IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 August 2022, accepted 1 September 2022, date of publication 8 September 2022, date of current version 16 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3205124

RESEARCH ARTICLE

Reinforcement Learning-Based Adaptive Optimal Fuzzy MPPT Control for Variable Speed Wind Turbine

NGA THI-THUY VU[1], HA DUC NGUYEN[1], AND ANH TUAN NGUYEN[2]
[1]School of Electrical and Electronics Engineering, Hanoi University of Science and Technology, Hanoi 11615, Vietnam
[2]Faculty of Electrical and Electronic Engineering, Phenikaa University, Ha Dong, Hanoi 12116, Vietnam

Corresponding author: Nga Thi-Thuy Vu (nga.vuthithuy@hust.edu.vn)

This work was supported by the Ministry of Education and Training of Vietnam (MOET-VN) under Grant B2022-BKA-05.

ABSTRACT A reinforcement learning-based adaptive optimal fuzzy controller is proposed for maximum power point tracking (MPPT) control of a variable-speed permanent magnet synchronous generator-based wind energy generation system. The algorithm consists of a critic, an adaptive optimal fuzzy controller, and an adaptive optimal fuzzy estimator. The critic is built based on an adaptive neuro-fuzzy inference system (ANFIS) network instead of the neural network as normal to reduce the computation. The error between the system output and the estimator output is used as the input of the critic. In addition, the critic is used to calculate the update law for the parameters of the adaptive optimal fuzzy controller and adaptive optimal fuzzy estimator based on minimizing the input error function. Moreover, the proposed control scheme is output feedback instead of state feedback, which does not require a system model as well as system parameters, so the system is robust to uncertainties and external disturbances. Besides, the stabilization proof is accomplished by using the Lyapunov stability theorem for the closed-loop system and the convergence of the update law. Finally, the effectiveness of the proposed reinforcement learning-based adaptive optimal fuzzy control scheme is verified through simulation with various scenarios such as step wind speed, random wind speed, and system parameter variations. Also, the comparisons with other control schemes in the state-of-art (neural network reinforcement learning based adaptive optimal fuzzy controller, PI controller) are executed to demonstrate the advantages of the proposed control scheme.

INDEX TERMS Adaptive, fuzzy logic, optimal, MPPT, permanent magnet synchronous generator, reinforcement learning, wind turbine.

I. INTRODUCTION

Nowadays, among fuel sources, wind energy has become an attractive and competitive clean renewable energy source because of its reliable operation and sustainable development [1], [2]. A wind energy conversion system (WECS) is a combination of aerodynamic, mechanical, and electrical components, so its operation mainly depends on considerable factors such as the type of generator, the wind speed, the installed location, etc. In these factors, the wind speed changes during the working period while other elements are determined carefully in the design and construction processes. In terms of wind speed variation, the main operation of the WECS is divided into two modes: maximum power point tracking (MPPT) mode in the region below the rated speed [3] and output power limitation mode in the region above the rated speed [2]. For above the rated wind speed operation region, the pitch angle controllers are used to limit the rotor speed to the rated value [4]. Meanwhile, below the rated speed region, the efficiency of the system depends on the working point that is controlled to track the maximum power point (MPP) in the power-speed curves [2].

Over the last years, many researchers have focused on developing advanced algorithms for MPPT control. Most MPPT control methods typically include two phases: the first is the MPP searching process, and the second is the control of the working point to follow this MPP. In particular, due to system nonlinearities, system parameter uncertainties,

The associate editor coordinating the review of this manuscript and approving it for publication was Zhong Wu.

and the influences of external disturbances (e.g., mechanical torque), the controller design for the machine side inverter is challenged in the second phase. To cope with these problems, numerous control strategies have been introduced such as PI control [5], sliding mode control (SMC) [6], [7], [8], [9], adaptive control [10], [11], and model predictive control (MPC) [12]. In [5], a PI controller is employed for the current control loop of the multi-motor wind turbine system. However, this PI controller cannot ensure good performance under condition variations due to the nonlinearity of WEGS and the affection of the environment (e.g., wind speed, air density). Next, SMC is considered as the nonlinear control technique used to deal with the parametric uncertainties and disturbances of WECS for the MPPT control. In [6] and [7], the high-order SMCs are applied to improve the performance of the WECS by reducing the chattering phenomenon with the continuous control input. However, the fluctuation of the output voltage and power is still high [7]. In [8], an enhanced reaching law-based SMC method is introduced for the MPPT control of offshore WECS, which consists of two loops: the current control loop with a conventional PI controller and the speed control loop with a finite time reaching SMC, and a mechanical torque observer. This control strategy significantly improves the ability of the WEGS to resist uncertainties and disturbances. In [9], a fixed-time fractional-order SMC is designed for both rotor side converter (RSC) and grid side converter (GSC) to improve the power quality of the PMSG wind turbine system. In this strategy, the friction-order expresses the controller in the continuous form while finite time stability guarantees that the system is stable within a given time. In parallel with sliding mode control, adaptive control is also a favorite strategy to deal with uncertainties and disturbances. In [10] and [11], the PMSG-based wind turbine system is controlled by a nonlinear adaptive control technique. The nonlinearities, the uncertainties, and the external disturbances are estimated by a high-gain observer then they are compensated to provide better performances for the system. The simulation and experimental results of this proposed controller are compared with other different observer-based control methods as well as other conventional vector control methods to verify the effectiveness. In [12], an MPC scheme is presented for PMSG-based WECS to obtain a fast dynamic response time in case of overmodulation. However, the steady-state error issue can exist in the normal modulation when the effort in reducing the sampling frequency is carried out.

Recently, the learning control methods such as fuzzy logic, neural network, or reinforcement learning (RL) become active solutions to handle the problems of nonlinear and uncertain systems [13], [14], [15], [16]. In [13] and [14], fuzzy wavelet networks are utilized for nonlinear dynamic systems. These algorithms guarantee predefined performances and reject the requirement of system dynamics. However, the optimal performance indexes are not mentioned in these works. A single-network adaptive critic fuzzy optimal tracking control is introduced for hypersonic flight vehicles in [15]. In this scheme, the unknown functions are approximated by fuzzy networks and the critic is estimated by an adaptive law. The optimal controller in this work is obtained by solving the Hamilton–Jacobi–Bellman (HJB) equation so the dynamic model should be partially known. Another adaptive optimal fuzzy control algorithm based on RL is addressed in [16] for nonlinear uncertain systems. In this control scheme, the critic is approximated by a neural network while the actor is built from the Takagi-Sugeno (TS) fuzzy system. The requirement of a dynamic model is removed; however, the computation burden of the neural network is heavy.

