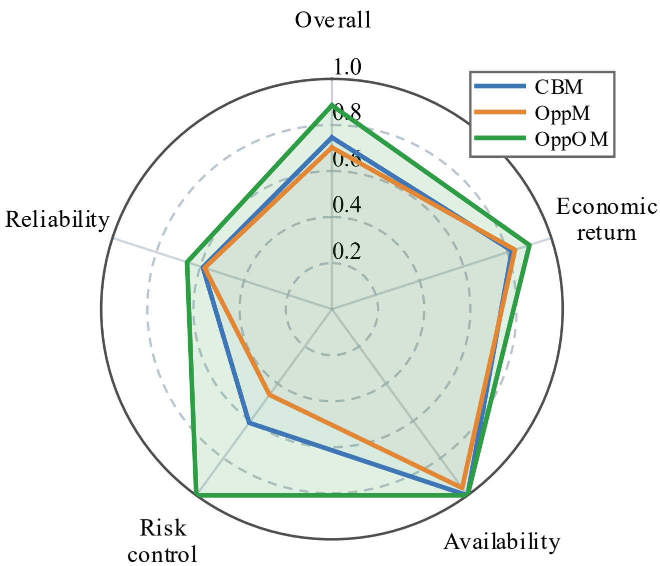


**Figure 5: Full-lifecycle comprehensive performance comparison of different O&M strategies**

### 4.4 Reliability Evolution and Maintenance-Decision Analysis

#### 4.4.1 Evolution of Turbine Health and Critical-Component Reliability

**Figure 6** illustrates the evolution of the turbine health indicator and reliabilities of the generator, gearbox, main bearing, and pitch system over the 19-year service life under the RL-OppOM strategy. As the operating time increased, the reliability of each component gradually decreased because of load exposure and cumulative degradation. Following maintenance intervention, the reliability of the maintained components was restored, accompanied by a corresponding recovery of the turbine health indicator. Over the entire lifecycle, the mean and minimum turbine health indicators were 0.6413 and 0.4491, respectively. The turbine health indicator remained above the high-risk threshold of 0.35 throughout the lifecycle, and no component failure occurred.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


At the component level, the four critical components exhibited distinct degradation patterns because of differences in lifetime parameters, load sensitivity, and post-maintenance restoration states. Some components experienced faster reliability degradation and entered the maintenance-candidate region earlier, whereas others exhibited relatively gradual degradation. The turbine health indicator was determined by the component with the lowest current reliability and therefore characterized the bottleneck health condition that limited safe turbine operation while identifying the weakest components at different stages of service life.

Maintenance actions were concentrated mainly in the periods when the reliability of the bottleneck component continued to decline or approached the maintenance-candidate region. By identifying health differences among components, the policy selectively determined the maintenance targets and maintenance intensity, facilitating timely reliability restoration of relatively degraded components and improving the turbine health.

