

Article

Optimization Solution for Unit Power Generation Plan Based on the Integration of Constraint Identification and Deep Reinforcement Learning

Dan Li ¹ , Lei Zhang ^{1,*}, Ning Mi ² and Hailiang Zhong ²

¹ College of Electrical and New Energy, Three Gorges University, Yichang 443002, China; 202308080011015@ctgu.edu.cn

² State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750002, China; 18371709188@163.com (N.M.); 18742519093@163.com (H.Z.)

* Correspondence: zhanglei3188@ctgu.edu.cn; Tel.: +86-15102713663

Abstract

In response to the complexity of renewable energy and the numerous safety constraints in actual power grid scenarios, which result in a large model size and difficulties in developing rapid solutions, this paper proposes an accelerated algorithm for solving the optimization of large-scale unit generation plans by combining deep reinforcement learning and security constraint identification. Firstly, this paper constructs an optimization model of a unit generation plan and incorporates conditional risk values to quantify the risk cost caused by operational uncertainty. Secondly, this paper uses a stacked noise-reduction automatic encoder to identify the effective constraint set in the optimization model of the power generation plan. Then, this paper transforms the model into Markov decision processes, designs a reward mechanism with the identified constraints, and uses the proximal policy optimization algorithm to solve it. Finally, this paper takes IEEE30 and a regional power grid in northwest China as examples and performs simulation analyses in various scenarios. The results show that it can greatly reduce the model training time, and the application effect on large-scale systems is obvious. In particular, the online solution time is effectively reduced by 15,837.09 s.

Keywords: deep reinforcement learning; safety constraint identification; power generation plan optimization; conditional value at risk; stacked de-noising autoencoder; proximal policy optimization algorithm



Academic Editor: Jiaqiang E

Received: 9 October 2025

Revised: 19 November 2025

Accepted: 20 November 2025

Published: 22 November 2025

Citation: Li, D.; Zhang, L.; Mi, N.; Zhong, H. Optimization Solution for Unit Power Generation Plan Based on the Integration of Constraint Identification and Deep Reinforcement Learning. *Processes* **2025**, *13*, 3778. <https://doi.org/10.3390/pr13123778>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As the scale of the power grid continues to expand, the number of generating units and transmission lines in the power system has increased sharply. The optimization and dispatching of power generation plans not only need to take into account the operational constraints of various types of resources but also need to consider network structure constraints, section constraints, and safety constraints. Additionally, with the large-scale integration of renewable energy sources, the problem of unit commitment in power generation planning has become more difficult to solve due to the significant fluctuations and randomness involved [1]. Meanwhile, with the increasing requirements for power optimization dispatching, daily and intraday generation plans need to be solved in a shorter period of time. The current simple invocation of commercial mathematical optimization software can no longer meet the requirements of industrial applications, and traditional

methods have difficulties achieving adaptive optimization. Thus, the current methods are unable to effectively support the needs of future practical engineering. Therefore, in the design of the algorithm process, human intelligent thinking patterns must be deeply integrated to avoid blindness in the optimization process, and to enhance the efficiency of solving large-scale power generation plans through intelligence.

Traditional solution methods include heuristic algorithms, mathematical optimization algorithms, and intelligent optimization algorithms [2], but the convergence efficiency for large-scale problems still needs to be improved. To further enhance solution efficiency, the researchers dealt with a large number of redundant constraints in the optimization model. In addition to the traditional manual experience-based presetting and iterative addition, the current research has combined deep learning methods to achieve more efficient and accurate constraint identification. For instance, reference [3] proposed a method for identifying active safety constraints based on graph neural networks. Reference [4] presented a Gaussian classifier with characteristics suitable for small-sample learning. Reference [5] employed a combined neural network to identify and reduce redundant constraints. However, the above methods still employ traditional approaches for the solution process, without deeply exploring the collaborative control of various scheduling resources in the power system. Moreover, their solution adaptability is poor, and it becomes difficult to establish an accurate rapid solution model as system complexity and uncertainty increase.

In recent years, reinforcement learning has gained more attention from researchers for solving complex, high-dimensional, and nonlinear problems [6]. Reference [7] established an optimization scheduling model applicable to the combined electric and thermal system according to the deep reinforcement learning algorithm. Reference [8] solved the dynamic economic operation model under the conditions of an unknown generation cost function by using distributed reinforcement learning. Reference [9] adopts deep reinforcement learning for optimization, achieving multi-objective collaborative optimization under the condition of the synchronous control of different devices. Reference [10] utilized a multi-agent deep reinforcement learning approach to optimize the dispatch of distributed power sources, achieving the adaptive handling of uncertainty in source-load conditions. However, traditional reinforcement learning methods have poor decision-making performances and the problem of dimensionality disaster in solving power generation plans, which increases the difficulty of model convergence. Moreover, the aforementioned studies have not further considered the feature extraction and learning optimization issues in the power generation plan optimization model.

Thus, this paper proposes a fast solution method for large-scale unit power generation plan optimization models by integrating deep reinforcement learning and effective constraint identification. Firstly, this paper constructs a two-layer power generation plan optimization model considering the conditional value at risk (CVaR); then, it uses a stacked de-noising auto-encoder (SDAE) to achieve the identification of the model's operational constraints, and combines the identified safety constraints with the dispatch cost and risk cost to design a reward function mechanism; subsequently, based on the proximal policy optimization (PPO) algorithm, this paper guides the intelligent agent to generate a dispatch strategy that satisfies both economic and safety requirements. Finally, this paper conducts a case simulation analysis using the IEEE30 system and an actual equivalent system in northwest China to verify the effectiveness of the strategy.

2. A Joint Optimization Model Considering the Output Limitations of Thermal Power Units and the Fluctuation Risks of Wind and Solar Energy

2.1. Joint Optimization Model

This paper constructs a joint optimization model. The upper layer is a hypothetical scenario unit combination aimed at minimizing the operating costs of the units and the energy storage systems. The lower layer is an error scenario backup optimization decision model aimed at minimizing the backup costs, the power-off cost, and the penalty costs for lost loads. Moreover, this paper introduces the conditional value at risk (CVaR) to quantify the risk costs generated by operational uncertainties [11], thereby formulating a more comprehensive, economical, and safe unit power generation optimization scheduling plan. The optimization objective of this model is shown in Figure 1, as follows, and it includes the total scheduling cost and the total risk cost.

$$\begin{aligned} \min F &= (1 - \lambda)F^{OPE} + \lambda F^{RSK} \\ F^{OPE} &= f^{uc} + f^{bd} \\ F^{RSK} &= k_1 \times CVaR_{WR} + k_2 \times CVaR_{PL} \end{aligned} \quad (1)$$

Here, F^{OPE} represents the total dispatching cost, which includes the operating cost of the units and the optimization cost of the standby; F^{RSK} represents the total risk cost, including the risk value under the condition of power outages and load shedding; k_1 and k_2 are, respectively, the risk coefficients for power outages and load loss; λ is the total risk penalty coefficient, with a value ranging from 0 to 1. The larger the value, the higher the risk penalty.

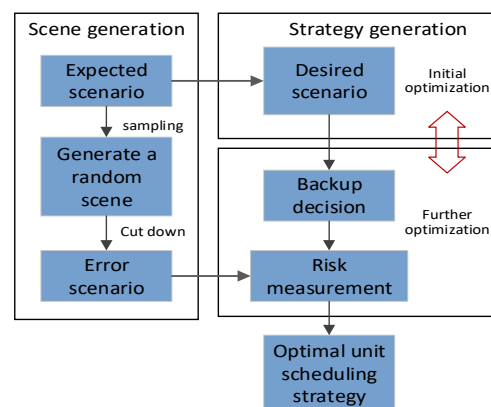


Figure 1. Joint optimization model.

2.1.1. Unit Combination Optimization

In this model, the unit combination optimization part aims to obtain the conventional units output plan based on the predictive scenarios.

1. The objective function for the unit combination optimization is as follows:

$$f^{uc} = \sum_t \left(\sum_g \left(C_g P_{g,t} + C_g^U I_{g,t}^U + C_g^D I_{g,t}^D \right) + \sum_h \left(C_h P_{h,t} \right) + \sum_s \left(C_s^c P_{s,t}^c + C_s^d P_{s,t}^d \right) \right) \quad (2)$$

In the formula, C_g , C_h represent the marginal power generation cost quotations for the thermal power unit g and the hydroelectric power unit h , respectively; $P_{g,t}$, $P_{h,t}$ are the operating outputs of g and h at time t ; C_g^U , C_g^D are the start-up and shutdown costs of g ; $I_{g,t}^U$, $I_{g,t}^D$ are the start-up and shutdown variables of g . C_s^c , C_s^d are the unit costs for the charging and discharging of energy storage s ; $P_{s,t}^c$, $P_{s,t}^d$ are the charging and discharging powers of s .

2. The constraints for the unit combination optimization are as follows:

(1) The power balance constraint

$$\sum_g P_{g,t} + \sum_h P_{h,t} + \sum_w P_{w,t} + \sum_v P_{v,t} + \sum_s P_{s,t}^d - \sum_s P_{s,t}^c = \sum_e P_{e,t}^L \quad (3)$$

In the formula, $P_{w,t}$, $P_{v,t}$ respectively represent the predicted power output of wind power and photovoltaic power at time t ; $P_{e,t}^L$ is the load demand.

