Load Frequency Control Strategy for Islanded Microgrid Based on SCQ(λ) Algorithm

Qiang Wang, China Three Gorges University, China

Zhenwei Huang, China Three Gorges University, China*

ABSTRACT

In the evolving landscape of power grids, where green transportation and intermittent clean energy play a crucial role, ensuring the security and reliability of the urban network is of utmost importance. However, the increasing volatility associated with these new energy sources poses a challenge to the traditional control methods. The large-scale integration of new energy in microgrids often leads to frequency instability and deviation in control performance standards. Addressing these issues, this paper introduces the SCQ(λ) algorithm, which accurately estimates the system's state to enhance controller capabilities. To evaluate the effectiveness of the proposed SCQ(λ) algorithm, the authors employ a load frequency control model in our simulation. In this model, they introduce various load change disturbances, including sine waves, square waves, and step disturbances to simulate realistic scenarios encountered in power systems. Throughout the simulation, they observe a significant reduction in frequency deviation in the case of step perturbation, with the deviation value decreasing by 0.0096.

KEYWORDS

Enhanced Learning, Isolated Microgrids, Load Frequency Control, Renewable Energy

INTRODUCTION

Currently, with the escalating global resource and environmental challenges, countries worldwide are increasingly embracing "dual-carbon" policies and initiatives. The emphasis on clean energy and electric vehicles signifies the prevailing trend towards upgrading the power supply infrastructure. To achieve the dual carbon objective, China is committed to constructing a new power system primarily reliant on new energy sources. However, the integration of a higher proportion of renewable energy sources brings about greater volatility and uncertainty in grid operations (Lam et al., 2020), which seriously affects the grid frequency stability and control performance standards (CPS). Microgrids offer a solution by enhancing the utilization rate of distributed new energy and effectively addressing electricity consumption challenges in remote areas, deserts, or islands. Moreover, microgrids provide a crucial avenue for integrating electric vehicles (EVs) and diverse forms of distributed green energy (Fan et al., 2022).

DOI: 10.4018/IJGCMS.339198

*Corresponding Author

In the context of load frequency control and regulation within microgrids, energy storage units play a vital role. In recent years, with the rapid popularization of EVs, these vehicles can be leveraged as controllable loads and distributed energy storage units (Chae et al., 2020; Iqbal et al., 2020). Through vehicle-to-grid (V2G) technology, EVs can absorb or transmit power back to the grid to regulate system frequency deviations in the event of grid disturbances or faults (Ziras et al., 2019; Chae et al., 2020). Literature (Li et al., 2019; Karkevandi et al., 2018) has established single-area and multi-area power system load frequency control models incorporating EVs. Various control strategies have been proposed, investigating the dynamic characteristics of system frequency control under these strategies. Simulation results indicate that EV involvement in power system frequency regulation can significantly enhance regulatory performance. However, the reliance on trial-and-error parameter adjustments in the Proportional-Integral control method employed in these models makes it challenging to achieve optimal control performance. Additionally, with the pursuit of the dual-carbon goal, extensive development and integration of new and clean energy sources have emerged as focal points of China’s energy landscape (Chen et al., 2020). As a result, the existing controller outlined in the aforementioned literature needs to be enhanced to address the stochastic disturbances stemming from large-scale new energy integration in islanded microgrids.

The rapid development of emerging machine learning techniques such as reinforcement learning and deep learning in recent years has provided new methods and ideas to solve the above problems. Barbalho et al. (2022) designed a microgrid controller based on the DDPG algorithm to achieve microgrid frequency stabilization by changing the output power of energy storage elements. Fan et al. (2022) proposed a DQN-based load frequency control strategy for microgrid with electric vehicle islanding, which effectively solves the microgrid frequency fluctuation problem under wind disturbance. Wang et al. (2018, 2021) proposed the design of load frequency controllers based on Q-learning and deep Q-learning. It is worth noting that the above methods are derived from classical Q-learning, which updates the Q-function by approximating the maximum desired action value. However, a major drawback of this approach is the overestimation of action values, leading to suboptimal results due to local optimization. In order to address this issue, Hasselt et al., (2015) proposed a Double Q-learning (DQL) algorithm. DQL improves upon classical Q-learning by decoupling the action and state of the Q-function and can effectively solve the overestimation problem of action value in Q-learning. However, the method is not completely unbiased and may introduce an underestimation bias in action values while solving the overestimation problem. This bias could potentially hinder intelligent agents from exploring optimal strategies in the stochastic environment.

In order to solve the problem of overestimation and underestimation, the SCQ algorithm introduces a self-correcting estimator. The main concept behind this estimator is to utilize prior information to correct the estimation process, thereby enhancing the estimation accuracy. By employing the self-correcting estimator, the SCQ algorithm is better equipped to select the most appropriate value estimator. Thus, it can effectively avoid the problems of overestimation and underestimation and improve the estimation accuracy. However, directly adopting the self-correcting estimator for every update may lead to computational inefficiencies and waste of computational resources. Moreover, this approach fails to provide adequate convergence conditions for SCQ. Considering that recalculating the entire model for each update can result in prolonged training times and slow convergence, a solution is needed to overcome these challenges. In order to tackle the above problem, the eligibility trace participation update estimator is introduced in the paper. This estimator leverages the eligibility trace decay coefficient, denoted as λ, to track the involvement level of state-action pairs, associating past state-action pairs with the current update process. Additionally, the cumulative rewards of previous state-action pairs are also taken into account, enabling a more comprehensive evaluation of the value of each state-action pair. To balance past and current learning, a decay factor is employed to diminish the influence of earlier state-action pairs, and the trace value is periodically cleared. This approach facilitates improved convergence speed, as the decay factor governs the rate at which past state-action pairs decay, while resetting trace values prevents the enduring impact of outdated experiences on the learning process.