In the field of WECS, the control schemes based on learning techniques are also widely popular [17], [18], [19], [20], [21], [22], [23], [24], [25]. These intelligent methodologies can be applied for wind speed estimation [17], [18], MPP searching algorithms [19], [20], or MPP tracking methods [21], [22], [23], [24], [25]. In the MPP tracking category, the nonlinearities in the dynamic model and disturbances are identified by the fuzzy logic system then they are compensated by the adaptive fuzzy controller [21] or sliding mode controller [22]. In [23], the current loop is controlled directly by a fuzzy logic controller with input tracking errors to overcome the disadvantages of the PI controller. The fuzzy logic technique is also employed in [24] to calculate the electromagnetic torque for current loop control using the error of mechanical speed and the change of this error as the inputs. This control structure enhances the performance of the WT under oscillation wind speed and guaranteed efficient and reliable grid integration of the wind turbine. In [25], adaptive dynamic programming (ADP) is employed for the MPPT control scheme of a WECS. Firstly, the dynamic model of the system is reconstructed to the data-driven model by the recurrent neural network using available input-output data then the adaptive optimal controller is designed based on the ADP technique. This configuration stabilizes the working point of the system near the optimal point; also, the dynamic responses and the robustness of the system are enhanced significantly.

All the above strategies are presented to face the challenges of the MPPT control for wind turbine systems such as nonlinearities, uncertainties, and disturbances by introducing suitable techniques. However, most of the introduced algorithms are state feedback controllers that need a dynamic model of the system as well as the complete measurement. Moreover, the mentioned methods just focus on improving the dynamic responses and reducing steady-state error for wind turbine systems, while the optimal performance is not investigated. In this paper, an RL-based adaptive optimal fuzzy control method for MPPT control of a PMSG-based WECS is introduced. The proposed control system is composed of the critic, the adaptive optimal fuzzy controller, and the adaptive optimal fuzzy estimator. First, the critic is formed based on the ANFIS technique with a hybrid update rule to lower the computation and storage capacity as compared to the neural network-based critic. Next, the adaptive optimal fuzzy controller and the adaptive optimal fuzzy estimator

are established by using the T-S fuzzy system. Besides, the parameters of both adaptive optimal fuzzy controller and adaptive optimal fuzzy estimator are updated by the critic which is subjected to minimizing the input error function. The stability of the overall system and the convergence of updated parameters are proven by using the Lyapunov stable theory. The advantages of the proposed control scheme are listed as the following:

- Unlike the existing control algorithm for MPPT control of WECS [5], [6], [7], [8], [9], [10], [11], [12] and [21], [22], [23], [24], [25], the proposed controller is output feedback and does not use the system dynamic model as well as system parameters. Also, the requirement for a mechanical torque observer is omitted.
- In comparison with [13], [14], this proposed control scheme guarantees optimal performance along with the static and transient responses.
- The fuzzy optimal tracking control developed in [15] is limited to affine systems. Meanwhile, the newly developed controller in this study is capable of handling non-affine systems.

The effectiveness of the proposed RL-based adaptive optimal fuzzy control scheme is validated by simulation results in MATLAB/Simulink on a 5-kW PMSG-based WECS using the direct-drive configuration with a back-to-back (BTB) converter. Various scenarios are investigated to illustrate the efficacy of the proposed method such as the step wind speed, random wind speed, nominal system parameters, and varying system parameters. In each scenario, the dynamic responses of the proposed control system are compared with the ones of the neural network RL-based adaptive optimal fuzzy system and the PI control system.

II. SYSTEM MODELING

In this paper, a direct-drive PMSG-based WECS configuration is used to investigate the proposed control method. The electric power produced by the PMSG is transferred to the grid through a BTB structure which includes a machine-side converter (MSC) and a grid-side converter (GSC). The MSC is controlled by appropriate duty cycles to extract the maximum power based on an MPPT algorithm when the wind speed is lower than the rated speed, or to limit the maximum extracted power at the rated value for speeds above the rated speed. Besides, the GSC is controlled to regulate the active and reactive power of the grid-connected WECS via a direct-current (DC) link regulator or satisfy the requirements about quality of voltage of the distribution feeders [26]. Within this context, the proposed method focuses on controlling the MSC to achieve the MPPT control capability, while assuming that another controller is adopted for the GSC control. Figure 1 shows the typical configuration of a direct-drive PMSG-based WECS using the BTB-connected.

A. WIND TURBINE AERODYNAMIC MODEL

In a practical WECS, the mechanical power that a wind turbine extracted from the wind is represented

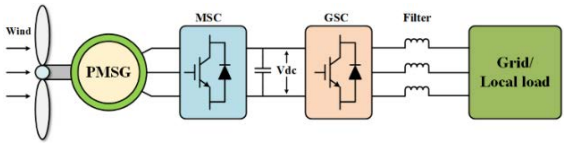

FIGURE 1. Configuration of direct-drive PMSG-based WECS using BTB-connected.

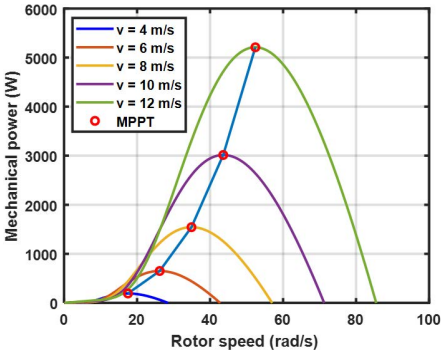

FIGURE 2. Wind turbine power-rotor speed characteristic curves for different wind speeds.

as the following [3]:

$$P_m = \frac{1}{2}\pi\rho R^2 V^3 C_p(\lambda, \beta) \tag{1}$$

where P_m is the mechanical power (in Watts), ρ is the air density (in kg/m3), R is the radius of the turbine blade (in m), V is the wind speed (in m/s), and Cp is the power coefficient which is identified by the tip speed ratio λ and the blade pitch angle β. It is noted that the pitch angle β of the blades is always constant during the MPPT control process in the event of the wind speed being below the rated value. The value of C_p is calculated as the following [24]:

$$C_p = 0.5176(116/\lambda_t - 0.4\beta - 5)e^{-21/\lambda_t} + 0.0068\lambda \tag{2}$$

in which:

$$\frac{1}{\lambda_t} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1}, \tag{3}$$

$$\lambda = \frac{\omega_t R}{V}, \tag{4}$$

where ω_t is rotor speed.

Fig. 2 shows the typical wind turbine power-rotor speed characteristic curves for different wind speeds. At each value of wind speed, the extracted mechanical power is maximized by controlling the rotational speed of the generator at the optimal value [24]:

$$\omega_{opt} = \frac{V}{R}\lambda_{opt} \tag{5}$$

where λ_{opt} depends on each certain wind turbine and is given by the manufacturer. This is the principle of the MPPT control algorithm in this work.