(2) The operating constraints of thermal power units

$$\begin{cases} u_{g,t} P_{g,\min} \leq P_{g,t} \leq u_{g,t} P_{g,\max} \\ P_{g,t} - P_{g,t-1} + I_{g,t-1}^{on} (P_{g,\min} - r_g^u P_{g,\max}) + I_{g,t}^{on} (P_{g,\max} - P_{g,\min}) \leq P_{g,\max} \\ P_{g,t-1} - P_{g,t} + I_{g,t}^{on} (P_{g,\min} - r_g^d P_{g,\max}) + I_{g,t-1}^{on} (P_{g,\max} - P_{g,\min}) \geq P_{g,\min} \\ I_{g,t}^U - I_{g,t}^D = I_{g,t}^{on} - I_{g,t-1}^{on}, I_{g,t}^U + I_{g,t}^D \leq 1 \\ I_{g,t}^{on} - I_{g,t-1}^{on} \leq I_{g,r}^{on} \leq 1 + I_{g,t}^{on} - I_{g,t-1}^{on} \\ C_{g,t}^U = \lambda_g^U I_{g,t}^U, C_{g,t}^D = \lambda_g^D I_{g,t}^D \end{cases} \quad (4)$$

In the formula, $u_{g,t}$ is the indicator variable for the operation of g . When it is 0, it indicates a fault; when it is 1, it indicates normal operation. r_g^u , r_g^d represent the upward and downward ramp rates of g correspondingly; $I_{g,r}^{on}$ is the minimum interval between unit start-up and shutdown; $I_{g,t}^{on}$ represents the start-up flag of unit g ; λ_g^U , λ_g^D represent the unit start-up and shutdown costs of g respectively.

(3) New constraints for thermal power units

1) The fluctuation constraints on the output range of the unit.

Considering the future fluctuations in the power generation capacity of units, this paper sets the upper and lower boundaries of the output of units within a certain fluctuation range [12], as shown in the following equation.

$$\begin{aligned} \lambda_{floor}^{\max} \times P_{g,t}^{\max} &\leq P_{g,\max} \leq \lambda_{cei}^{\max} \times P_{g,t}^{\max} \\ \lambda_{floor}^{\min} \times P_{g,t}^{\min} &\leq P_{g,\min} \leq \lambda_{cei}^{\min} \times P_{g,t}^{\min} \end{aligned} \quad (5)$$

In the formula, $P_{g,\min}$, $P_{g,\max}$ represent the actual minimum and maximum output values of g respectively; $\lambda_{floor}^{\min}$, $\lambda_{floor}^{\max}$, $\lambda_{cei}^{\min}$, $\lambda_{cei}^{\max}$ respectively represent the allowable fluctuation range of the ratio. The specific values can be obtained based on actual data.

2) Non-planned outage constraints of units.

Due to the long-term operation of units and the need to constantly adjust their operating status to meet dispatch requirements, this exhausting state can lead to certain potential faults during operation [13]. Therefore, this paper, considering the different outage accident probabilities of generators of different capacities in the power system, proposes unplanned outage constraints and linearizes them [14], resulting in the following equation.

$$FT_G = \begin{cases} \sum_{g=1}^G u_{g,t} \geq NG - KG, u_{g,t} \in \{0,1\}, \forall t \\ \sum_{g=1}^{NG} u_{g,t} \times (\ln(1 - T_{g,t}) - \ln(T_{g,t})) \geq \ln(Th_f) \end{cases} \quad (6)$$

In the formula, NG represents the number of units; KG is the maximum number of units that can be shut down; Th_f represents the set non-shutdown threshold. $T_{g,t}$ is the unplanned shutdown probability of g . Correspondingly, the occurrence of unplanned shutdowns simultaneously for various units can be adjusted by setting Th_f . As the value increases, the situations where multiple units with lower non-shutdown probabilities simultaneously experience unplanned shutdowns become less frequent.

(4) Operation constraints of hydropower units.

The output constraints of hydroelectric units, as well as the upper/lower ramp constraints, are similar to those of thermal power units and are described as above. Additionally, it also includes the conversion constraints of hydropower and the water volume constraints, as shown in the following equation.

$$\begin{cases} P_{h,t} = 0.00981 \eta^H H_{h,t} Q_{h,t} \\ Q_{h,\min} \leq \sum_h \sum_t Q_{h,t} \leq Q_{h,\max} \end{cases} \quad (7)$$

In this formula, $H_{h,t}$ is the corresponding water head height for power generation; $Q_{h,t}$ is the corresponding power generation flow rate; η^H is the power generation efficiency coefficient; $Q_{h,\min}$ and $Q_{h,\max}$ are the smallest and largest limits of the capacity of the hydroelectric generator.

(5) Power output constraint.

$$\begin{cases} 0 \leq P_{w,t} \leq P_{w,\max} \\ 0 \leq P_{v,t} \leq P_{v,\max} \end{cases} \quad (8)$$

In the formula, $P_{w,\max}$ and $P_{v,\max}$ is the maximum output of wind and photovoltaic power.

(6) Energy storage operation constraints.

$$\begin{cases} S_{s,t} = S_{s,t-1}(1 - \delta) + (P_{s,t}^c \eta_c - \frac{P_{s,t}^d}{\eta_d}) \times \Delta t \\ S_{\min} \leq S_{s,t} \leq S_{\max} \\ u_t^c P_{s,\min}^c \leq P_{s,t}^c \leq u_t^c P_{s,\max}^c \\ u_t^d P_{s,\min}^d \leq P_{s,t}^d \leq u_t^d P_{s,\max}^d \\ u_t^c + u_t^d < 1 \end{cases} \quad (9)$$

In the formula, $S_{s,t}$ represents the charge state of s ; δ represents the self-discharge rate; η_c, η_d are the charging and discharging efficiencies; $S_{\min}, S_{\max}$ are the maximum and minimum capacities of s ; u_t^c, u_t^d represent the charging and discharging states of s , with charging being 0 and discharging being 1; $P_{s,\min}^c, P_{s,\min}^d$ are the minimum values of the charging and discharging powers.

(7) Grid safety constraint.

$$\begin{cases} P_{l,\min} \leq P_{l,t}^0 \leq P_{l,\max}, l \in N \\ P_{l,\min} \leq P_{l,t}^k \leq P_{l,\max}, l \in N \\ P_{l,t} = \sum_i S_i \times (P_{i,t} - P_{e,t}^L) \end{cases} \quad (10)$$

In this formula, $P_{l,t}$ represents the active power transmitted by line l ; $P_{l,\max}$ and $P_{l,\min}$ represent the maximum and minimum transmission power; $P_{l,t}^0$ and $P_{l,t}^k$ represent the active power of line l under the base state and the k th ($N - 1$) fault event; and N is the number of system branches. S_i is the sensitivity coefficient of the injected power of node i with respect to the branch [15].

2.1.2. Backup Decision Optimization

The simulation system of error scenarios is the base for accounting for uncertainties [16], incorporating thermal power, hydropower, and energy storage for providing backup, and incorporating penalties for discarding power and losing load to further optimize the scheduling strategy.

1. The objective function of the backup decision is as follows:

$$f^{bd} = \sum_t \sum_m \pi_m \left(\begin{aligned} & \sum_g (C_g^{Ru} R_{g,t}^{Ru} + C_g^{Rd} R_{g,t}^{Rd}) + \sum_h (C_h^{Ru} R_{h,t}^{Ru} + C_h^{Rd} R_{h,t}^{Rd}) \\ & + \sum_s (C_s^c R_{s,t}^{Rc} + C_s^d R_{s,t}^{Rd}) + \sum_w \tilde{C}_w^c P_{wc,t,m} \\ & + \sum_v \tilde{C}_v^c P_{vc,t,m} + \sum_e \tilde{C}_e^L P_{ec,t,m}^L \end{aligned} \right) \quad (11)$$

In the formula, π_m represents the probability of error scenario; m is the number of error scenarios; C_g^{Ru} , C_g^{Rd} and C_h^{Ru} , C_h^{Rd} are the positive and negative standby capacity costs of thermal power unit g and hydropower unit h , respectively; $R_{g,t}^{Ru}$, $R_{g,t}^{Rd}$ and $R_{h,t}^{Ru}$, $R_{h,t}^{Rd}$ are the upper and lower standby capacities reserved by g and h in scenario m , respectively; $R_{s,t}^{Rc}$ and $R_{s,t}^{Rd}$ are the standby capacity of energy storage charging and discharging; $\tilde{C}_w^c$ and $\tilde{C}_v^c$ are the load curtailment costs of wind and solar, respectively; $P_{w,m,t}^{wc}$ and $P_{v,m,t}^{vc}$ are the load curtailment quantities of wind and solar, respectively; $\tilde{C}_e^L$ is the load loss penalty; $\Delta P_{e,t}^L$ is the load loss amount.