Therefore, this paper proposes a novel algorithm called multi-step self-correcting Q-learning with eligibility traces (SCQ(λ)) which aims to address the issues of overestimation in traditional Q-learning and underestimation in double Q-learning. By achieving a balance between these two extremes, the SCQ(λ) algorithm enables the intelligent agent to select and explore action values with moderate optimism, thereby improving the control accuracy of the unit. Additionally, the incorporation of eligibility traces in the SCQ(λ) algorithm enhances the convergence speed, allowing the controller to effectively handle the time delay between command signals and unit responses. This reduction in time delay impact results in a more reliable and efficient system. To evaluate the performance of the SCQ(λ) controller, a two-area load frequency control model is constructed, considering factors such as electric vehicles, wind power, and energy storage systems. A sinusoidal wave perturbation is applied for pre-learning, and the results show that the SCQ(λ) controller has a faster convergence speed. In order to simulate the system facing a strong stochastic perturbation problem, the step and square wave perturbations are applied to the model, and the simulation results show that the SCQ(λ) controller has higher CPS performance, smaller frequency deviation, reduced regional control error, and faster convergence.

MICROGRID CONTROL MODEL WITH EVS

Load Frequency Control Model for Micro Gas Turbine

Micro gas turbine (MT) is a kind of small thermal generator with high reliability and safety, which is characterized by high energy conversion rate, low emission, and environmental friendliness. These qualities make them a popular choice in microgrids (Hongxin et al., 2021). Recognizing their significance, this paper focuses on MTs as the primary frequency modulation unit. The dynamic characteristic functions of the fuel system and turbine system of the micro gas turbine can be described as follows:

$$f_{m1} = \frac{1}{1 + T_f s} \tag{1}$$

$$f_{m2} = \frac{1}{1 + T_t s} \tag{2}$$

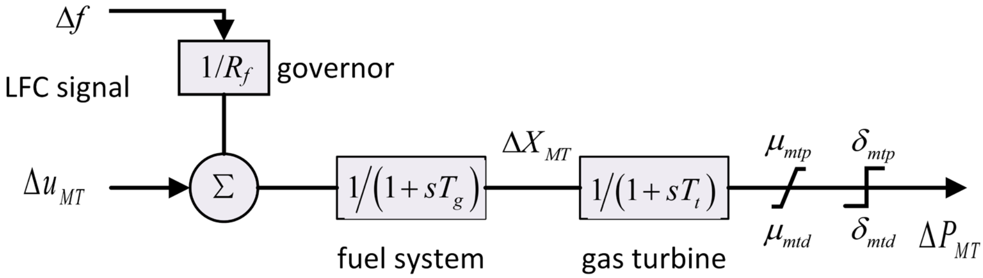
where T_f is the time constant of the fuel system and T_t is the time constant of the turbine system.

The continuous time transfer function model of the MT is shown in Figure 1. Δf is the system frequency change value; Δu_{MT} is the load frequency control signal received by the micro combustion engine; ΔX_{MT} is the valve position change signal received by the turbine system; R is the governor system parameter; ΔP_{MT} is the micro combustion engine power output signal.

Load Frequency Control Model for Energy Storage Systems

In this paper, we address the issue of the stochastic and fluctuating nature of renewable energy access systems. We recognize that MTe alone may not be sufficient to provide enough power for maintaining the active balance of such systems. Therefore, we propose the integration of a battery energy storage system (BESS) to compensate for the power fluctuations and reduce frequency fluctuations. Huang et al. (2015) found that the participation of BESS systems in frequency modulation is superior to other types of energy storage systems, such as flywheel energy storage systems, superconducting electromagnetic energy storage systems, and capacitor energy storage systems.

Figure 1. Load Frequency Control Model for MT



Huang et al. (2015) studied the battery energy storage power model for grid frequency regulation and proposed a simulation model featuring a well-structured framework that effectively addresses both primary and secondary frequency regulation requirements of the system. In this paper, we utilize the energy storage battery simulation model introduced in their literature to represent the BESS simulation model. The BESS frequency response model contains various components such as the power conversion system (PCS), response delay-time conversion, energy storage link, and charge limiting. As shown in Figure 2, $\Delta P_{\text{ord-BESS}}$ is the power output value, where positive values indicate the discharge of energy from the BESS to the system and negative values signify the charging of energy from the system to the BESS. When the BESS receives the power command from the load frequency controller, the signal undergoes processing through the PCS response delay-time conversion link, and the system generates the power demand ΔP_{req}. Subsequently, this power demand value passes through the storage link to produce the output power ΔP_{out} as well as the charging state ΔQ_{soc}. The power output value of the BESS corresponds the power output value of the system, and the power output value of the system is the power demand value of the system.

The BESS needs to satisfy the following constraints during operation:

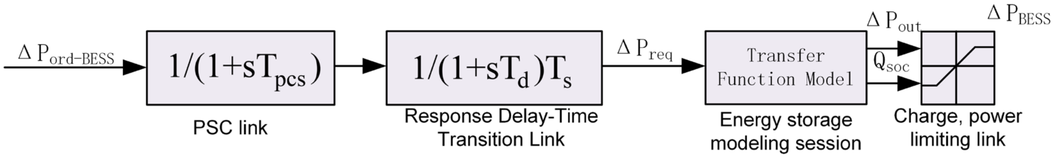
$$0 \le P_c \le \eta P_c^{max} \tag{3}$$

$$0 \le P_d \le \left(1-\eta\right) P_d^{max} \tag{4}$$

P_c and P_d denote the charging and discharging power of the BESS, respectively, η and 1-η denotes the charging and discharging state variable of the BESS, respectively, with the value of 1 indicating charging and 0 indicating discharging. P_c^{max} and P_d^{max} represent the maximum and minimum charge and discharge power, respectively.

$Q_{\text{soc,max}}$ and $Q_{\text{soc,min}}$ are the upper and lower boundaries in the limiting link, and the magnitude of their values determines whether ΔP_{out} can really be output or not. When Q_{soc} crosses the limit, the limiting link avoids the BESS from generating charging or discharging behaviors, and the upper and lower boundaries of the limiting link in this paper are set to 20% and 80%, respectively. The expression for ΔP_{BESS} is as follows:

Figure 2. Load Frequency Control Model for BESS



$$\Delta P_{BESS} = \begin{cases} \Delta P_{out} & Q_{SOC,min} < Q_{SOC} \, and \, Q_{SOC} < Q_{SOC,max} \\ 0 & Q_{SOC} \geq Q_{SOC,max} \, and \, Q_{SOC} \leq Q_{SOC,min} \end{cases} \tag{5}$$