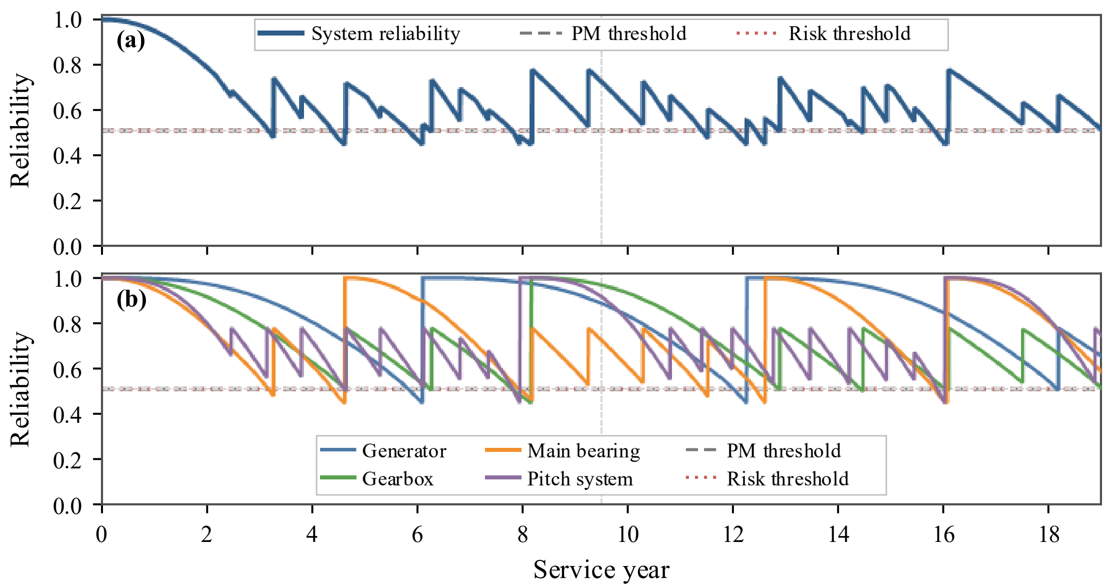


**Figure 6: Evolution of the turbine health indicator and critical-component reliability under the RL-OppOM strategy**

### 4.4.2 Lifecycle Maintenance Scheduling and Opportunistic-Maintenance Analysis

**Figure 7** illustrates the operating modes, maintenance actions, annual actual energy production, and energy losses of the RL-OppOM strategy over the 19-year service life. Maintenance activities were distributed nonuniformly throughout the lifecycle. During the early service period, the component health remained relatively high and maintenance actions were infrequent. As degradation accumulated, minor and major maintenance actions became more frequent in the middle and later stages, indicating that maintenance decisions were dynamically adapted to component health states and lifecycle stages, rather than being executed at fixed maintenance intervals.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


Power DR is mainly performed during periods of high load, health degradation, or approaching maintenance intervention to mitigate component degradation and provide temporal flexibility for maintenance scheduling. When multiple components satisfy the maintenance-candidate conditions and maintenance was feasible, the policy could coordinate them within the same maintenance opportunity for joint maintenance, thereby allowing the associated fixed access cost and downtime to be shared, thereby reducing the additional losses caused by repeated maintenance interventions.

Overall, RL-OppOM coordinated the DR intensity, maintenance timing, and component combinations to maintain relatively stable annual energy production under long-term degradation and maintenance interventions. These results demonstrated that the coordination between power DR and opportunistic maintenance could jointly balance component health preservation, maintenance-resource utilization, and long-term energy production.

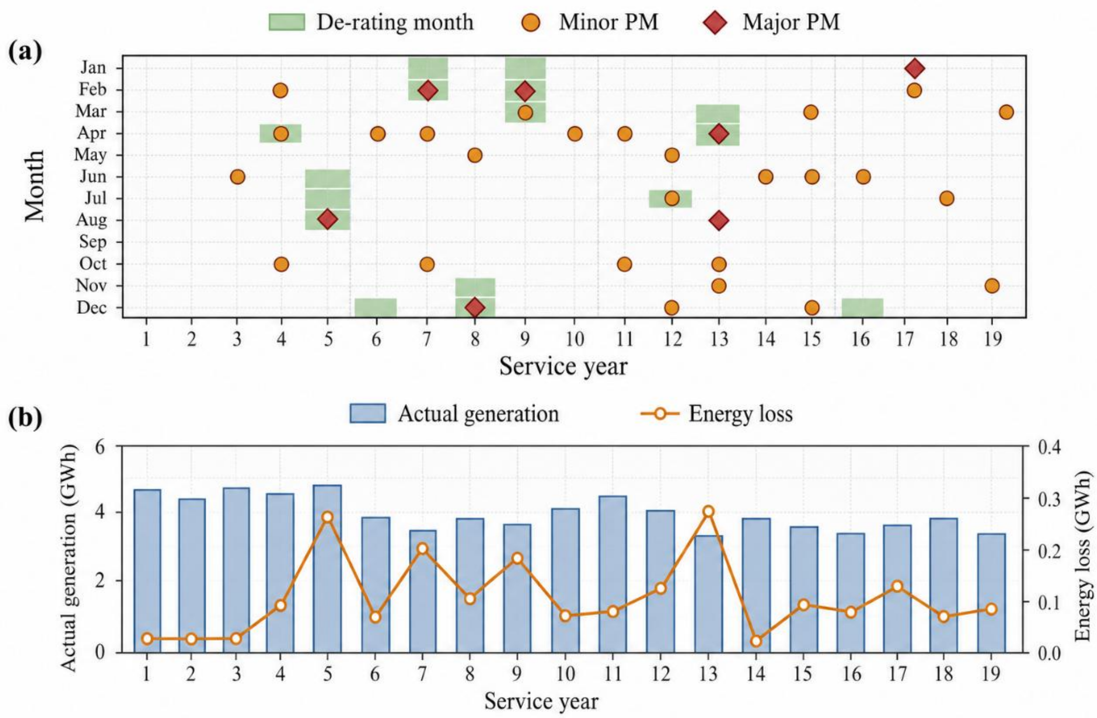


**Figure 7: Full-lifecycle operational performance of the RL-OppOM strategy: (a) full-lifecycle operation and maintenance scheduling; (b) annual actual energy production and energy loss**

### 4.5 Coordination Mechanism between Power DR and Maintenance

**Figure 8** illustrates the operating-control actions, maintenance interventions, and evolution of the turbine health indicator within a representative local time window under the RL-OppOM strategy. This window corresponded to a period over which the bottleneck component continued to degrade but had not yet entered the high-risk region. It was used to examine how the strategy coordinated short-term power DR with subsequent maintenance interventions.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


At the beginning of the window, the turbine health indicator was 0.4686. During mild DR, the indicator continued to decrease and reached a minimum of 0.4495, while remaining above the high-risk threshold of 0.35. Subsequently, the policy executed an appropriate maintenance action according to the component health states and maintenance conditions, restoring the turbine health indicator to 0.7799, which remained at 0.7714 at the end of the window. This process demonstrated that power DR mitigated the premaintenance degradation accumulation by reducing the operating load, whereas maintenance interventions restored component health states.

