Renewable Energy 215 (2023) 118893

Contents lists available at ScienceDirect

Renewable Energy

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

Data-driven torque and pitch control of wind turbines via reinforcement learning☆

Jingjie Xie, Hongyang Dong, Xiaowei Zhao *

Intelligent Control & Smart Energy (ICSE) Research Group, School of Engineering, University of Warwick, Coventry, CV4 7AL, UK

ARTICLE INFO

Keywords:
Wind turbine control
Reinforcement learning
Deep neural network
Model predictive control

ABSTRACT

This paper addresses the torque and pitch control problems of wind turbines. The main contribution of this work is the development of an innovative reinforcement learning (RL)-based control method targeting wind turbine applications. Our RL-based control framework synergistically combines the advantages of deep neural networks (DNNs) and model predictive control (MPC) technologies. The proposed control strategy is data-driven, adapting to real-time changes in system dynamics and enhancing control performance and robustness. Additionally, the incorporation of an MPC structure within our design improves learning efficiency and reduces the high computational complexity typically found in deep RL algorithms. Specifically, a DNN is designed to approximate the wind turbine dynamics based on a continuously updated dataset composed of state and action measurements taken at specified sampling intervals. The real-time control policy is generated by integrating the online trained DNN into an MPC architecture. The proposed method iteratively updates the DNN and control policy in real-time to optimize performance. As a primary result of this work, the proposed method demonstrates superior robustness and control performance compared to commonly-employed MPC and other baseline wind turbine controllers in the presence of uncertainties and unexpected actuator faults. This effectiveness is showcased through simulations with a high-fidelity wind turbine simulator.

1. Introduction

The worldwide wind energy capacity has been growing drastically in recent years [1,2]. The control system is the core of wind turbine operations. In general wind turbine control tasks, there are typically four working regions based on the ranges of inflow wind speeds: two shutdown regions and two operating regions [3]. No power generation process is involved in the two shutdown regions, i.e., Region I: below cut-in wind speed and Region IV: above cut-out wind speed. In Region II (above cut-in speed and below rated speed), torque control is utilized to extract as much power as possible. In Region III (above-rated speed and below cut-out speed), pitch control is employed to maintain the power generation at the rated level. Therefore, regions II & III are essential in the maximum power point tracking (MPPT) control tasks of wind turbines.

In practical applications, proportional–integral–derivative (PID) controllers are widely adopted in wind turbine control tasks. For example, Ref. [4] achieved MPPT for wind energy conversion systems via a PID controller. A hybrid PID-fuzzy pitch controller was presented in Ref. [5] for wind turbines to optimize energy extraction. PID controllers are easy to implement given their simple structures, but their control performances are limited [6]. Other advanced methods such as adaptive fuzzy control [7], sliding-model control [8], H-∞ control [9] were investigated to improve wind turbine control performance. These methods have the potential to achieve better performance than the conventional PID controllers. But they rely on analytical wind turbine models. They are sensitive to unmodeled dynamics and lack robustness to real-time model changes (such as the changes induced by faults). Another challenge for these wind turbine control methods is the limited constraints handling abilities w.r.t state & input constraints, such as the generator speed and torque restrictions.

As a cutting-edge optimal control method, the model predictive control (MPC) technique has drawn wide attention and has been applied to wind turbine control [10,11]. The fundamental advantages of MPC strategies are their ability to handle constraints and achieve optimal control under multiple objectives [12–14]. For example, Ref. [15] proposed a linearized MPC algorithm to achieve multi-objective wind turbine control. A further study using this method was presented in Ref. [16]. An economic nonlinear MPC scheme for wind turbine control was designed in Ref. [17] to improve the control performance of traditional nonlinear MPC. Other examples of MPC methods for wind

☆ This work has received funding from the UK Engineering and Physical Sciences Research Council (grant number: EP/T021713/1).
* Corresponding author.
E-mail addresses: jingjie.xie@warwick.ac.uk (J. Xie), hongyang.dong@warwick.ac.uk (H. Dong), xiaowei.zhao@warwick.ac.uk (X. Zhao).

https://doi.org/10.1016/j.renene.2023.06.014
Received 25 October 2022; Received in revised form 14 May 2023; Accepted 6 June 2023
Available online 8 June 2023

Table 1
Summary of existing studies about RL-based wind turbine control.

Ref.	Method	Property	Goals	Simulator	Advantages
[21]	Q-learning	Model-free	Torque control	WECS	Online learning. Adaptability to changes. Low memory requirements.
[22]	Q-learning	Model-free	Torque control	WECS	Working in real-time. with varying wind speed Small computational time.
[23]	Artificial neural network and Q-learning	Model-free	Torque control	WEC	Improved learning efficiency. Online learning can be triggered.
[24]	ADP	Model-based	Pitch control	WT Dynamics	Enhanced performance. Reduced fluctuations.
[25]	RL-enhanced PID	Model-free	Pitch control	WT model	Enhanced learning convergence. Reduced power oscillations.
[26]	Actor-Critic RL	Model-free	Torque & pitch control	OpenFAST	Whole system optimization. Improved power efficiency.
[27]	RL-based adaptive controller using Monte-Carlo method	Model-free	Torque & pitch control	FAST	Multi-objective: maximizing power and minimizing unwanted forces.
[28]	Markov chain Monte Carlo-based RL	Model-free	Torque & pitch control	WECS	Addressing uncertainties. Improving energy efficiency.

turbines to operate in regions II & III include Refs. [18,19]. However, analytical wind turbine models are an inevitable prerequisite to implementing these model-based MPC controllers, rendering them sensitive to modeling errors and unmodeled dynamics. In addition, faults in real-time operations may significantly degrade the performance of MPC methods.

Reinforcement learning (RL) is a promising new technique that has the ability to overcome the above-mentioned challenges in wind turbine control. It aims to learn an optimal policy by maximizing a long-time reward through interactions between the agent and environment [20]. Attempts have been made to apply RL on wind turbine control [21–28]. In Ref. [21], an RL-based algorithm was proposed for variable-speed wind energy conversion systems, achieving higher efficiency than conventional MPPT methods. Ref. [22] introduced a Q-Learning-enhanced MPPT method, while Ref. [23] developed an adaptive-network-based RL approach for wind turbines. To particularly enhance the performance of pitch control, Ref. [24] proposed an adaptive dynamic programming (ADP) method while Ref. [25] developed an RL-enhanced PID controller. Furthermore, Ref. [26] introduced an actor-critic RL wind turbine control method to globally optimize the entire system performance. Additionally, RL-based adaptive controllers utilizing the Monte Carlo (MC) method [27] and a Markov chain-based MC method [28] had also been developed for both wind turbine torque and pitch control. Based on these advances, Table 1 summarizes the recent RL-based wind turbine control methods. This table emphasizes the properties, differences, and advantages of various methods to offer a clearer overview of existing RL-based wind turbine control technologies.

Mainstream RL methods usually can be categorized into two typical structures, as shown in Fig. 1 (in which the plants are illustrated by wind turbines). The first structure (Fig. 1(a)) does not estimate the system model. Instead, it directly evaluates the long-term reward function and the optimal control policy based on the estimated reward function. In contrast, the other structure (Fig. 1(b)) employs additional deep neural network (DNN) modules to capture the key system information, which forms a surrogate model for the system. Then the reward function and control policy are learnt with the help of the surrogate model. Based on whether a surrogate model is employed or not, the former is commonly referred to as model-free RL while the latter is usually called "model-based" RL [29,30]. But it should be emphasized that both model-free and "model-based" RL are data-driven and independent of analytical system models. Here the term "model-based" is just to indicate that data-driven surrogate models via DNN are employed. To the authors' best knowledge, all existing RL methods for wind turbine control are model-free RL. On the one hand, they inherit the advantages of RL, such as strong robustness