LOAD FREQUENCY CONTROL MODEL FOR ELECTRIC VEHICLES

In the context of China's goal of carbon peaking and carbon neutrality, wind power and photovoltaic are expected to become the main source of clean energy, gradually replacing current auxiliary energy sources. However, the variability of weather conditions poses a challenge, leading to the increasing issue of wind and solar energy curtailment. To address this problem, EVs offer several advantages, including fast response, flexible dispatching, and the ability to serve as both energy sources and storage units. One promising application of EVs is their potential to contribute to peak shaving, valley filling, and auxiliary frequency and voltage regulation through vehicle-to-grid (V2G) technology (Boglou et al., 2023; Boglou et al., 2022). Khokhar & Parmar (2022) designed a new adaptive intelligent model predictive control scheme to regulate the system frequency by managing the state of charge of EV batteries. This control approach enables EVs to participate in the primary frequency regulation of the power system.

$$\Delta P_m - \Delta P_1 = \left(M_s + D\right)\Delta f + \frac{K_E}{1 + sT_E}\Delta f \tag{6}$$

where ΔP_1 is the load disturbance, ΔP_m is the total generator output power, Ms is the inertia constant, D is the damping coefficient, T_E is the time constant, K_E is the sag control parameter, and Δf is the system frequency deviation. In order to ensure the longevity and performance of EV batteries, it is important to minimize the frequency of charging and discharging operations. This is due to the inherent characteristics of batteries, as frequent fluctuations in their charge levels can negatively impact their health. As a solution to address this issue, a dead band module is incorporated into the EV model, allowing it to refrain from participating in frequency regulation when the system frequency experiences minor fluctuations.

The charging and discharging power of the EV is as follows:

When the frequency deviation $\Delta f \leq \left|f_{dz}\right|$:

$$P_{EV} = C_{EV} \frac{SOC_e - SOC}{t_{out} - t_{in}} \tag{7}$$

When the frequency deviation $\Delta f \leq -f_{dz}$:

$$P_{EV} = \begin{cases} k_d \Delta f & k_d \Delta f > -P_{max} \\ -P_{max} & k_d \Delta f < -P_{max} \end{cases} \tag{8}$$

When the frequency deviation $\Delta f > f_{dz}$:

$$P_{EV} = \begin{cases} k_d \Delta f & k_d \Delta f > -P_{max} \\ -P_{max} + k_c \Delta f & k_d \Delta f \le -P_{max} \end{cases} \tag{9}$$

In the formula, f_{dz} is the dead zone of frequency modulation. If the system frequency deviation Δf is within the dead zone range of $\left[-f_{dz}, f_{dz}\right]$, the EV will not participate in frequency modulation service and focus solely on charging to fulfill the power requirements of their users. $P_{\max}$ is the maximum charging and discharging power of an EV, C_{EV} is the rated capacity of an EV, SOC_{e} is the expected SOC value of an EV, t_{in} and t_{out} are the moments when an EV enters into the grid to participate in frequency modulation and exits from the grid.

Wind Power Model

Wind power generation is mainly affected by wind speed, which is characterized by intermittency and volatility. Currently, the Weibull distribution is widely used to model wind speed magnitude (Zhao, et al., 2018), and its probability density function expression is as follows:

$$f\left(x; \lambda, k\right) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}} \tag{10}$$

Where x is the random variable; λ is the scale factor and k is the shape parameter.

The output power of a wind turbine is related to the wind speed as follows:

$$P_{wt} = \begin{cases} P_r \dfrac{v_t^3 - v_c^3}{v_r^3 - v_c^3} & v_c < v_t < v_r \\ P_r & v_r < v_t < v_f \end{cases} \tag{11}$$

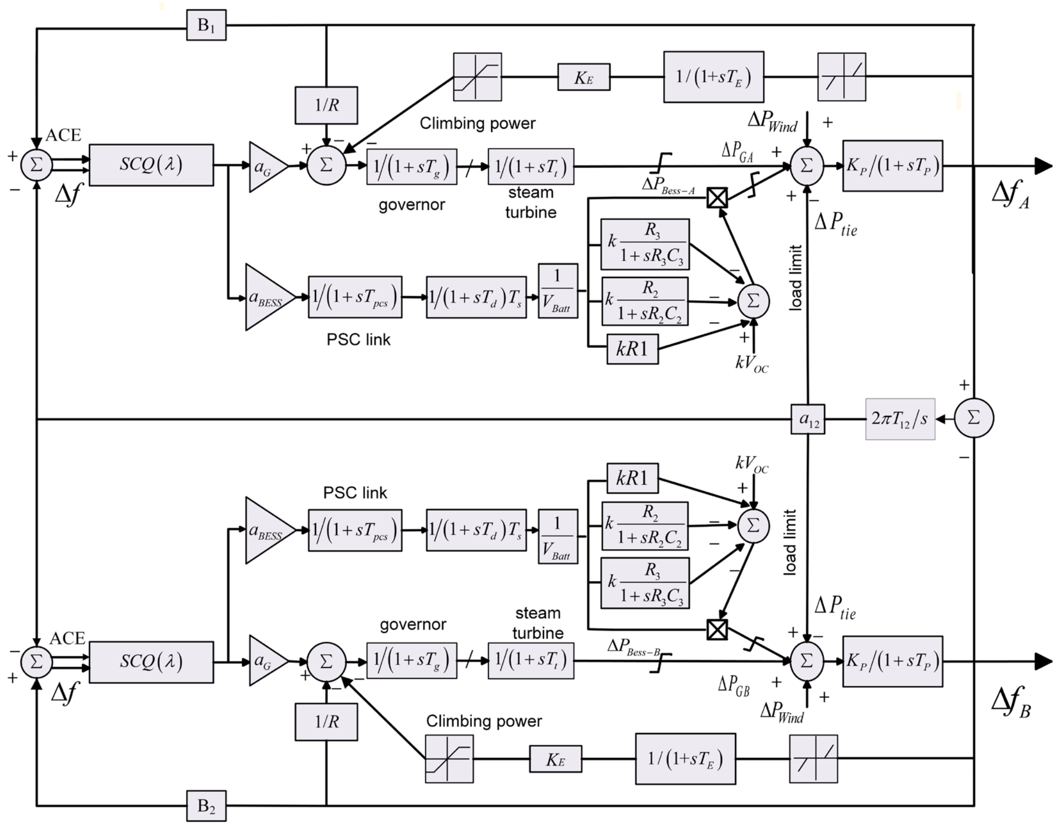
where V_{t} is the wind speed at moment t, V_{c} is the fan cut-in wind speed, taken as 4m/s, V_{r} is the rated wind speed of the fan, taken as 12m/s, V_{f} is the fan cut-out wind speed, taken as 24m/s, P_{r} is the rated output power of the fan, taken as 20kW.