Therefore, the coordination of RL-OppOM was reflected in the temporal coupling between the operating control and maintenance intervention. When maintenance conditions were not yet satisfied, the strategy applied moderate power DR to control the operational risk, and subsequently, performed the appropriate maintenance action once maintenance became feasible, thereby avoiding sole reliance on either premature shutdown or delayed maintenance.

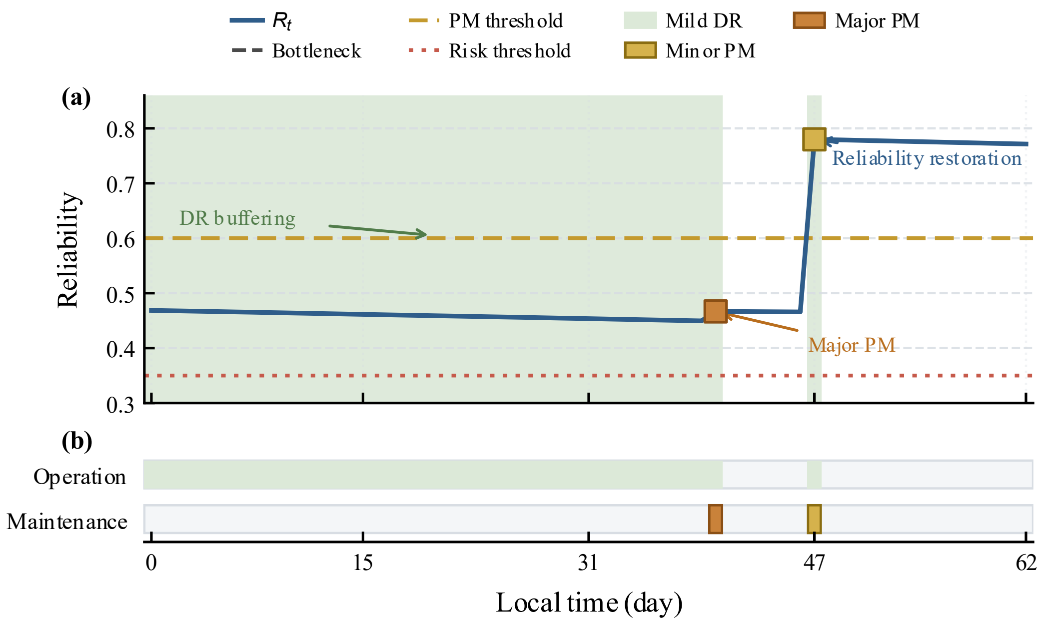


**Figure 8 : Coordinated power-derating and maintenance decisions with turbine-health evolution within a representative time window**

### 4.6 Energy-Loss and Cost-Structure Analysis

**Figure 9** compares the full-lifecycle energy losses and cost structures of CBM, OppM, and RL-OppOM. The total energy losses of the three strategies were 2.01, 1.77, and 1.83 GWh, respectively. For RL-OppOM, the downtime, DR, and health-related energy losses were 1.12, 0.35, and 0.37 GWh, respectively. Compared with CBM and OppM, RL-OppOM reduced the health-related energy loss by 57.95% and 66.36%, respectively. However, owing to the additional energy loss introduced by active power DR, its total energy loss remained slightly higher than that of OppM.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


In terms of cost, the direct maintenance cost and operational-loss cost of RL-OppOM were 5.93 and 1.39 M CNY, respectively, resulting in a full-lifecycle comprehensive cost of 7.32 M CNY. In comparison, the comprehensive costs of CBM and OppM were 8.49 and 8.29 M CNY, respectively. RL-OppOM, therefore, reduces the comprehensive cost by 13.78% and 11.70%, respectively, relative to the two benchmark strategies.

These results demonstrated that RL-OppOM coordinated the trade-offs among DR loss, health-related energy loss, and maintenance expenditure. Although active power DR introduced a certain amount of additional energy loss, it substantially reduced the health-related energy loss and direct maintenance expenditure, thereby achieving a lower full-lifecycle comprehensive cost.