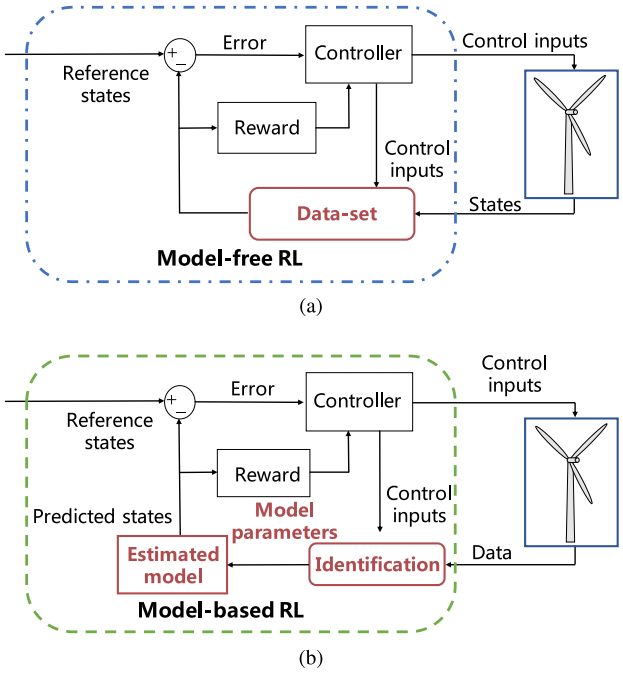

Fig. 1. Two structures of RL methods applied to wind turbine control. (a) Model-free RL. (b) "Model-based" RL.

to environments, modeling errors, and uncertainties, showing superior performance to conventional wind turbine control methods. On the other hand, they suffer from the limitations of model-free RL, including the requirement of a huge amount of data in training (which can lead to incredible computational loads), relatively low learning efficiency, and high sample complexity [31]. These issues can potentially degrade their applicability in practical wind turbine operations.

This paper designs a novel "model-based" RL-based wind turbine control method subject to uncertainties, unmodeled nonlinear dynamics, and state & input constraints to address the challenges mentioned above in existing wind turbine control approaches. The proposed method has a hybrid mechanism that combines the advantages of deep neural networks (DNN) and MPC technologies. Specifically, we design a DNN structure to function as an information processor and learn the key information of the wind turbine control system. Its training is carried out by a real-time updated dataset consisting of state & action measurements in specified sampling intervals. Based on the

DNN structure, an MPC architecture is employed to learn the optimal control policy with respect to user-defined performance metrics. It is noteworthy that the DNN and the optimal control policy are updated by turns online, adapting to the real-time changes of system dynamics and enhancing the control performance & robustness. The FAST (Fatigue, Aerodynamics, Structures, and Turbulence) simulator designed by NREL (National Renewable Energy Laboratory) is utilized to validate the effectiveness of our control method. The main contributions of this paper, in comparison with relevant studies, are in order:

(1) Compared with conventional wind turbine control approaches, the proposed RL method is data-driven. It is able to learn optimal control policies without relying on analytical wind turbine models. The system information is captured and updated in real-time via DNN. This feature enables our method to have strong data mining ability and adaptability.

(2) Existing RL-based wind turbine control methods are commonly based on model-free RL. Most of those methods are based on the famous deep deterministic policy gradient (DDPG) [32] algorithm and its derivatives. Those designs can effectively handle complex tasks and have a simplified structure compared to "model-free" RL. However, they commonly suffer from high computational complexity and low learning efficiency. Moreover, because they do not embed surrogate models into the learning algorithm, their adaptability and robustness to real-time model & environment changes (such as sudden actuator faults) are still limited. A novel wind turbine control approach based on "model-based" RL is developed in this paper to address these limitations. To the best of the authors' knowledge, this is for the first time such technology has been applied to wind turbine control tasks. As we mentioned before, our method does not require the analytical dynamics of turbines. Here the term "model-based" indicates that we utilize real-time data to construct and update a DNN-based surrogate model. It renders the learning process for the reward function and optimal control policy more stable and effective.

(3) The proposed DNN structure, serves as an information processor, is updated by a real-time dataset consisting of current state & action measurements. This structure allows our algorithm to learn the key information of the wind turbine control system, especially capture the potential online changes of wind turbine models under the unexpected fault occurs. Such a design is beneficial in recovering the control performance when unknown uncertainties and/or unanticipated faults happen online, eventually improving the reliability and applicability of the whole wind turbine control system.

The rest of this paper is organized as follows. Section 2 presents the preliminaries and problem formulation of wind turbine control to lay a foundation for the developed strategy. Then, the proposed RL-based wind turbine control method is introduced in Section 3. Numerical simulations and case studies under different controllers considering faults and uncertainties are illustrated in Section 4. Finally, a conclusion is drawn in Section 5.

2. Preliminaries and problem formulation

The preliminaries related to wind turbine dynamics and the formulation of wind turbine control problems are introduced in this section.

2.1. Preliminaries

The commonly used nonlinear wind turbine dynamics for torque control and pitch control can be expressed as

$$\dot{x} = f(x,u) \tag{1}$$

where $x = \begin{bmatrix} \theta & w_r & w_g & T_g & \beta \end{bmatrix}^T$ denotes the integrated wind turbine state, θ is the torsion angle, w_r is the rotor speed, w_g is the generator speed, T_g is the generator torque, β is the pitch angle. In addition, $u = \begin{bmatrix} T_{g,ref} & \beta_{ref} \end{bmatrix}^T$ denotes the control input, $T_{g,ref}$ and β_{ref} are the generator torque and pitch angle control signals. Considering the wind turbine rotor, drive train, and generator dynamics, the specific form of Eq. (1) can be formulated as [33,34]

$$\begin{bmatrix} \dot{\theta} \\ \dot{w}_r \\ \dot{w}_g \\ \dot{T}_g \\ \dot{\beta} \end{bmatrix} = \begin{bmatrix} w_r - \frac{1}{N_g} w_g \\ -\frac{K_s}{J_r}\theta - \frac{D_s}{J_r} w_r + \frac{D_s}{J_r N_g} w_g + \frac{T_r}{J_r} \\ \frac{K_s}{J_g N_g}\theta + \frac{D_s}{J_g N_g} w_r - \frac{D_s}{J_g N_g^2} w_g - \frac{1}{J_g} T_{g,ref} \\ -\frac{1}{\tau_g} T_g + \frac{1}{\tau_g} T_{g,ref} \\ -\frac{1}{\tau_\beta}\beta + \frac{1}{\tau_\beta}\beta_{ref} \end{bmatrix} \tag{2}$$

where J_r is the rotor's moment of inertia, N_g is the gear ratio, J_g is the generator's moment of inertia, K_s is the spring constant, D_s is the torsion damping coefficient, T_r is the rotor torque, satisfying $T_r = \frac{1}{2}\rho\pi R^2 v_w^3 C_p(\lambda,\beta)/w_r$, where ρ is the air density, R is the wind turbine radius, v_w is the wind speed, C_p is the power coefficient, which is a nonlinear function of tip speed ratio λ and the pitch angle β. τ_g and τ_β are time constants.

The mechanical power production P captured by a wind turbine follows

$$P = \eta w_g T_g \tag{3}$$

where η is the power transfer efficiency.

In conventional model-based methods, the wind turbine model in Eqs. (1)–(3) is directly utilized to design controllers. However, as mentioned in the introduction, those designs are sensitive to modeling errors and lack robustness to real-time dynamics changes. In this work, we aim to develop a data-driven RL strategy that does not require the analytical model in (1)–(3), mitigating the limitations of conventional approaches.

