Aalborg Universitet

Novel adaptive stability enhancement strategy for power systems based on deep reinforcement learning

Zhao, Yincheng; Hu, Weihao; Zhang, Guozhou; Huang, Qi; Chen, Zhe; Blaabjerg, Frede

Published in:
International Journal of Electrical Power & Energy Systems

DOI (link to publication from Publisher):
10.1016/j.ijepes.2023.109215

Creative Commons License
CC BY-NC-ND 4.0

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Zhao, Y., Hu, W., Zhang, G., Huang, Q., Chen, Z., & Blaabjerg, F. (2023). Novel adaptive stability enhancement strategy for power systems based on deep reinforcement learning. *International Journal of Electrical Power & Energy Systems*, *152*, 1-13. Article 109215. https://doi.org/10.1016/j.ijepes.2023.109215

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: December 05, 2025

Electrical Power and Energy Systems 152 (2023) 109215

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Novel adaptive stability enhancement strategy for power systems based on deep reinforcement learning

Yincheng Zhao [a], Weihao Hu [a,*], Guozhou Zhang [a], Qi Huang [a], Zhe Chen [b], Frede Blaabjerg [b]

[a] School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
[b] Department of Energy Technology, Aalborg University, Aalborg, Denmark

ARTICLE INFO

Keywords:
Wind energy
Static var compensator
Additional damping controller
Generalized regression neural network
Low-frequency oscillation
Deep reinforcement learning

ABSTRACT

As the access rate of wind energy in a power system has significantly increased, stabilizing the power system has become challenging. Among these challenges, low-frequency oscillation is one of the most harmful problems, effectively resolved by adding a damping controller according to the relevant properties of the low-frequency oscillation. However, the controller often fails to adapt to the constantly changing wind energy system owing to the lack of a targeted dynamic change strategy. Thus, to address this issue, an adaptive stabilization strategy that uses a static var compensator with an additional damping controller structure is proposed. Specifically, the entire power system is equivalently represented as a generalized regression neural network, with a deep reinforcement learning algorithm called soft actor-critic introduced to train the agent based on the generalized regression neural network model. After the training process, the agent can provide additional efficient static var compensator damping controller parameters under different operating conditions, vastly improving the system stability. Simulation results verify the improved performance using the proposed strategy compared to other optimization methods, regardless of whether the low-frequency oscillations were suppressed in the time or frequency domains.

1. Introduction

An increasing number of wind farms (WFs) are being installed in power systems to alleviate energy shortages [1], with increased global installed wind energy capacity in certain areas [2]. For example, during 2007–2013, the United States had an average annual increase of 7.1 GW in installed wind power capacity [3]. China may augment its 2030 renewable energy capacity target from 20% to 35% by 2030; meanwhile, wind energy plays an important role in planning [4]. Global average annual wind capacity additions are planned to be within the scope of 50 GW from 2017 to 2040 [5]. Studies have confirmed that renewable energy utilisation will increase by 28% globally between 2015 and 2040 [6]. Using renewable energy, such as wind power, in power systems meets the electricity demand [7] and reduces carbon dioxide emissions [8], thus indicating that wind energy is considered an important energy source to support the operation of a sustainable society.

However, owing to the uncertain nature of wind speed, a low-frequency oscillation (LFO) arises when several wind farms are integrated into power systems [9]. LFO severely limits the energy transmission ability and disrupts grid stability in certain situations. If the LFO cannot be suppressed for a period of time, generators and transmission lines are tripped and cut off [10,11]. Thus, LFO is a major issue in power systems. Therefore, an effective method should be explored to suppress this phenomenon and ensure the stability of a power system.

A series of measures have been proposed to search for a valid suppression method. Flexible alternative current transmission system (FACTS) devices are currently the most popular controlling devices for these measures. In a previous study [12], certain types of FACTSs were used for damping and suppressing oscillations in power systems. These types of devices can improve the damping performance of the system by increasing the transient stability limit and improving damping for synchronising power flow oscillations in the working process. In another study [13], a new SVC structure was proposed according to the fuzzy logic control theory. The designed SVC can flexibly change the electrical susceptance to improve the system stability. However, the effects of FACTS are limited, considering that the structure of FACTS is simple. To overcome these drawbacks of FACTS, an ADC is applied to further

[*] Corresponding author.
E-mail addresses: yczloveln@163.com (Y. Zhao), whu@uestc.edu.cn (W. Hu), zgz@std.uestc.edu.cn (G. Zhang), hwong@uestc.edu.cn (Q. Huang), zch@et.aau.dk (Z. Chen), fbl@et.aau.dk (F. Blaabjerg).

https://doi.org/10.1016/j.ijepes.2023.109215
Received 11 December 2022; Received in revised form 3 March 2023; Accepted 27 April 2023
Available online 11 May 2023
0142-0615/© 2023 Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

improve damping. Various intelligent algorithms have been utilised to enhance the performance of ADC. An optimal method based on the particle swarm optimisation (PSO) algorithm was used to design an ADC to suppress LFO [14]. The eigenvalue analysis method and ADC tuning based on the residues method have been proposed to improve the system damping according to the position of the eigenvalues in a complex plane [15]. A particle swarm optimisation method with an oscillating exponential decay (PSO-OED) algorithm was previously proposed to design the parameters of a power system stabiliser and an automatic voltage regulator [16]. In addition, researchers applied genetic algorithms (GA) to obtain optimal coordination between the PSSs and unified power flow controller (UPFC)-based stabiliser parameter [17]. Controller parameters using these methods can significantly promote the damping of a power system. However, the controllers using these algorithms are not sufficiently affected because traditional optimisation algorithms are limited in optimising the process, possibly resulting in limited system damping lifting in certain scenarios.

Robust optimisation methods have been proposed in several studies to ensure the effectiveness of ADC under different conditions. A robust and decentralised Takagi Sugeno (TS) fuzzy controller was utilised to design a damping controller to manage oscillations, while the parameters of the controller were tuned by using the probability collocation method (PCM) [18]. A robust internal model control proportional integral derivative double derivative (IMC-PIDD2)-enabled controller was designed for suppressing active power oscillations and frequency oscillations [19]. An H_∞ robust control strategy was investigated in the presence of system uncertainties [20]. A robust type-2 fuzzy-based fractional-order PID control method has been proposed [21]. Moreover, a dynamic genetic algorithm and bacteria foraging algorithm were combined for tuning these types of controller parameters. The robust optimisation method is more generalised than traditional control methods because different scenarios and states are considered in the optimisation process. In contrast, the optimization process considering extreme cases provides conservative optimization results.

In recent years, the adaptive control method has gradually become popular. Owing to its dynamic characteristics, controllers designed by the adaptive control method can obtain a better suppression effect despite undergoing extreme scenarios. For example, an adaptive optimal damping control architecture was proposed for mitigating oscillations in renewable energy systems [22]. Analogously, a novel online and adaptive wide-area damping controller (WADC) was used to achieve maximum damping of the inter-area oscillations [23]. A proportional resonance (PR) controller was designed based on an adaptive method to maintain a stable frequency [24]. In addition, the adaptive control method can achieve more potent results when combined with other control methods. For instance, a method combined with the H_∞ robust optimised method and adaptive control method was presented [25] to enhance the transient stability of a power system in reference [26]. Similarly, an adaptive fuzzy logic controller was proposed to smoothen the grid power curve and enhance the power quality when the wind system operates at different wind speeds [27]. Apparently, the adaptive control method is advanced for diverse variations in new energy systems.

Most of the aforementioned power system transient stability problems rely on the mathematical model of the system to be solved. However, considering controller design problems, most of the mathematical expressions of the system model are unknown, which highly complicates solving the controller design problem. Fortunately, artificial intelligence (AI) technology provides a new method for suppressing power system oscillations. Deep learning (DL) can fit the complex nonlinear relationships of multiple characteristic variables based on a large amount of data and deep network structure. Reinforcement learning (RL) enables an agent to learn based on trial and error, guiding behaviour through rewards obtained by interacting with the environment, with the goal of maximising rewards for the agent. Deep reinforcement learning (DRL) combines the advantages of DL and RL to enable prediction and optimisation in a continuous action space and has been successfully applied in power systems. The proximal policy optimisation (PPO) DRL method was used to tune the parameters for a novel multiband power system stabiliser to control multimode low-frequency oscillations [28]. A DRL algorithm asynchronous advantage actor-critic (A3C) was applied to train an agent to provide optimal parameter settings for PR-PSS [29]. The oscillation damping problem was solved using a novel faster exploration strategy-based deep deterministic policy gradient (DDPG) algorithm [30].

Inspired by the aforementioned research, this study proposes a data-driven adaptive control method for self-tuning the parameters of the SVC-ADC, considering the uncertain characteristics of the wind farm system. The aims of this study are summarised as follows:

1) Build a high-accuracy real-time dynamic system model for adaptive control, recursive least squares (RLS) [31], which have higher accuracy than the conventional Prony [32] and Steiglitz-McBride (SM) methods [33], are used to identify the equivalent transfer function of wind power integrated into the power system in a single scenario. Based on this process, the GRNN is used to learn the mapping relationship between the system transfer function parameters and the system state to obtain a variational model of the dynamic characteristics of the variable wind speed of a wind farm system in continuous state space for adaptive control.
2) To remove the deficiencies of the traditional controller for the suppression performance of multi-oscillation modes, an SVC additional dual-channel damping controller structure is proposed for a scenario in which the dominant mode of LFO is not unique when the wind farm is connected to the power system. The proposed structure, based on system linearization and residue analysis, allows for a more targeted suppression based on the dominant mode of LFO for the system and significantly improves system stability.
3) Convert the SVC-ADC parameter tuning problem into the maximum entropy Markov decision process (MDP) according to the actual wind farm access to the power system change. Meanwhile, an SAC algorithm is also introduced to solve the MDP by training an agent capable of achieving an optimal adaptive control strategy for the dynamic changes in the system for adaptive control.

In Section 2, the required mathematical models of the dynamic power systems under naturally varying wind speeds are obtained. Section 3 introduces the architecture of the SVC with an additional dual-channel damping controller. Section 4 presents the controller tuning problem formulated as an MDP and demonstrates the SAC-based controller tuning process. Section 5 presents the results of the proposed strategy, finally followed by the conclusions in Section 6.