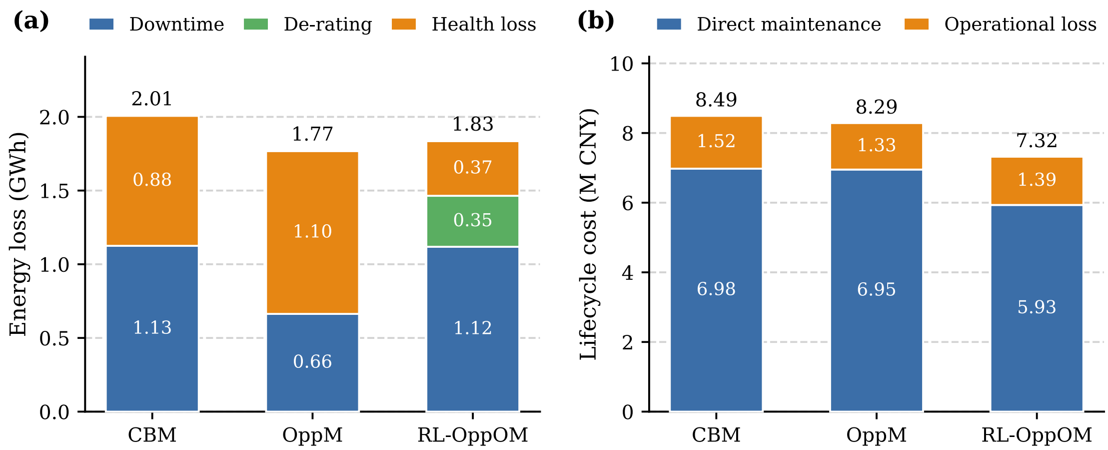


**Figure 9 : Full-lifecycle energy-loss composition and cost structures of different O&M strategies**

### 4.7 Multiscenario Sensitivity and Robustness Analysis

#### 4.7.1 Sensitivity Analysis of Multiobjective Preference Weights

**Figure 10** compares the normalized performance of RL-OppOM under three preference-weight configurations: baseline weighting, cost-weight–enhanced weighting, and health-risk–priority weighting. Under the baseline configuration, the scores for Overall performance, Economic return, Availability, Reliability, and Risk control were 0.99, 0.98, 0.99, 0.98, and 1.00, respectively. The five indicators remained relatively balanced, and the Overall performance was the highest among that of the three configurations, indicating that the baseline weights provided a favorable trade-off among the economic performance, operational availability, system reliability, and risk control.

Under the cost-weight–enhanced configuration, the Economic return and Availability scores both increased to 1.00, while Risk control remained at 1.00. However, the Reliability score decreased from 0.98 to 0.95, and the Overall performance slightly decreased from 0.99 to 0.98. These results indicated that increasing the emphasis on economic objectives could improve economic return and availability by reducing some maintenance interventions with limited economic benefit; however, this improvement is accompanied by a modest reduction in system reliability.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


Under the health-risk–priority configuration, the Reliability score increased to 1.00, whereas the Economic return and Risk control scores decreased to 0.94 and 0.60, respectively. Availability remained relatively high at 0.99, whereas the Overall performance decreased to 0.88. This result indicated that excessive emphasis on health-related objectives could improve the average system reliability, but alter the allocation and timing of minor and major maintenance actions, leading to greater exposure of the bottleneck component to high-risk operating conditions. Consequently, the improvement in system reliability did not necessarily translate into better risk-control performance. Overall, baseline weighting provided the most balanced full-lifecycle performance among the three configurations.