2. The alternative decision constraints are as follows:

(1) Power balance constraint under error scenarios.

$$\left\{ \begin{aligned} & \sum_g P_{g,t,m} + \sum_h P_{h,t,m} + \sum_s (P_{s,t,m}^c - P_{s,t,m}^d) + \sum_w (P_{w,t,m} - P_{wc,t,m}) \\ & + \sum_v (P_{v,t,m} - P_{vc,t,m}) = \sum_L (P_{e,t,m}^L - P_{ec,t,m}^L) \\ & p_{g,t,m} = P_{g,t} + R_{g,t}^{Ru} - R_{g,t}^{Rd} \\ & p_{h,t,m} = P_{h,t} + R_{h,t}^{Ru} - R_{h,t}^{Rd} \\ & p_{s,t,m}^c = P_{s,t}^c + R_{s,t}^{Rc} \\ & p_{s,t,m}^d = P_{s,t}^d + R_{s,t}^{Rd} \end{aligned} \right. \quad (12)$$

In this formula, $P_{g,t,m}$ and $P_{h,t,m}$, respectively, represent the power of g and h under the error scenario m ; $P_{w,t,m}$ and $P_{v,t,m}$, respectively, represent the power generation of w and v under the error scenario m ; $P_{s,t,m}^c$ and $P_{s,t,m}^d$, respectively, represent the charging and discharging of s under scenario m ; $P_{wc,t,m}$, $P_{vc,t,m}$, and $P_{ec,t,m}^L$ represent the power losses due to wind power rejection, solar power rejection, and load loss under scenario m .

(2) Operating constraints for thermal power plants, hydropower plants, and energy storage units.

$$\left\{ \begin{aligned} & u_{g,t} P_{gc,\min} \leq P_{g,t,m} \leq u_{g,t} P_{gc,\max} \\ & u_{h,t} P_{h,\min} \leq P_{h,t,m} \leq u_{h,t} P_{h,\max} \\ & u_t^c P_{s,\min}^c \leq P_{s,t,m}^c \leq u_t^c P_{s,\max}^c \\ & u_t^d P_{s,\min}^d \leq P_{s,t,m}^d \leq u_t^d P_{s,\max}^d \end{aligned} \right. \quad (13)$$

In the formula, $u_{g,t}$, $u_{h,t}$ and u_t^c , u_t^d , respectively, represent the 0/1 variables for thermal power units, hydroelectric power units, energy storage charging, and energy storage discharging. While 1 indicates operation, 0 indicates shutdown.

(3) Climbing speed constraint of power generation units.

$$\left\{ \begin{aligned} & -r_g^d \leq P_{g,t,m} - P_{g,t-1,m} \leq r_g^u \\ & -r_h^d \leq P_{h,t,m} - P_{h,t-1,m} \leq r_h^u \end{aligned} \right. \quad (14)$$

In the formula, r_g^u , r_g^d and r_h^u , r_h^d represent the upward and downward climbing rates of thermal power and hydroelectric power plants, respectively.

(4) Capacity constraints of backup power generation units.

$$\left\{ \begin{aligned} & 0 \leq R_{g,t}^{Ru} \leq Rup_g \\ & 0 \leq R_{g,t}^{Rd} \leq Rdn_g \\ & 0 \leq R_{h,t}^{Ru} \leq Rup_h \\ & 0 \leq R_{h,t}^{Rd} \leq Rdn_h \\ & R_{s,t}^{Rc} \geq 0, R_{s,t}^{Rd} \geq 0 \end{aligned} \right. \quad (15)$$

In the formula, Rup_g , Rdn_g , Rup_h and Rdn_h represent the maximum upward and downward standby capacities of g and h, respectively.

(5) Wind, solar power outage, and load shedding constraints.

$$\begin{cases} 0 \leq P_{wc,t,m} \leq P_{w,t,m} \\ 0 \leq P_{vc,t,m} \leq P_{v,t,m} \\ 0 \leq P_{ec,t,m}^L \leq P_{e,t,m}^L \end{cases} \quad (16)$$

(6) The grid safety constraint is the same as described above.

2.2. Risk Measurement Based on CVaR

Due to the uncertainty of renewable energy, the predicted output value for the current scheduling period is significantly different from the actual output value [17]. This can lead to insufficient reserve regulation or transmission line congestion, and there is a certain risk of power curtailment and load shedding, which is especially true in the northwest region.

This paper employs the CVaR method to calculate the risk losses of power curtailment and load loss caused by wind and solar energy uncertainties, as well as limited and non-stop thermal power output. The CVaR theory is derived from the VaR model, which refers to the conditional expected value when the loss exceeds the value at the risk VaR level at a certain confidence level [18]. It can better reflect the potential risks of the decision-making plan. By definition, the CVaR at the confidence level β is expressed as the following [19]:

$$\begin{aligned} CVaR_{\beta}(x) &= VaR_{\beta}(x) + E[f(x, y) - VaR_{\beta}(x) | f(x, y) \geq VaR_{\beta}(x)] \\ &= E[f(x, y) - VaR_{\beta}(x) | f(x, y) \geq VaR_{\beta}(x)] \\ &= \frac{1}{1-\beta} \int_{f(x,y) \geq VaR_{\beta}(x)} f(x, y) p(y) dy \end{aligned} \quad (17)$$

Since CVaR can be obtained from historical data or Monte Carlo simulation calculations, assuming that the random variable y simulates q sample values $[y^1, y^2, \dots, y^q]$ and the corresponding loss $[f(x, y^1), (x, y^2), \dots, (x, y^q)]$ through the Monte Carlo method, the calculation formula of CVaR under the decision variable x and the confidence level can be derived from its definition and the definition of conditional probability [20]:

$$\begin{aligned} CVaR_{\beta}(x) &= VaR_{\beta}(x) + E[f(x, y) - VaR_{\beta}(x) | f(x, y) \geq VaR_{\beta}(x)] \\ &= VaR_{\beta}(x) + \sum_{k=1}^q [f(x, y^k) - VaR_{\beta}(x)]^+ \frac{Prob(y^k)}{Prob(f(x, y) \geq VaR_{\beta}(x))} \\ &= VaR_{\beta}(x) + \frac{1}{q(1-\beta)} \sum_{k=1}^q [f(x, y^k) - VaR_{\beta}(x)]^+ \end{aligned} \quad (18)$$

In the formula, $[f(x, y^k) - VaR_{\beta}(x)]^+$ represents $\max\{0, f(x, y^k) - VaR_{\beta}(x)\}$, and $Prob(y^k)$ represents the probability of y^k occurring.

Therefore, referring to the above calculation process, the risk measurement process is as follows:

(1) Firstly, this paper randomly generates the probability density functions of the fluctuations in each random variable at time t based on the cumulative distribution functions of each random variable [21]. These include the fluctuation in the power output of the wind and photovoltaic stations $P_{w,t,m}^{\Delta}$ and $P_{v,t,m}^{\Delta}$, the fluctuation in the power output range of the thermal power units and unplanned outages $P_{g,t,m}^{\Delta}$ and $P_{FT,t}^{\Delta}$.

(2) Then, this paper separately calculates the wind and solar power curtailment quantities $P_{WR,t}$ and curtailment losses $f_{WR,t}(P_{WR,t})$, and load shedding quantities $P_{PL,t}$ and load shedding losses $f_{WR,t}(P_{WR,t})$, at each moment under the influence of random variable uncertainty, as shown in the following formula:

$$\begin{aligned}
P_{WR,t} &= \sum_g P_{g,\min} I_{gt}^{on} - P_{e,t}^L + P_{w,t} + P_{v,t} - P_{FT,t}^\Delta \\
&\quad - P_{g,t,m}^\Delta + P_{w,t,m}^\Delta + P_{v,t,m}^\Delta \\
f_{WR,t}(P_{WR,t}) &= [K_{WR} P_{WR,t}]^+
\end{aligned} \tag{19}$$

$$\begin{aligned}
P_{PL,t} &= P_{e,t}^L - P_{v,t} - P_{w,t} - \sum_g P_{g,\max} I_{gt}^{on} + P_{g,t,m}^\Delta \\
&\quad - P_{w,t,m}^\Delta - P_{v,t,m}^\Delta + P_{FT,t}^\Delta \\
f_{PL,t}(P_{PL,t}) &= [K_{PL} P_{PL,t}]^+
\end{aligned} \tag{20}$$

Here, K_{WR} represents the unit cost of lost electricity and K_{PL} represents the unit cost of lost load. $[K_{WR} P_{WR,t}]^+$, $[K_{PL} P_{PL,t}]^+$ represents the maximum value among them, that is, $\max\{0, K_{WR} P_{WR,t}\}$ and $\max\{0, K_{PL} P_{PL,t}\}$.

(3) This paper conducts Monte Carlo simulation sampling based on the calculated electricity waste and load loss amounts in (2) [20], thereby obtaining multiple scenarios of electricity waste loss $[f_{WR,t}(1), f_{WR,t}(2), \dots, f_{WR,t}(k_{WR,t})]$ and load loss $[f_{PL,t}(1), f_{PL,t}(2), \dots, f_{PL,t}(k_{PL,t})]$ within T .