2. Model of the dynamics power system under varying wind speeds

This section briefly demonstrates the mathematical models of the dynamic power system acquisition schemes based on RLS with GRNN. Specifically, the controlled autoregressive and moving average (CARMA) model of the power system is first introduced, after which a model identification method for a single case using the RLS algorithm is proposed. Subsequently, based on the above, the GRNN is used to fit the nonlinear mapping between the wind speed and the power system transfer function to obtain the dynamic equivalent model.

2.1. Power system model with pending parameters

The complexity and nonlinearity of power systems significantly increase owing to the integration of wind turbines, making obtaining models with full dynamic characteristics difficult. In the ADC design problem, low-order equivalent models with a fixed structure and parameters that vary with the operating conditions are sufficient for the

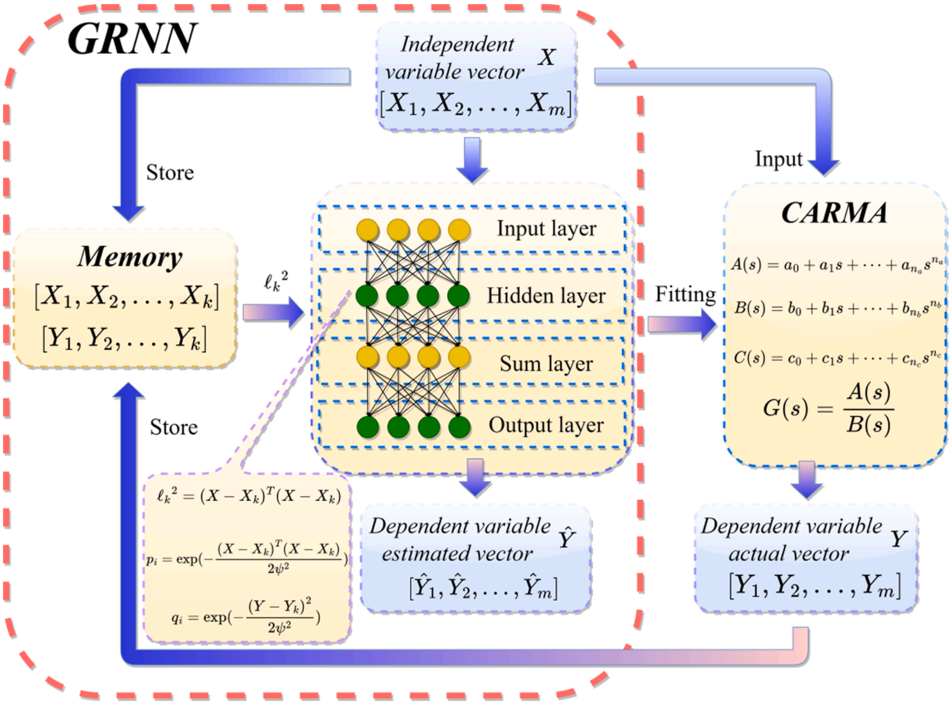
GRNN
Independent variable vector X
$[X_1, X_2, \ldots, X_m]$
Store
Input
Memory
$[X_1, X_2, \ldots, X_k]$
$[Y_1, Y_2, \ldots, Y_k]$
$\ell_k{}^2$
Input layer
Hidden layer
Sum layer
Output layer
Fitting
CARMA
$A(s) = a_0 + a_1 s + \cdots + a_{n_a} s^{n_a}$
$B(s) = b_0 + b_1 s + \cdots + b_{n_b} s^{n_b}$
$C(s) = c_0 + c_1 s + \cdots + c_{n_c} s^{n_c}$
$G(s) = \frac{A(s)}{B(s)}$
Store
$\ell_k{}^2 = (X - X_k)^T (X - X_k)$
$p_i = \exp(-\frac{(X - X_k)^T (X - X_k)}{2\psi^2})$
$q_i = \exp(-\frac{(Y - Y_k)^2}{2\psi^2})$
Dependent variable estimated vector $\hat{Y}$
$[\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_m]$
Dependent variable actual vector Y
$[Y_1, Y_2, \ldots, Y_m]$

Fig. 1. GRNN fitting fundamentals for the CARMA model.

analysis. To characterise the dynamic properties of power systems, the CARMA model is employed to describe the relationship between the input and output signals, which is expressed as follows:

$$A(s)y(t) = B(s)u(t) + C(s)e(t) \tag{1}$$

where $y(t)$ and $u(t)$ are the specific output and input sequences of the system, respectively. $A(s)$ and $B(s)$ are the undetermined coefficient sequences of the CARMA model: $A(s)\in\Re[s,n_a]$, $A(s) = a_0 + a_1s+\ldots+ a_{na}s^{na}$ where $a_0 = 1$, $B(s) \in\bullet\Re[s,n_b]$, and $B(s) = b_0 + b_1s+\ldots+b_{nb}s^{nb}$. The system order n_b needs to be set as large as possible to include the first zero value owing to the non-zero dead time. For simplification, it is assumed that $e(t)$ is white noise, providing $C(s) = 0$.

2.2. Single case model parameter identification

Under a single operating condition, the model parameters of the CARMA wind power system are static, which can be estimated by an RLS algorithm with a variable forgetting factor. Introducing an appropriate forgetting factor can significantly improve the recognition accuracy when applying the RLS algorithm to the parameters. To identify the CARMA model parameters shown in (1), the algorithm can be expressed as follows:

$$\begin{aligned} \eta(t) &= y(t) - \lambda^T(t)\gamma(t-1) \\ K(t) &= P(t-1)\lambda(t)/[1+\lambda^T(t)P(t-1)\lambda(t)] \\ \gamma(t) &= \gamma(t-1) + K(t)\eta(t) \\ \mu(t) &= 1 - [1-\lambda^T(t)K(t)]\eta(t)^2/\Sigma_0 \\ P(t) &= [I - K(t)\lambda^T(t)]P(t-1)/\lambda(t) \end{aligned} \tag{2}$$

where $\gamma(t) = [a_1, \cdots, a_{n_a}, b_0, \cdots, b_{n_b}]^T$ denotes the vector of the requested CARMA model parameter sequence, $\lambda(t) = [-y(t-1), \ldots, -y(t- \ n_a),$ $u(t-1), \ldots, u(t-n_b)]^T$ expresses the input–output time series, $P(t)$ denotes the covariance matrix, I is the identity matrix, $K(t)$ is the adjustment gain, $0 < \mu < 1$ is the forgetting factor used to gradually reduce the impact of past results, and $\sum_0$ is constantly prepared. $\eta(t)$ is used as an intermediate variable to support the implementation of the RLS process. Owing to the various possible sudden disturbances in a power system, modelling may present significant errors, which were prevented by introducing moving boundaries for each parameter. The high and low boundaries for each parameter are defined as follows:

$$\begin{aligned} B_h &= B + \tau|B| \\ B_l &= B - \tau|B| \end{aligned} \tag{3}$$

where $0 < \tau < 1$; the larger the value τ, the more likely the parameters vary. B represents the estimated value of the parameter obtained in each iteration. B_h and B_l represent the upper and lower boundaries of the estimated parameters, respectively.

2.3. Model of the wind speed variation system based on GRNN

As a single-pass learning neural network, GRNN has a particular training strategy where the network memorises each input and target data, achieving a better performance in the fitting task. After the GRNN is adequately trained, it can generalise and receive the output according to the memory of the GRNN. In addition, the training strategy is rapid and precise given that the GRNN is trained with a significant amount of information in its memory, and overfitting or underfitting will not occur.

Suppose $f(x, y)$ is a given joint continuous probability density function (PDF) of a stochastic independent variable vector x and stochastic scalar dependent variable y; if x is a specific measured vector of the random variable x, then the regression of y on X can be expressed as follows:

$$E[y|X] = \frac{\int_{-\infty}^{\infty} yf(X,y)\partial y}{\int_{-\infty}^{\infty} f(X,y)\partial y} \tag{4}$$

In most cases, PDF $f(x, y)$ is unknown; however, the relationship can be approximated by the actual observations of x and y. The estimation formula can be given as follows:

$$F(X,Y) = \frac{\frac{1}{m}\sum_{k=1}^{m}\exp\left(\frac{-(X-X_k)^T(X-X_k)}{2\psi^2}\right)\exp\left(\frac{-(Y-Y_k)^2}{2\psi^2}\right)}{(2\pi)^{\frac{u+1}{2}}\psi^{u+1}} \tag{5}$$

where $F(X, Y)$ is the estimating equation of PDF $f(x, y)$ obtained by using the Parzen consistent estimator. X_k and Y_k denote the actual sampling point for the stochastic independent variable x vector and stochastic

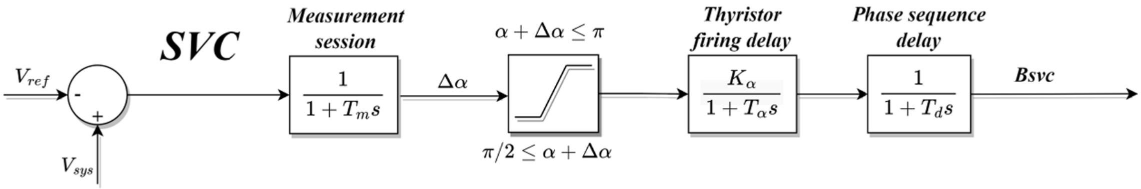

Fig. 2. Configuration of the SVC.