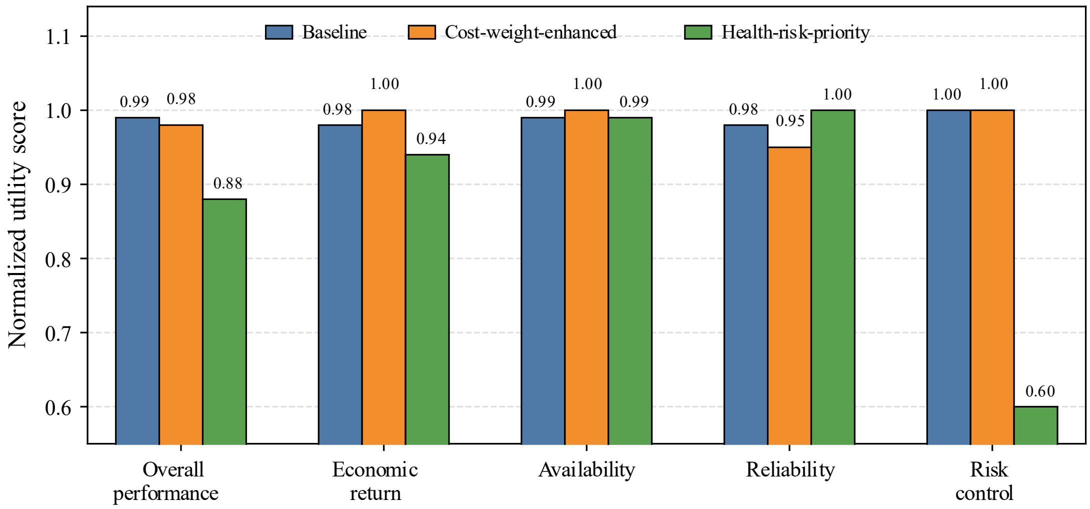


**Figure 10 : Sensitivity of comprehensive performance to different multiobjective preference weights**

### 4.7.2 Sensitivity Analysis of Power-DR Benefit Parameters

**Figure 11** presents the performance of RL-OppOM under multiple power-DR parameter settings. Three scenarios with low, baseline, and high DR benefits were considered, with the corresponding parameter combinations set to (0.5, 1.3), (1.0, 1.0), and (1.5, 0.7), respectively. These scenarios were used to quantitatively investigate how the trade-off between the benefit of power DR and the associated generation loss affected operation and maintenance decisions.

From the full-lifecycle performance results, the net profits of the three scenarios remained at comparable high levels. The low-DR-benefit scenario yielded a net profit of 50.82 M CNY, the baseline scenario achieved the highest net profit of 50.89 M CNY, and the high-DR-benefit scenario exhibited a slightly decreased net profit of 50.67 M CNY. The mean system reliability increased monotonically from 0.638 to 0.641 and 0.651 across the low, baseline, and high DR-benefit scenarios, respectively, indicating that greater emphasis on DR improved the long-term system reliability. Moreover, no high-risk operating periods or component failures occurred in any of the three scenarios, demonstrating that RL-OppOM maintained effective operational-risk control across different parameter settings.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


The preventive-maintenance statistics further revealed that the DR parameters substantially affected maintenance frequency. The low-DR-benefit scenario resulted in 51 preventive maintenance actions, including 45 minor and 6 major maintenance actions. The baseline scenario required the fewest maintenance actions—37 in total, comprising 29 minor and 8 major maintenance actions. Under the high-DR-benefit scenario, the maintenance frequency increased to 62 actions, including 56 minor and 6 major maintenance actions. These results indicated that variations in the DR parameters altered the intensity and allocation of maintenance interventions, while all three scenarios avoided unexpected component failures.

Overall, RL-OppOM maintained stable economic performance and effective risk control under different DR-benefit settings, whereas the system reliability and maintenance intensity varied with the parameter configuration. The baseline setting achieved the highest net profit with the lowest maintenance frequency, whereas the high-DR-benefit setting provided the highest mean system reliability. These results demonstrated the adaptability of the proposed strategy to variations in the DR-benefit parameters and revealed the trade-off among economic return, system reliability, and maintenance-resource utilization.

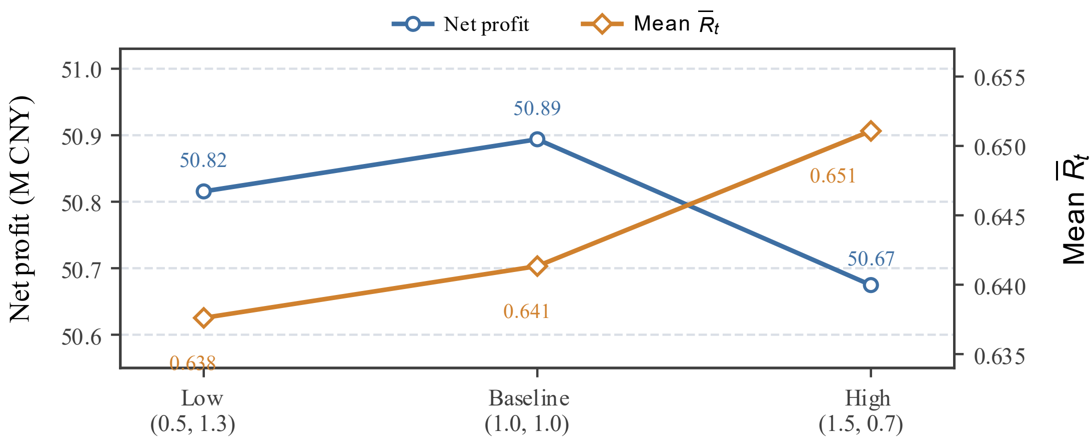


**Figure 11. Performance response and policy adaptation of RL-OppOM under different power-DR parameter settings**

## 5 Conclusions

This study proposed an RL-based collaborative optimization method for wind-turbine power DR and opportunistic maintenance. Graded power DR, multicomponent maintenance combinations, reliability evolution, and field operational constraints were integrated into a unified sequential decision-making framework. An FAM-DC-PPO algorithm was employed to solve the high-dimensional joint O&M optimization problem, producing a long-term dynamic balance among the power-generation revenue, O&M expenditure, equipment health, and operational risk. The main conclusions were as follows:

(1) A physics-driven simulation environment integrating continuous wind conditions, graded power DR, heterogeneous component degradation, and overall turbine reliability was developed. The environment characterized the lifecycle evolution

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026




of multicomponent health states and system-level risks, providing a unified basis for training and evaluating collaborative O&M strategies.