2.2. Problem formulation

The wind turbine control problems considered in this paper include power generation maximization (or maximum power point tracking) in Region II and power maintenance in Region III. In Region II, we aim to extract as much power as possible by adjusting the generator torque. This task can be considered as an optimal control problem with respect to a long-term objective function as in the following equation. (Note that the control schemes in this work are designed in the discrete-time domain for real-time implementation.)

$$\min_u J_{T_g}(t) = \sum_t^\infty [(T_g - T_g^*)^T Q_T (T_g - T_g^*) + (w_g - w_g^*)^T Q_{w_g}(w_g - w_g^*) + \Delta u_t^T R \Delta u_t] \tag{4}$$

subject to

$$w_r \le 1.1 w_r^{rated} \tag{5}$$

$$w_g \le 1.1 w_g^{rated} \tag{6}$$

$$T_g^{min} \le T_{g,ref} \le T_g^{max} \tag{7}$$

$$\beta^{min} \le \beta_{ref} \le \beta^{max} \tag{8}$$

$$\Delta T_g^{min} \le \Delta T_{g,ref} \le \Delta T_g^{max} \tag{9}$$

$$\Delta\beta^{min} \le \Delta\beta_{ref} \le \Delta\beta^{max} \tag{10}$$

where T_g^* and w_g^* are the generator torque and generator speed references, $\Delta u_t = u_t - u_{t-1}$ denote the change of control input, Q_T, Q_{w_g} and R are weight matrices. When the generator speed w_g and torque T_g are controlled to track the reference values, the mechanical power $P = \eta w_g T_g$ can be maintained at the optimal value to achieve the maximum power tracking. Eqs. (5) and (6) represent the constraints

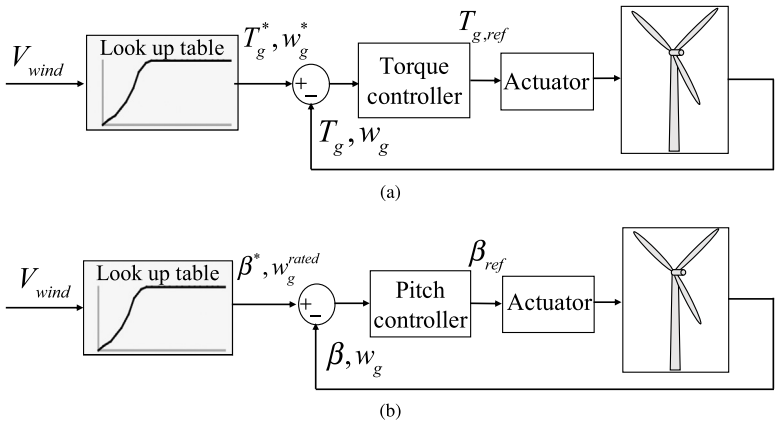

Fig. 2. The block diagram of the wind turbine control system. (a) Torque control. (b) Pitch control.

on system states, w_r^{rated} and w_g^{rated} denote the rated rotor speed and generator speed, respectively. They are restricted to certain ranges, ensuring the safety of the wind turbine. In addition, the limits of control inputs are presented in Eqs. (7) and (8). T_g^{min}, T_g^{max}, β^{min}, and β^{max} are the lower & upper bounds of the generator torque and pitch angle, respectively. Note that T_g^{max} is usually set to be 10% above the rated torque value ($T_g^{max} = 1.1T_g^{rated}$) to prevent overloading. In addition, the constraints on actuators' changing rates are shown in Eqs. (9) and (10). Specifically, ΔT_g^{min}, ΔT_g^{max}, $\Delta\beta^{min}$, and $\Delta\beta^{max}$ are the lower & upper bounds of the changing rates of torque and pitch angle, respectively.

For the pitch control task in Region III, the objective is to keep the output power and generator speed at their rated values. The performance metric is defined as

$$\min_u J_\beta(t) = \sum_t^\infty [\left(w_g - w_g^{rated}\right)^T Q'_{w_g} \left(w_g - w_g^{rated}\right) + \left(\beta - \beta^*\right)^T Q_\beta \left(\beta - \beta^*\right) + \Delta u_t^T R \Delta u_t] \tag{11}$$

where w_g^{rated} is the rated values of generator speed, β^* is the pitch angle reference, Q'_{w_g} and Q_β are weight matrices. The block diagrams of torque and pitch control systems are demonstrated in Figs. 2(a) and 2(b), respectively. The reference signals T_g^*, w_g^*, w_g^{rated}, and β^* are derived according to the wind speed by the lookup table scheme [35].

3. Model-based reinforcement learning method for wind turbine control

The proposed reinforcement learning (RL) technique for wind turbine control is introduced in detail in this section.

3.1. Model-based reinforcement learning method

The RL built upon the Markov decision process [21] aims to learn an optimal policy by optimizing designed long-time reward for an agent through interactions with the environment. Specifically, at current time t, whose state is denoted by $s_t \in S$. It takes the action $a_t \in A$, and then receives the reward $r_t \in R$. Here S, A, and R denote the state, action, and reward spaces, respectively. Based on the transition function $f : S \times A \to S$, the next state $s_{t+1} \in S$ at time $t+1$ is given by $s_{t+1} = f\left(s_t, a_t\right)$. The goal of this RL agent is to learn an effective policy $\pi(s) : S \to A$ that optimizes the long-time reward, formulated as $\sum_{t=t'}^\infty \gamma^{t-t'} r_t\left(s_t, a_t\right)$, where $\gamma \in (0, 1]$ is a discount factor, and $r_t\left(s_t, a_t\right)$ is the reward at s_t after taking a_t [36,37].

Particularly, the mainstream RL methods commonly employ a critic structure to evaluate the long-term objective function and an actor structure to learn the optimal control policy given the result from the critic — which can be a quite complex process and require a large data set for training. Distinct from the mainstream actor-critic structure, we utilize a deep neural network (DNN) structure as an information processor to estimate the system dynamics, which is updated in real-time to enhance the control robustness. Built upon this DNN, we fit our actor to an MPC framework to learn the optimal control policy. Such a design is much more flexible for online applications and renders our RL-based wind turbine controller quickly adapt to unexpected model changes, such as the ones induced by sudden actuator faults. The details of this RL technique are given in the following.

3.1.1. System dynamics approximation

A DNN structure is employed in the designed RL method to approximate system dynamics. A DNN typically consists of input, hidden, and output layers, which are connected by neurons and weights. We denote the approximated system dynamics as $\hat{f}_\theta\left(s_t, a_t\right)$. For the training at a time step t, the state s_t and the action a_t are the inputs of DNN, while the next state s_{t+1} is taken as the output. We denote the measurement data set as D with a size of N_D. Therefore, D should contain a state set $\left(s_0, s_1, \ldots, s_{N_D-1}\right)$, an action set $\left(a_0, a_1, \ldots, a_{N_{n-1}}\right)$, and a successor state set $\left(s_1, s_2, \ldots, s_{N_D}\right)$. Note that if the difference between a state s_t and its successor s_{t+1} is near zero, the change between s_t and s_{t+1} (denoted as Δs_t) can be employed to be the output of $\hat{f}_\theta\left(s_t, a_t\right)$. In this case, one has $\Delta s_t = s_{t+1} - s_t = \hat{f}_\theta\left(s_t, a_t\right)$.

In addition, pre-processing is carried out for the data set D. Specifically, the z-score method is employed for normalization purposes. Moreover, to improve the learning robustness, noises obeying zero-mean Gaussian distribution are added to the data set. It is noteworthy that the data in D is updated in real-time to capture the potential online changes of system dynamics. Based on D, DNN is trained via a supervised learning manner, aiming to minimize the following loss function

$$\xi(\theta) = \sum_{t=0}^{N_D} \frac{1}{2} \left\| s_{t+1} - \hat{f}_\theta\left(s_t, a_t\right) \right\|^2 \tag{12}$$

If the changes of states are used as outputs, then the loss function is

$$\xi(\theta) = \sum_{t=0}^{N_D} \frac{1}{2} \left\| \left(s_{t+1} - s_t\right) - \hat{f}_\theta\left(s_t, a_t\right) \right\|^2 \tag{13}$$

Remark 1. It should be emphasized again that though our method can be categorized as a "model-based" RL approach, it does not require any information from the analytical wind turbine model. Instead, it employs the DNN structure to build a surrogate model and updates it online with real-time measurements.