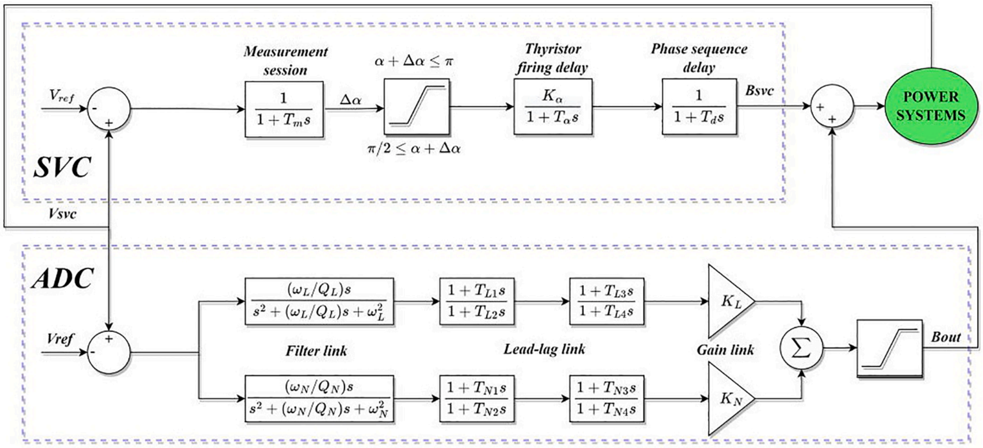
V_{ref}
SVC
Measurement session
$\frac{1}{1+T_m s}$
$\Delta\alpha$
$\alpha+\Delta\alpha\le\pi$
$\pi/2\le\alpha+\Delta\alpha$
Thyristor firing delay
$\frac{K_\alpha}{1+T_\alpha s}$
Phase sequence delay
$\frac{1}{1+T_d s}$
Bsvc
POWER SYSTEMS
Vsvc
ADC
Vref
$\frac{(\omega_L/Q_L)s}{s^2+(\omega_L/Q_L)s+\omega_L^2}$
$\frac{1+T_{L1}s}{1+T_{L2}s}$
$\frac{1+T_{L3}s}{1+T_{L4}s}$
K_L
Filter link
Lead-lag link
Gain link
Σ
Bout
$\frac{(\omega_N/Q_N)s}{s^2+(\omega_N/Q_N)s+\omega_N^2}$
$\frac{1+T_{N1}s}{1+T_{N2}s}$
$\frac{1+T_{N3}s}{1+T_{N4}s}$
K_N

Fig. 3. Structure of the novel static var compensator additional damping controller.

scalar dependent variable y. Y is the corresponding target value of the specific measurement value vector X, m is the total number of samples in the network memory, u is the total dimension of the input vector, and ψ is the smoothing factor for the estimation process.

According to equations (3) and (4), mapping between X and Y can be expressed as follows:

$$\widehat{Y}(X) = \frac{\sum_{k=1}^{M} Y_k \exp(\frac{-\ell_k^2}{2\psi^2})}{\sum_{k=1}^{M} \exp(\frac{-\ell_k^2}{2\psi^2})} \tag{6}$$

$$\ell_k^2 = (X - X_k)^T (X - X_k) \tag{7}$$

where ℓ_k denotes the difference between the input vector X and the stored data X_k, called the distance. Apparently, only one parameter ψ needs to be determined in the GRNN model; this parameter determines the learning ability of the model.

The basic structure and fundamentals of the mapping fitting process of the GRNN are demonstrated in Fig. 1. The input and output data of the CARMA system are used as the memory of the GRNN. Subsequently, the results of the independent variables are used as the input, and the predicted output is obtained by the GRNN using equations (5) and (6) based on the memory.

3. Model of SVC-ADC

In this section, specific models of the SVC and ADC are constructed. Subsequently, a novel SVC-ADC is proposed.

3.1. SVC transfer function configuration

As the most common FACTS device, the SVC primarily maintains the busbar voltage. Moreover, an ADC is always added to the SVC to further improve its performance. The configuration of the SVC transfer function model is shown in Fig. 2.

Here, V_{ref} is the reference voltage, V_{sys} is the power system voltage, T_m is the gain constant of the regulator, α is the thyristor-firing angle, and $\Delta\alpha$ is the change in the thyristor-firing angle. K_α and T_α are the gain constants of the thyristor firing delay, and T_d is the gain constant of the phase-sequence delay.

3.2. Design of novel SVC-ADC

Owing to the access of wind power in a power system, which may significantly increase instability, more than one oscillating mode may be generated considering the control theory, for which the traditional single-branch controller cannot meet the requirements. Therefore, a new controller should be designed to suppress the two oscillation modes. The structure of the new controller proposed in this study is shown in Fig. 3.

Each branch of the additional damping controller proposed in this study consists of a filter link, lead-lag link, limiting link, and gain link. As shown in Fig. 3, the transfer function of one branch of the SVC-ADC is given as follows:

$$F(s) = \frac{(\omega_L/Q_L)s}{s^2 + (\omega_L/Q_L)s + \omega_L^2} \frac{1+T_{L1}s}{1+T_{L2}s} \frac{1+T_{L3}s}{1+T_{L4}s} K_L \tag{8}$$

where K_L is the gain link constant, Q_L is the filter link time constant, T_{L1}, T_{L2}, T_{L3}, T_{L4} are the constants of the lead-lag links, and ω_L is the selected oscillation mode frequency of the filtering link. The input of this controller is the deviation between the voltage at the output of the SVC and the reference voltage, and the output is the reactance of the SVC.

4. Problem converted as an MDP and the SAC-based solutions

In this study, an MDP is introduced and used for the controller parameter tuning. The MDP can be described as follows:

$$P(s^{'}, s, a) = P\{s(t) = s^{'} | s(t-1) = s, a(t-1) = a\}, s^{'}, s \in S, a \in A \tag{9}$$

The SAC algorithm, a DRL algorithm, is used to solve the SVC-ADC parameter tuning problem.

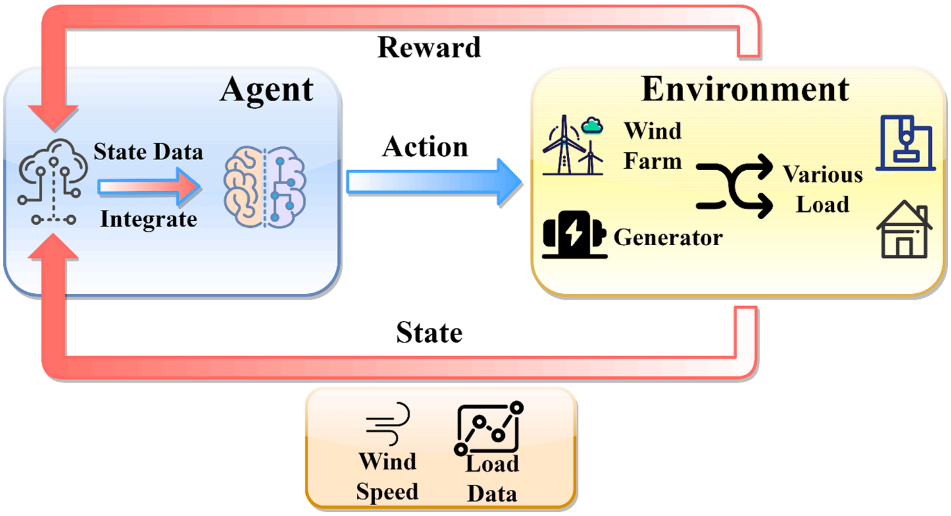

Fig. 4. RL framework of the proposed method to the additional damping controller designing.

4.1. Conversion of the controller parameter tuning to MDP

To solve the controller parameter tuning problem using deep reinforcement learning algorithms, it is first necessary to build an MDP model based on this problem, which is given as follows:

1) State S: State s_t includes wind speeds W_{spi} and the steady-state voltage of the power system node U_{nj}.
2) Action A: Action a_t at time step t is defined as ($K_{L/N}$,$T_{L/N}$, $Q_{L/N}$), where $K_{L/N}$,$T_{L/N}$, and $Q_{L/N}$ denote the parameters of the SVC-ADC.
3) Reward R: Reward r_t at time step t is obtained when the agent performs the action a_t based on state s_t. This variable visually expresses the validity of the action. In this study, the reward is used to evaluate the impact of the current controller parameters on the system in real-time.
4) Transition P: The transition probability $P_{s_{t+1}|s_t,a_t}$ explicitly defines the transition between the states; when s_t is determined, s_{t+1} and a_t are determined along with it.

The entire MDP framework is composed based on these four variables. The action agent a_t is derived from state s_t and policy π. Action a_t is applied to the environment, and the response of the entire system is evaluated using the reward function to derive the performance of the action. The agent maximises the rewards by exploring the state space. The aforementioned RL framework is presented in Fig. 4.

The reward for the RL task should be defined according to practical application scenarios and optimisation objectives. To obtain the performance of the controller parameters, a model performance index is introduced to measure the comprehensive performance.

To indicate the relationship between the input and output, the linearised model of the entire system can be expressed as follows:

$$\begin{cases} \dot{x} = Ax + Bu_{in} \\ y_{out} = Cx \end{cases} \tag{10}$$

where x is the system state vector, A represents the state matrix, B and C are the state input and output matrices, respectively; and u_{in} and y_{out} are the input and output vectors of the power system, respectively.

If the form of x is changed to $x = Pz$, the linearised model of the entire power system can be rewritten as follows:

$$\begin{cases} \dot{z} = P^{-1}APz + P^{-1}Bu_{in} = \Lambda z + Q^T Bu_{in} \\ y_{out} = CPz = C\sum_{j=1}^{n} p_j z_j \end{cases} \tag{11}$$

where P is the right modal matrix of z, Q is the left modal matrix of z, and Λ is the diagonal matrix, which can be expressed as follows:

$$\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_l); \lambda_l = \sigma_l + j\omega_l \tag{12}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_l$ are the eigenvalues of A, and σ_l and ω_l are the real and imaginary parts of λ_l, respectively.