LOAD FREQUENCY CONTROL MODEL FOR MICROGRIDS

The interconnected power system, composed of multiple regions, relies on an effective control method to ensure stability. One commonly used approach is tie line bias control (TBC). TBC takes into consideration both the system frequency deviation and the contact line power deviation, making it a comprehensive representation of power system frequency stability. In addition, TBC allows each control region to account for load changes in neighboring regions, enabling the whole network to achieve optimal power-frequency dynamic performance. To enhance the existing model, this paper incorporates EVs and energy storage modules into the IEEE two-region standard model (Mansour et al., 2022). The modified model is depicted in Figure 3.

In Figure 3, Δf_{A} and Δf_{B} are the frequency deviation of region 1 and region 2, respectively, B_1 and B_2 represent the frequency deviation coefficients, ACE is the regional control deviation, R is the speed control coefficient, Tg is the regulator time constant, Tt is the microfuel time constant, ΔP_{GA} and ΔP_{GB} symbolize the changes in the power output of the unit, and K_{P} is the active frequency

Figure 3. Two-Region LFC Model



conversion factor of the system. T_P is the time constant of the power system, T_{12} is the synchronization coefficient of the contact line, and a_{12} signifies the power conversion coefficient.

The simulation model parameters are shown in Table 1.

Modeling Of SCQ(λ) Algorithm for Microgrid Load Frequency Control

Self-Correcting Q Learning

Temporal-difference learning is an advanced algorithm used in reinforcement learning. It is a model-free approach that estimates the value function, which is a crucial component in decision-making processes. TD learning can be divided into two main categories: single-estimator and dual-estimator methods.

Q-learning algorithm is widely recognized as a prominent single-estimator method in reinforcement learning. The term "single-estimation" refers to the practice of using the maximum estimate from a set of estimates to approximate the expected value. Introduced by Watkins in 1989, Q-learning demonstrates exceptional self-learning capabilities, aiming to discover state-action optimal policies that maximize cumulative rewards (Watkins & Dayan, 1992). Nonetheless, the Q-learning algorithm is subject to maximization bias. This bias emerges from the tendency to use the maximum action value as an approximation for the maximum desired action value. Consequently, action values may be overestimated, resulting in overly optimistic estimates (Wu et al., 2021).

To solve this problem, Hasselt (2010) proposed the Double Q-learning algorithm, which utilizes two Q-functions as part of the learning process. The key feature of this algorithm lies in its ability to

Table 1. Parameters of Two-region LFC Model

Parameters	Area 1	Area 2
n_{EVs}	100	100
C_E	32 kW•h	32 kW•h
SOC_{max}	0.9	0.9
SOC_{min}	0.1	0.1
SOC_{in}	N(0.4,0.01)	N(0.4,0.01)
SOC_p	N(0.8,0.01)	N(0.8,0.01)
η	0.92	0.92
Pmax	7kW	7kW
R	0.05	0.062
B	20.6	16.9

sample actions by incorporating the values from both Q-functions. By doing so, the overestimation problem of action values can be effectively mitigated. However, it's worth noting that this approach may result in an underestimation of action values. Overall, the Double Q-learning algorithm stands out as a prominent double estimator method in the field.

To address the challenge of balancing overestimation and underestimation of action values, recent research by Zhu & Rigotti (2020) proposes a novel self-correcting estimator. This estimator is designed to estimate the maximum expected value, effectively mitigating the issues caused by overestimation with a single estimator and underestimation with a double estimator.

According to equation 14, two independent unbiased estimation sets $\mathrm{E}\left[Q_n\left(s',\mathrm{a}\right)\right]$ as well as $Q_{n+1}\left(s',\mathrm{a}\right)$ of $Q_{n+1}^{\beta}\left(s',\mathrm{a}\right)$ are utilized to construct another unbiased estimation set:

$$Q_n\left(s',a\right)=\tau Q_{n+1}\left(s',a\right)+\left(1-\tau\right)Q_{n+1}^{\beta}\left(s',a\right) \tag{12}$$

where $\tau \in \left(0,1\right)$ denotes the degree of correlation between $Q_{n+1}\left(s',\mathrm{a}\right)$ and $Q_{n+1}^{\beta}\left(s',\mathrm{a}\right)$. The smaller τ is, the lower the degree of correlation is. When τ tends to zero it is not correlated at all.

The bias of the self-correcting estimator $Q_n\left(s',\mathrm{a}\right)$ is always between the positive bias of the single estimator and the negative bias of the dual estimator, which can be achieved by balancing the overestimation due to the single estimator as well as the underestimation due to the dual estimator. By choosing the appropriate parameter β, the maximum bias can be eliminated completely. Compared to the Double Q-learning algorithm, the SCQ algorithm chooses value functions with different time steps, which eliminates the need for updating two value functions simultaneously and reduces computational as well as memory costs.

Construction of The SCQ(λ) Algorithm

In this paper, a novel multi-step fast convergence algorithm SCQ(λ) is proposed based on the SCQ algorithm incorporating eligibility traces using discrete-time Markov decision process as the mathematical basis to solve the time confidence allocation problem of SCQ and improve the convergence speed of the algorithm (Sutton, 1988; Sutton & Barto, 1998) . Eligibility traces play a crucial role in propagating error updates in reinforcement learning algorithms. They act as a form of memory, preserving information about past events and gradually decaying over time. When a state is accessed or an action is performed, the eligibility traces remember these events and allow the system to update the credit value or assign errors specifically to relevant states or actions. By utilizing eligibility traces, the algorithm can focus its updates on the states and actions that directly contributed to the TD error. This selective approach avoids unnecessary updates to unrelated elements and accelerates the convergence of the algorithm. The errors are propagated throughout the entire state space, ensuring that only the accessed states and actions are influenced during the Q-value update process (Fu et al., 2013).