B. PERMANENT MAGNET SYNCHRONOUS GENERATOR MODEL

In this paper, the surface-mounted PMSG is adopted because of the uniform air gap and high power density due to the

greater flux linkage between the permanent magnets and the stator windings. The mathematic model of PMSG in the dq-reference frame is expressed as the following [3]:

$$\begin{cases} \dfrac{d\omega_r}{dt} = \dfrac{1}{J}(T_m - B\omega_r - T_e) \\ \dfrac{di_q}{dt} = \dfrac{-R_s}{L_s} i_q - p_n \omega_r i_d - \dfrac{\psi_m p_n}{L_s}\omega_r + \dfrac{1}{L_s} v_q \\ \dfrac{di_d}{dt} = \dfrac{-R_s}{L_s} i_d + p_n \omega_r i_q + \dfrac{1}{L_s} v_d \end{cases} \tag{6}$$

where ω_r is the rotor speed of PMSG, T_m is the mechanical torque of the wind turbine; T_e is the electromagnetic torque of PMSG; i_d and i_q are d-axis and q-axis currents, respectively; L_s is the stator inductance; R_s is the stator resistance; p_n is the number of pole pairs; ψ_m is the magnet flux linkage; J is the equivalent rotor inertia; B is the equivalent viscous friction coefficient. The electromagnetic torque can be calculated by the following equation:

$$T_e = \frac{3}{2}\psi_m p_n i_q \tag{7}$$

It should be noted that the direct-drive PMSG-based WECS configuration is adopted in this study. Thus, the turbine rotating speed (ω_t) is the same as the rotor speed (ω_r) [27].

III. REINFORCEMENT LEARNING BASED ADAPTIVE OPTIMAL FUZZY CONTROLLER DESIGN

Equation (6) can use the defined state variable vector $x = [x_1 x_2 x_3]^T$ to establish the new error dynamics, where $x_1 = \omega_r - \omega_{ref}$; $x_2 = i_q - i_{qref}$; $x_3 = i_d - i_{dref}$. In which $\omega_{ref} = \omega_{opt}$ is the desired rotor speed, $i_{qref} = \frac{2}{3}\frac{1}{p_n\psi_m}(T_m - B\omega_{ref} - J\frac{d\omega_{ref}}{dt})$ is q-axis desired current, and $i_{dref} = 0$ is d-axis desired current.

It is noted that when the system works at the optimal point we have:

$$T_m = T_m^{opt} = \frac{1}{2}\rho\pi R^5 \frac{C_p^{\max}}{\lambda_{opt}^3}\omega_{opt}^2 = K_{opt}\omega_{opt}^2 \tag{8}$$

Then

$$i_{qref} = \frac{2}{3}\frac{1}{p_n\psi_m}\left(K_{opt}\omega_{ref}^2 - B\omega_{ref} - J\frac{d\omega_{ref}}{dt}\right) \tag{9}$$

The time derivative of x using (6) is obtained by the following:

$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{pmatrix} \dfrac{1}{J}(T_m - B\omega_r - T_e) - \dot{\omega}_{ref} \\ -\dfrac{R_s}{L_s} i_q - p_n\omega_r i_d - \dfrac{\psi_m p_n}{L_s}\omega_r - \dot{i}_{qref} + \dfrac{1}{L_s} v_q \\ -\dfrac{R_s}{L_s} i_d + p_n\omega_r i_q - \dot{i}_{dref} + \dfrac{1}{L_s} v_d \end{pmatrix}. \tag{10}$$

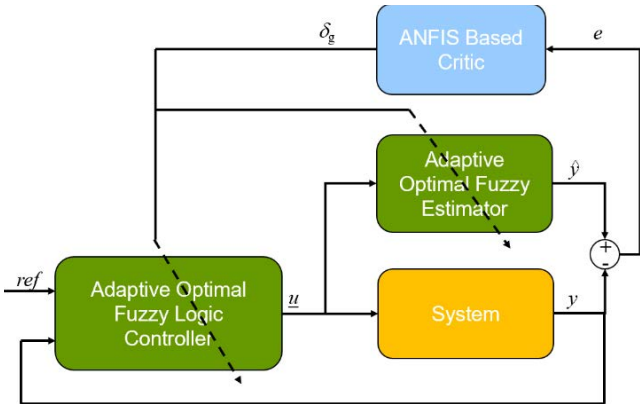

FIGURE 3. Block diagram of the proposed algorithm.

Dynamic model (10) can be rewritten as:

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= Cx \end{aligned} \tag{11}$$

where

$$f(x,u) = \begin{pmatrix} \dfrac{1}{J}(T_m - B\omega_r - T_e) - \dot{\omega}_{ref} \\ -\dfrac{R_s}{L_s} i_q - p_n\omega_r i_d - \dfrac{\psi_m p_n}{L_s}\omega_r - \dot{i}_{qref} + \dfrac{1}{L_s} v_q \\ -\dfrac{R_s}{L_s} i_d + p_n\omega_r i_q - \dot{i}_{dref} + \dfrac{1}{L_s} v_d \end{pmatrix}; \quad u = \begin{bmatrix} v_q \\ v_d \end{bmatrix}; \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{12}$$

In this section, an adaptive optimal controller based on reinforcement learning is proposed for the system (11). The process of designing consists of the following steps:

- Step 1: Design the system model estimator and system controller using the T-S fuzzy technique.
- Step 2: Critic network design based on ANFIS technique.
- Step 3: Update the parameters of the fuzzy estimator and fuzzy controller based on the optimal rule.

The block diagram of the proposed scheme is presented in Fig. 3.

A. T-S FUZZY CONTROLLER AND ESTIMATOR

In this section, the model of the system will be identified by fuzzy technique for each local working point. For each local model, a local controller is designed based on this estimated model. Finally, the global model and controller are achieved by the Weighted Average defuzzification.

The j^{th} rule for estimator and controller has the following form [28]:

R_j: IF $y(k)$ is Y_1^j and $y(k-1)$ is Y_2^j THEN

$$\hat{y}^j(k+1) = \begin{bmatrix} \hat{y}_1^j(k+1) \\ \hat{y}_2^j(k+1) \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^j(k)\, y_1(k) + a_{12}^j(k)\, y_1(k-1) \\ \quad + b_{11}^j(k)\, u_1(k) + b_{12}^j(k)\, u_1(k-1) \\ a_{21}^j(k)\, y_2(k) + a_{22}^j(k)\, y_2(k-1) \\ \quad + b_{21}^j(k)\, u_2(k) + b_{22}^j(k)\, u_2(k-1) \end{bmatrix} \tag{13}$$

$$\begin{aligned} u^j(k) &= \begin{bmatrix} u_1^j(k) \\ u_2^j(k) \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{b_{11}^j(k)} \left(a_{11}^j(k)\, y_1(k) + a_{12}^j(k)\, y_1(k-1) \right. \\ \quad \left. + b_{12}^j(k)\, u_1(k-1) \right) \\ -\frac{1}{b_{21}^j(k)} \left(a_{21}^j(k)\, y_1(k) + a_{22}^j(k)\, y_1(k-1) \right. \\ \quad \left. + b_{22}^j(k)\, u_2(k-1) \right) \end{bmatrix} \end{aligned} \tag{14}$$

where $a^j(k) = [a_{11}^j(k),\ a_{12}^j(k),\ a_{21}^j(k),\ a_{22}^j(k)]^T$ and $b^j(k) = [b_{11}^j(k),\ b_{12}^j(k),\ b_{21}^j(k),\ b_{22}^j(k)]^T$ are parameters that will be updated after each step.