When a unit pulse input is applied to the system, the system output can be expressed as follows:

$$y_{out} = Cp_j z_j^{im} = \sum_{l=1}^{n} R_{ilj}\exp(\lambda_l t) \tag{13}$$

where R_{ilj} is the residue of the LFO mode and can be defined as follows:

$$R_{ilj} = \mu_{il}\nu_{lj} \tag{14}$$

$$U = CP, V = Q^T B \tag{15}$$

Here μ_{il} represents the ith row and lth column element of U, and ν_{lj} represents the lth row and jth column element of V. Based on the aforementioned study, the effect of the kth LFO mode on the output can be summarised as follows:

$$F_k = \sqrt{\varpi(\sigma_k, \omega_k)\sum_{i=1}\sum_{j=1}|R_{ilj}|^2} \tag{16}$$

$$\varpi(\sigma_k, \omega_k) = \frac{\sigma_k^2 + \omega_k^2}{2\sigma_k^3}\left(\exp\left[\frac{2\sigma_k^3\tau_r}{\sigma_k^2 + \omega_k^2}\right] - 1\right) \tag{17}$$

where τ_r is the response time of the system model with a unit pulse as the input excitation signal. To optimise all the target LFO modes, the reward function can be defined as follows:

$$r = \sum_{l=1}\vartheta_l F_l \tag{18}$$

where ϑ_l is the weight coefficient of the lth LFO mode, which is equally distributed to each LFO mode to ensure that each mode is efficiently optimised.

4.2. Controller parameter tuning based on SAC

Generally, the DRL algorithm was introduced to solve the MDP. Among them, the PPO algorithm is the most common DRL algorithm; however, it is an on-policy algorithm, which has the disadvantage of low

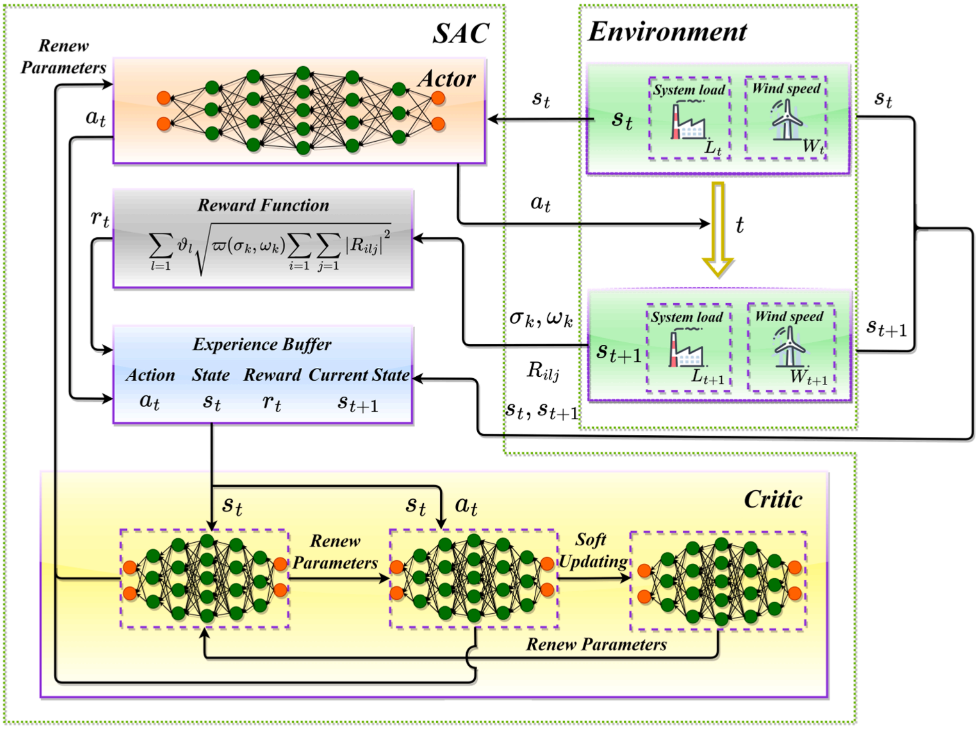

Fig. 5. Structure of the SAC algorithm.

Table 1
Controller parameter tuning based on the SAC algorithm.

```
Input: Wind speeds and node steady-state voltage states of the power system: $(W_{spi},U_{nj})$
Output: SVC-ADC parameter of the controller: $(K_{L/N},T_{L/N},Q_{L/N})$
Initialize
Randomly initialize the parameter critic and actor networks with $\phi,\varphi$, and $\delta$, respectively
Randomly initialize critic network and actor networks with weight $\phi' \leftarrow \phi$ and $\delta' \leftarrow \delta$, respectively
Initialize the size of the experiences buffer
for episode do
  Initialize the noise signal $n$ obeying the standard normal distribution for exploration
  Initialize the state of the text system $s_t$
  Observe the initial state of the test system $s_1$
  for step do
    Select action $a_t = f_\delta(n_t, s_t)$
    Action $a_t$ is executed to the text system
    Obtain reward $r_t$ and new state $s_{t+1}$
     Store the information vector $(s_t,a_t,r_t,s_{t+1})$ in the experiences buffer
     If the experiences buffer is full of information:
     Randomly sample the batch of information from the experiences buffer
     Update Critic networks based on (22)-(24)
      Update Actor networks based on (26)-(30)
      Soft update based on (25)
  end
end
```

sampling efficiency and requires a large amount of sampling data to learn the rule of different types of variates. DDPG is an off-policy algorithm aimed at continuous control, which is more efficient than PPO in sampling, but has the problem of sensitivity to its hyper-parameters and poor convergence. The SAC algorithm is an off-policy algorithm developed for maximum-entropy reinforcement learning. Unlike DDPG, SAC adopts a stochastic policy that enables it to explore more efficiently than deterministic policies during the training process. In contrast, the SAC algorithm distributes almost an equal probability for actions with a close value based on the maximum entropy, which avoids entrapment in the same action leading to sub-optimal conditions via maximising the reward to give up low-reward actions.

4.2.1. Soft value function structure

The SAC algorithm makes decisions based on the maximum entropy stochastic strategy. Entropy can also be part of the reward to appraise the appropriateness of the current action in the strategy. The Bellman equation of maximum entropy reinforcement learning is introduced into the reward to encourage exploration, which is given as follows:

$$H(\pi(s|a)) = -\log\pi(s|a) \tag{19}$$

Meanwhile, the reward function, which contains a description of the SAC performance, is defined as follows:

$$R(s_t, a_t) = \sum_{i=0}^{K-t} \gamma^i [r_{t+i} - \lambda \log\pi(s_t|a_t)] \tag{20}$$

where $\gamma \in [0,1]$ is the discount coefficient. λ regulates the stochastic degree of the policy that determines the relative importance of the entropy term for the reward. To evaluate the effectiveness of the entire optimisation flow, the algorithm introduces a Q-value function as follows:

$$Q^\pi(s_t, a_t) = \mathbb{E}_\pi[\sum_{\mathrm{i=0}}^{\mathrm{K-t}} \gamma^{\mathrm{i}} [\mathrm{r_{t+i}} - \lambda \log\pi(\mathrm{s_t}|\mathrm{a_t})]] \tag{21}$$

4.2.2. Critic model structure

In the RL task, the agent learns by trial-and-error to receive the greatest rewards, with critics and actors introduced in the optimisation process to reach this goal. The critic assumes the function of evaluating the policy, whereas the actor makes improvements to the policy based on the results of the critic's evaluation.

In this study, a novel method based on a deep neural network (DNN) was used in the evaluation process because a DNN has a stronger effect on the fit of the value function. In this part, the output value $Q^\phi(s,a)$ of the DNN is used to approximate the soft Q-value by extensive training. However, DNN training is based on a large amount of data. To meet the

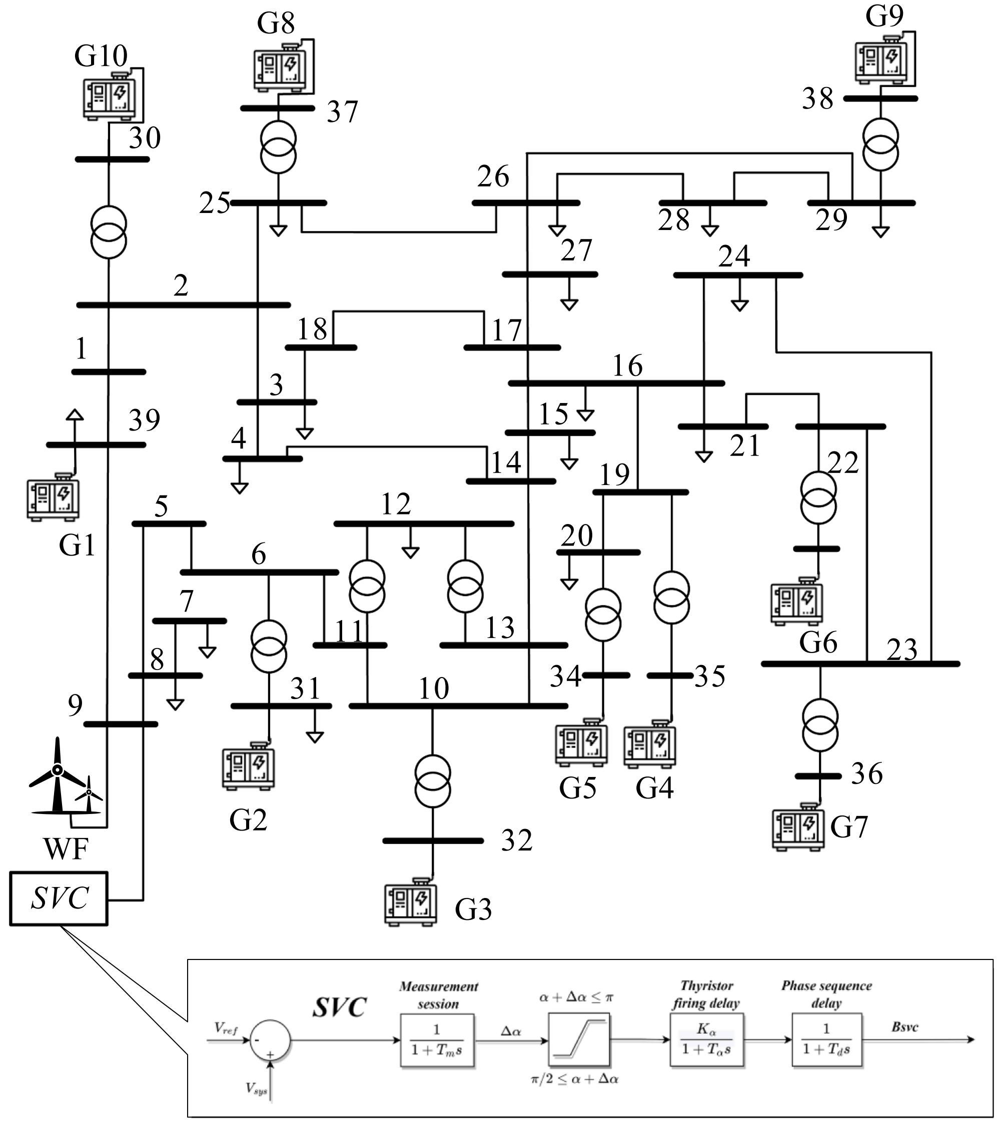

Fig. 6. Topology of the 10-machine 39-bus text system.