The commonly used eligibility traces are cumulative traces, substitution traces, and Dutch traces, among which substitution traces have a wide range of applications in reinforcement learning. The updated formula of the algorithm is as follows:

$$e_n\left(S,A\right) = \begin{cases} 1 & S = s, A = a \\ \gamma\lambda e_{n-1}\left(S,A\right) + 1 & S \neq s, A \neq a \end{cases} \tag{13}$$

$$\delta_n = r\left(s,a\right) + \gamma \max_{a_g \in A} Q_n\left(s', a_g\right) - \max_{a_g \in A} Q_n\left(s, a_g\right) \tag{14}$$

$$Q_{n+1}\left(S,A\right) = Q_n\left(S,A\right) + \alpha\delta_n e_n\left(S,A\right) \tag{15}$$

$$Q_{n+1}\left(s,a\right) = Q_{n+1}\left(s,a\right) + \alpha\left(r + \gamma \max_{a_g \in A} Q_n\left(s', a_g\right) - Q_n\left(s,a\right)\right) \tag{16}$$

where *δ* is the error in the *Q*-value function; e_t denotes the eligibility of the state-action pair, indicating its contribution to the generation of *δ*. The magnitude of the eligibility determines the extent of the updates made, with larger eligibility values assigned a higher level of temporal credence. Conversely, state-action pairs with smaller eligibility values receive a lower temporal credence, implying that they are subject to reduced rewards or penalties from the current event. Here, *γ* is the discount factor, where $0 \leq \gamma \leq 1$; *λ* is the decay factor, where $0 \leq \lambda \leq 1$. After a number of iterative updates, the value matrix converges to the optimal *Q*-value matrix with a probability of 1.

Defining $\beta = 1 / \left(1 - \tau\right)$, equation 14 can be rewritten as:

$$Q_{n+1}^{\beta}\left(s', a\right) = Q_{n+1}\left(s', a\right) - \beta\left[Q_{n+1}\left(s', a\right) - Q_n\left(s', a\right)\right] \tag{17}$$

However, since $Q_{n+1}\left(s', \mathrm{a}\right)$ is not available at the moment of time step n, to solve this problem, we utilize $Q_{n-1}\left(s', \mathrm{a}\right)$ and $Q_n\left(s', \mathrm{a}\right)$ from the previous update step instead of $Q_n\left(s', \mathrm{a}\right)$ and $Q_{n+1}\left(s', \mathrm{a}\right)$:

$$Q_{n+1}^{\beta}\left(s', a\right) = Q_{n+1}\left(s', a\right) - \beta\left[Q_n\left(s', a\right) - Q_{n-1}\left(s', a\right)\right] \tag{18}$$

The inclusion of value functions with varying time steps in the updating process signifies a progressive self-adjustment of the estimator's bias. This occurs as the disparity between the value function at time step n and the value function at time step n-1 diminishes progressively.

After that we use $\hat{a}^{\beta} = argmaxQ_n^{\beta}\left(s', \mathrm{a}\right)$ for action selection and $Q_n^{\beta}\left(s', \hat{a}^{\beta}\right)$ to estimate the value of the next step:

$$Q_{n+1}\left(s,a\right) = Q_n\left(s,a\right)\alpha\left[r + \gamma Q_n\left(s', \hat{a}^{\beta}\right) - Q_n\left(s,a\right)\right] \tag{19}$$

State Space and Action Space Definitions

In terms of grid frequency performance evaluation, the current evaluation standards implemented in China's power grid are based on the control performance standard (CPS) 1 and 2 used in the North American power grid.

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn. CPS1 and CPS2 are key performance indicators used in power system control. CPS1 assesses the long-term performance by measuring the ACE and the frequency deviation of the control area every minute. On the other hand, CPS2 focuses on the short-term performance by measuring the average ACE value during 10-minute intervals. The criteria for CPS1 and CPS2 are outlined below:

CPS1≥100%,CPS2≥90%。

In an automated power generation system, the smart controller takes input from the actual grid dispatch in the form of total power generation regulation commands. The state space encompasses grid frequency deviation and CPS dataset, while the action space comprises power fluctuation dataset. The controller processes this information and updates the action set, optimizing it based on the state value and reward value. Ultimately, the controller outputs an optimal control signal to ensure the power system's frequency dynamic balance is maintained.

Reward Function Definition

To ensure the stabilization of power fluctuations within an acceptable range and optimize the long-term benefits of CPS, this paper proposes a reward function that integrates ACE and CPS1 through a linearly weighted combination. The reward function for each regional grid is defined as shown in equation 22.

$$R\left(s_{n-1}, s_{n}, a_{n-1}\right) = -\eta\left[ACE\left(t\right)\right]^{2} - \frac{\left(1-\eta\right)\left[CPS1\left(t\right)\right]}{1000} \tag{20}$$

where *ACE(t)* is the instantaneous value of ACE of the system at time t, *CPS1(t)* is the value of CPS1 of the system at time t, η is the weight value of ACE, $1\text{-}\eta$ is the weight value of CPS1, and the value of η is taken as 0.5 in this paper.

Parameter Setting

The system parameters need to be set appropriately, among other things:

Learning rate α

In order to enhance the stability of the algorithm, this parameter is utilized where the velocity of updating the value function is adjusted based on the value of α. Specifically, when α is large, the speed of updating the value function is increased, and vice versa when α is small. This approach effectively enhances the overall stability of the system. To strike a balance between learning speed and stability, a number of simulations have demonstrated that setting α to 0.1 yields optimal performance for the control system.

Iterative strategy learning factor β

It is used to measure the degree of influence of the action selection strategy on the iterative strategy update, where a larger β speeds up convergence and a smaller β ensures that the system is able to fully explore other actions in the space. In this paper, β is chosen to be 0.5.

Attenuation factor for eligibility traces λ

It is used to assign credits to state-action pairs. A smaller value of λ results in a lower assignment of reputation, while a larger value of λ leads to a slower decay of reputation for the previous action. In our study, we have set λ equal to 0.95.

Discount factor γ

In order to strike a balance between reward weights, the parameter γ is employed. When γ approaches 1, the intelligent system places more emphasis on long-term rewards. Conversely, when γ approaches 0, the system prioritizes immediate rewards. For the purposes of this study, we have set γ at a value of 0.9.

SIMULATION ANALYSIS

Pre-Learning Phase

In reinforcement learning, it is essential for intelligent systems to undergo a preliminary phase known as randomized trial and error pre-learning. This pre-learning phase, which is the focus of this paper, involves the introduction of a sinusoidal load perturbation. Specifically, a sinusoidal load perturbation with a period of 1200 seconds, an amplitude of 1000kW, and a duration of 20000 seconds is implemented in order to train SCQ(λ) and facilitate its convergence towards the optimal policy.