The global output of the estimator and controller obtained from local rules using Weighted Average defuzzification have the following form:

$$y(x(k)|\,\delta) = \sum_{j=1}^{N} h^j(k) \odot \hat{y}^j(k) \tag{15}$$

$$u(k) = \sum_{j=1}^{N} h^j(k) \odot u^j(k) \tag{16}$$

where N is the number of rules, $\odot$ is the Hadamard product,

$$\begin{aligned} h^j(k) &= \begin{bmatrix} h_1^j(k) & h_2^j(k) \end{bmatrix}^T \\ &= \begin{bmatrix} \frac{\mu_1^j(k)}{\sum_{j=1}^N \mu_1^j(k)} & \frac{\mu_2^j(k)}{\sum_{j=1}^N \mu_2^j(k)} \end{bmatrix}^T \end{aligned} \tag{17}$$

in which $\mu_i^j(k)$ $(i = 1, 2; j = 1, 2, \ldots, N)$ is the membership function corresponding with the rule j^{th},

$$\begin{aligned} \mu^j(k) &= \begin{bmatrix} \mu_1^j \\ \mu_2^j \end{bmatrix} \\ &= \begin{bmatrix} \exp\left(-\frac{1}{2} \left(\frac{y_1(k) - c_{11}^j(k)}{\sigma_{11}^j(k)} \right)^2 \right) \\ \exp\left(-\frac{1}{2} \left(\frac{y_1(k-1) - c_{12}^j(k)}{\sigma_{12}^j(k)} \right)^2 \right) \\ \exp\left(-\frac{1}{2} \left(\frac{y_2(k) - c_{21}^j(k)}{\sigma_{21}^j(k)} \right)^2 \right) \\ \exp\left(-\frac{1}{2} \left(\frac{y_2(k-1) - c_{22}^j(k)}{\sigma_{22}^j(k)} \right)^2 \right) \end{bmatrix}, \end{aligned} \tag{18}$$

$\delta = \{a^j, b^j, c^j, \sigma^j\}$ is the set of system parameters.

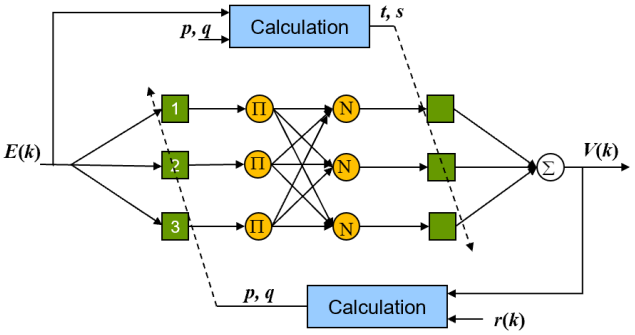

FIGURE 4. Structure of the Critic network.

B. THE CRITIC NETWORK DESIGN

In this work, the Critic network is designed based on an adaptive neuro-fuzzy inference system (ANFIS). The structure of the Critic network is illustrated in Fig. 4.

The input of the Critic network is the error function which is defined as the following:

$$E(k) = \|e(k)\| \tag{19}$$

where $e(k)$ is the error between the outputs of the system and the fuzzy estimator.

The output $V(k)$ of the Critic network is used to approximate the value:

$$V(k-1) = r(k) + \alpha V(k) \tag{20}$$

where $\alpha \in (0, 1)$ is the discount factor, $r(k)$ is the reward at the iteration k. The reward function is defined by the designer and depends on the specific problem. In this work, the rework function is defined as the following:

$$r(k) = \begin{cases} 0 & if\ E(k) \le \varepsilon \\ 1 & \text{others} \end{cases} \tag{21}$$

in which ε is the small enough scalar.

The output of the Critic network is the approximated value function which is determined as follows:

$$V(k) = \sum_{j=1}^{N} m^j(k) V^j(k) \tag{22}$$

where $m^j(k) = l^j(k) \Big/ \sum_{j-1}^{N} l^j(k)$ in which $l^j(k)$ $(j = 1, 2, \ldots, N)$ is the membership function which is chosen as bell-shaped function:

$$l^j(k) = \exp\left(-\left(\frac{E(k) - p^j}{q^j} \right)^2 \right). \tag{23}$$

The value function of each rule is approximated as:

$$V^j(k) = t^j(k) + s^j(k) E(k) \tag{24}$$

In equations (23) and (24), $\{p^j, q^j\}$ and $\{t^j, s^j\}$ are ANFIS parameter sets that will be updated online during the training process by using the Hybrid Learning Rule (HLB) in two phases:

- In the forward phase: consequent parameters $\{t^j, s^j\}$ are updated based on the Least Squares method [29]. The updating formula has the following form:

$$\begin{cases} \zeta(k+1) = \zeta(k) \\ \quad + \nu(k+1)d(k+1)\left(g^T(k+1) - d^T(k+1)\zeta(k)\right) \\ \nu(k+1) = \frac{1}{\lambda}\left[\nu(k) - \frac{\nu(k)d(k+1)d^T(k+1)\nu(k)}{\lambda + d^T(k+1)\nu(k)d(k+1)}\right] \end{cases} \tag{25}$$

where $\zeta(k) = [t^j(k) \quad s^j(k)]^T$, $g(k)$ is desired output vector, $\lambda \in (0, 1)$ is a scalar, and $d(k) = [m^j(k) \quad m^j(k)E(k)]^T$.

- In the backward phase: premise parameters $\rho^j = \{p^j, q^j\}$ are updated by the following rule:

$$\rho^j(k+1) = \rho^j(k) + \eta_1 \frac{v_1 e_c(k)}{2v_2} \frac{\partial e_c(k)}{\partial \rho^j(k)} \tag{26}$$

where v_1 and v_2 are positive scalars, η_1 is the learning rate, and $e_c(k)$ is the prediction error of the critic network. The $e_c(k)$ is defined as follows:

$$e_c(k) = V(k-1) - r(k) - \alpha V(k). \tag{27}$$

Remark 1: In comparison with the neural network-based adaptive critic, the proposed ANFIS-based critic significantly reduces the computation and storage. Thoroughly, in [16], a feedforward neural network with a hidden layer of 20 neurons is addressed to deal with the nonlinear optimal problems. More specifically, this neural network has 40 scalars that need updating: 20 parameters for the weight of the output layer and 20 parameters for the weight of the hidden layer. Meanwhile, to build the ANFIS network, we only need to update 12 parameters: 6 premise parameters $\{p_m, q_m\}, m = 1, 2, 3$ in the first layer and 6 consequent parameters $\{t_m, s_m\}$, $m = 1, 2, 3$ in the fourth layer. That is why the proposed ANFIS network reduces the computational cost.

C. CONTROLLER AND ESTIMATOR PARAMETERS UPDATE

The parameters of two T-S fuzzy systems are updated based on the principle that minimizes the error function:

$$E_a = \frac{1}{2}e_a^2 \tag{28}$$

where e_a is the error between the output of approximation and its desired value.

$$e_a(k) = V(k) - V^*(k) \tag{29}$$

In this work, the desired value $V^*(k) = 0$.