(4) Finally, based on the scenarios of (3), this paper calculates the conditional value at risk of power outage $CVaR_{WR,t}$ and load loss $CVaR_{PL,t}$ at each time point t , as well as the total conditional value at risk of power outage $CVaR_{WR}$ and load loss $CVaR_{PL}$. The calculation formulas are as follows:

$$\begin{aligned}
CVaR_{WR,t} &= VaR_{WR,t} + \frac{1}{k_{WR,t}(1-\beta)} \sum_{k_{WR,t}} [f_{WR,t}(k) - VaR_{WR,t}]^+ \\
CVaR_{WR} &= \sum_{NT} CVaR_{WR,t}
\end{aligned} \tag{21}$$

$$\begin{aligned}
CVaR_{PL,t} &= VaR_{PL,t} + \frac{1}{k_{PL,t}(1-\beta)} \sum_{k_{PL,t}} [f_{PL,t}(k) - VaR_{PL,t}]^+ \\
CVaR_{PL} &= \sum_{NT} CVaR_{PL,t}
\end{aligned} \tag{22}$$

In these formulas, β represents the set confidence level. Because this paper adopts a moderate conservative risk-aversion attitude, aligned with similar literature analyses [18,19], it also takes 0.95. $VaR_{WR,t}$ and $VaR_{PL,t}$, respectively, denote the power-off and load loss risk values at time t .

3. Model Solution Based on Deep Reinforcement Learning and Constraint Identification

3.1. The Constraint Identification Method Based on the SDAE

In view of the situation that conventional active constraint identification takes too long and manual selection results are inaccurate, this paper proposes a guaranteed strategy for active constraint identification and incorporates it into the reinforcement learning model calculation framework to further improve the solution efficiency of the model. We consider that system node power fluctuations are coupled with each other in a complex and nonlinear manner, have high-dimensional nonlinearity, and the measurement data contains noise, and the fluctuations in renewable energy and loads themselves also have uncertainties, etc. However, the stacked de-noising auto-encoder (SDAE) [22] not only has the ability to learn deep nonlinear features and automatically extract advanced features but can also use the deep stack structure and encoding and decoding functions to mine the nonlinear relationship between system conditions and constraints to achieve the efficient identification of the model active's constraints. Moreover, it has good robustness. For power system data with noise and uncertainty, it can still learn the essential distribution behind the data. Therefore, combining the task and model characteristics, this paper chooses SDAE for identification. The model structure is shown in Figure 2.

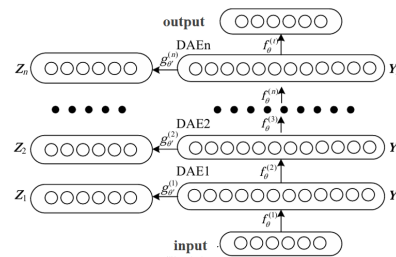


Figure 2. Model architecture of stacked noise-reduction automatic encoder model.

Considering that the load prediction accuracy is relatively high, the injected power at the nodes is thus selected as the input feature vector, which effectively reflects the fluctuations. However, the set of effective constraints accounts for a relatively small proportion in the N-1 safety constraint set [23]. If the effective constraints are used as the output feature, it is likely to cause the problem of sample imbalance. Therefore, in this paper, the generator output power is considered as the output vector. The transmission power of the line is obtained through the following formula, and based on whether it exceeds the limit, whether the corresponding line safety constraint is an effective constraint can be determined. Finally, by using the predicted generator output values, we can conduct N-1 analysis to obtain the initial active constraint sets.

$$-\alpha P_{l,\max} \leq S_i \times (P_{l,t}^k - P_{e,t}^L) \leq \alpha P_{l,\max} \quad (23)$$

In the formula, α is a relaxation factor of less than 1, which can be used to narrow the range and improve the accuracy.

The calculation formula for implementing the constraint identification through deep neural networks is as follows:

$$\begin{aligned} Y &= f^\theta(X_{in}) \\ f &= R(WX_{in} + \tau) \end{aligned} \quad (24)$$

In the formula, X_{in} is the input feature vector of the system operating conditions (normalized); Y is the output vector of the generator output (normalized); f is the feedforward function of the SDAE neural network; R is the activation function; and θ is the neural network parameters, where W is the coefficient vector of the neurons and τ is the bias vector of the neurons.

The learning algorithm adopts the performance-optimized RMSProp algorithm, and the intermediate layer of the neural network adopts the widely used ReLU function [24]. The activation function connecting the neurons of the last layer of the deep neural network is designed as a linear function, enabling the deep neural network to capture broader numerical features. The training process of SDAE can be found in reference [25]. The overall process of using deep neural networks for identifying the activation constraints is shown in Figure 3.

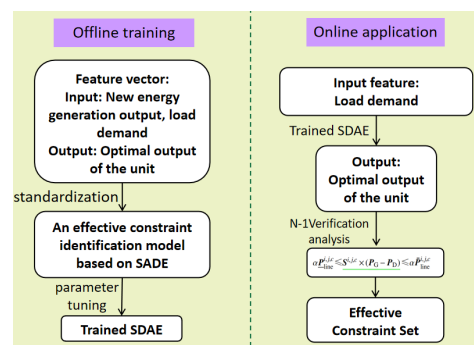


Figure 3. Solving the process of constraint identification.

3.2. Markov Decision Process Modeling

The typical reinforcement learning decision process can be described by a five-tuple, $\langle S, A, P, R, \gamma \rangle$ [26] where S is the state space; A is the action space; P is the state transition probability determined by the environment; R is the reward space, which is the immediate reward provided by the environment to the decision-maker based on the feedback of S and A , and is used to evaluate the quality of option A ; and γ is the discount factor.

Considering that traditional reinforcement learning methods have limited ability to solve decision-making problems with continuous action control and high safety requirements [27], this paper redesigns the S , A , R , and forms a deep reinforcement learning algorithm suitable for solving power generation plan optimization. The specific design is as follows:

State Space S : The state space should be selected to include the factors that have an impact on decision-making. Considering that the intelligent agent makes the optimal scheduling decision for the output of each unit in the power system by observing the operating state of the power system, the observed states therefore select the output of conventional units in the previous scheduling period, the actual output of w and v , the charging and discharging power of s , and the state of energy storage, as well as the start-stop and standby status of conventional units. Thus, it can be established as follows:

$$S = \{S_G, S_W, S_V, S_S, S_C, S_D, T_G\} \quad (25)$$

In this formula, S_G is the output power set of conventional units; S_W and S_V represent the output power set of wind and photovoltaic; S_S represents the charge and discharge power of s ; S_C represents the state of s ; S_D represents the system load set; and T_G represents the start-stop status set of conventional units. Within each scheduling period t , a corresponding one-dimensional state vector can be determined from S .

Action A : The action space is composed of the decision quantities. Considering the time-series coupling characteristics of the output power of each unit under different time periods, this paper sets the incremental output of the current moment's conventional units in the next moment as the decision variable A_G , thereby achieving decoupling of the decision variables over multiple time periods and effectively reducing the complexity of model training. Meanwhile, the standby capacity for upward and downward adjustments are, respectively, set as A_{Ru} , A_{Rd} , and the actual regulation output of the energy storage is set as A_S . Thus, it can be established as follows:

$$A = \{A_G, A_{Ru}, A_{Rd}, A_S\} \quad (26)$$

In this formula, A_G represents the set of output increments of conventional units; A_{Ru} and A_{Rd} represent the backup capacity for the unit; and A_S is the output set of s .

Based on the power output $P_{G,n,t}$ of each unit, the operating status $T_{G,n,t}$, and the power increment $A_{G,n,t}$, the power output of each unit in the next time step can be calculated using the following formula:

$$P_{G,n,t+1} = P_{G,n,t} + A_{G,n,t} \quad (27)$$

After updating the generator output, combining the wind-solar-load prediction values can obtain the state variables for the next time step. Therefore, the corresponding state transition probability is 1.

Reward R : This is the key to gradually adjusting the power generation plan to the target state. Consider treating the negative value of the objective function as part of the reward function, including system operation costs, backup costs, and risk costs. Furthermore, the safety constraints are added to the reward function in the form of a penalty function that

violates the power flow operation constraints. Combining Equations (1) and (23), the final reward function can be expressed as follows:

$$\begin{cases} r_t = \varepsilon_0(r_1 + r_2 + r_3) \\ r_1 = -(f^{\text{uc}} + f^{\text{bd}}) \\ r_2 = -\delta(CVaR_{WR} + CVaR_{PL}) \\ r_3 = -\lambda(|P_{l,\min} - P_{l,t}^k| + |P_{l,t}^k - P_{l,\max}|) \end{cases} \quad (28)$$

In the formula, r_t indicates the immediate reward that the scheduling planning agent can obtain after choosing a_t in s_t ; ε_0 is the scaling coefficient of the reward function; δ represents the additional penalty coefficient for system operation risks; and λ represents the penalty coefficient for violating safety constraints.

Then, the cumulative reward function for the entire scheduling period T can be expressed as follows:

$$R_t = \sum_{t'=t}^T (\gamma^{t'-t} r_{t'}) \quad (29)$$

where $\gamma^{t'-t}$ represents the discount factor for the $t' - t$ period, and its value ranges from 0 to 1.

3.3. Optimal Strategy Solution

The reinforcement learning algorithm solves the optimization problem by finding a strategy π that enables the agent to obtain the maximum reward value [28]. This paper uses the action-value function $Q_\pi(s, a)$ and the state-value function $V_\pi(s)$ to evaluate the quality of the strategy π . The calculation is as follows:

$$\begin{aligned} Q_\pi(s, a) &= E_\pi \left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} \mid S = s, A = a \right] \\ V_\pi(s) &= E_\pi \left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} \mid S = s \right] \end{aligned} \quad (30)$$

In the formula, $E_\pi[\cdot]$ represents the expected reward value under strategy π ; s, a represent the state and action within that state.