Figure 4 shows the control performance curve of the SCQ(λ) controller during the pre-learning phase in region A. Figure 4(a) shows the load perturbation curve, from which it can be seen that the SCQ(λ) controller can basically track the upper load perturbation around 1300s. Figure 4(b) shows the average value of 10min CPS1, which is 195.5494% in region A of the figure, and the 10min assessment criterion of CPS1 is kept above 185%. Figure 4(c) shows the frequency variation curve under sinusoidal disturbance, and the maximum power deviation in the region is 0.0056 Hz, which is much smaller than the actual requirement of 0.2 Hz. Figure 4(d) indicates the 10min average curve of the ACE with a value of 1.7448kW, and the simulation results show that the 10min assessment index value of the ACE always stays within 2kW, which shows that the designed controller has strong stability.

The 2-parameter $| \left| Q_{ik}\left(s,a\right) - Q_{k-1}\left(s,a\right) \right| |^{2} \leq$ of the Q matrix (ς=0.0001 is the specified criterion) is chosen in this paper as the termination criterion for pre-learning to reach the optimal policy. Figure 5 shows the convergence effect of the SCQ(λ) algorithm in region A in comparison to the Q, SCQ, Q(λ) intelligent algorithms. As depicted in the figure, the SCQ(λ) algorithm can significantly improve the convergence speed.

In summary, extensive training explorations have demonstrated that the SCQ(λ) controller has successfully approximated the optimal CPS control strategy. As a result, the SCQ(λ) controller is ready to be effectively deployed in real-world environments.

Simulation Analysis for Different Relative Perturbation Intensities

During online operation s, it is essential to simulate a sudden increase in load in the grid and introduce a step load perturbation to the two-region model. This paper proposes a controller design that utilizes four algorithms: Q, Q(λ), SCQ, and SCQ(λ). The objective is to compare and analyze the performance of the two-region model under different controllers by applying a step load perturbation with an

Figure 4. Pre-Learning Effects of the SCQ(λ) Controller

(**a**) *Load Disturbance Curve,* (**b**) *Mean value curve of 10minCPS1,* (**c**) *Frequency variation curves under sinusoidal perturbation,* (**d**) *10min mean curve of ACE.*

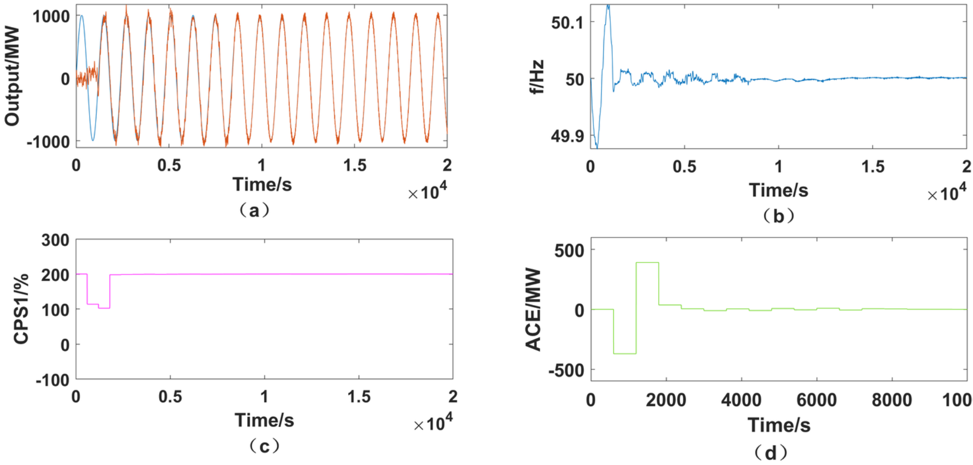
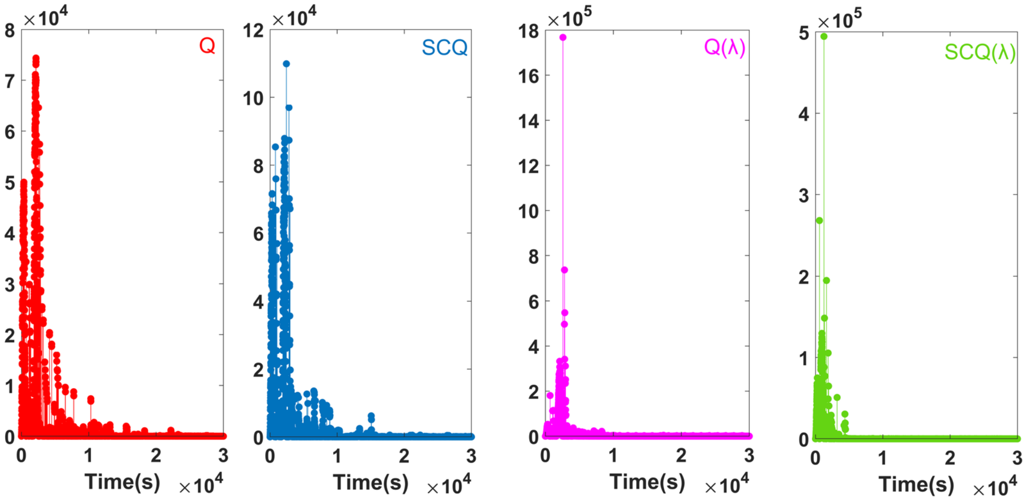


Figure 5. Comparison of Convergence Effect of Pre-Learning Stage of Each Algorithm