Table 2
System state parameters in different test cases.

Case	Parameter	Value	Case	Parameter	Value
Case 1	WF	49.8 MW	Case 2	WF	53.6 MW
	Load 8	1046 MW		Load 8	1098 MW
	Load 25	564 MW		Load 25	678 MW
	Load 39	284 MW		Load 39	314 MW
Case 3	WF	47.8 MW	Case 4	WF	57.8 MW
	Load 8	923 MW		Load 8	891 MW
	Load 25	712 MW		Load 25	495 MW
	Load 39	223 MW		Load 39	244 MW
Case 5	WF	44.3 MW	Case 6	WF	42.1 MW
	Load 8	1106 MW		Load 8	1138 MW
	Load 25	596 MW		Load 25	469 MW
	Load 39	236 MW		Load 39	303 MW
Case 7	WF	54.9 MW	Case 8	WF	41.2 MW
	Load 8	1036 MW		Load 8	1178 MW
	Load 25	612 MW		Load 25	533 MW
	Load 39	261 MW		Load 39	271 MW

training data requirements, a memory buffer has been proposed to store the historical data of the system status, action, and reward in the process of the interaction between the agent and environment. The parameter of DNN ϕ is renewed based on randomly batch-sampled data from the memory buffer.

To describe the fitting effect of the DNN, the mean square error (MSE) between the output value $Q^{\phi}(s_t, a_t)$ and Q target value $Q^{\pi}(s_t, a_t)$ is applied in the performance evaluation. The MSE formulation can be expressed as follows:

$$L(\phi) = E\left[\frac{1}{2}\left(Q^{\phi}(s_t, a_t) - Q^{\pi}(s_t, a_t)\right)^2\right] \tag{22}$$

The parameters of the DNN ϕ can be updated by using gradient descent during the iterative process to ensure that the output MSE is gradually reduced. The updated value is closely related to the loss and can be expressed as follows:

$$\phi_{t+1} \leftarrow \phi_t - \alpha \nabla_{\phi} L(\phi) \tag{23}$$

where α is the learning rate of the critic network and $\nabla_{\phi} L(\phi)$ is the loss function gradient.

To achieve better evaluation results, the SAC algorithm added a new DNN to assess the policy for avoiding the overestimation of values. The DNN can be used as a state estimation function by training, and the evaluation index is MSE, which can be expressed as follows:

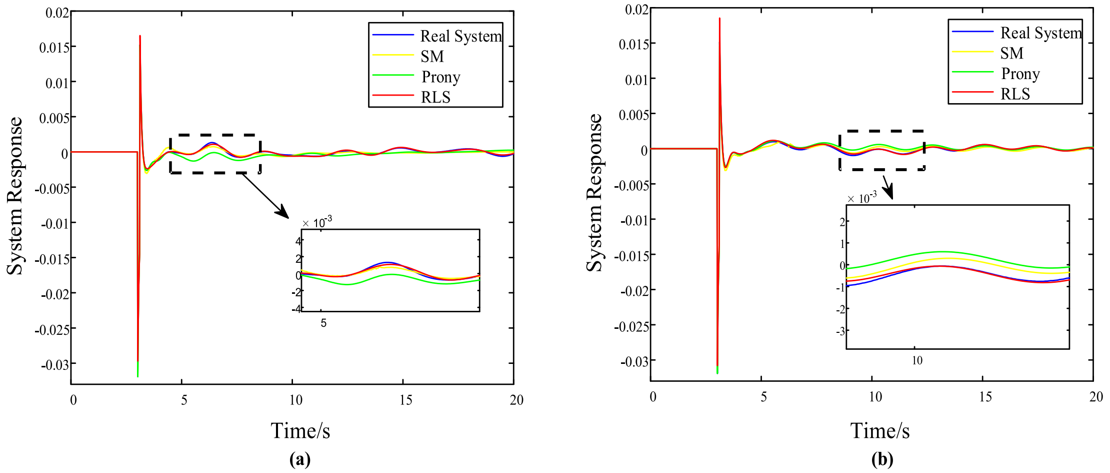

Fig. 7. Comparison of the identification method results: (a) Case 1; (b) Case 2.

Table 3
Identification average MAE of the three methods in multiple scenarios.

Method	Variance	Average MAE
SM	1.48892E-05	1.14371E-04
Prony	6.61147E-06	7.84344E-05
RLS	3.17327E-06	4.29994E-05

$$L(\varphi) = E\left[\frac{1}{2}\left(Q^{\varphi}(s_t, a_t) - Q^{\pi}(s_t, a_t) - \lambda\log\pi(s_t|a_t)\right)^2\right] \tag{24}$$

Similarly, Q^{φ} is the output of the new DNN, and φ is the parameter of the new DNN.

To stabilise the effective update of the critic, SAC introduces a third DNN to estimate the current state value, and the update formula of its parameters can be expressed as follows:

$$\overline{\varphi}_{t+1} \leftarrow \rho\varphi_{t+1} + (1-\rho)\overline{\varphi}_t \tag{25}$$

where $\rho \in (0,1)$ is the soft update factor.

4.2.3. *Actor model structure*

The critic part is responsible for the evaluation, while the actor part is responsible for making sense of the policy. Thus, a DNN is introduced to approximate the policy function $\pi(s|a)$ for each state because the state space is continuous, and KL divergence is used to measure the difference between the policy function $\pi_{\delta}(s|a)$ and the optimal policy $\pi^{*}(s|a)$ in this study:

$$L(\delta) = E[D_{kl}(\pi_{\delta}(s_t|a_t)|\pi^{*}(s_t|a_t))] \tag{26}$$

where δ is the network parameter of the DNN and D_{kl} represents the KL divergence. The action loss $L(\delta)$ represents the difference between the current DNN-fitted policy $\pi_{\delta}(s|a)$ and policy $\pi^{*}(s|a)$ that can produce the desired action. The larger the action loss $L(\delta)$, the further it is from the desired direction. The ultimate optimisation goal is to minimise the action loss $L(\delta)$ by the iteration update policy $\pi_{\delta}(s|a)$. The process of the policy update iteration can be expressed as follows:

$$\pi' = \mathrm{argmin}E[D_{kl}(\pi_{\delta}(s_t|a_t)|\pi^{*}(s_t|a_t))] \tag{27}$$

As the number of iterations continues to increase, the policy will continue to be updated in a better direction. The essence of this policy is to derive the current action based on the current state; thus, it can be considered as a function fitted by the DNN from another perspective. The process of obtaining the action of the fitter based on the state can be expressed as follows:

$$a_t = f_{\delta}(n_t, s_t) \tag{28}$$

Based on the aforementioned study, policy iteration process can be rewritten as follows:

$$\pi' = \mathrm{argmin}E[\lambda\log\pi(s|f_{\delta}(n_t, s_t)) - Q^{\pi}(s|f_{\delta}(n_t, s_t))] \tag{29}$$

Similar to the parameter update of the critic, a gradient descent is used as the parameter update method for the actor DNN:

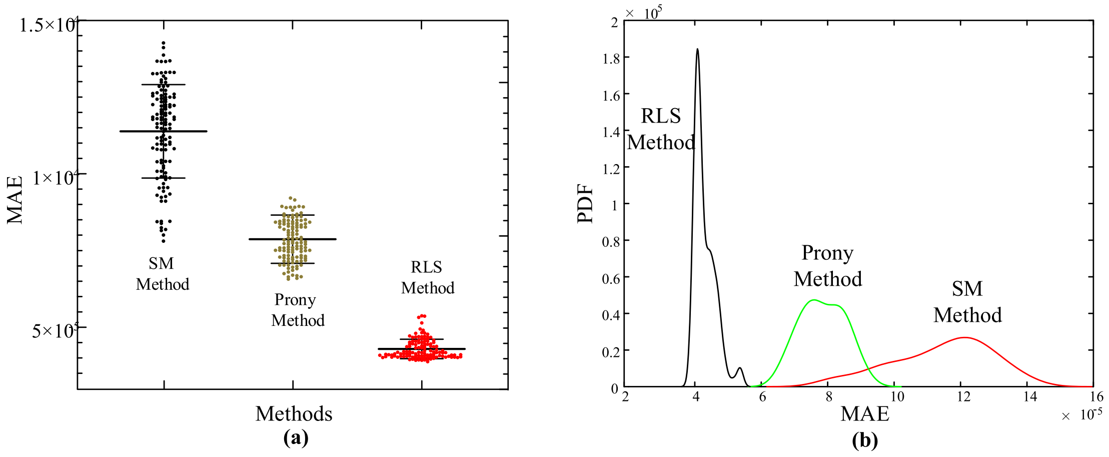

Fig. 8. (a) MAE Distribution of the three methods; (b) PDF curves of the three methods.

Table 4
Identification result of RLS.

Case	Parameter	value	Parameter	value
Case 1	a_0	−0.00235	b_0	1
	a_1	−0.01707	b_1	12.877
	a_2	−0.26336	b_2	156.33
	a_3	−0.75789	b_3	974.39
	a_4	−3.2009	b_4	2576.03
	a_5	−4.5133	b_5	7591.34
	a_6	2.4652	b_6	6293.99
	a_7	−10.369	b_7	3315.37
	a_8	1.6323	b_8	1.224e-8
Case 2	a_0	1.0373	b_0	1
	a_1	36.558	b_1	9.242
	a_2	87.924	b_2	18.071
	a_3	524.207	b_3	91.680
	a_4	203.094	b_4	95.551
	a_5	2088.109	b_5	248.364
	a_6	−1126.824	b_6	154.571
	a_7	433.238	b_7	119.573
	a_8	−8.116	b_8	39.690

$$\delta_{t+1} \leftarrow \delta_t - \beta \nabla_\delta L(\delta) \tag{30}$$

where β is the learning rate of the actor network and $\nabla_\delta L(\delta)$ is the gradient of the actor loss function.