The update law of the system parameter set is given by the following:

$$\delta^j(k+1) = \delta^j(k) + \eta_2 \frac{v_1 e_a(k)}{2v_2} \frac{\partial e_a(k)}{\partial \delta^j(k)} \tag{30}$$

The stability of the system as well as the convergence of the updated parameters are provided in the next section.

Remark 2: The action component (i.e., the controller) of the algorithm which is based on reinforcement learning is obtained by solving the HJB equation [15] so the system model is partially known. Thoroughly, with the given system $\dot{x} = f(x) + g(x)u$, the adaptive optimal controller has the formula $u = -\frac{1}{2}R^{-1}g(x)\frac{\partial V}{\partial x}$ where R is the positive definite matrix, and V is the value function which is approximated by the critic network. This means that the $g(x)$ component should be known. In our work, the dynamic model is considered in the general nonaffine form as (11), the controller is addressed by the T-S fuzzy controller (14) and the parameters are optimized by solving (28); therefore, the dynamic model is not required in this work.

D. STABILITY ANALYSIS

Define general variables $\delta_g = \{a^j, b^j, c^j, \sigma^j, p^j, q^j\}$, $e_g = \{e_c, e_a\}$, $\eta_g = \{\eta_1, \eta_2\}$. Using this definition, the update laws (26) and (30) are rewritten in the following general form:

$$\delta_g(k+1) = \delta_g(k) + \eta_g \frac{v_1 e_g(k)}{2v_2} \frac{\partial e_g(k)}{\partial \delta_g(k)}. \tag{31}$$

Choose the Lyapunov function as follows:

$$L(k) = \frac{v_1}{2}(e_g(k))^2 + \frac{v_2}{2}\left\|\Delta\delta_g(k)\right\|^2. \tag{32}$$

The time derivative of the Lyapunov function has the following formula:

$$\Delta L(k) = \frac{v_1}{2}\Delta(e_g(k))^2 + \frac{v_2}{2}\Delta\left\|\Delta\delta_g(k)\right\|^2. \tag{33}$$

Each component in equation (33) is analyzed as the following

$$\frac{v_1}{2}\Delta(e_g(k))^2 = \frac{v_1}{2}(e_g(k+1))^2 - \frac{v_1}{2}(e_g(k))^2. \tag{34}$$

Using Taylor series expansion [30] for the first term of (34), the following is obtained:

$$\frac{v_1}{2}(e_g(k+1))^2 = \frac{v_1}{2}(e_g(k))^2 + \frac{\partial\left(\frac{v_1}{2}(e_g(k))^2\right)}{\partial\delta_g(k)}\Delta\delta_g(k) + HOC \tag{35}$$

where the symbol "HOC" represents the higher-order components which can be omitted.

Also, applying Taylor series expansion for $e_g(k+1)$, the following result is obtained:

$$e_g(k+1) = e_g(k) + \frac{\partial e_g(k)}{\partial \delta_g(k)}\Delta\delta_g(k) \tag{36}$$

or

$$\frac{\partial e_g(k)}{\partial \delta_g(k)}\Delta\delta_g(k) = e_g(k+1) - e_g(k) = \Delta e_g(k). \tag{37}$$

Replace (35) and (37) into (34) yields:

$$\frac{v_1}{2}\Delta(e_g(k))^2 = v_1 e_c(k)\Delta e_c(k). \tag{38}$$

Similarly, the second component of (33) is shortened as the following:

$$\frac{v_2}{2}\Delta\left(\|\Delta\rho(k)\|^2\right) = v_2\|\Delta\rho(k)\|^2. \tag{39}$$

Substituting (38) and (39) into (33), the derivative of the Lyapunov function is obtained as:

$$\Delta L(k) = v_1 e_g(k)\Delta e_g(k) + v_2 \left\|\Delta\delta_g(k)\right\|^2. \tag{40}$$

Equation (40) can be rewritten as:

$$v_2 \left\|\Delta\delta_g(k)\right\|^2 + v_1 e_c(k)\frac{\Delta e_c(k)}{\left\|\Delta\delta_g(k)\right\|}\left\|\Delta\delta_g(k)\right\| - \Delta L(k) = 0. \tag{41}$$

Equation (41) is the second-order equation of $\left\|\Delta\delta_g(k)\right\|$. This equation has a unique solution if the following is satisfied:

$$v_1^2\left(e_g(k)\frac{\Delta e_g(k)}{\|\Delta\rho(k)\|}\right)^2 + 4v_2\Delta L(k) = 0. \tag{42}$$

Equation (42) leads to the following result:

$$\Delta L(k) = -\frac{v_1^2}{4v_2}\left(e_g(k)\frac{\Delta e_g(k)}{\|\Delta\rho(k)\|}\right)^2 \le 0. \tag{43}$$

Inequation (43) implies that the closed system is stable, and the estimation errors converge to zero according to the Lyapunov theorem.

IV. SIMULATION RESULT AND DISCUSSION

In this section, the correctness and feasibility of the proposed adaptive optimal control scheme are verified through Matlab/Simulink. The parameters of the wind turbine and PMSG are given in Table 1 [3].

TABLE 1. PMSG and wind turbine nominal parameters.

Parameters	Value
Rated Power (P_{rated})	5 kW
Stator Resistance (R_s)	0.3676 Ω
Stator Inductance (L_s)	3.55 mH
Magnet Flux Linkage (ψ_m)	0.2867 Wb
Inertia (J)	7.856 kg.m^2
Viscous Friction Coefficient (B)	0.002 kg.m^2/s
Number of Pole Pairs (P)	14
Rotor Radius of Blades (R)	1.84 m
Air Destiny (ρ)	1.25 kg/m^3
Optimal tip-speed ration λ_{opt}	8.1

The validation of the proposed control algorithm is executed under various conditions such as step wind speed, random wind speed, nominal system parameters, and uncertain system parameters. Also, in each condition, the performances of the proposed ANFIS-based reinforcement learning adaptive optimal fuzzy control scheme (ANFIS-RL) are compared with PI controller and neural network-based reinforcement learning adaptive optimal fuzzy control scheme (NN-RL).

The configuration of the neural network for the neural network-based reinforcement learning control scheme is as the following:

- The number of layers: three layers (input layer, hidden layer, output layer).

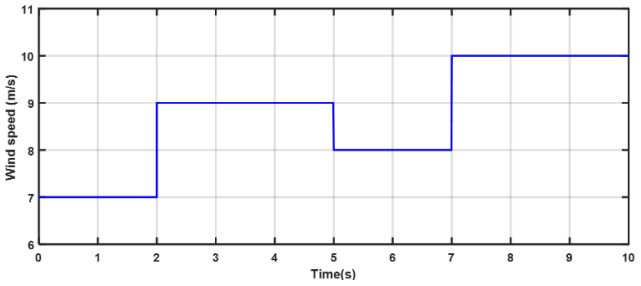

FIGURE 5. Step wind speed.