The optimal strategy based on PPO is to learn the π^* through the interaction between the agent and environment to maximize the $V_\pi(s)$, and then meet the requirements of the objective function [29]. Therefore, π^* and $V_\pi(s)$ are specifically expressed as follows:

$$\begin{aligned} \pi^* &= \operatorname{argmax} V_\pi(s) \\ V_\pi(s) &= E_\pi \left[\sum_{k=t}^T \gamma^{k-t} R(s_k, a_k) \mid s = s_t \right] \end{aligned} \quad (31)$$

3.4. Solution Architecture

For the solution of the joint optimization model, firstly, the paper solves the unit commitment problem to obtain the initial day-ahead unit commitment status and output plan of the units based on the envisioned scenario. Secondly, based on the error scenarios, the backup decision-making is optimized and further combined with the calculated risk cost for optimization solution. In the solution process, this paper introduces the PPO method. At the same time, the safety constraints are identified based on the SDAE to calculate the safety cost, which is then embedded as a reward into the PPO model. During the offline training phase, this paper uses historical data to train the models. In the subsequent online stage, adaptive learning scheduling strategies can be decided. The overall architecture is shown in Figure 4.

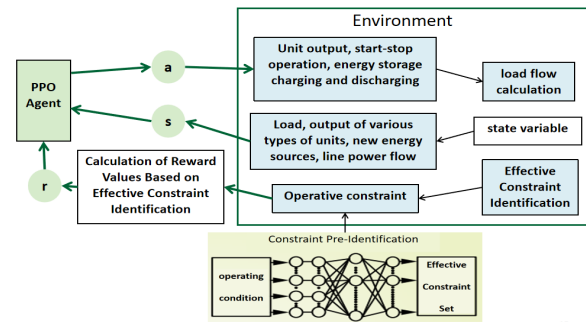


Figure 4. Model solution framework.

The training process of PPO involves the following: firstly, reading the environmental parameters and initialization and setting the training parameters N , T , T_{max} . Then, by obtaining the current load demand, the agent takes action a_t based on s_t and receives a reward value R_t . Subsequently, it stores these data in the experience pool as $\langle s_t, a_t, R_t \rangle$. Then, through multiple training processes to collect these data until the required number of experiences N has been reached and simultaneously update the network parameters. When the training round reaches T_{max} , save the network parameters and end the training.

During the actual decision-making stage, the scheduling decision of the optimization model is solely based on the Actor network trained previously and does not require the participation of the Critic network. When a task arrives, the agent selects a_t according to the policy $\pi_{\theta}^*(a_t, s_t)$. Simultaneously, calculate R_t and transfer to s_{t+1} until a total of T periods is completed.

Among them, the Actor network and Critic network used in this paper are both composed of an input layer, three hidden layers, and an output layer. The output layer of the Actor network adopts the hyperbolic tangent function ($\tanh$), which limits the output actions within the range of $[-1, 1]$. Additionally, both employ the Adam optimization method, with the learning rates set at $\alpha = 0.0001$ and $\beta = 0.0002$. Other training parameters are set as follows: $\gamma = 0.95$, $\varepsilon_0 = 0.001$, $\delta = 10$, $\lambda = 100$, $N = 96$, and $T_{max} = 5000$.

4. Case Analysis

4.1. Introduction to Practical Examples

This paper validates the effectiveness and adaptability of this method using the IEEE 30 system and an equivalent IEEE 118 system of a certain regional power grid [30]. The IEEE 30 system consists of 6 conventional generating units, 41 lines, and a thermal power capacity of 335 MW. The wind and photovoltaic stations are separately installed at nodes 10, 27 and 15, 20. Their installed capacity are, correspondingly, 60 MW and 50 MW each, and the penetration rate of new energy is 39.64%. These systems are as shown in Figure 5.

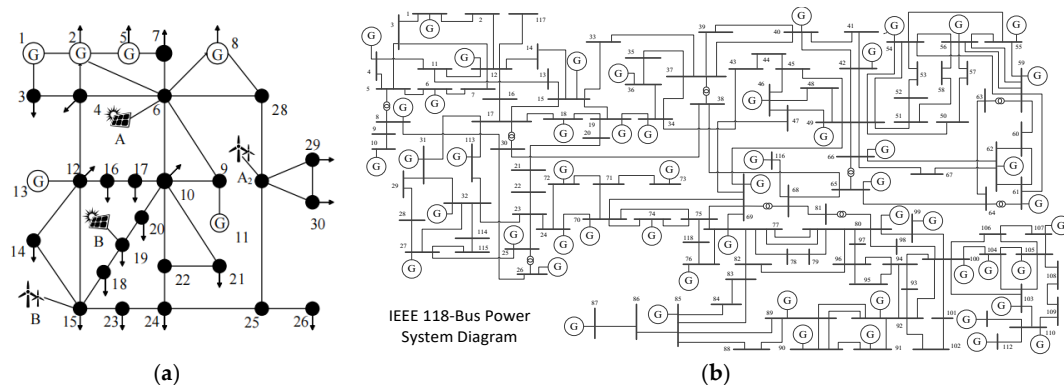


Figure 5. Test node example. (a) IEEE 30 test system; (b) IEEE 118 test system.

After equivalent aggregation, the actual power grid includes 76 thermal power units, 15 hydropower units, 130 wind power units, 226 photovoltaic units, and 37 energy storage units. Their total installed capacities are 29,710 MW, 422 MW, 15,015 MW, 21,830 MW, and 3563 MW, respectively, and the renewable energy capacity penetration rate is about 52%. The actual power grid can be equivalent to an 118-node example, as shown in Figure 5b, where wind turbines are added at nodes 34, 36, 40, 45, 49, 54, 59, and 63, and photovoltaics are added at nodes 72, 75, 78, 82, 85, 88, 92, 95, 98, and 102, respectively. The unit price of power for the energy storage system is CNY 1500/kW, the capacity unit price is CNY 2500/(kwh), and the operating cost unit price is CNY 0.05/(kwh). The penalty coefficients for abandoning wind and light are both CNY 500/(MWh), and the penalty coefficient for loss of load is CNY 1000/(MWh) [31]. The predicted curve for node load demand is shown in Figure 6.

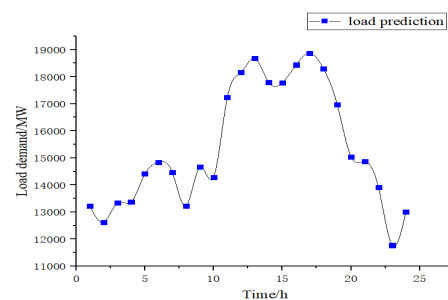


Figure 6. Load demand forecasting.

4.2. Uncertainty Handling

Based on the renewable energy forecast outputs of a certain provincial power grid, this paper separately fits the Frank copula function to the wind and photovoltaic power output data for 24 time periods. Then, it uses Latin hypercubic sampling (LHS) to generate scenarios and finally uses the backward reduction method to obtain the reduced wind and solar power prediction scenarios, as shown in Figure 7. It can be seen that during the adjustment process of the error scenarios, the fluctuations in renewable energy output are relatively large.

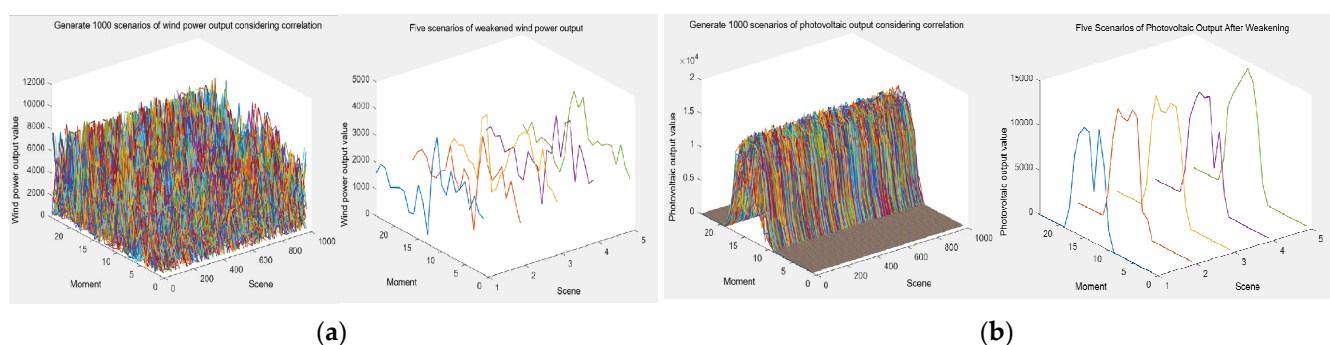


Figure 7. Generation and reduction in wind/photovoltaic output scenes. (a) The generation and reduction in wind and (b) the generation and reduction in photovoltaic.

This paper considers the generation of typical scenarios for the uncertainties of wind and photovoltaic power outputs, with a total of 25 scenarios. The scenario probability is the product of the probabilities of the corresponding output scenarios. In Figure 7a, the corresponding probabilities of the wind power output scenarios are 0.252, 0.138, 0.181, 0.247, and 0.182. In Figure 7b, the corresponding probabilities of the photovoltaic power output scenarios are 0.212, 0.202, 0.226, 0.198, and 0.162.