4.2.4. *Soft actor-critic controller parameter tuning*

As described above, the entire algorithmic structure of SAC consists of the actor, critic, and experience buffer parts. The actor part is used to fit the nonlinear relationship between the state and action vectors. The critic part is responsible for estimating the state-action pair, while the experience buffer stores data from the exploration stage as empirical information to prepare for training the DNN. The structure of the entire SAC algorithm is shown in Fig. 5.

The actor part consists of a parameterised DNN, while the critic part consists of three similar DNNs. A matrix of information data serves as the experience buffer. Certain re-parameterisation conditions, including the sampled noise signal obeying the standard normal distribution $N(0,1)$, are used to solve the sampling problem of the model before the approximation policy $\pi(s|a)$ and expand the degree of exploration of the agent in the continuous state space to achieve higher reward possibilities. Considering the current environmental state s_t, the actor obtains and performs an action a_{t+1} according to the policy $\pi(s|a)$. When the reward r_t and s_{t+1} are returned, the information vector (s_t, a_t, r_t, s_{t+1}) is stored in the experience buffer part. Critics are randomly sampled from the experience buffer to train the critic network to break the correlation with different types of input variables. By interacting with the environment to collect information based on experiences by an agent, the key parameters of the DNN are updated. The parameter-tuning process of using the algorithm is presented in Table 1.

5. Case study

In this section, the effects of the strategies presented in this study in

Table 5
Key hyperparameters of the SAC algorithm.

Parameter	Value
Memory size	10,000
Discount factor	0.99
Mini-batch size	128
Training episodes	9000
Step in each episode	10
Soft update coefficient	0.01
Learning rate for critic network	6E-3
Learning rate for actor network	2E-3
Entropy regularization coefficient	0.2

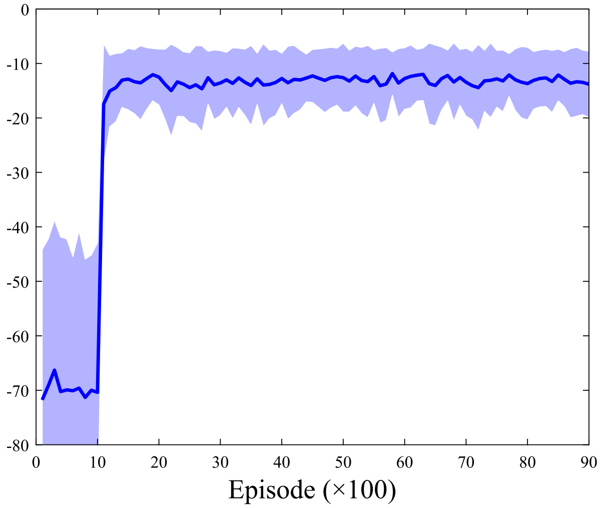

Fig. 10. Cumulative reward mean values and standard deviation for Case 3.

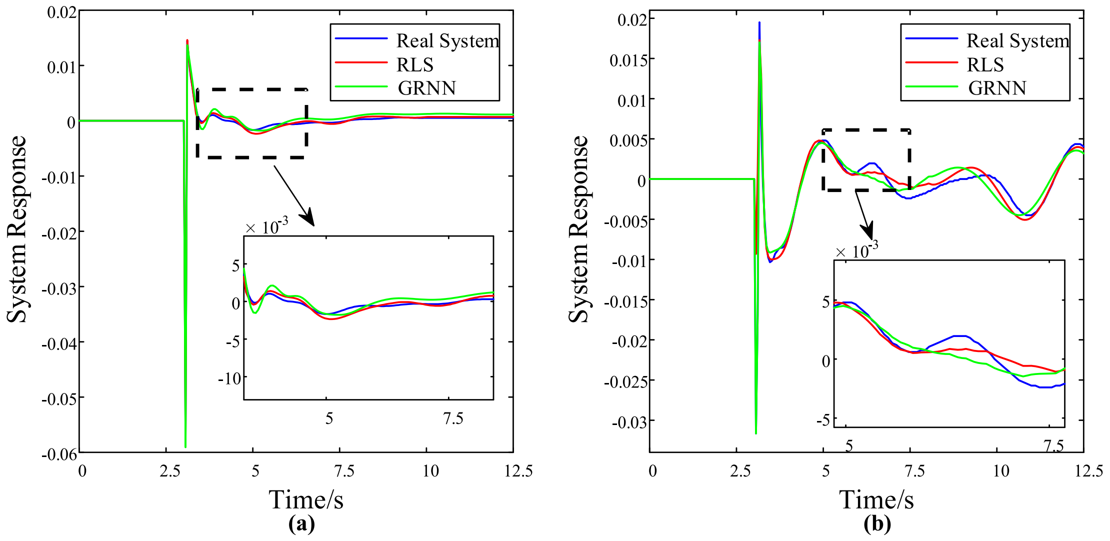

Fig. 9. Comparison of the identification result: (a) Case 3 and (b) Case 4.

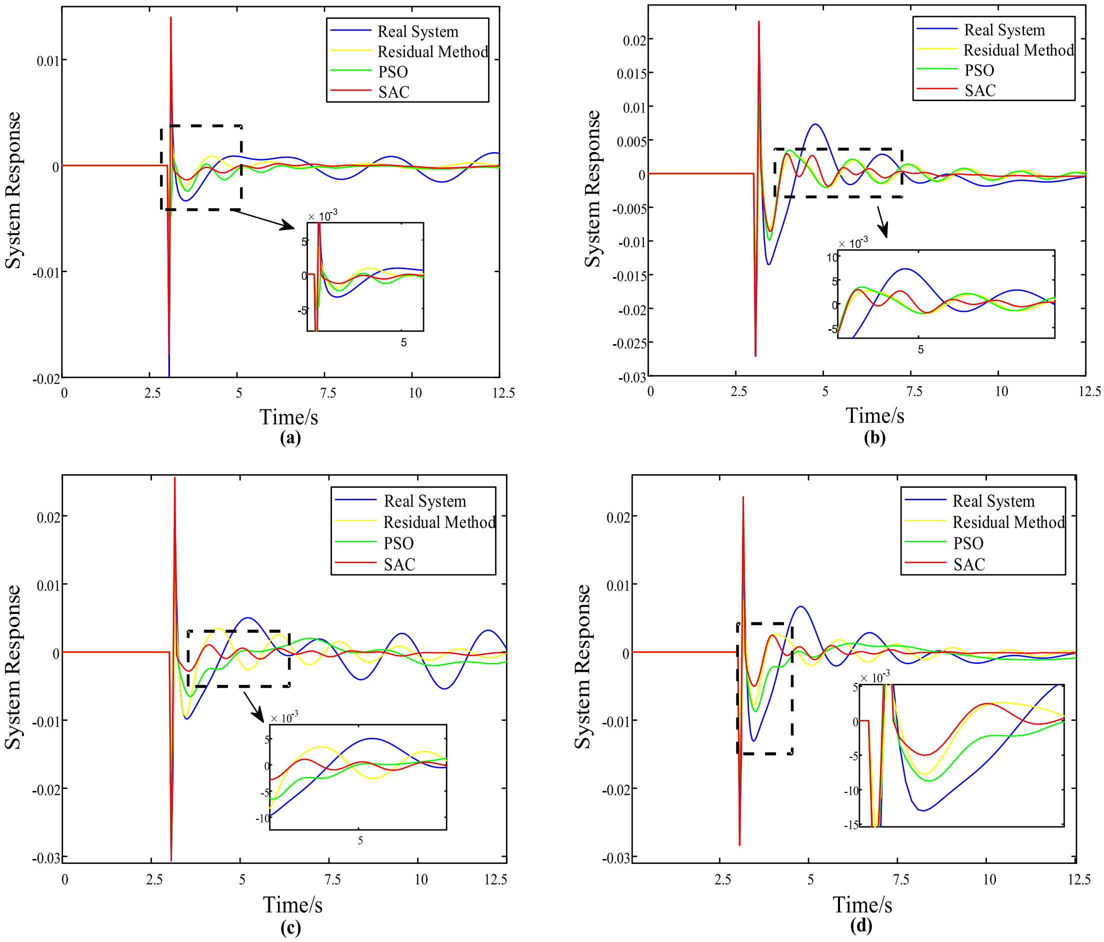

Fig. 11. Time-domain system response of different methods: (a) Case 5, (b) Case 6, (c) Case 7, and (d) Case 8.

various scenarios are discussed. The actual situation regarding the entire power system connected to wind power is described in Subsection 5.1. The complete implementation process of the entire optimisation strategy is detailed in Subsection 5.2. The identification method of the system combining RLS and GRNN and their full descriptions are presented in Subsection 5.3 based on an example. The training process of SAC, key parameters, and comparison of different optimisation methods are elaborated in Subsection 5.4. Finally, Subsections 5.5 and 5.6 demonstrate a comparison of the proposed controller with other controllers and the discrepancy in the effectiveness between the SAC optimisation approach and other optimisation methods.

5.1. Introduction to experimental power systems

A modified IEEE 10-machine 39-bus system was used as a test system to investigate the performance of the proposed strategy. The topology of the test system is shown in Fig. 6. The total amount of generator and load connected to the system was 5620 MW and 6140.3 MW, respectively. Random wind speeds and partially variable loads were introduced to the test system to mimic a real-time system change situation. In this study, to demonstrate the effectiveness of the control strategy, several typical operational conditions were selected for the test cases, as presented in Table 2.

5.2. System transfer function identification

In this study, to characterise the dynamic properties of the test system and perform highly targeted oscillation suppression based on the properties, the first step is to identify the system to obtain a high-precision equivalent model of the system.