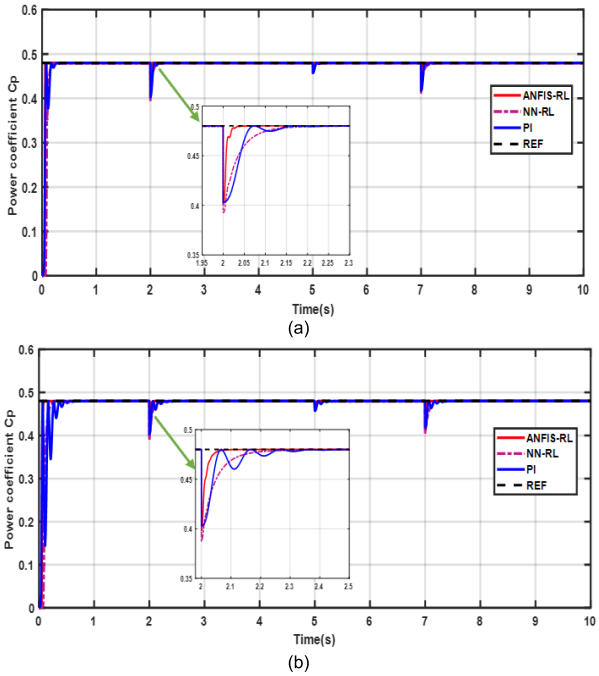

FIGURE 6. Comparative results of power coefficient with a random wind speed of three controllers. (a) Nominal parameters. (b) Variation parameters (i.e., +50% variations of R_s, L_s, B, J, ψ_m).

- The number of neurons: 1 for the input layer, 20 for the hidden layer, and 1 for the output layer.
- The active function: *tansig* function.

A. STEP WIND SPEED

The profile of wind speed is given in Fig. 5 whereas the wind speed is steeply changed in the range of 7−10m/s. The simulation results for the step wind speed condition are presented in Figs. 6−8. In each figure, the response of the proposed ANFIS-based reinforcement learning (ANFIS-RL) control scheme, NN-based reinforcement learning (NN-RL) control scheme, and PI control scheme are compared in both nominal parameter and varying parameters (i.e., +50% variations of R_s, L_s, B, J, ψ_m) conditions.

In the case of nominal parameters, the steady-state performances of the three controllers are almost the same for power coefficient response, mechanical response, and rotor speed response. Thoroughly, in Fig. 6(a), the power coefficient response of the three control schemes has the same value as the ideal power coefficient (0.49) at a steady- state. Also, in Figs.7(a) and 8(a), the difference in response of the three

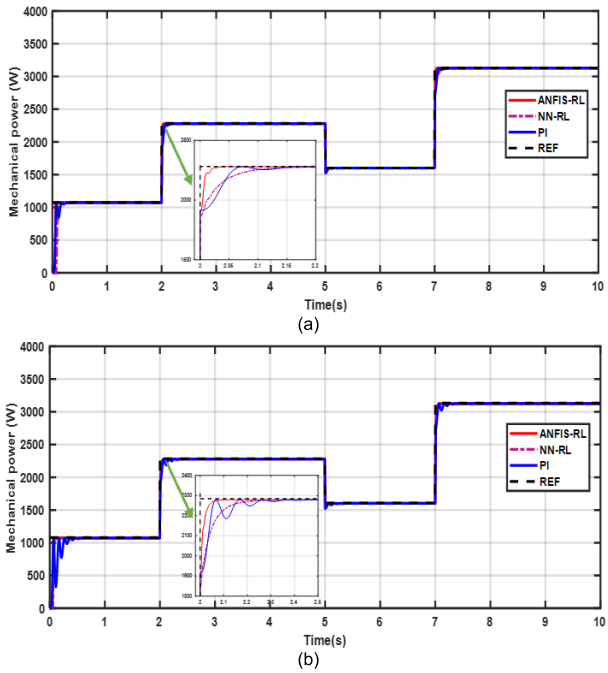

FIGURE 7. Comparative results of mechanical power with a step wind speed of three controllers. (a) Nominal parameters. (b) Variation parameters (i.e., +50% variations of R_s, L_s, B, J, ψ_m).

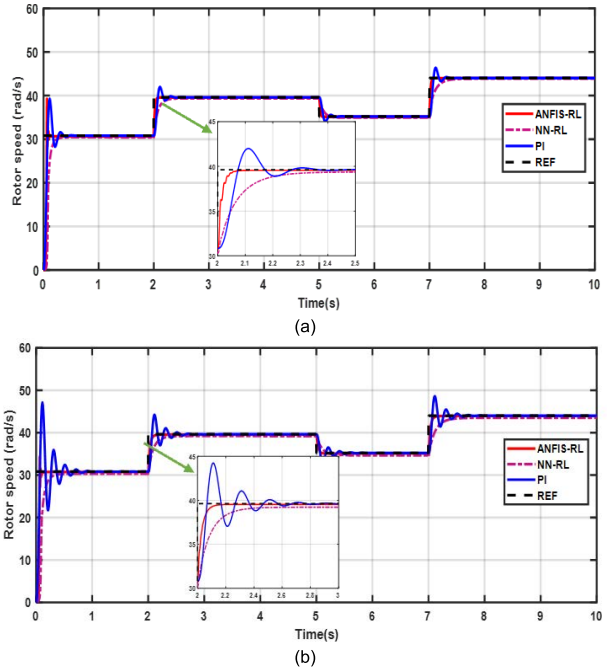

FIGURE 8. Comparative results of rotor speed with a random wind speed of three controllers. (a) Nominal parameters. (b) Variation parameters (i.e., +50% variations of R_s, L_s, B, J, ψ_m).

controllers is not countable even though the wind speed changes. However, the transient responses of the proposed ANFIS-based reinforcement controller, the NN- based reinforcement controller, and the PI controller are much different. In Figs. 6(a) and 7(a), the transient time of the presented ANFIS-RL controller is about 0.02s, meanwhile, the transient response of the NN-RL controller and PI controller is about 0.1s. Moreover, the NN-RL controller response has

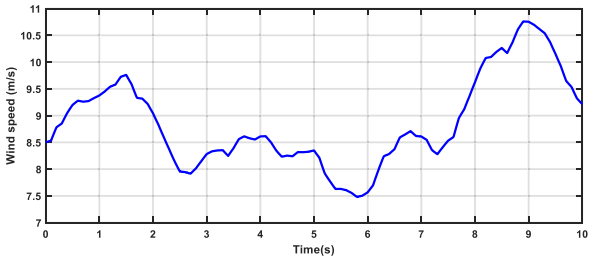

FIGURE 9. Random wind speed.