(2) FAM-DC-PPO integrated policy factorization, action masking, and dual clipping to solve the joint decision-making problem while enforcing the feasible action space. The training performance improved progressively with policy updates and approached a stable state after approximately 700 updates.

(3) RL-OppOM achieved a full-lifecycle comprehensive cost of 7.320 M CNY, representing reductions of 13.80% and 11.65% compared with that of CBM and Rule-based OppM, respectively, while increasing the lifecycle net profit by 2.36% and 1.93%. The mean and minimum turbine health indicators reached 0.6413 and 0.4491, respectively, and the number of high-risk exposure days was reduced to zero. These results demonstrated that the proposed strategy effectively balanced lifecycle economic performance, health preservation, and operational risk control.

The proposed RL-OppOM method coordinated power DR with multicomponent opportunistic maintenance, allowing operating control to actively participate in suppressing component degradation while optimizing maintenance timing according to equipment states and operational conditions. Future work will extend the collaborative optimization framework into a digital-twin system capable of iterative updates synchronized with turbine operating states. Particular attention will be paid to uncertainties in equipment-health assessment and wind-condition forecasting, coupled degradation mechanisms involving corrosion fatigue and random shocks, and resource constraints associated with maintenance personnel, spare parts, and maintenance equipment, with the aim of improving the applicability and decision reliability of the proposed method under complex operating environments.

## Code and data availability

The SCADA data used in this study are not publicly available due to confidentiality restrictions but may be made available upon reasonable request, subject to permission from the data owner. The code and simulation data supporting the findings of this study are available from the corresponding author upon reasonable request.

## Author contributions

**Jing Shi:** Writing – original draft, Validation, Project administration, Methodology, Formal analysis, Data curation. **Wei Chen:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.**Jie Lin:** Writing – review & editing, Supervision. **Jianghao Zhu:** Writing – review & editing, Supervision.**Xubin Li:** Writing – review & editing, Supervision.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

## Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors acknowledge the financial support from the National Natural Science Foundation of China (Grant No 52667011 and 62666044), Gansu Province Joint Research Foundation Major Program (25JRRA1143), and Gansu Province Key Research and Development Program (25YFGA032). The authors would also like to thank the colleagues for valuable suggestions during this research.

## Financial support

This work was supported by the National Natural Science Foundation of China (Grant No 52667011 and 62666044), Gansu Province Joint Research Foundation Major Program (25JRRA1143), and Gansu Province Key Research and Development Program (25YFGA032).

## Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

## References

Aafif, Y., Chelbi, A., Mifdal, L., Dellagi, S., and Majdouline, I.: Optimal preventive maintenance strategies for a wind turbine gearbox, Energy Reports, 8, 803–814, https://doi.org/10.1016/j.egyr.2022.07.084, 2022.

Badihi, H., Zhang, Y., Jiang, B., Pillay, P., and Rakheja, S.: A Comprehensive Review on Signal-Based and Model-Based Condition Monitoring of Wind Turbines: Fault Diagnosis and Lifetime Prognosis, Proceedings of the IEEE, 110, 754–806, https://doi.org/10.1109/JPROC.2022.3171691, 2022.

Capaldo, M. and Mella, P.: Damping analysis of floating offshore wind turbines (FOWTs): a new control strategy reducing the platform vibrations, Wind Energy Science, 8, 1319–1339, https://doi.org/10.5194/wes-8-1319-2023, 2023.

Chatterjee, J. and Dethlefs, N.: Scientometric review of artificial intelligence for operations & maintenance of wind turbines: The past, present and future, Renewable and Sustainable Energy Reviews, 144, 111051, https://doi.org/10.1016/j.rser.2021.111051, 2021.

Chen, Q., Cao, J., and Zhu, S.: Data-Driven Monitoring and Predictive Maintenance for Engineering Structures: Technologies, Implementation Challenges, and Future Directions, IEEE Internet of Things Journal, 10, 14527–14551, https://doi.org/10.1109/JIOT.2023.3272535, 2023.

Coraddu, A., Oneto, L., Walker, J., Patryniak, K., Prothero, A., and Collu, M.: Floating offshore wind turbine mooring line sections health status nowcasting: From supervised shallow to weakly supervised deep learning, Mechanical Systems and Signal Processing, 216, 111446, https://doi.org/10.1016/j.ymssp.2024.111446, 2024.