3.1.2. Control policy optimization

With $\hat{f}_\theta\left(s_t, a_t\right)$, we can recursively predict the future states based on the candidate actions and choose the optimal control policy $\pi\left(s_t \to a_t\right)$

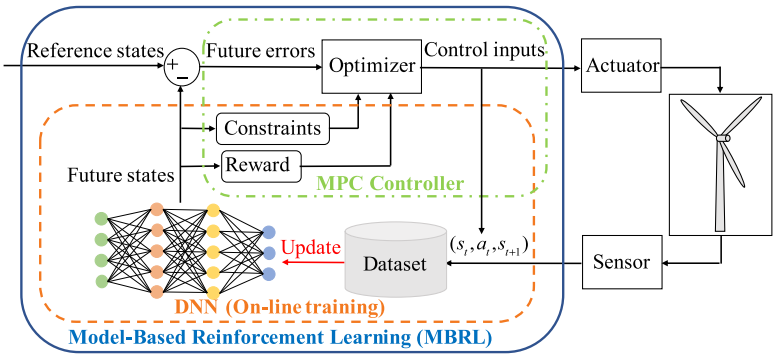

Fig. 3. The framework of the proposed RL-based wind turbine control technique.

with the highest reward. The MPC structure is embedded into our proposed RL control method to solve the optimal closed-loop control problem described in Eqs. (4)–(11). Compared to other controllers, the MPC method possesses the fundamental advantages of realizing multi-objective optimal control and handling the state and control constraints [38]. Specifically, in each receding-horizon (we use T_s to denote the prediction horizon), the DNN-approximated system dynamics are employed to predict the future states. Then, we aim to find an optimal control solution $\boldsymbol{a}^* = \left(a_{t'}, a_{t'+1}, \ldots, a_{t'+T_s-1}\right)$, such that

$$\boldsymbol{a}^* = \arg\max_{a} \sum_{t=t'}^{t'+T_s-1} \gamma^{t-t'} r_t\left(\hat{s}_t, a_t\right) \tag{14}$$

$$\hat{s}_{t'} = s_{t'}, \hat{s}_{t+1} = \hat{f}_\theta\left(\hat{s}_t, a_t\right) \tag{15}$$

subject to the constraints in Eqs. (5)–(10).

The optimal sequence of actions is solved by this MPC scheme, and the first action in the sequence is applied to the system for the next state generation. After executing that on a specified sampling interval T_{int}, the generated on-policy data $\left(s_t, \ldots, s_{T_{\text{int}}}, a_t^*, \ldots, a_{T_{\text{int}}}^*, s_{t+1}, \ldots, s_{T_{\text{int}}+1}\right)$ is added into the data set D. Here D has a first-in–first-out structure, allowing most recent data to replace the oldest stored data. Meanwhile, the approximation of system dynamics (i.e. $\hat{f}$) is updated based on the updated data set D. The whole process described above is carried out recursively until the control task is accomplished. Note that the prediction horizon T_s in each interval should satisfy $T_s \leq T_{\text{int}}$.

3.2. RL-based wind turbine control

The optimal control policy should be acquired by optimizing the long-term reward, which is paramount for achieving satisfactory control performance. For the proposed RL-based wind turbine control method, the specific forms of rewards are defined in the objective function in (4) for Region II and in (11) for Region III.

Based on all these designs, we summarize the structure of our RL-based wind turbine control scheme in Fig. 3. It can be seen that the proposed control scheme has a DNN-based surrogate model that is updated according to the real-time measured data. This design is of significance to the complicated wind turbine systems since it is difficult to obtain the accurate turbine models, especially under uncertainties and faults. Then, the detailed implementation steps of our proposed RL-based wind turbine control method is summarized in Algorithm 1, where i_{max} is the total simulation step.

Note that some uncertainties and unpredictable faults are inevitable in practical applications, and those issues can potentially degrade control performance. One merit of our "model-based" RL structure is the strong robustness and adaptability to uncertainties and unpredictable faults. We take the generator actuator fault as an example, which can be caused by a fault in converter electronics or an offset in generator torque estimation [39]. Such a fault would cause the wind turbine model changes, resulting in the degradation of the control performance and large tracking errors.

ALGORITHM 1 RL-based Wind Turbine Control Algorithm

```
Collect initial data, including (s_0, s_1, ..., s_{N_D} − 1),
(a_0, a_1..., a_{N_D} − 1), and (s_1, s_2, ..., s_{N_D}).
Pre-process initial data and add to the dataset D
for i = 0 to i_max do
    Train the wind turbine dynamics f̂_θ(s_t, a_t) with the dataset
    D
    for t = iT_int to iT_int + T_int do
        Execute control policy learning based on the trained
        model f̂_θ(s_t, a_t)
        Optimize the objective function shown in Eq. (4) for
        torque control or in Eq. (11) for pitch control.
        As a result, obtain a sequence of control actions
        (torque/pitch command), denoted by (a*_t, a*_{t+1}..., a*_{t+T_s−1})
        Apply the first control action a*_t to the system and get
        s_{t+1}
        Store (s_t, a*_t, s_{t+1}) into D (which has a first-in first-out
        structure)
    end for
    i = i + 1
end for
```

Promisingly, with the help of the proposed DNN-based surrogate model structure, the developed RL method can better deal with these uncertainties and unpredictable faults without any prior knowledge of the uncertainties or faults. Specifically, when the fault occurs, the current measured data that reflect the key information of wind turbine models will be stored into the dataset D, which has a first-in–first-out structure, and replace the oldest stored data. According to the new data, the real-time updated DNN-based surrogate model contains the potential changes of wind turbine models in the existence of faults. This design renders the proposed RL method has strong robustness that can alleviate the undesirable effects on power production caused by potential uncertainties and faults.

Remark 2. Our RL-based wind turbine control method takes advantage of both DNN and MPC technologies. On the one hand, distinct from conventional MPC methods, we update the approximated system dynamics online. Such a design allows our controller to mitigate the mismatch between the learned dynamics and the true environment. The online updated system dynamics approximator also enables our method to have enhanced learning & data mining efficiency and to adapt to the real-time changes of system dynamics. On the other hand, different from the mainstream model-free deep RL methods such as DDPG, we utilize an MPC structure to replace the actor-critic DNN modules. This mechanism allows our method to have a reduced overall computational complexity.

Table 2
Parameters of the 5 MW wind turbine.

Variable	Description	Value
R	Rotor radius	63 m
J_r	Rotor moment of inertia	3.8768×10^7 kg · m^2
J_g	Generator moment of inertia	534, 116 kg · m^2
K_s	Drive train spring constant	8.67637×10^8 N · m/rad
D_s	Torsion damping coefficient of drive train	6.215×10^6 N · m/(rad/s)
N_g	Gear ratio	97
η	Power transferring efficiency	94.4%
τ_g	Generator time constant	0.01
τ_β	Pitch actuator time constant	0.01

Table 3
State and input constraints on the 5 mw wind turbine.

Variable	Description	Value
w_r^{rated}	Rated rotor speed	1.2671 rad/s
w_g^{rated}	Rated generator speed	122.9096 rad/s
T_g^{rated}	Rated generator torque	43,093,55 N · m
T_g^{max}/T_g^{min}	Maximum/Minimum generator torque	47,403 N · m, 200 N · m,
β^{max}/β^{min}	Maximum/Minimum pitch angle	90°/0°
$\Delta T_g^{max}/\Delta T_g^{min}$	Maximum/Minimum generator torque rate	1500 N · m/s/1500 N · m/s
$\Delta\beta^{max}/\Delta\beta^{min}$	Maximum/Minimum pitch angle rate	8 deg/s/−8 deg/s

4. Numerical simulations

In this section, simulations with the FAST (Fatigue, Aerodynamics, Structures, and Turbulence) simulator [40] are conducted to verify our method's effectiveness. FAST is a comprehensive aeroelastic simulator developed by NREL (National Renewable Energy Laboratory). We consider the NREL 5 MW variable-speed variable-pitch wind turbine dynamics in FAST. Its parameters are listed in Table 2 and its state & input constraints are presented in Table 3. Different wind profiles with a wind speed of [9 19] m/s are used in the validations, generated by the NREL software TurbSim [41].

4.1. Simulations under different controllers

For performance comparison, not only the proposed RL method but also the baseline control and the vanilla MPC strategy are employed in this subsection to carry out numerical simulations. (1) Baseline control: For the torque control, the reference generator torque command is obtained from the kw_r^2 law [42], which is proportional to the square of the filtered rotor speed [43]. For the pitch control, The gain-scheduled proportional–integral (PI) controller is commonly utilized to minimize the rotor speed or generator speed error between their measured and rated values [42]. (2) Model predictive control (MPC): The wind turbine dynamics are utilized to predict the future states in a finite prediction horizon. Based on these future states, the optimal control sequence is solved by optimizing the objective under various constraints.

The simulation time for a single run is 200 s, the predictive horizon T_s is 10 s, and the time interval T_{int} is 20 s. Note that the proposed RL method will be denoted as 'RL' thereafter.

Remark 3. It should be emphasized that the MPC strategy employed in our case studies directly utilizes the FAST model to carry out control. It does not suffer from any modeling errors or uncertainties under normal working conditions, rendering it a benchmark and an excellent approach to compare with our RL method.