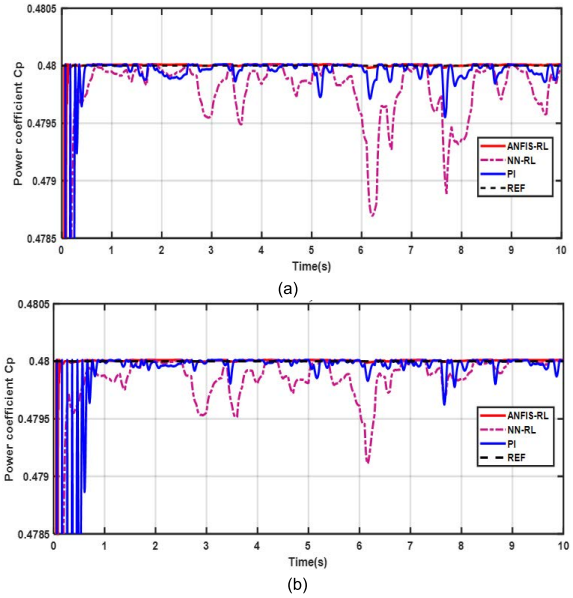

FIGURE 10. Comparative results of power coefficient with a random wind speed of three controllers. (a) Nominal parameters. (b) Variation parameters (i.e., +50% variations of R_s, L_s, B, J, ψ_m).

the highest undershoot (about 22.5% for ANFIS-RL and PI controllers, about 25% for NN-RL controller). In Fig. 8(a), the transient time of the ANFIS-RL control scheme is 0.05s and the response has no overshoot, however, the response time of the NN-RL and PI control schemes is much slower (0.35s) and the response of the PI controller has overshoot of about 7.5%.

For the system parameters change case, the simulation results are depicted in Figs. 6(b)−8(b). Similar to the case of nominal parameters, the introduced ANFIS-RL control algorithm has the best responses in both transient and steady states with the fastest response time (about 0.05s), no overshoot, and zero steady-state error. The response of the NN-RL controller is still good with no overshoot, and zero steady-state error, but the transient time is longer (0.2s for power coefficient and mechanical power responses, 0.4s for rotor speed response). The response of the PI controller is worst when the system parameters increase by 50%, i.e., the transient responses of all variables are fluctuation, high overshoot (about 10% in Fig. 6(b), and long transient time (0.35s in Figs. 7(b) and 0.6s in Fig. 8(b).

B. RANDOM WIND SPEED

The profile of random wind speed tested in this work is shown in Fig. 9, the dynamic responses of the WECS

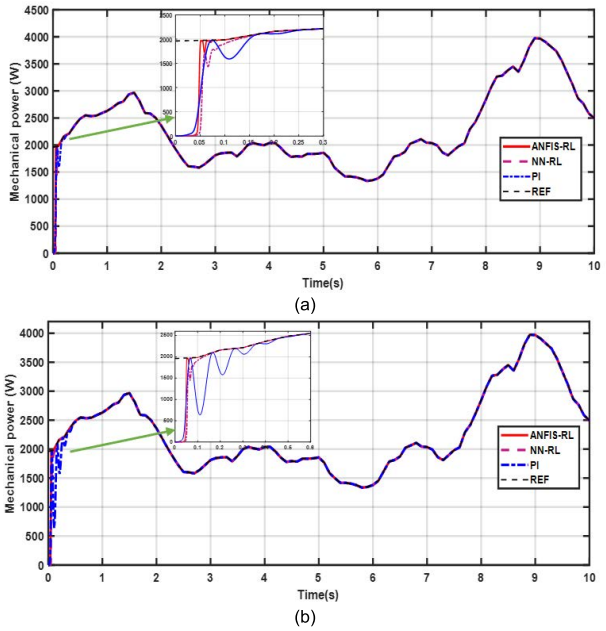

FIGURE 11. Comparative results of mechanical power with a random wind speed of three controllers. (a) Nominal parameters. (b) Variation parameters (i.e., +50% variations of R_s, L_s, B, J, ψ_m).

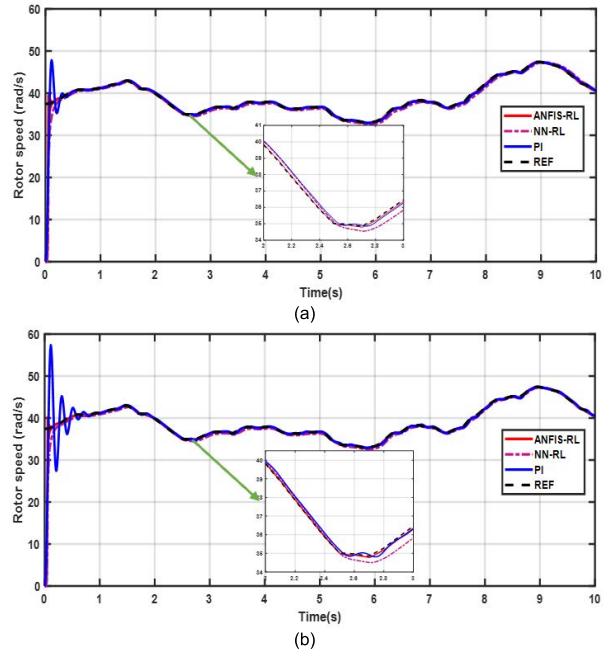

FIGURE 12. Comparative results of rotor speed with a random wind speed of three controllers. (a) Nominal parameters. (b) Variation parameters (i.e., +50% variations of R_s, L_s, B, J, ψ_m).

with different controllers are presented in Figs. 10−12. Also, the dynamic responses of each controller under conditions of 50% increased system parameters are illustrated in each figure.

As shown in Figs. 10−12, when the wind speed changes randomly, the proposed ANFIS-RL controller still remains the best dynamic response for both nominal system parameters and varying system parameters, i.e., fast and no overshoot transient response and almost zero steady-state error. For the NN-RL controller, the dynamic responses are also good, but the response time is longer and the steady-state error is larger than the responses of the proposed ANFIS-RL controller. In the case of the PI controller, the dynamic responses are good at steady-state with almost zero error. However, the overshoot and transient times are high, and these responses become worse when the system parameters increase by 50%.

V. CONCLUSION

In this paper, the reinforcement learning-based adaptive optimal fuzzy control method is proposed for MPPT control of a direct drive PMSG-based WECS. The control system consists of three parts: the critic, the adaptive fuzzy controller, and the adaptive fuzzy estimator. The critic is designed based on the ANFIS technique with a hybrid update rule to reduce both the computational burden and storage capacity in comparison with the NN-based critic. The adaptive fuzzy controller and the adaptive fuzzy estimator are built using the T-S fuzzy system. The parameters of both the controller and estimator are updated from the critic with the optimal rule so that the input error function is minimized. The superiority of the proposed control algorithm is the employing of only output feedback, so the knowledge about the system dynamic model as well as system parameters is omitted. The stability of the control system and the convergence of updated parameters are proven via the Lyapunov stable theory. The efficiency of the proposed control scheme is evaluated via simulation with various scenarios: step wind speed, random wind speed, nominal system parameters, and varying system parameters. In each scenario, the dynamic responses of the proposed control system are compared with the ones of the NN-RL adaptive optimal fuzzy control system and the PI control system. Finally, the simulation results show that the proposed ANFIS-RL adaptive optimal fuzzy control scheme guarantees the best performances in all cases with fast response, no overshoot, zero steady-state error, and robustness against the system parameter variations. In the future, the value iteration can be employed to find the optimal value function $V(k)$ instead of approximation. In this trend, the convergence conditions and convergence speed are still issues that need to be investigated.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] Z. Yang and Y. Chai, "A survey of fault diagnosis for onshore grid-connected converter in wind energy conversion systems," *Renew. Sustain. Energy Rev.*, vol. 66, pp. 345–359, Dec. 2016.