Before the identification begins, the test system excitation signal and corresponding response signal data are acquired using simulations. Next, the order of the transfer function model and number of undetermined parameters for the model are determined. The identification of the system transfer function in a single scenario is implemented based on the RLS, which identifies the undetermined parameters based on the excitation signal data, response signal data, and equation (1) to derive the mathematical model of the equivalent transfer function of the system in the scenario.

The advancement of the proposed identification method was verified by comparing it with the SM and Prony methods. The SM method proposed by Steiglitz and McBride in 1965 is an algorithm for estimating the transfer function with the prescribed time-domain impulse response and has been widely applied in filter design and system identification. Prony is a method that identifies signals composed of complex exponentials, essentially based on the singular value decomposition (SVD) of a specially constructed Hankel matrix. Signal processing techniques have been widely used in time series analysis and forecasting. To further demonstrate the identification results of the three methods, the time-domain dynamic responses of the three identification methods under Cases 1 and 2 are shown in Fig. 7. The proposed method clearly has strong robustness and identification accuracy.

Moreover, to further demonstrate the identification results of the three methods, the identification MAE and its distribution of the three methods in different cases are shown in Table 3 and Fig. 8, respectively. The SM method achieves an average MAE of 1.14E-4, while the Prony method achieves an average MAE of 7.84E-5 when performing multiple scenarios for identification. The RLS method used in this study has an

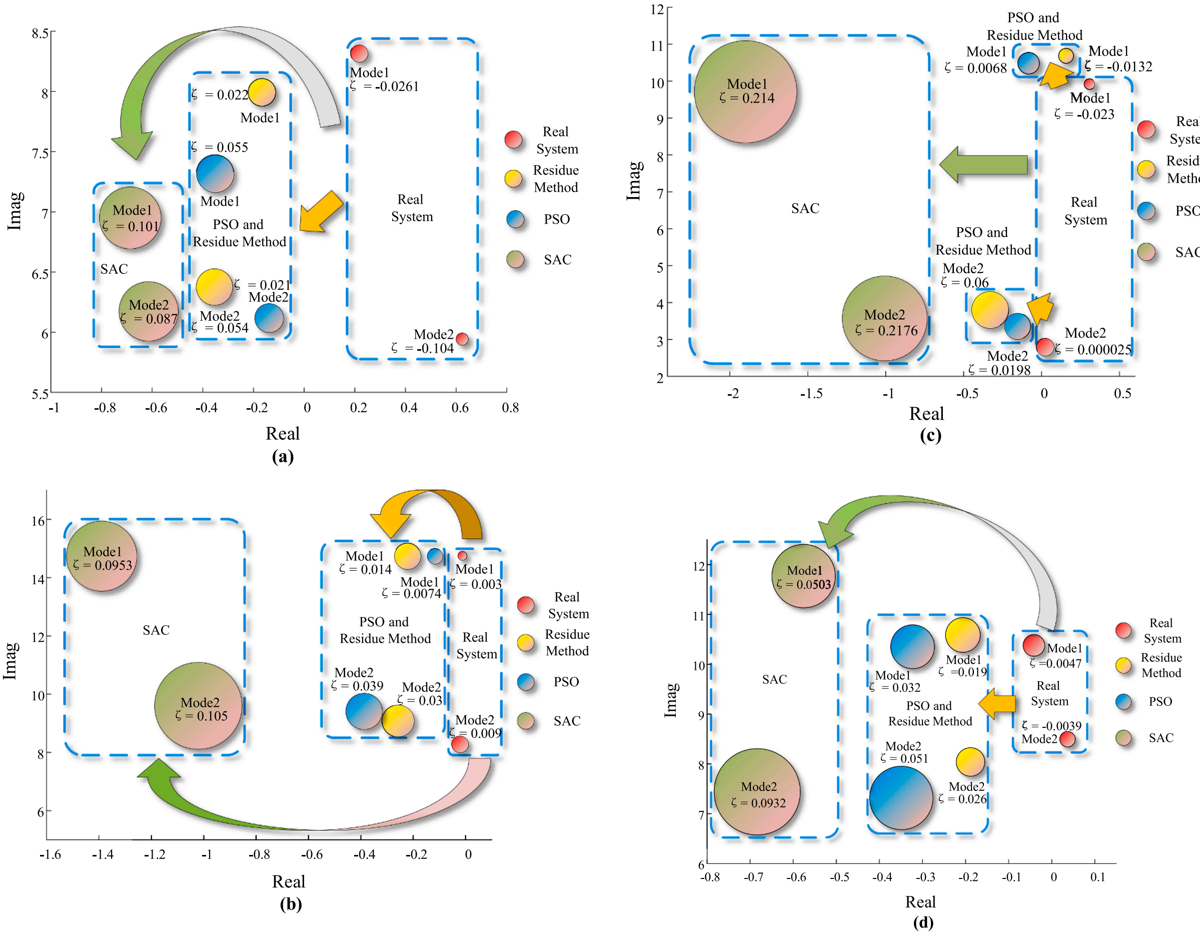

Fig. 12. Distribution of modes in a complex plane: (a) Case 5, (b) Case 6, (c) Case 7, and (d) Case 8.

average MAE of 4.30E-5, also shown in Table 3. The overall distribution of MAE in multiple scenarios also demonstrates that the RLS method presents a strong robustness compared to the other two methods shown in Fig. 8. The results of the RLS identification of the transfer function parameters in specific scenarios are presented in Table 4.

However, the dynamic characteristics of power systems change in real-time, making the system transfer function identification in a single scenario insufficient for meeting the controller needs for adaptive control, for which GRNN will be used to characterise the dynamic system based on the RLS identification results in this study. The network training was based on the system parameter data, with 8500 sets of data used for training and 1000 sets for validating the model accuracy in this study. The time-domain responses of the dynamic system characterised by the RLS and GRNN are compared with the time-domain response of the original system, as shown in Fig. 9.

5.3. Agent training process

The SAC-based agent training process requires determining its agent structure and key hyperparameters before initialisation. The overall structure of the neural network has three hidden layers, each with 256 neurons, whether it is an actor or critic. The agent containing the actor and critic performs the training process by interacting with the environment represented by the dynamic system. Environmental changes are represented as real-time changes in wind speed and load. The change range of the load and wind speed variation is [0.8,1.2] in this study. During the training process, the gain and time parameters of the additional damping controller are selected in a continuous state space in the range of [0,60], [0.02, 30], respectively, to maximise the simulation of the actual controller parameter selection. The key hyperparameters involved are listed in Table 5.

The cumulative reward, calculated using equation (17), is an important indicator of the performance of the training process in the reinforcement learning process. The cumulative reward of the agent during the training process in Case 3 is shown in Fig. 10, which demonstrates that the mean value and standard deviation of the reward values are in a state of fluctuation when the number of episodes is between 0 and 1000. This phenomenon arises because the agent is exploring the continuous state space to gain experience for subsequent agent learning. When the number of iterations is between 1000 and 1322, the mean value curve significantly rises, while the standard deviation region converges, indicating that the agent is currently entering the learning status. When the number of iterations is between 1322 and 9000, the mean value curve and standard deviation domain remain stable, indicating that the agent has found the optimal solution for the current case.

5.4. Analysis of the comparative results

To measure the effectiveness of the proposed strategies, the PSO algorithm and residue method are proposed to compare the effectiveness of the proposed strategy. Multiple cases (Cases 5, 6, 7, and 8) are used as test cases, among which PSO is one of the representatives of the swarm intelligence optimisation algorithm, which uses the idea of mutual cooperation to find the optimal solution to the problem and has high search efficiency and accuracy. The residue method reflects the sensitivity of the controller output signal to the system characteristic root The calculation process does not involve the controller structure and is

Table 6
Comparison of the dominant mode distribution and damping using different methods.

Case	Strategy	Modes	Real	Imag	Damping
Case 5	Real System	Mode 1	0.217	8.31858	−0.026131
		Mode 2	0.623999	5.94272	−0.104428
	Residue Method	Mode 1	−0.178826	7.97642	0.022414
		Mode 2	−0.353087	6.34878	0.055529
	PSO	Mode 1	−0.397183	7.34329	0.054009
		Mode 2	−0.132206	6.11713	0.021607
	SAC	Mode 1	−0.712108	6.99081	**0.101339**
		Mode 2	−0.540381	6.1581	**0.087415**
Case 6	Real System	Mode 1	−0.042334	14.7569	0.002869
		Mode 2	−0.07366	8.14997	0.009038
	Residue Method	Mode 1	−0.205315	14.632	0.014031
		Mode 2	−0.275951	9.06737	0.030419
	PSO	Mode 1	−0.109457	14.7364	0.007427
		Mode 2	−0.354621	9.19108	0.038555
	SAC	Mode 1	−1.43455	14.9779	**0.095342**
		Mode 2	−0.963797	9.09675	**0.10536**
Case 7	Real System	Mode 1	0.2300	9.9224	−0.023171
		Mode 2	−0.000071	2.79703	0.000025
	Residue Method	Mode 1	0.139256	10.5718	−0.013171
		Mode 2	−0.227165	3.71863	0.060975
	PSO	Mode 1	−0.071476	10.4502	0.00684
		Mode 2	−0.05929	2.99799	0.019773
	SAC	Mode 1	−2.10909	9.64594	**0.213604**
		Mode 2	−0.713727	3.20162	**0.217586**
Case 8	Real System	Mode 1	0.033303	8.49559	−0.00392
		Mode 2	−0.048464	10.3168	0.004698
	Residue Method	Mode 1	−0.203327	7.90806	0.025703
		Mode 2	−0.202968	10.4737	0.019375
	PSO	Mode 1	−0.34914	6.83689	0.051001
		Mode 2	−0.322907	10.068	0.032056
	SAC	Mode 1	−0.69611	7.44028	**0.093153**
		Mode 2	−0.569107	11.2885	**0.050351**

suitable for the LFO damping controller design based on the relationship between the system residue number, controller transfer function, and system characteristic root. The comparison results are shown in Fig. 11.