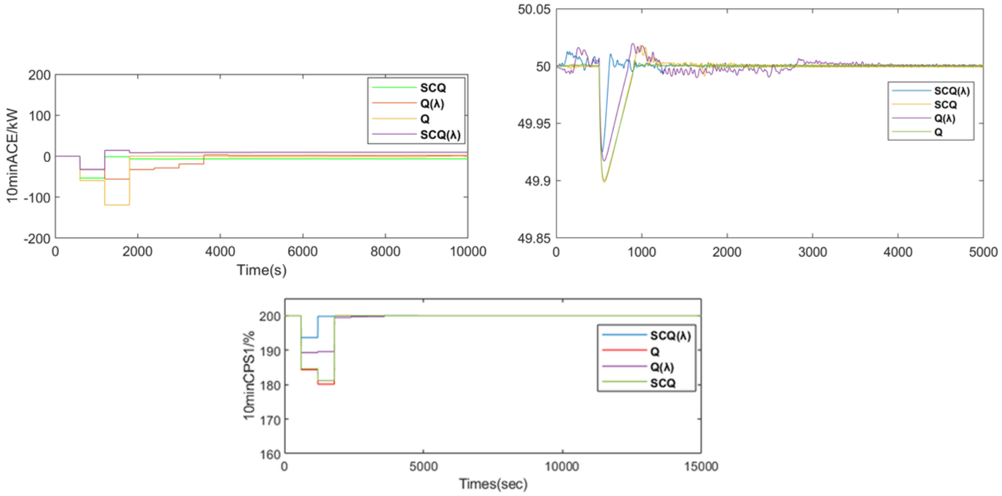


amplitude of 1000kW. The obtained control performance curves for the A-region, based on various algorithmic controllers, are illustrated in Figure 6.

In Figure 6(a), the comparative effect of frequency change curves for the four intelligent algorithms is displayed. The absolute change in frequency, |Δf|, for each algorithm is recorded as follows: Algorithm 1 yields 0.0110Hz, Algorithm 2 yields 0.0030Hz, Algorithm 3 yields 0.0055Hz, and Algorithm 4 yields 0.0014Hz. Notably, SCQ(λ) exhibits a significantly reduced |Δf| compared to the other algorithms. Figure 6(b) presents the average values of 10-minute ACE (Area Control Error). The respective values for each algorithm are as follows: Algorithm 1 yields 19.2539kW, Algorithm 2 yields 4.9626kW, Algorithm 3 yields 7.3493kW, and Algorithm 4 yields 3.8191kW. Figure 6(c) illustrates the average variation curve of 10-minute CPS1 (Control Performance Score

Figure 6. Simulation Curves of Different Algorithms Under Step Load Disturbance

Note. (**a**) *Frequency change curve,* (**b**) *Mean change curve of 10minACE,* (**c**) *Curve of change in mean value of 10minCPS1*



1) with the values derived from the four intelligent algorithms: 199.297%, 199.675%, 199.552%, and 199.876%, respectively. Table 2 displays the control performance indexes of the four algorithms under step load perturbation. It is evident that the SCQ(λ) algorithm outperforms the other algorithms in all performance indexes, showcasing its exceptional control performance.

An amplitude of 1000 kW square wave load disturbance is applied for a duration of 24 hours to simulate extreme cases of load surge and smoothness, which is done in order to further validate the control performance of the proposed algorithm. To evaluate the performance of the controllers for the four algorithms (Q, Q(λ), SCQ, and SCQ(λ)), a 5-hour load perturbation is introduced as the evaluation period. Table 3 presents the performance assessment indexes of these four algorithms in region A. The simulation results demonstrate that the SCQ(λ) controller maintains stable control effectiveness even when subjected to random load fluctuations.

DISCUSSION

In order to address the issue of frequency fluctuation in microgrids resulting from the integration of large-scale new energy sources into the power grid, this research introduces a reinforcement learning-based SCQ(λ) control algorithm. Additionally, a two-region load frequency control model is developed, which includes electric vehicles, wind power, MTs (microturbines), and battery energy

Table 2. Different Algorithms Control Performance Based on Step Perturbation

Algorithms	|Δf|/Hz	ACE/kW	CPS1/%
Q	0.0110	19.2539	199.297
Q(λ)	0.0030	4.9626	199.675
SCQ	0.0055	7.3493	199.552
SCQ(λ)	0.0014	3.8191	199.876

Table 3. Different Algorithms Based on Square Wave Perturbation Control Performance

Algorithms	|Δf|/Hz	ACE/kW	CPS1/%
Q	0.02725	67.163	180.21
Q(λ)	0.01383	39.517	191.84
SCQ	0.01345	39.426	192.29
SCQ(λ)	0.01123	32.291	193.89

storage. Conventional reinforcement learning approaches often suffer from the problems of state action function value overestimation and underestimation. To overcome these limitations, this study proposes the utilization of a self-correcting estimator which enhances the accuracy compared to the Q-learning algorithm. Moreover, considering the inefficiency of the self-correcting estimator when updated on a per-pass basis, this paper incorporates the eligibility traces strategy in combination with the self-correcting estimator. Consequently, the SCQ(λ) algorithm is introduced, significantly improving the convergence performance of the overall control algorithm.

In the simulation analysis, sine wave, step perturbation, and square wave perturbation are used for simulation experiments in this paper. According to the simulation results, the SCQ(λ) controller shows excellent exploratory capability, which can effectively cope with the load frequency control problem of the microgrid under strong stochastic conditions and significantly improve the frequency control rate and effect.

Specifically, when subjected to step perturbation, the SCQ(λ) controller achieves a frequency deviation of only 0.0014 Hz. Furthermore, it demonstrates an average 10-minute ACE (Area Control Error) of 3.8191 kW and a corresponding 10-minute CPS1 (Control Performance Standard 1) value of 199.876%. When exposed to square wave perturbation, the SCQ(λ) controller outperforms the Q controller by achieving a reduction of 58.7% in frequency deviation. Additionally, it exhibits a decrease of 51.9% in the 10-minute ACE and an increase of 13.68% in the 10-minute CPS1 value compared to the Q controller. These simulation results demonstrate that the SCQ(λ) controller surpasses other controllers in all performance indices when enhanced random perturbation is applied, thus effectively ensuring system frequency stabilization and dynamic balance of power exchange in the contact line.

The simulation model includes the integration of electric vehicle and wind power generation models, among others. However, it is worth noting that the model is not connected to the actual power grid. Future research will focus on the development of relevant techniques to address this limitation.