For the torque control task in Region II, a wind profile with the mean speed of 9 m/s is employed, as shown in Fig. 4. The power responses under all the three methods are depicted in Fig. 5, in which the dashed blue line denotes the desired reference power trajectory. From Fig. 5, it can be observed that the power outputs under our RL-based method and the MPC method can precisely follow the desired optimal power trajectory, showing superior performance compared with the baseline control scheme. In addition, the variations of generator torques under three different controllers are presented in Fig. 6. One can see that our RL method and the MPC method lead to good generator torque tracking performance, while the baseline controller fails to accurately track the optimal generator torque.

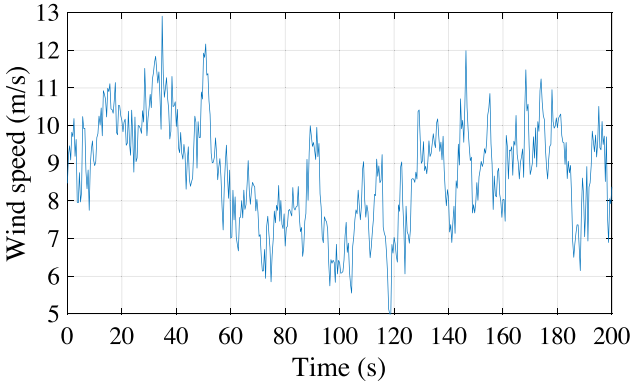

Fig. 4. Wind speed profile with the mean of 9 m/s.

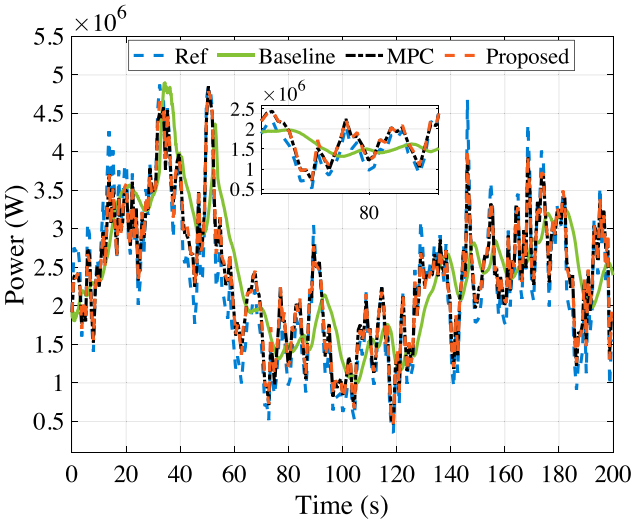

Fig. 5. Simulation results of power production in Region II under the RL technique, MPC method, and the baseline controller.

As for the pitch control in Region III, a wind profile with the mean speed of 19 m/s is employed, as shown in Fig. 7. Fig. 8 illustrates simulation results of power production under different pitch controllers. Compared with the baseline controller, the amplitude of power fluctuations with the designed RL method and the MPC method is remarkably reduced and thus the power quality is improved. Furthermore, the generator speeds in Region III using the baseline, MPC, and our RL methods are displayed in Fig. 9. One can see that all the controllers can achieve the desired objectives, maintaining the generator speed around the rated value. However, the generator speed under the baseline controller experiences larger oscillation around the rated generator speed. In comparison, considerable fluctuation reduction is achieved by the RL and MPC. Besides, the pitch angle's histories are given in Fig. 10. One can see that under RL and MPC, the pitch angle is closer to the reference value compared with the traditional baseline controller, which exhibits some apparent deviations. In summary, compared to the conventional baseline controller, the RL and MPC controllers are capable of improving the power quality and leading to superior tracking performance.

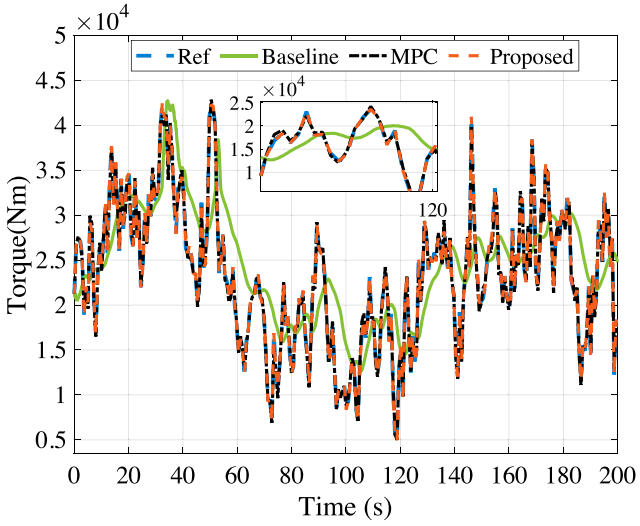

Fig. 6. The variation of generator torque in Region II with the RL technique, MPC method and baseline controller.

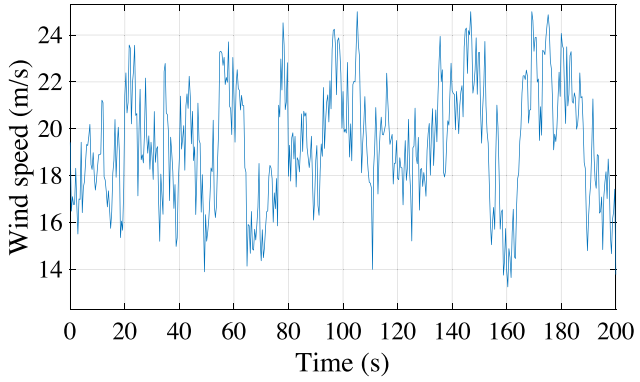

Fig. 7. Wind speed profile with the mean of 19 m/s.

It should be emphasized again that the MPC controller serves as a benchmark in this case study. It directly employs the model embedded in FAST to carry out control without being influenced by modeling errors and uncertainties under normal working conditions. Therefore, the results of MPC can be regarded as optimal in this case study. Here our RL controller leads to similar performance as MPC, showing that our controller can successfully approximate the optimal one. It is noteworthy that, distinct from MPC, our RL method is data-driven and can quickly adapt to system dynamics changes with the proposed DNN structure. To show that, the control performance of RL and MPC under faulty conditions will be tested in the following subsection.

4.2. Simulation results under torque actuator faults

To further evaluate the performance of the RL method, the scenario considering the generator actuator fault with torque offset is conducted. The generator actuator fault caused by a fault in converter electronics or an offset in generator torque estimation can significantly degrade the control performance. The performance of RL and MPC are re-evaluated under the torque offset fault in Region II. The generator torque with 4000 Nm and 7000 Nm offset faults from 60 s to 120 s is considered. The actuator faults will lead to unexpected changes in the generator power outputs. This is shown in Fig. 11. From Fig. 11(b), it can be seen that large power tracking errors are induced during the faulty period. Notably, it seems that the positive offset of the torque actuator leads to increased power, but this is not desired. Specifically, it can cause

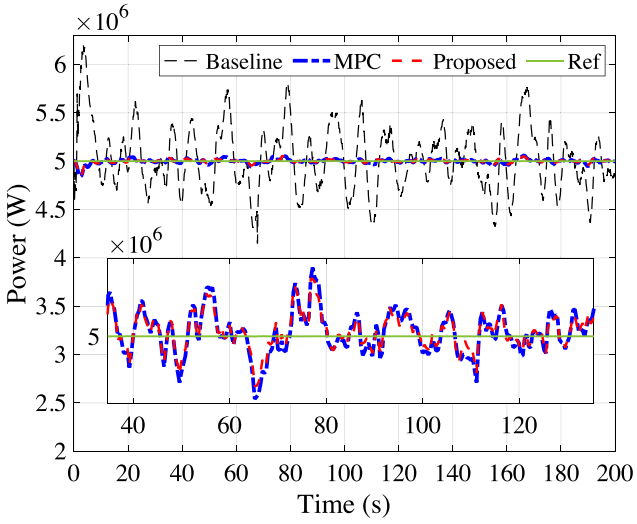

Fig. 8. Simulation results of power production in Region III under the RL technique, MPC method and baseline controller.