Furthermore, this paper simulates actual unplanned outage by setting different non-stop thresholds and the number of units with non-stop. It can be found that when $kG > 0$, with the increase in the non-stop threshold, the operating cost of the system decreases. When the non-stop threshold is 10^{-4} , considering that the smaller the value is, the poorer the robustness will be [14], so this paper takes the non-stop threshold as $Th_f = 10^{-4}$.

4.3. Comparison of Model Training Effects

This paper uses an IEEE30-node example to generate 10,000 different running scenarios and randomly sets the training set, verification set, and test set at 2:1:1, with 100 iterations, and records the accuracy and loss of each iterative training. The training results are shown in Figure 8, where, with the increase in iteration times, the convergence speed of network training is faster and, finally, lower loss values and higher accuracy are obtained. Therefore, the trained model can meet the recognition requirements well.

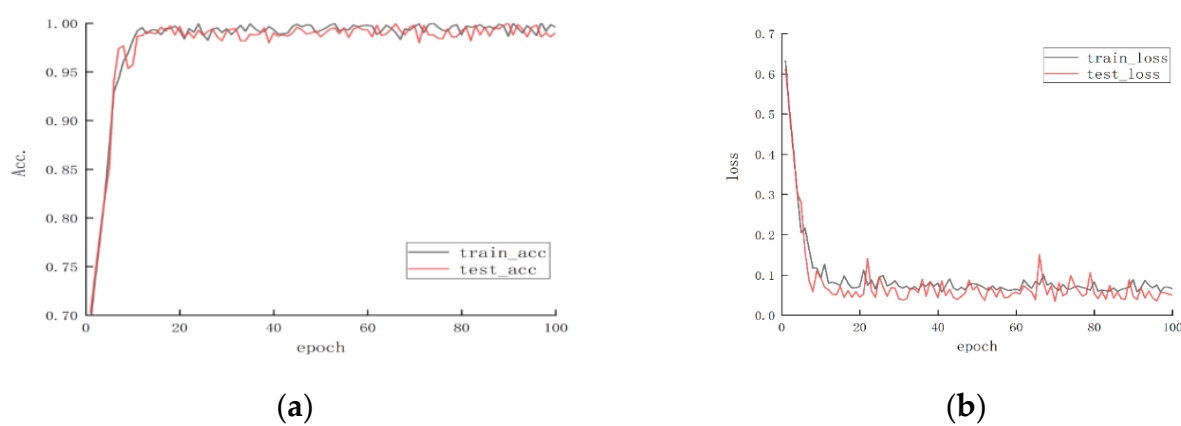


Figure 8. Performance of the SDAE Model. (a) The changes in accuracy and (b) the changes in loss.

Furthermore, to reflect the ablation of the SDAE model, this paper compares the identification performance of different neural networks, as shown in Table 1.

Table 1. Test performances of different recognition models.

Model	ACC	Recall	Performance Decline
SDAE	98.8%	97.5%	<2%
Non-de-noising AE	93.2%	91.4%	~10%
Shallow AE	89.5%	85.7%	~15%
MLP	90.7%	88.2%	~12%

It can be seen that compared with the simplified AE, the accuracy and robustness of the SDAE's test results are obviously improved, which shows that it is necessary to add de-noising and multi-layer settings. Compared with the MLP, the SDAE has more obvious advantages in accuracy and robustness, which also shows that the SDAE has more advantages in identifying complex nonlinear constraints.

Meanwhile, this paper assesses the changes in the reward values during the training stage. As shown in Figure 9, it is found that the proposed method has a relatively large fluctuation in the reward curve in the first 2000 events, and there is a certain convergence after approximately 6000 training events.

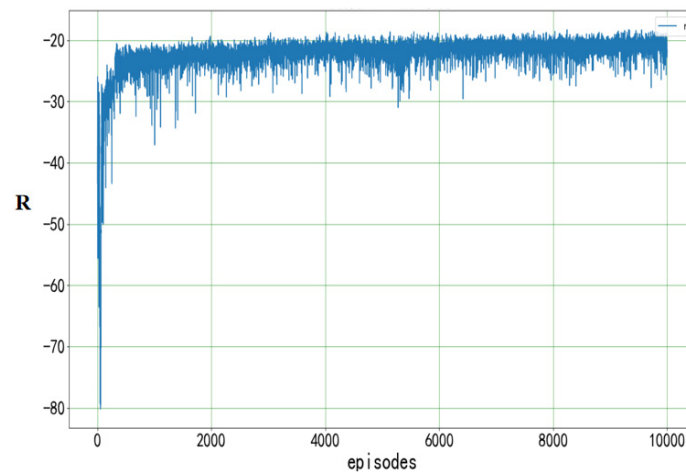


Figure 9. Changes in training process reward.

This paper compares the results obtained by solving the PPO optimization model with the method considering effective constraint identification and the method not considering effective constraint identification, and analyzes the iterative performance in different scenarios, as shown in Table 2. It can be observed that, compared with applying PPO, DDPG alone, and using the Gurobi solver to directly solve MILP, the PPO solution considering constraint identification can ensure accuracy and significantly reduce the time.

Table 2. The effects under different solution methods.

Method	Total Cost (CNY 10,000)	Risk Cost (CNY 10,000)	Training Time (h)	Calculation Time (s)
PPO + SDAE	6775.39	6521.23	3.3	2.77
PPO	6771.62	6521.21	14	16.39
DDPG	6765.62	6515.61	9	127.54
MILP(Gurobi)	6768.45	6510.78	--	15,839.86

Furthermore, this paper conducts a test and comparison between a 30-node system and the actual node system. It is found that as the system scale increases, the solving time of MILP grows exponentially, while the online reasoning time of PPO remains almost unchanged and it can meet the time requirements in industrial practical applications.

4.4. Analysis of Optimization Results Comparison

This paper sets up four schemes for comparative analysis to examine the impact of adding different constraints to the actual system's operation.

Scheme 1: it does not consider the limitation constraint of thermal power units and does not include risk measurement.

Scheme 2: it incorporates the limitation constraint of units but does not include risk measurement.

Scheme 3: it does not include the limitation constraint of units but includes risk measurement.

Scheme 4: it considers the limitation constraint of thermal power units and includes risk measurement.

The results are presented in Table 3, and the optimization strategies of Scheme 1 and Scheme 4 are, respectively, depicted in Figure 10.

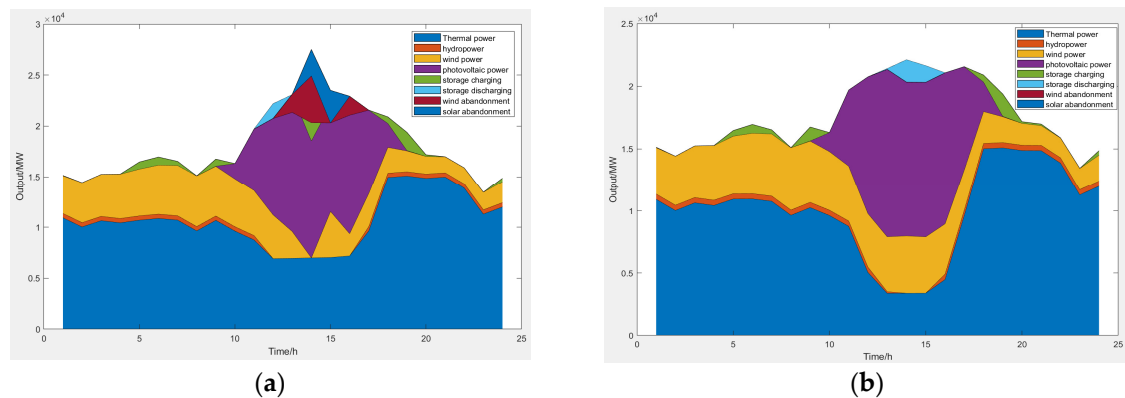


Figure 10. Optimization scheduling strategies under different conditions. (a) The optimized scheduling strategy of Scheme 1 and (b) the optimized scheduling strategy of Scheme 4.

As shown in the figure above, during periods when the output of new energy sources is low, thermal power plants have borne the main load demand. By starting and stopping the units and adjusting their output, they can better adapt to the fluctuations in the output of new energy sources. Additionally, the integrated energy storage has significantly improved the peak–valley difference in the load, reduced the number of unit start-ups and shutdowns, and minimized the losses from new energy outages. During the period of abundant new energy generation, priority is given to generating electricity from new energy sources. The thermal power plants then quickly reduce their output to the minimum, in an effort to fully absorb as much renewable energy as possible. However, in the first scheme, there was a certain degree of power curtailment during the period of abundant new energy power generation. This indicates that the optimized solution obtained through the aforementioned method can effectively meet the energy supply and consumption requirements of the system operation and achieve optimal scheduling among multiple units, but if risk assessment is not included during peak wind and solar power generation periods, it will affect the possibility of wind and solar power being discarded.

Table 3. Optimization results of the power system in four schemes.