Fig. 11 demonstrates that the proposed method clearly has a stronger suppression amplitude and speed for the time-domain system response oscillations compared to the PSO and residue methods. The result indicates a more reasonable method for the parameter tuning of the controller proposed in this study, which presents a better optimisation effect.

The variation in the distribution of the dominant modes and their damping in the complex plane are presented in Fig. 12 and Table 6 for four cases (Cases 5, 6, 7, and 8) To characterise the optimisation effectiveness and self-adaptive capability of the proposed strategy for dynamic systems.

In the real system, most modes are located on the right side of the complex plane, and the damping (expressed as the size of a circle) is low; thus, the system generates an intense LFO. As shown in Fig. 12, the PSO and residue methods play a suppression role for both dominant modes; however, their effect is not apparent, and the damping of the dominant modes does not reach the requirement of 5%, insufficient for making the system reach a steady state. The proposed strategy can significantly improve the damping of the dominant mode and adaptively change its parameters according to the system state to achieve a better performance compared to the traditional strategy. Considering the modal distribution shown in Fig. 12 and Table 6, it appears that after the optimisation of the proposed strategy, the damping of the dominant modes is improved to more than 5%, and the system stability requirement is achieved.

The root cause of this phenomenon can be attributed to the limitations of the algorithmic mechanism. The PSO method can better suppress the modes in a single scenario. However, the PSO method lacks a better generalisation capability owing to the limitation of the algorithm mechanism, which cannot be adaptively adjusted. Similarly, the disadvantage of the residue method is that the designed damping controller is not sufficiently robust. When multiple oscillation modes exist in the system or the system operation state changes, such as in the cases proposed in this study, the system linearization result also changes, resulting in the design parameters not achieving the expected effect of suppressing the LFO.

6. Conclusion

In this study, an adaptive control strategy based on a generalized regression neural network and deep reinforcement learning is proposed to achieve an optimal parameter tuning process of the static var compensator additive damping controller. The strategy employs recursive least squares to identify the system transfer function parameters and is combined with a generalized regression neural network to achieve a dynamic system fitting in a continuous state space, based on which the soft actor-critic algorithm is used to train the agent to achieve the adaptive adjustment of the static var compensator additional damping controller parameters with an improved inhibitory effect and robustness. To verify the effect of the strategy, recursive least squares parameter identification, generalized regression neural network system dynamic fitting, and deep reinforcement learning dynamic parameter adjustment of the strategy are executed and compared with traditional methods using different cases. The results demonstrate that the proposed strategy achieves both adaptive parameter tuning and significant robustness considering the dynamic changes in the system. The mean absolute error of the identification process is reduced by 30% compared with the traditional method. The damping of each oscillation mode in the frequency domain is increased to more than 5%. The oscillation amplitude in the time domain is significantly reduced, while the dominant modes are located in the stability region, meeting the system stability requirements.

Declaration of Competing Interest

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

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (52277083).

References

[1] Luo S, Hu W, Liu W, Liu Z, Huang Qi, Chen Z. Flexibility enhancement measures under the COVID-19 pandemic – a preliminary comparative analysis in Denmark, the Netherlands, and Sichuan of China. Energy 2022;239:122166.
[2] Swenson L, Gao L, Hong J, Shen L. An efficacious model for predicting icing-induced energy loss for wind turbines. Appl Energy 2022;305:117809.
[3] Li L, Zhang S. Peer-to-peer multi-energy sharing for home microgrids: An integration of data-driven and model-driven approaches. Int J Electr Power Energy Syst 2021;133:107243.
[4] Keighobadi J, Mohammadian KhalafAnsar H, Naseradinmousavi P. Adaptive neural dynamic surface control for uniform energy exploitation of floating wind turbine. Appl Energy 2022;316:119132.
[5] Xu X, Hu W, Cao Di, Liu W, Huang Qi, Hu Y, et al. Enhanced design of an offgrid PV-battery-methanation hybrid energy system for power/gas supply. Renew Energy 2021;167:440–56.
[6] Xu X, Hu W, Cao Di, Huang Qi, Liu Z, Liu W, et al. Scheduling of wind-battery hybrid system in the electricity market using distributionally robust optimization. Renew Energy 2020;156:47–56.
[7] Li J, Wang Ni, Zhou D, Hu W, Huang Qi, Chen Z, et al. Optimal reactive power dispatch of permanent magnet synchronous generator-based wind farm considering levelised production cost minimisation. Renew Energy 2020;145:1–12.

[8] Luo S, Hu W, Liu W, Xu X, Huang Qi, Chen Z, et al. Transition pathways towards a deep decarbonization energy system—a case study in Sichuan, China. Appl Energy 2021;302:117507.

[9] Khalil AM, Iravani R. Impact of high-depth penetration of wind power on low-frequency oscillatory modes of interconnected power systems. Int J Electr Power Energy Syst 2019;104:827–39.

[10] Hoffmann ABG, Beuter CH, Pessoa ALS, Ferreira LRA, Oleskovicz M. Techniques for the diagnosis of oscillatory transients resulting from capacitor bank switching in medium voltage distribution systems. Int J Electr Power Energy Syst 2021;133:107198.

[11] Chau TK, Yu SS, Fernando TL, Iu H-C, Small M. A novel control strategy of DFIG wind turbines in complex power systems for enhancement of primary frequency response and LFOD. IEEE Trans Power Syst 2018;33(2):1811–23.

[12] Li Z, Mehmood K, Xie K. Magnetically controllable reactor based multi-FACTS coordination control strategy. Int J Electr Power Energy Syst 2021;133:107272.

[13] Naderipour A, Abdul-Malek Z, Heidari Gandoman F, Nowdeh SA, Shiran MA, Hadidian Moghaddam MJ, et al. Optimal designing of static var compensator to improve voltage profile of power system using fuzzy logic control. Energy 2020;192:116665.

[14] Sui X, Tang Y, He H, Wen J. Energy-storage-based low-frequency oscillation damping control using particle swarm optimization and heuristic dynamic programming. IEEE Trans Power Syst 2014;29(5):2539–48.

[15] Zhang J, Chung CY, Han Y. A novel modal decomposition control and its application to PSS design for damping interarea oscillations in power systems. IEEE Trans Power Syst 2012;27(4):2015–25.

[16] Rodrigues F, Molina Y, Silva C, Ñaupari Z. Simultaneous tuning of the AVR and PSS parameters using particle swarm optimization with oscillating exponential decay. Int J Electr Power Energy Syst 2021;133:107215.

[17] Hassan LH, Moghavvemi M, Almurib HAF, Muttaqi KM. A coordinated design of PSSs and UPFC-based stabilizer using genetic algorithm. IEEE Trans Ind Appl 2014;50(5):2957–66.

[18] Krishnan VVG, Srivastava SC, Chakrabarti S. A robust decentralized wide area damping controller for wind generators and FACTS controllers considering load model uncertainties. IEEE Trans Smart Grid 2018;9(1):360–72.

[19] Kumar M, Hote YV. Robust PIDD2 controller design for perturbed load frequency control of an interconnected time-delayed power systems. IEEE Trans Control Syst Technol 2021;29(6):2662–9.

[20] Moradi H, Vossoughi G. Robust control of the variable speed wind turbines in the presence of uncertainties: A comparison between H_∞ and PID controllers. Energy 2015;90:1508–21.

[21] Kuttomparambil Abdulkhader H, Jacob J, Mathew AT. Robust type-2 fuzzy fractional order PID controller for dynamic stability enhancement of power system having RES based microgrid penetration. Int J Electr Power Energy Syst 2019;110:357–71.

[22] Ogundairo O, Kamalasadan S, Nair AR, Smith M. Oscillation damping of integrated transmission and distribution power grid with renewables based on novel measurement-based optimal controller. IEEE Trans Ind Appl 2022;58(3):4181–91.

[23] Ghosh S, El Moursi MS, El-Saadany EF, Hosani KA. Online coherency based adaptive wide area damping controller for transient stability enhancement. IEEE Trans Power Syst 2020;35(4):3100–13.

[24] Seifi K, Moallem M. An adaptive PR controller for synchronizing grid-connected inverters. IEEE Trans Ind Electron 2019;66(3):2034–43.

[25] Chang L, et al. Design of adaptive H_∞ controller for power system based on prescribed performance. ISA Trans 2020;100:244–50.

[26] Yin Y, Liu J, Luo W, Wu L, Vazquez S, Leon JI, et al. Adaptive control for three-phase power converters with disturbance rejection performance. IEEE Trans Syst Man Cybernet: Syst 2021;51(2):674–85.

[27] Zamzoum O, Derouich A, Motahhir S, El Mourabit Y, El Ghzizal A. Performance analysis of a robust adaptive fuzzy logic controller for wind turbine power limitation. J Clean Prod 2020;265:121659.

[28] Zhang G, Hu W, Zhao J, Cao Di, Chen Z, Blaabjerg F. A novel deep reinforcement learning enabled multi-band pss for multi-mode oscillation control. IEEE Trans Power Syst 2021;36(4):3794–7.

[29] Zhang G, Hu W, Cao Di, Huang Qi, Yi J, Chen Z, et al. Deep reinforcement learning-based approach for proportional resonance power system stabilizer to prevent ultra-low-frequency oscillations. IEEE Trans Smart Grid 2020;11(6):5260–72.

[30] Hashmy Y, Yu Z, Shi Di, Weng Y. Wide-area measurement system-based low frequency oscillation damping control through reinforcement learning. IEEE Trans Smart Grid 2020;11(6):5072–83.

[31] Zhang G, Hu W, Cao Di, Huang Qi, Chen Z, Blaabjerg F. A novel deep reinforcement learning enabled sparsity promoting adaptive control method to improve the stability of power systems with wind energy penetration. Renew Energy 2021;178:363–76.

[32] Zhang X, Lu C, Liu S, Wang X. A review on wide-area damping control to restrain inter-area low frequency oscillation for large-scale power systems with increasing renewable generation. Renew Sustain Energy Rev 2016;57:45–58.

[33] Ma Yu, Sclavounos PD, Cross-Whiter J, Arora D. Wave forecast and its application to the optimal control of offshore floating wind turbine for load mitigation. Renew Energy 2018;128:163–76.