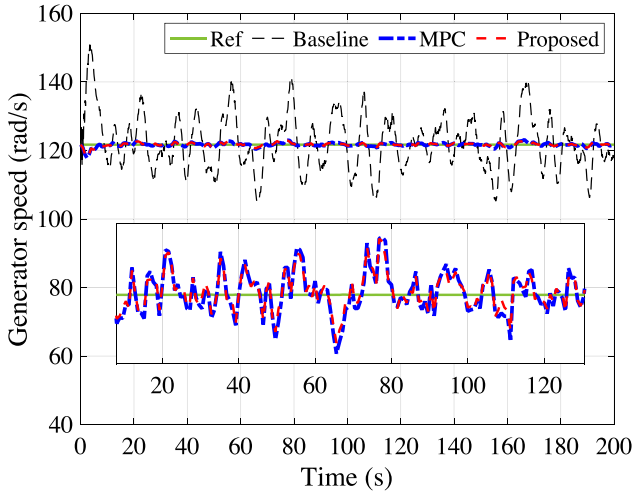

Fig. 9. The variation of generator speed in Region III with the RL technique, MPC method and baseline controller.

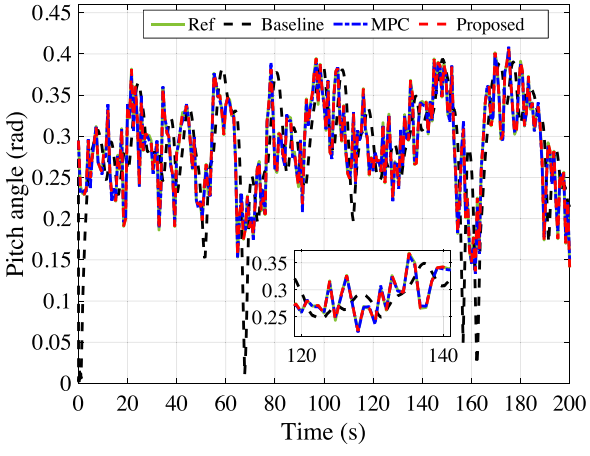

Fig. 10. The variation of pitch angle in Region III with the RL technique, MPC method and baseline controller.

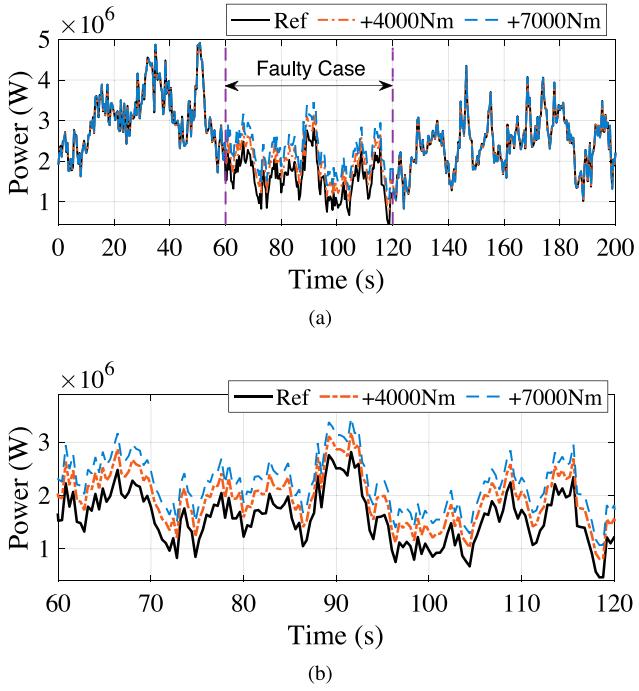

Fig. 11. Simulation results of power production by the RL without fault-tolerant scheme. (a) Simulation results during time 0–200 s. (b) The local enlarged figure during time 60 s–120 s.

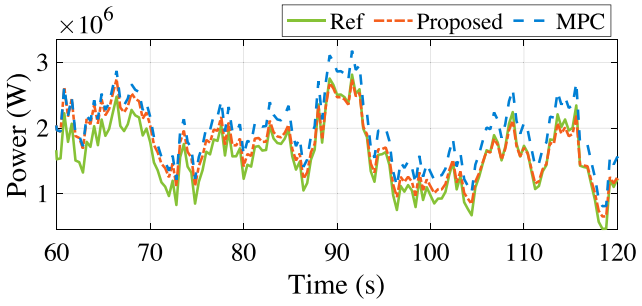

Fig. 12. Simulation results of power response under different controllers considering the torque actuator offset fault with +4000 Nm during time 60 s–120 s.

significant power quality problems, aggravate fatigue loads, and even lead to further torque actuator failure [43].

Fig. 12 depicts the power response of different controllers considering the torque actuator offset fault with +4000 Nm. As expected, the generator power produced by the proposed control method stays close to the reference power even under the faulty condition, which indicates that the effects caused by the torque actuator fault are compensated. In contrast, the power generation produced by the MPC controller cannot accurately follow the power reference. In addition, Fig. 13 shows the power responses of different controllers considering a faulty condition with the offset to be +7000 Nm. It can be observed that the proposed method still performs better in restoring the generator power to the optimal reference compared with the MPC controller. Although there exists slight power tracking accuracy deterioration during the initial fault period, our method can gradually achieve performance recovery and eventually meet the desired power output. These results show that our RL-based wind turbine controller has the strong fault-tolerant ability and online learning ability in comparison with the MPC method.

Additional simulation cases with the torque actuator offsets to be +3000 Nm, +4000 Nm, +5000 Nm, +6000 Nm, and +7000 Nm are also

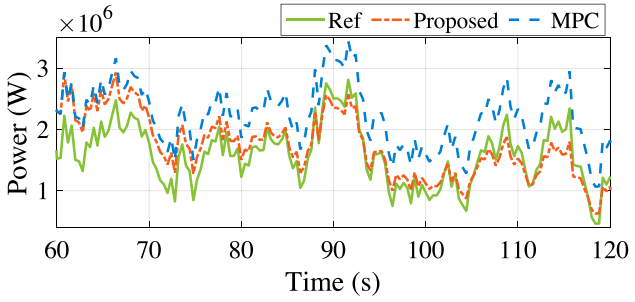

Fig. 13. Simulation results of power response under different controllers considering the torque actuator offset fault with +7000 Nm during time 60 s–120 s.

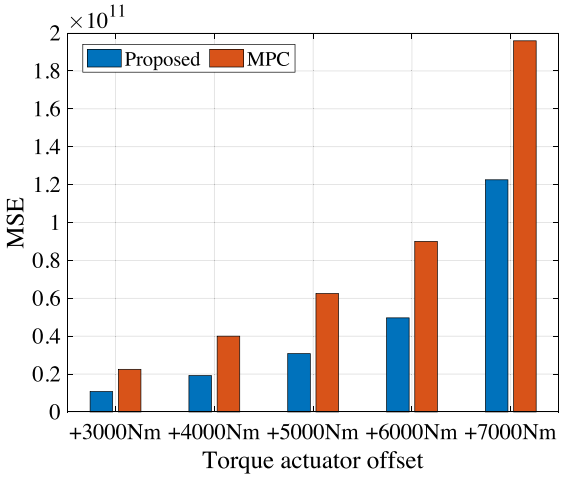

Fig. 14. Quantitative comparison results of MSE under the RL and MPC methods.

conducted. Specifically, the mean square error (MSE) in each case is compared, which is defined as

$$\mathrm{MSE} = \frac{1}{M}\sum_{k=1}^{M}\left(y_i - \hat{y}_i\right)^2 \tag{16}$$

where y_i denotes the data and M is the total number of data.

The corresponding results are summarized in Table 4 and Fig. 14, which shows the MSE values under RL and MPC. It can be observed that the MSE value of RL in each case is smaller than that of MPC, indicating better tracking performance under our RL controller in the presence of faults.

4.3. Simulation results under parameter uncertainties

Parameter uncertainties are considered in this subsection to further validate the adaptability and robustness of the proposed method. Specifically, the stiffness uncertainty, damping coefficient uncertainty, and moment of inertia uncertainty are considered. By varying the uncertain parameters from 10% to 30%, simulation results under the proposed method are shown in Fig. 15. It can be seen that the power outputs are rarely influenced when the perturbation level is 10%, and perform with acceptable deviations around the reference power when the perturbation level is 20% and 30%.