Scheme	Total Cost (CNY 10,000)	Running Cost (CNY 10,000)	Standby Cost (CNY 10,000)	Shutoff Power (MWh)	Load Shedding Power (MWh)	Risk Cost (CNY 10,000)
1	7222.76	6578.82	521.6	900	100	7406.27
2	6914.74	6596.78	406.08	200	0	6782.56
3	6890.24	6603.26	452.72	100	0	6641.63
4	6775.39	6612.94	423.51	0	0	6383.81

As can be seen in Table 3, compared with Scheme 1, Scheme 2 takes into account the actual fluctuation in the output of thermal power units and the non-fault conditions, which expands the adjustment range of units and enhances the start–stop regulation function. It reduces the system’s standby costs to a certain extent. Scheme 3 incorporates risk measurement, which further reduces the risks of power outage and load loss. However, it does not incorporate the actual fluctuation and the non-fault conditions, which to some extent increases the standby costs. Scheme 4 incorporates constraints on the output limitations of units and risk measurement, effectively reducing the system’s standby and risk costs. Moreover, the losses from power curtailment and load shedding have been reduced to zero, enabling the system to operate more reliably.

Furthermore, the plans for both Scheme 1 and Scheme 4 during each period are presented in Figure 11. It can be observed that, compared with Scheme 1, in the periods of low load and when wind power is at its peak (from 11 o’clock to 16 o’clock), Scheme 4 can

not only further increase the output of the generating units but can also actively reduce the output of the units, and even shut down some less economically viable units. For example, in Scheme 1, multiple units simultaneously experienced unplanned outages from 16:00 to 21:00. The units that were affected were those with higher economic efficiency, which led to significant load loss. In Scheme 4, the units that underwent start–stop operations were mainly during periods of abundant wind and solar power generation. The units that were shut down were mostly those with lower economic efficiency. In the subsequent periods, the units with higher economic efficiency were started. In addition, Scheme 4 considered the situation where the entire unit experiences unplanned shutdown simultaneously, and the risk of non-shutdown incidents for thermal power plants is reduced. Additionally, by combining with risk assessment, the costs of power abandonment and load loss were also decreased.

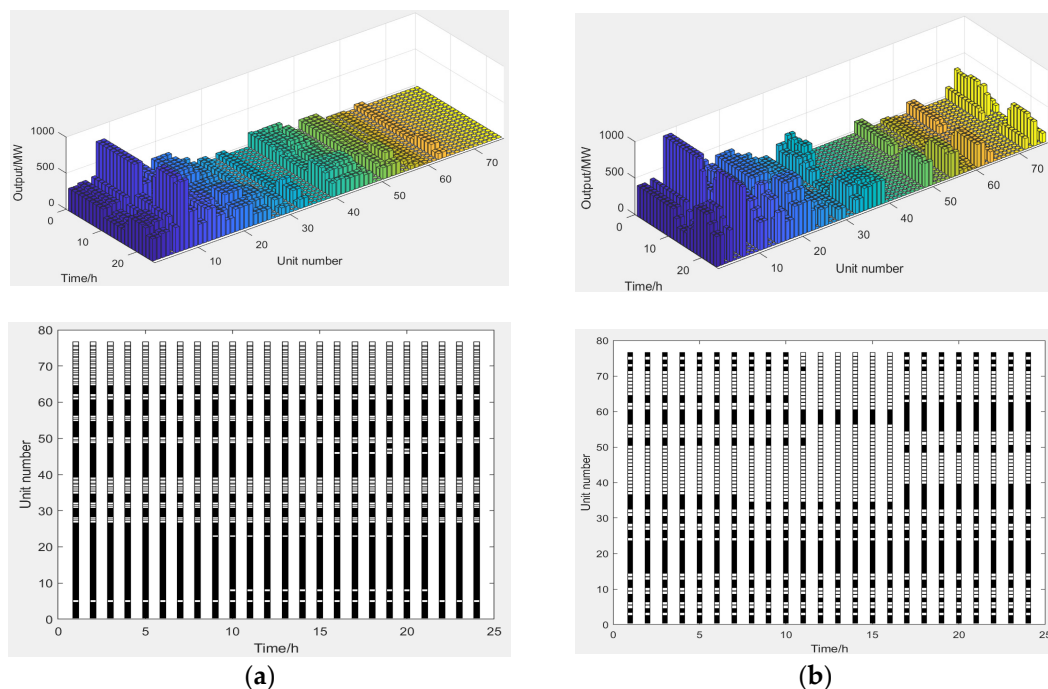


Figure 11. The output and start–stop of the unit. (a) The result of Scheme 1 and (b) the result of Scheme 4.

Overall, this paper incorporates new thermal power constraints and risk measurement, thereby increasing the operating costs by 5.07%. However, the standby capacity, power loss due to disuse, and load shedding losses have significantly decreased, meaning that the standby cost and risk cost have decreased significantly (by 15.44%), and the total cost has decreased by 6.19%. This indicates that incorporating constraints related to thermal power generation capacity limitations and risk assessment into the power generation planning process can reduce potential power outages and load shedding losses. The constructed two-layer optimization model can better ensure the economic and safety performance of the system operation.

4.5. Sensitivity Analysis of Different Risk Penalty Systems

To verify the reliability of the risk management in the optimized model due to the addition of risk measurement, this paper compares the optimized costs under different risk penalty coefficients. The results are presented in Table 4.

Table 4. Optimization results of the system under different risk penalty coefficients.

Risk Factor	Running Cost (CNY 10,000)	Standby Cost (CNY 10,000)	Risk Cost (CNY 10,000)	Total Cost (CNY 10,000)
0	6578.82	643.94	7406.27	7222.76
0.1	6442.37	664.82	7129.36	7109.41
0.2	6411.54	608.32	7002.15	7016.32
0.3	6520.97	513.97	6713.97	6938.65
0.4	6612.93	423.51	6383.81	6775.39
0.5	6980.47	498.15	6295.34	6886.98
0.6	7422.99	539.24	6175.63	6950.27
0.7	8646.20	602.56	6021.23	7129.49
0.8	11,584.86	634.99	5896.45	7401.13
0.9	21,273.33	699.13	5628.26	7622.68
1	6578.82	643.94	7406.27	7222.76

It can be observed that, as the risk penalty coefficient gradually increases, the impact of risk on the optimal dispatching scheme becomes increasingly significant, and the risk costs of curtailed power and load loss decrease accordingly. When the risk penalty coefficient is small, the risk acceptance level is higher. As λ increases, the risk cost decreases rapidly. When λ is greater than 0.4, and as it continues to increase, the risk cost decreases more gradually, the total cost increases rapidly, and the risk acceptance level gradually decreases. Meanwhile, compared with not incorporating risk measurement, the strategy risk cost in Scheme 4 can be reduced by CNY 1022.46 million, and the total cost decreased by CNY 186.31 million. This indicates that incorporating risk measurement can better ensure the economic operation of the system while meeting the security requirements. Additionally, it can also provide more practical reference information for system dispatching decision-makers with different risk requirements.

5. Conclusions

5.1. Summary of Work

Due to the increased complexity and risk levels of system operation, the difficulty of solving the unit commitment problem in the power generation plan also rises. Especially in regions where wind energy and solar energy account for a relatively high proportion, thermal power generation, which serves as the “ballast stone” for power supply security, urgently needs to fully consider its actual output characteristics, accurately assess the actual power generation capacity, and formulate more secure and feasible optimization scheduling strategies and their solution methods. In this regard, the work is as follows:

(1) The output constraint and risk measurement of thermal power units are included in the optimization model, which effectively reduces the system standby and risk costs, and the loss of power abandonment and load shedding is reduced to 0.

(2) This paper employs the SDAE for effective safety constraint identification, achieving an identification accuracy rate of up to 98.8% with a performance fluctuation in less than 2%. Compared to ordinary auto-encoders and MLP, it offers better accuracy and robustness for constraint identification in actual power grids.

(3) This paper uses SDAE as the safety feature perception component of the PPO model and integrates it into the reward function. This significantly enhances the efficiency and quality of the PPO agent in learning safety strategies, effectively reducing the training time of the deep reinforcement learning model by 10.7 h.

(4) This paper has redesigned the decision-making quintuple and developed a DRL algorithm that is suitable for solving the security optimization scheduling of power generation plans. The algorithm is solved using the PPO + SDAE method. Compared to

traditional methods, it can significantly reduce the computational time of actual large-scale systems, requiring only 2.77 s.

5.2. Limitations and Future Work

The algorithm proposed in this paper can conduct targeted training for intelligent agents and quickly obtain the optimal strategy based on changes in the scenario, meeting the real-time requirements of power grid operation. However, there are certain limitations in its application process, including the following:

(1) Insufficient power grid simulation data: as PPO and SDAE are algorithms based on deep neural networks, they both require a large amount of learning simulation data. In the future, a benchmark testing platform for power systems based on digital twins needs to be built, and the accumulation of digital resources should be strengthened to provide a foundation for applications.

(2) System scalability: the single-day optimization mentioned in this study involves shifting the computational burden from online solving to offline training. For multi-day optimization, transfer learning can be utilized. Pre-training can be conducted on a single-day model, followed by fine-tuning with multi-day data, and a “24+24” hour rolling optimization framework can be adopted. For larger-scale systems, a hierarchical distributed architecture can be employed, decomposed by voltage level or geographical region, and pre-training can be conducted on small systems while fine-tuning is performed on large systems. However, verification is still required in conjunction with actual large power grid systems.

(3) Interpretability: because PPO is based on RL, which is not conducive to its interpretability, in the future, interpretable strategic actions can be given by combining interpretable machine learning to improve the interpretability of the results.