REFERENCES

Barbalho, P. I. N., Lacerda, V. A., Fernandes, R. A. S., & Coury, D. V. (2022). Deep reinforcement learning-based secondary control for microgrids in islanded mode. *Electric Power Systems Research*, *11*, 1–7. doi:10.1016/j.epsr.2022.108315

Boglou, V., Karavas, C., Karlis, A., Arvanitis, K. G., & Palaiologou, I. (2023). An optimal distributed RES sizing strategy in hybrid low voltage networks focused on EVs' integration. *IEEE Access : Practical Innovations, Open Solutions*, *11*, 16250–16270. doi:10.1109/ACCESS.2023.3245152

Boglou, V., Karavas, C. S., Karlis, A., & Arvanitis, K. (2022). An intelligent decentralized energy management strategy for the optimal electric vehicles' charging in low-voltage islanded microgrids. *International Journal of Energy Research*, *46*(3), 46. doi:10.1002/er.7358

Chae, S. H., Kim, G. H., Choi, Y. J., & Kim, E. H. (2020). Design of isolated microgrid system considering controllable EV charging demand. *Sustainability (Basel)*, *12*(22), 9746. Advance online publication. doi:10.3390/su12229746

Chen, X., Shuai, C., Wu, Y., & Zhang, Y. (2020). Analysis on the carbon emission peaks of China's industrial, building, transport, and agricultural sectors. *The Science of the Total Environment*, *709*, 135768. doi:10.1016/j.scitotenv.2019.135768 PMID:31884279

Fan, P., Ke, S., Kamel, S., Yang, J., Li, Y., Xiao, J., Xu, B., & Rashed, G. I. (2022). New findings reported from Wuhan University describe advances in electronics (A frequency and voltage coordinated control strategy of island microgrid including electric vehicles). *Electronics (Basel)*, *11*(1), 17. doi:10.3390/electronics11010017

Fan, P., Yang, J., & Xiao, J. (2022). Load frequency control strategy based on deep Q learning for island microgrid with electric vehicles. *Electric Power Construction*, *43*, 91–99.

Fu, Q., Liu, Q., Xiao, F., & Chen, G. (2013). The second order temporal difference error for Sarsa (λ). *Proceedings of the IEEE Symposium on Adaptive Dynamic Programming & Reinforcement Learning*.

Hasselt, H. V. (2010). Double Q-learning. *Advances in Neural Information Processing Systems*, *23*, 2613–2621.

Hasselt, H. V., Guez, A., & Silver, D. (2015). Deep reinforcement learning with Double Q-learning. *AAAI Conference on Artificial Intelligence; Innovative Applications of Artificial Intelligence Conference*.

Hongxin, L., Feifei, X., Peixiao, F., Linping, L., Hongjun, W., Yi, Q., Song, K., Yonghui, L., & Jun, Y. (2021). Load frequency control strategy of island microgrid with flexible resources based on DQN. *The 2021 IEEE Sustainable Power and Energy Conference (iSPEC)*, 632-637. doi:10.1109/iSPEC53008.2021.9735574

Huang, J., Li, X., Cao, Y., Zhang, Q., & Liu, W. (2015). Battery energy storage power supply simulation model for power grid frequency regulation. *Dianli Xitong Zidonghua*, *39*, 20–24, 74.

Iqbal, S., Xin, A., Jan, M. U., Salman, S., Zaki, A. U., Rehman, H. U., Shinwari, M. F., & Abdelbaky, M. A. (2020). V2G strategy for primary frequency control of an industrial microgrid considering the charging station operator. *Electronics (Basel)*, *9*(4), 549. Advance online publication. doi:10.3390/electronics9040549

Karkevandi, A. E., Daryani, M. J., & Usta, Ö. (2018). ANFIS-based intelligent PI controller for secondary frequency and voltage control of microgrid. *Proceedings of the 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. doi:10.1109/ISGTEurope.2018.8571748

Khokhar, B., & Parmar, K. P. S. (2022). A novel adaptive intelligent MPC scheme for frequency stabilization of a microgrid considering SoC control of EVs. *Applied Energy*, *309*, 118423. doi:10.1016/j.apenergy.2021.118423

Lam, Q. L., Bratcu, A. I., Riu, D., Boudinet, C., Labonne, A., & Thomas, M. (2020). Primary frequency H∞ control in stand-alone microgrids with storage units: A robustness analysis confirmed by real-time experiments. *International Journal of Electrical Power & Energy Systems*, *115*, 105507. doi:10.1016/j.ijepes.2019.105507

Li, J., Ai, X., & Hu, J. (2019). Supplementary frequency regulation modeling and control strategy with electric vehicle. *Power System Technology*, *43*, 495–503. doi:10.13335/j.1000-3673.pst.2018.1784

Mansour, S., Badr, A. O., Attia, M. A., Sameh, M. A., Kotb, H., Elgamli, E., & Shouran, M. (2022). Fuzzy logic controller equilibrium base to enhance AGC system performance with renewable energy disturbances. *Energies*, *2022*(15), 6709. doi:10.3390/en15186709

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction.* Academic Press.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), 9–44. doi:10.1007/BF00115009

Wang, C., Yu, H., Chai, L., Liu, H., & Zhu, B. (2021). Emergency load shedding strategy for microgrids based on dueling deep Q-learning. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 19707–19715. doi:10.1109/ACCESS.2021.3055401

Wang, P., Tang, H., & Lv, K. (2018). Simulation model for the AGC system of isolated microgrid based on Q-learning method. *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS).*

Watkins, C. J. C. H., & Dayan, P. (1992). Technical note: Q-learning. *Machine Learning*, *8*(3-4), 279–292. doi:10.1007/BF00992698

Wu, R., Gong, J., Tong, W., & Fan, B. (2021). Network attack path selection and evaluation based on Q-learning. *Applied Sciences (Basel, Switzerland)*, *11*(1), 285. doi:10.3390/app11010285

Zhao, M., Zhang, J., & Ren, K. (2018). Load frequency control of interconnected power system with wind power based on active disturbance rejection control. *Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. doi:10.1109/IAEAC.2018.8577941

Zhu, R., & Rigotti, M. (2020). *Self-correcting Q-learning*. Academic Press.

Ziras, C., Marinelli, M., Prostejovsky, A., & Bindner, H. W. (2019). Decentralized and discretized control for storage systems offering primary frequency control. *Electric Power Systems Research*, *177*, 106000. Advance online publication. doi:10.1016/j.epsr.2019.106000

Qiang Wang, adjunct professor, master's degree, graduated from North China Electric Power University in 2002. Worked in China Three Gorges University. His research interests include Power system power electronic device, distribution network power quality, new energy power generation and grid-connected technology.

Zhenwei Huang, master degree candidate, graduated from Harbin University of Science and Technology in 2016. His research interests include Power system operation and control.