Remark 4. Extensive simulation results indicate that our RL method is effective for wind turbine control tasks. In normal cases, the control results of RL are highly close to the optimal solution (i.e. the one from MPC). In faulty conditions, our RL method has much better performance than MPC. Compared with MPC, the wind turbine dynamics are

Table 4
Quantitative comparison of generator power output under different torque actuator offset.

Case	Reference	+3000 Nm		+4000 Nm		+5000 Nm		+6000 Nm		+7000 Nm	
		Proposed	MPC	Proposed	MPC	Proposed	MPC	Proposed	MPC	Proposed	MPC
MSE	2,381,032	2,378,897	2,463,107	2,372,811	2,490,466	2,362,519	2,517,824	2,342,070	2,545,183	2,193,692	2,572,541

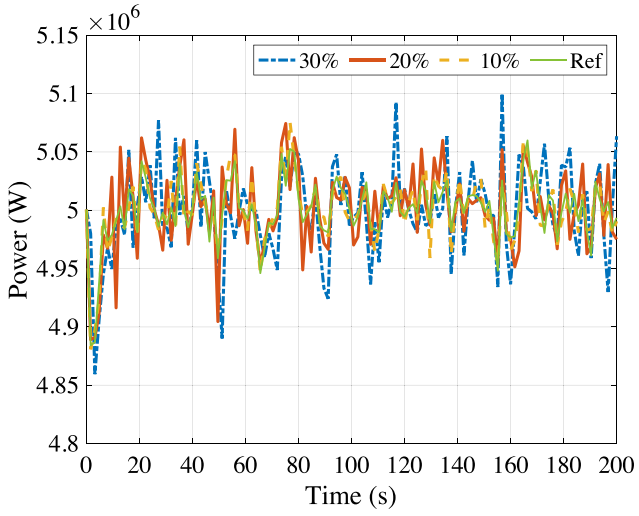

Fig. 15. Simulation results of generator power under the proposed methods in the presence of uncertainties.

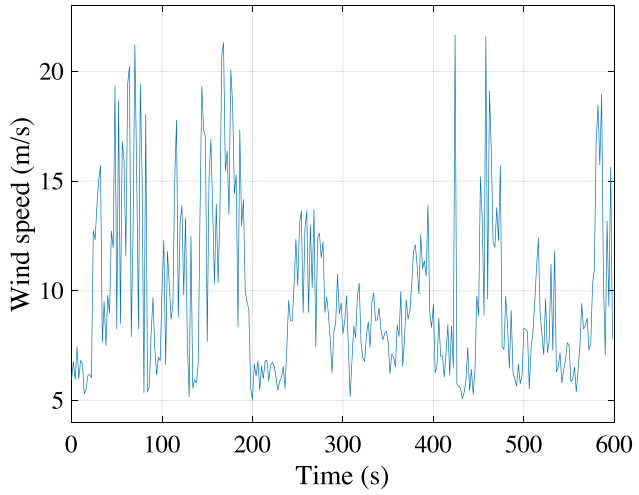

Fig. 16. Wind speed profile.

learnt and updated online in our RL method. In addition, unlike the conventional RL approaches, our RL algorithm has lower computational complexity and higher online learning efficiency.

4.4. Simulation results under real wind speed data

The aforementioned simulations were conducted using wind speed data generated by the NREL software TurbSim [41]. To provide a more comprehensive analysis, we also employ real wind speed data, obtained from the Xinglong Mountain wind farm in China, to further investigate the effectiveness of our proposed method in real-world conditions.

Specifically, the wind speed profile with a 1 s interval is shown in Fig. 16, encompassing inflow wind speeds both below and above the rated wind speeds. Simulation results for power production using the baseline, MPC, and the proposed method are presented in

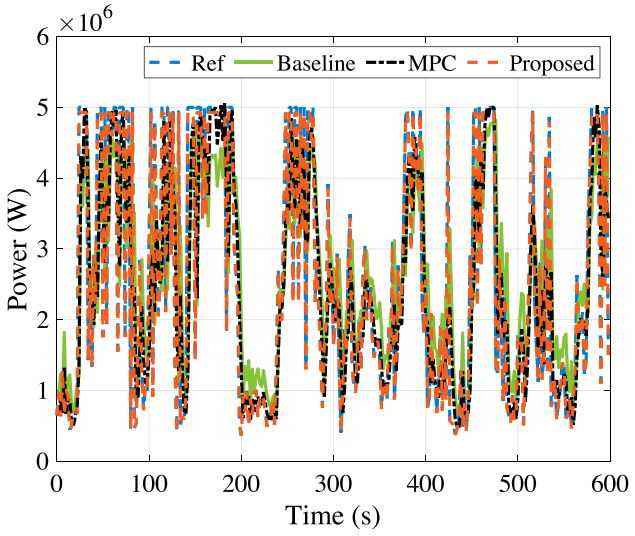

Fig. 17. Power production under different methods with the real wind speed.

Fig. 17, demonstrating the effectiveness of our controller when applied to real-world data. Furthermore, these comparative results highlight the superior performance of our proposed method and the MPC strategy compared to the baseline controller.

In summary, all simulation results demonstrate the effectiveness and robustness of the proposed method. The outcomes under different controllers reveal the improvement in power quality and superior tracking performance of our method. The results under torque actuator faults indicate that the proposed approach exhibits stronger robustness than both the MPC and baseline methods, verifying the advantages of the designed DNN that learns wind turbine models in real-time. Furthermore, the results under parameter uncertainties show that the proposed method is adept at handling uncertainties with impressive adaptability. Finally, simulations with real-world wind speed data further corroborate the effectiveness of the designed method.

5. Conclusion

In this paper, we proposed a data-driven algorithm for wind turbine torque and pitch control, developed through a synergistic integration of deep neural network (DNN), reinforcement learning (RL) and model predictive control (MPC) technologies. Our design captures and updates wind turbine dynamics online using a special DNN structure, while an MPC strategy, assisted by the surrogate DNN-based dynamics, generates the optimal control policy. The proposed method exhibits strong adaptability and robustness, effectively handling uncertainties and unexpected faults and leading to superior performance than commonly-employed MPC and other baseline methods. It is able to deal with wind control problems in both operating regions II & III under state and input constraints. High-fidelity simulations using the FAST (Fatigue, Aerodynamics, Structures, and Turbulence) simulator verified the effectiveness and advantages of the proposed method. Particularly, the results under parameter uncertainties and unexpected faults illustrated our method's impressive adaptability and robustness. The effectiveness of our method was also proven under real-world wind speed data. Future research will explore the integration of LIDAR data

and wind turbine control using machine learning algorithms, aiming to further enhance power generation efficiency and effectively respond to sudden events, such as gusts.

CRediT authorship contribution statement

Jingjie Xie: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Hongyang Dong:** Conceptualization, Methodology, Investigation, Software, Formal analysis, Data curation, Writing – review & editing. **Xiaowei Zhao:** Conceptualization, Investigation, Formal analysis, Funding acquisition, Resources, Supervision, Writing – review & editing.

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.

References

[1] A.M. Update, Global Wind Report, Global Wind Energy Council, 2017.

[2] X. Wang, D.W. Gao, J. Wang, W. Yan, W. Gao, E. Muljadi, V. Gevorgian, Implementations and evaluations of wind turbine inertial controls with FAST and digital real-time simulations, IEEE Trans. Energy Convers. 33 (4) (2018) 1805–1814.

[3] A.D. Wright, L. Fingersh, Advanced Control Design for Wind Turbines; Part I: Control Design, Implementation, and Initial Tests, Tech. Rep., National Renewable Energy Lab. (NREL), Golden, CO (United States), 2008.

[4] H.T. Do, T.D. Dang, H.V.A. Truong, K.K. Ahn, Maximum power point tracking and output power control on pressure coupling wind energy conversion system, IEEE Trans. Ind. Electron. 65 (2) (2017) 1316–1324.

[5] O.F. Alarcón, B.I. Velásquez, A.R. Hunter, L.B. Pavez, R. Moncada, Hybrid PID-fuzzy pitch control for wind turbines, in: 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies, CHILECON, IEEE, 2017, pp. 1–6.

[6] R. Kandiban, R. Arulmozhiyal, Speed control of BLDC motor using adaptive fuzzy PID controller, Procedia Eng. 38 (2012) 306–313.