(4) Robustness: in practical engineering, the real-time relevant parameters input by the system may differ from the training samples, which will affect the accuracy of the calculation and may also involve issues such as measurement noise and communication delay. In the future, we can monitor the model’s performance and improve the robust DRL method.

Author Contributions: Conceptualization, D.L.; methodology, D.L. and L.Z.; validation, L.Z. and H.Z.; investigation, N.M.; resources, L.Z. and H.Z.; data curation, D.L. and N.M.; writing—original draft preparation, D.L.; writing—review and editing, D.L.; visualization, L.Z.; supervision, L.Z.; project administration, N.M. and H.Z.; funding acquisition, N.M. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Ning Mi and Hailiang Zhong were employed by the company State Grid Ningxia Electric Power Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

PPO	Proximal Policy Optimization
SDAE	Stacked De-Noising Auto-Encoder
AE	Auto-Encoder
CVaR	Conditional Value at Risk
LHS	Latin Hypercubic Sampling

DRL	Deep Reinforcement Learning
MLP	Multi-Layer Perceptron
MDP	Markov Decision Process
DDPG	Deep Deterministic Policy Gradient
MILP	Mixed-Integer Linear Programming

References

- Lu, Z.X.; Xu, X.; Yan, Z.; Wu, J.; Sang, D.; Wang, S. Overview on data-driven optimal scheduling methods of power system in uncertain environment. *Autom. Electr. Power Syst.* **2020**, *44*, 172–183. [\[CrossRef\]](#)
- Li, J.H.; Xie, Y.T.; Zeng, H.Y. A Review of Uncertain Optimal Scheduling Research and Its Application in the New Power System. *High Volt. Technol.* **2022**, *48*, 3447–3464. [\[CrossRef\]](#)
- Jiang, W.; Feng, B.; Guo, X.Z. An Identification Method for Effective Safety Constraints Based on Graph Neural Network. *Electr. Autom.* **2023**, *45*, 106–108.
- Zhu, Z.C.; Yang, Z.F.; Yu, J. A Data-driven Fast Calculation Method for Safety-Constrained Economic Dispatch in Small Sample Scenarios. *Proc. CSEE* **2022**, *42*, 4430–4440. [\[CrossRef\]](#)
- Wang, K.; Chen, S.Y.; Xu, J. Robust Decision-making Method for Forward Dispatch of Power System Based on Redundancy Constraint Fast Identification. *Autom. Electr. Power Syst.* **2025**, *10*, 1–14. Available online: <https://link.cnki.net/urlid/32.1180.TP.20250107.1436.002> (accessed on 6 November 2025).
- Fan, S.X.; Li, L.X.; Wang, S.Y. Application analysis and exploration of artificial intelligence technology in power grid dispatch and control. *Power Syst. Technol.* **2020**, *44*, 401–411. [\[CrossRef\]](#)
- Dong, L.; Liu, Y.; Qiao, J. Optimal dispatch of combined heat and power system based on multi-agent deep reinforcement learning. *Power Syst. Technol.* **2021**, *45*, 4729–4737. [\[CrossRef\]](#)
- Dai, P.C.; Yu, W.W.; Wen, G.H. Distributed reinforcement learning algorithm for dynamic economic dispatch with unknown generation cost functions. *IEEE Trans. Ind. Inform.* **2020**, *16*, 2258–2267. [\[CrossRef\]](#)
- Shen, R.; Zhong, S.; Wen, X. Multi-agent deep reinforcement learning optimization framework for building energy system with renewable energy. *Appl. Energy* **2022**, *312*, 118724. [\[CrossRef\]](#)
- Zhang, J.Y.; Pu, T.J.; Li, Y. Distributed Generation Optimal Scheduling Strategy Based on Multi-Agent Deep Reinforcement Learning. *Power Syst. Technol.* **2022**, *46*, 3496–3504. [\[CrossRef\]](#)
- Rockafellar, R.T.; Uryasev, S. Optimization of conditional value-at-risk. *J. Risk* **2000**, *2*, 21–42. Available online: <https://api.semanticscholar.org/CorpusID:854622> (accessed on 19 November 2025). [\[CrossRef\]](#)
- Luo, J.S.; Tian, X.Q.; Wang, Y. Research on the Output Range of Thermal Power Units under Heating Conditions Based on Data-driven Approach. *Energy Conserv. Technol.* **2024**, *42*, 39–43.
- Tian, X.Q.; Luo, J.S.; Yang, L. A Method for Judging Coal Quality Changes and Their Impact Based on Operating Data of Thermal Power Plants. *Energy Conserv. Technol.* **2024**, *42*, 73–76+92.
- Chen, Y.; Zhang, Z.; Liu, Z.; Zhang, P.; Ding, Q.; Liu, X.; Wang, W. Robust N-k CCUC model considering the fault outage probability of units and transmission lines. *IET Gener. Transm. Distrib.* **2019**, *13*, 3782–3791. [\[CrossRef\]](#)
- Zhang, B.M.; Chen, S.S.; Yan, Z. *Advanced Power Network Analysis*; Tsinghua University Press: Beijing, China, 2007.
- Li, J.Z.; Xie, M.; Li, S. Consider CVaR unit combination and scene decision alternatives more joint optimization. *J. South. Energy Constr.* **2021**, *8*, 16. [\[CrossRef\]](#)
- Wu, W.C.; Xu, S.W.; Yang, Y. Probabilistic scheduling of high-proportion new energy power systems based on risk quantification. *Autom. Electr. Power Syst.* **2023**, *47*, 3–11. [\[CrossRef\]](#)
- Liu, Y.J. Short-Term Optimization and Risk Management of Renewable Energy Generation Under Different Operating Environments. Ph.D. Thesis, Shanghai Jiao Tong University, Shanghai, China, 2013.
- Artzner, P.; Delbaen, F.; Eber, J.-M.; David, D.H. Coherent measures of risk. *Math. Financ.* **1999**, *9*, 203–228. [\[CrossRef\]](#)
- Yang, Y.F. Optimization of Power System Generation Plan and Risk Assessment for Adapting to Large-Scale Integration of Renewable Energy Sources. Master's Thesis, Huazhong University of Science and Technology, Wuhan, China, 2015. [\[CrossRef\]](#)
- Song, Y.; Li, H. Generation of Wind and Light Output Scenarios Based on Kernel Density Estimation and Copula Function. *Electr. Technol.* **2022**, *23*, 56–63. [\[CrossRef\]](#)
- Wang, Y.L.; Zhou, T.; Chen, Z. Stepwise Inertial Intelligent Control for Wind Power Frequency Regulation Based on Stacked Denoising Autoencoder and Deep Neural Network. *J. Shanghai Jiao Tong Univ.* **2023**, *57*, 1477–1491. [\[CrossRef\]](#)
- Wu, Y.L.; Zhang, J.X.; Li, B. A Fast Clearing Method for Power Market Based on Deep Learning-Assisted Constraint Identification. *China Electr. Power* **2020**, *53*, 90–97+207. [\[CrossRef\]](#)
- Yu, J.; Yang, Y.; Yang, Z.F. A Probabilistic Energy Flow Fast Calculation Method Based on Deep Learning. *Proc. CSEE* **2019**, *39*, 22–30+317. [\[CrossRef\]](#)

25. Ou, J.Y.; Zhang, Y.; Xin, R. Distributed Cooperative Regulation Strategy for Power Quality in Low-Voltage Distribution Networks Based on Multi-Agent Deep Reinforcement Learning. *Proc. CSEE* **2025**, *12*, 1–15. Available online: <https://link.cnki.net/urlid/11.2107.TM.20241227.1319.011> (accessed on 19 November 2025).
26. Chen, Z.; Pan, Y.; Fan, S.X. Research on Unit Commitment Optimization Method Based on Deep Reinforcement Learning. *Electr. Power Inf. Commun. Technol.* **2023**, *21*, 33–40.
27. Feng, B.; Hu, Y.J.; Huang, G. Review of New Dispatching Optimization Methods for Power Systems Based on Deep Reinforcement Learning. *Autom. Electr. Power Syst.* **2023**, *47*, 187–199. [[CrossRef](#)]
28. Lin, W.S.; Wang, X.J.; Sun, Q.K. Research on Deep Reinforcement Learning Optimization Scheduling Strategy for Integrated Energy System Considering Safety Constraints. *Power Syst. Technol.* **2023**, *47*, 1970–1983. [[CrossRef](#)]
29. Peng, L.Y.; Sun, Y.Z.; Xu, J.; Liao, S.Y.; Yang, L. Adaptive Uncertain Economic Dispatching Based on Deep Reinforcement Learning. *Power Syst. Autom.* **2020**, *44*, 33–42. [[CrossRef](#)]
30. Yang, Z.X.; Ren, Z.Y.; Sun, Z.Y. Security-Constrained Economic Dispatch Method for New Energy Power Systems Based on Proximal Policy Optimization Algorithm. *Power Syst. Technol.* **2023**, *47*, 988–998. [[CrossRef](#)]
31. Li, T.; Li, Z.W.; Yang, J.Y. Complementary and Coordinated Optimal Dispatch of Multi-energy Systems Considering Peak-Shaving Activeness of Wind, Solar, Hydro and Thermal Power Sources. *Power Syst. Technol.* **2020**, *44*, 3622–3630. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.