Cuesta, J., Leturiondo, U., Vidal, Y., and Pozo, F.: A review of prognostics and health management techniques in wind energy, Reliability Engineering & System Safety, 260, 111004, https://doi.org/10.1016/j.ress.2025.111004, 2025.

Garcia-Teruel, A., Rinaldi, G., Thies, P. R., Johanning, L., and Jeffrey, H.: Life cycle assessment of floating offshore wind farms: An evaluation of operation and maintenance, Applied Energy, 307, 118067, https://doi.org/10.1016/j.apenergy.2021.118067, 2022.

Kim, S., Kim, D., and Kim, B.: Effect of multiple load reduction control systems on the ultimate load and fatigue load of 4 MW class wind turbine, Journal of Renewable and Sustainable Energy, 12, 2020.

Kipchirchir, E., Do, M. H., Njiri, J. G., and Söffker, D.: Prognostics-based adaptive control strategy for lifetime control of wind turbines, Wind Energy Science, 8, 575–588, https://doi.org/10.5194/wes-8-575-2023, 2023.

Lee, N., Woo, J., and Kim, S.: A deep reinforcement learning ensemble for maintenance scheduling in offshore wind farms, Applied Energy, 377, 124431, https://doi.org/10.1016/j.apenergy.2024.124431, 2025.

Li, M., Jiang, X., Carroll, J., and Negenborn, R. R.: A multi-objective maintenance strategy optimization framework for offshore wind farms considering uncertainty, Applied Energy, 321, 119284, https://doi.org/10.1016/j.apenergy.2022.119284, 2022.

Li, M., Jiang, X., Carroll, J., and Negenborn, R. R.: A closed-loop maintenance strategy for offshore wind farms: Incorporating dynamic wind farm states and uncertainty-awareness in decision-making, Renewable and Sustainable Energy Reviews, 184, 113535, https://doi.org/10.1016/j.rser.2023.113535, 2023.

Liu, M., Fang, S., Dong, H., and Xu, C.: Review of digital twin about concepts, technologies, and industrial applications, Journal of Manufacturing Systems, 58, 346–361, https://doi.org/10.1016/j.jmsy.2020.06.017, 2021.

McKenna, R., Pfenninger, S., Heinrichs, H., Schmidt, J., Staffell, I., Bauer, C., Gruber, K., Hahmann, A. N., Jansen, M., Klingler, M., Landwehr, N., Larsén, X. G., Lilliestam, J., Pickering, B., Robinius, M., Tröndle, T., Turkovska, O., Wehrle, S., Weinand, J. M., and Wohland, J.: High-resolution large-scale onshore wind energy assessments: A review of potential definitions, methodologies and future research needs, Renewable Energy, 182, 659–684, https://doi.org/10.1016/j.renene.2021.10.027, 2022.

McMorland, J., Collu, M., McMillan, D., and Carroll, J.: Operation and maintenance for floating wind turbines: A review, Renewable and Sustainable Energy Reviews, 163, 112499, https://doi.org/10.1016/j.rser.2022.112499, 2022.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


McMorland, J., Collu, M., McMillan, D., Carroll, J., and Coraddu, A.: Opportunistic maintenance for offshore wind: A review and proposal of future framework, Renewable and Sustainable Energy Reviews, 184, 113571, https://doi.org/10.1016/j.rser.2023.113571, 2023.

Merainani, B., Laddada, S., Bechhoefer, E., Chikh, M. A. A., and Benazzouz, D.: An integrated methodology for estimating the remaining useful life of high-speed wind turbine shaft bearings with limited samples, Renewable Energy, 182, 1141–1151, https://doi.org/10.1016/j.renene.2021.10.062, 2022.

Najafi, S. and Lee, C.-G.: A deep reinforcement learning approach for repair-based maintenance of multi-unit systems using proportional hazards model, Reliability Engineering & System Safety, 234, 109179, https://doi.org/10.1016/j.ress.2023.109179, 2023.

Njiri, J. G., Beganovic, N., Do, M. H., and Söffker, D.: Consideration of lifetime and fatigue load in wind turbine control, Renewable Energy, 131, 818–828, https://doi.org/10.1016/j.renene.2018.07.109, 2019.

Ogunfowora, O. and Najjaran, H.: Reinforcement and deep reinforcement learning-based solutions for machine maintenance planning, scheduling policies, and optimization, Journal of Manufacturing Systems, 70, 244–263, https://doi.org/10.1016/j.jmsy.2023.07.014, 2023.