[7] M.A. Soliman, H.M. Hasanien, H.Z. Azazi, E.E. El-Kholy, S.A. Mahmoud, An adaptive fuzzy logic control strategy for performance enhancement of a grid-connected PMSG-based wind turbine, IEEE Trans. Ind. Inform. 15 (6) (2018) 3163–3173.

[8] C.A. Evangelista, A. Pisano, P. Puleston, E. Usai, Receding horizon adaptive second-order sliding mode control for doubly-fed induction generator based wind turbine, IEEE Trans. Control Syst. Technol. 25 (1) (2016) 73–84.

[9] X. Yin, X. Tong, X. Zhao, A. Karcanias, Maximum power generation control of a hybrid wind turbine transmission system based on H∞ loop-shaping approach, IEEE Trans. Sustain. Energy 11 (2) (2019) 561–570.

[10] T.G. Hovgaard, S. Boyd, J.B. Jørgensen, Model predictive control for wind power gradients, Wind Energy 18 (6) (2015) 991–1006.

[11] B.M. Gavgani, A. Farnam, J.D. De Kooning, G. Crevecoeur, Efficiency enhancements of wind energy conversion systems using soft switching multiple model predictive control, IEEE Trans. Energy Convers. 37 (2) (2021) 1187–1199.

[12] H.-S. Yan, Z.-Y. Duan, Tube-based model predictive control using multidimensional Taylor network for nonlinear time-delay systems, IEEE Trans. Automat. Control 66 (5) (2020) 2099–2114.

[13] S. Zhan, J. Na, G. Li, B. Wang, Adaptive model predictive control of wave energy converters, IEEE Trans. Sustain. Energy 11 (1) (2018) 229–238.

[14] J. Liu, G. Li, H.K. Fathy, An extended differential flatness approach for the health-conscious nonlinear model predictive control of lithium-ion batteries, IEEE Trans. Control Syst. Technol. 25 (5) (2016) 1882–1889.

[15] A. Jain, G. Schildbach, L. Fagiano, M. Morari, On the design and tuning of linear model predictive control for wind turbines, Renew. Energy 80 (2015) 664–673.

[16] A. Koerber, R. King, Combined feedback–feedforward control of wind turbines using state-constrained model predictive control, IEEE Trans. Control Syst. Technol. 21 (4) (2013) 1117–1128.

[17] S. Gros, A. Schild, Real-time economic nonlinear model predictive control for wind turbine control, Internat. J. Control 90 (12) (2017) 2799–2812.

[18] M.D. Spencer, K.A. Stol, C.P. Unsworth, J.E. Cater, S.E. Norris, Model predictive control of a wind turbine using short-term wind field predictions, Wind Energy 16 (3) (2013) 417–434.

[19] M. Soliman, O.P. Malik, D.T. Westwick, Multiple model predictive control for wind turbines with doubly fed induction generators, IEEE Trans. Sustain. Energy 2 (3) (2011) 215–225.

[20] Z.-Y. Duan, H.-S. Yan, Adaptive dynamic programming based on multi-dimensional Taylor network for time-delay nonlinear system with uncertainties, in: 2020 2nd World Symposium on Artificial Intelligence, WSAI, IEEE, 2020, pp. 44–49.

[21] C. Wei, Z. Zhang, W. Qiao, L. Qu, Reinforcement-learning-based intelligent maximum power point tracking control for wind energy conversion systems, IEEE Trans. Ind. Electron. 62 (10) (2015) 6360–6370.

[22] A. Kushwaha, M. Gopal, B. Singh, Q-learning based maximum power extraction for wind energy conversion system with variable wind speed, IEEE Trans. Energy Convers. 35 (3) (2020) 1160–1170.

[23] C. Wei, Z. Zhang, W. Qiao, L. Qu, An adaptive network-based reinforcement learning method for MPPT control of PMSG wind energy conversion systems, IEEE Trans. Power Electron. 31 (11) (2016) 7837–7848.

[24] P. Chen, D. Han, F. Tan, J. Wang, Reinforcement-based robust variable pitch control of wind turbines, IEEE Access 8 (2020) 20493–20502.

[25] J.E. Sierra-Garcia, M. Santos, R. Pandit, Wind turbine pitch reinforcement learning control improved by PID regulator and learning observer, Eng. Appl. Artif. Intell. 111 (2022) 104769.

[26] B. Fernandez-Gauna, M. Graña, J.-L. Osa-Amilibia, X. Larrucea, Actor-critic continuous state reinforcement learning for wind-turbine control robust optimization, Inform. Sci. 591 (2022) 365–380.

[27] N. Tomin, V. Kurbatsky, H. Guliyev, Intelligent control of a wind turbine based on reinforcement learning, in: 2019 16th Conference on Electrical Machines, Drives and Power Systems, ELMA, IEEE, 2019, pp. 1–6.

[28] V.T. Aghaei, A. Ağababaoğlu, B. Bawo, P. Naseradinmousavi, S. Yıldırım, S. Yeşilyurt, A. Onat, Energy optimization of wind turbines via a neural control policy based on reinforcement learning Markov chain Monte Carlo algorithm, Appl. Energy 341 (2023) 121108.

[29] Q. Huang, Model-based or model-free, a review of approaches in reinforcement learning, in: 2020 International Conference on Computing and Data Science, CDS, IEEE, 2020, pp. 219–221.

[30] W. Sun, N. Jiang, A. Krishnamurthy, A. Agarwal, J. Langford, Model-based rl in contextual decision processes: Pac bounds and exponential improvements over model-free approaches, in: Conference on Learning Theory, PMLR, 2019, pp. 2898–2933.

[31] A. Nagabandi, G. Kahn, R.S. Fearing, S. Levine, Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning, in: 2018 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2018, pp. 7559–7566.

[32] T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, 2015, arXiv preprint arXiv:1509.02971.

[33] Y. She, X. She, M.E. Baran, Universal tracking control of wind conversion system for purpose of maximum power acquisition under hierarchical control structure, IEEE Trans. Energy Convers. 26 (3) (2011) 766–775.

[34] Z. Lin, Z. Chen, Q. Wu, S. Yang, H. Meng, Coordinated pitch & torque control of large-scale wind turbine based on Pareto efficiency analysis, Energy 147 (2018) 812–825.

[35] Y. Xia, K.H. Ahmed, B.W. Williams, Wind turbine power coefficient analysis of a new maximum power point tracking technique, IEEE Trans. Ind. Electron. 60 (3) (2012) 1122–1132.

[36] H. Dong, X. Zhao, Composite experience replay-based deep reinforcement learning with application in wind farm control, IEEE Trans. Control Syst. Technol. (2021).

[37] J. Xie, H. Dong, X. Zhao, A. Karcanias, Wind farm power generation control via double-network-based deep reinforcement learning, IEEE Trans. Ind. Inform. 18 (4) (2021) 2321–2330.

[38] R. Chai, A. Savvaris, A. Tsourdos, S. Chai, Y. Xia, Optimal tracking guidance for aeroassisted spacecraft reconnaissance mission based on receding horizon control, IEEE Trans. Aerosp. Electron. Syst. 54 (4) (2018) 1575–1588.

[39] P.F. Odgaard, J. Stoustrup, M. Kinnaert, Fault-tolerant control of wind turbines: A benchmark model, IEEE Trans. Control Syst. Technol. 21 (4) (2013) 1168–1182.

[40] J.M. Jonkman, M.L. Buhl, et al., FAST User's Guide, vol. 365, National Renewable Energy Laboratory, Citeseer, Golden, CO, 2005, p. 366.

[41] B.J. Jonkman, M.L. Buhl, TurbSim User's Guide, Tech. Rep., National Renewable Energy Lab. (NREL), Golden, CO (United States), 2006.

[42] J. Jonkman, S. Butterfield, W. Musial, G. Scott, Definition of a 5 MW Reference Wind Turbine for Offshore System Development, Tech. Rep., National Renewable Energy Lab. (NREL), Golden, CO (United States), 2009.

[43] H. Badihi, Y. Zhang, H. Hong, Wind turbine fault diagnosis and fault-tolerant torque load control against actuator faults, IEEE Trans. Control Syst. Technol. 23 (4) (2014) 1351–1372.