Oh, S. Y., Joung, C., Lee, S., Shim, Y.-B., Lee, D., Cho, G.-E., Jang, J., Lee, I. Y., and Park, Y.-B.: Condition-based maintenance of wind turbine structures: A state-of-the-art review, Renewable and Sustainable Energy Reviews, 204, 114799, https://doi.org/10.1016/j.rser.2024.114799, 2024.

Østergaard, P. A., Duic, N., Noorollahi, Y., and Kalogirou, S.: Advances in renewable energy for sustainable development, Renewable Energy, 219, 119377, https://doi.org/10.1016/j.renene.2023.119377, 2023.

Pacheco-Blazquez, R., Garcia-Espinosa, J., Di Capua, D., and Pastor Sanchez, A.: A Digital Twin for Assessing the Remaining Useful Life of Offshore Wind Turbine Structures, Journal of Marine Science and Engineering, 12, 573, https://doi.org/10.3390/jmse12040573, 2024.

Papadopoulos, P., Coit, D. W., and Ezzat, A. A.: Seizing Opportunity: Maintenance Optimization in Offshore Wind Farms Considering Accessibility, Production, and Crew Dispatch, IEEE Transactions on Sustainable Energy, 13, 111–121, https://doi.org/10.1109/TSTE.2021.3104982, 2022.

Pliego Marugán, A., Pinar-Pérez, J. M., and García Márquez, F. P.: A reinforcement learning agent for maintenance of deteriorating systems with increasingly imperfect repairs, Reliability Engineering & System Safety, 252, 110466, https://doi.org/10.1016/j.ress.2024.110466, 2024.

Ren, Z., Verma, A. S., Li, Y., Teuwen, J. J. E., and Jiang, Z.: Offshore wind turbine operations and maintenance: A state-of-the-art review, Renewable and Sustainable Energy Reviews, 144, 110886, https://doi.org/10.1016/j.rser.2021.110886, 2021.

Sahin, M. and Yavrucuk, I.: Adaptive envelope protection control of wind turbines under varying operational conditions, Energy, 247, 123544, https://doi.org/10.1016/j.energy.2022.123544, 2022.

Sarker, B. R. and Faiz, T. I.: Minimizing maintenance cost for offshore wind turbines following multi-level opportunistic preventive strategy, Renewable Energy, 85, 104–113, https://doi.org/10.1016/j.renene.2015.06.030, 2016.

https://doi.org/10.5194/wes-2026-157
Preprint. Discussion started: 10 September 2026


Sun, Y. and Yin, J.: Dynamic condition-based maintenance for shock systems based on damage evolutions using deep reinforcement learning, Reliability Engineering & System Safety, 261, 111095, https://doi.org/10.1016/j.ress.2025.111095, 2025.

Tang, H., Wang, H., and Li, C.: Time-varying cost modeling and maintenance strategy optimization of plateau wind turbines considering degradation states, Applied Energy, 377, 124464, https://doi.org/10.1016/j.apenergy.2024.124464, 2025.

Truong, H. V. A., Dang, T. D., Vo, C. P., and Ahn, K. K.: Active control strategies for system enhancement and load mitigation of floating offshore wind turbines: A review, Renewable and Sustainable Energy Reviews, 170, 112958, https://doi.org/10.1016/j.rser.2022.112958, 2022.

Valet, A., Altenmüller, T., Waschneck, B., May, M. C., Kuhnle, A., and Lanza, G.: Opportunistic maintenance scheduling with deep reinforcement learning, Journal of Manufacturing Systems, 64, 518–534, https://doi.org/10.1016/j.jmsy.2022.07.016, 2022.

VanDerHorn, E. and Mahadevan, S.: Digital Twin: Generalization, characterization and implementation, Decision Support Systems, 145, 113524, https://doi.org/10.1016/j.dss.2021.113524, 2021.

Xia, J. and Zou, G.: Operation and maintenance optimization of offshore wind farms based on digital twin: A review, Ocean Engineering, 268, 113322, https://doi.org/10.1016/j.oceaneng.2022.113322, 2023.

Zhao, C., Liu, H., Qu, X., Liu, M., Tang, R., and Xie, A.: Reliability analysis of floating offshore wind turbine generator based on failure prediction and preventive maintenance, Ocean Engineering, 288, 116089, https://doi.org/10.1016/j.oceaneng.2023.116089, 2023.

Zhao, X., Guo, B., and Chen, Y.: A condition-based inspection-maintenance policy for critical systems with an unreliable monitor system, Reliability Engineering & System Safety, 242, 109710, https://doi.org/10.1016/j.ress.2023.109710, 2024.

Zhu, J., Chen, W., Su, L., Lan, B., Pei, T., and Jin, L.: Predictive maintenance for wind turbines: A physics-driven reinforcement learning strategy with economic-reliability collaborative optimization, Energy and AI, 22, 100620, https://doi.org/10.1016/j.egyai.2025.100620, 2025.