[2] N. A. Orlando, M. Liserre, R. A. Mastromauro, and A. Dell'Aquila, "A survey of control issues in PMSG-based small wind-turbine systems," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1211–1221, Jul. 2013.

[3] D. Zholtayev, M. Rubagotti, and T. D. Do, "Adaptive super-twisting sliding mode control for maximum power point tracking of PMSG-based wind energy conversion systems," *Renew. Energy*, vol. 183, pp. 877–889, Jan. 2022.

[4] E. Cai, Y. Yan, L. Dong, and X. Liao, "A control scheme with the variable-speed pitch system for wind turbines during a zero-voltage ride through," *Energies*, vol. 13, no. 13, pp. 1–22, Jun. 2020.

[5] R. You, X. Yuan, and X. Li, "A multi-rotor medium-voltage wind turbine system and its control strategy," *Renew. Energy*, vol. 24, pp. 366–377, Aug. 2022.
[6] M. Karabacak, "A new perturb and observe based higher order sliding mode MPPT control of wind turbines eliminating the rotor inertial effect," *Renew. Energy*, vol. 133, pp. 807–827, Apr. 2019.
[7] E. H. Dursun and A. A. Kulaksiz, "Second-order sliding mode voltage-regulator for improving MPPT efficiency of PMSG-based WECS," *Int. J. Electr. Power Energy Syst.*, vol. 121, May 2021, 106149.
[8] L. Pan and C. Shao, "Wind energy conversion systems analysis of PMSG on offshore wind turbine using improved SMC and extended state observer," *Renew. Energy*, vol. 161, pp. 149–161, Dec. 2020.
[9] S. Huang, J. Wang, C. Huang, L. Zhou, L. Xiong, J. Liu, and P. Li, "A fixed-time fractional-order sliding mode control strategy for power quality enhancement of PMSG wind turbine," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107354.
[10] J. Chen, W. Yao, C.-K. Zhang, Y. Rend, and L. Jiang, "Design of robust MPPT controller for grid-connected PMSG-based wind turbine via perturbation observation based nonlinear adaptive control," *Renew. Energy*, vol. 134, pp. 478–495, Apr. 2019.
[11] B. Yang, T. Yu, H. Shu, Y. Han, P. Cao, and L. Jiang, "Adaptive fractional-order PID control of PMSG-based wind energy conversion system for MPPT using linear observers," *Int. Trans. Electr. Energy Syst.*, vol. 29, no. 1, p. e2697, Jan. 2019.
[12] M. E. Zarei, D. Ramirez, M. Prodanovic, and G. M. Arana, "Model predictive control for PMSG-based wind turbines with overmodulation and adjustable dynamic response time," *IEEE Trans. Ind. Electron.*, vol. 69, no. 2, pp. 1573–1582, Jan. 2022.
[13] X. Bu, Y. Xiao, and H. Lei, "An adaptive critic design-based fuzzy neural controller for hypersonic vehicles: Predefined behavioral nonaffine control," *IEEE/ASME Trans. Mech.*, vol. 24, pp. 1871–1881, Aug. 2019.
[14] X. Bu, Q. Qi, and B. Jiang, "A simplified finite-time fuzzy neural controller with prescribed performance applied to waverider aircraft," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 7, pp. 2529–2537, Jul. 2022.
[15] X. Bu and Q. Qi, "Fuzzy optimal tracking control of hypersonic flight vehicles via single-network adaptive critic design," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 1, pp. 270–278, Jan. 2022.
[16] A. A. Khater, A. M. El-Nagar, M. El-Bardini, and N. M. El-Rabaie, "Adaptive T-S fuzzy controller using reinforcement learning based on Lyapunov stability," *J. Franklin Inst.*, vol. 355, no. 14, pp. 6390–6415, 2018.
[17] R. Sitharthan, M. Karthikeyan, D. S. Sundar, and S. Rajasekaran, "Adaptive hybrid intelligent MPPT controller to approximate effectual wind speed and optimal rotor speed of variable speed wind turbine," *ISA Trans.*, vol. 96, pp. 479–489, Jan. 2020.
[18] J.-L. Fernando, K. Godpromesse, and L.-L. Francoise, "A novel online training neural network-based algorithm for wind speed estimation and adaptive control of PMSG wind turbine system for maximum power extraction," *Renew. Energy*, vol. 86, pp. 38–48, Feb. 2016.
[19] C. Wei, Z. Zhang, W. Qiao, and L. Qu, "An adaptive network-based reinforcement learning method for MPPT control of PMSG wind energy conversion systems," *IEEE Trans. Power Electron.*, vol. 31, no. 11, pp. 7837–7848, Nov. 2016.
[20] M. Bašić, M. Bubalo, D. Vukadinović, and I. Grgić, "Sensorless maximum power control of a stand-alone squirrel-cage induction generator driven by a variable-speed wind turbine," *J. Elect. Eng. Technol.*, vol. 16, pp. 333–347, Jan. 2021.
[21] N. Bounar, S. Labdai, A. Boulkroune, M. Farza, and M. M'Saad, "Adaptive fuzzy control scheme for variable-speed wind turbines based on a doubly-fed induction generator," *Iranian Jour. Sci. Tech., Trans. Electr. Eng.*, vol. 44, pp. 629–641, Jun. 2020.
[22] I. U. Haq, Q. Khan, I. Khan, R. Akmeliawati, K. S. Nisar, and I. Khan, "Maximum power extraction strategy for variable speed wind turbine system via neuro-adaptive generalized global sliding mode controller," *IEEE Access*, vol. 8, pp. 128536–128547, 2020.
[23] M. M. Mahmoud, M. M. Aly, H. S. Salama, and A.-M.-M. Abdel-Rahim, "A combination of an OTC based MPPT and fuzzy logic current control for a wind-driven PMSG under variability of wind speed," *Energy Syst.*, Aug. 2021.
[24] A. A. Salem, N. A. N. Aldin, A. M. Azmy, and W. S. E. Abdellatif, "Implementation and validation of an adaptive fuzzy logic controller for MPPT of PMSG-based wind turbines," *IEEE Access*, vol. 9, pp. 165690–165707, 2021.
[25] P. Chen, D. Han, and K.-C. Li, "Robust adaptive control of maximum power point tracking for wind power system," *IEEE Access*, vol. 8, pp. 214538–214550, 2020.
[26] R. You and X. Lu, "Voltage unbalance compensation in distribution feeders using soft open points," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 4, pp. 1000–1008, 2022.
[27] S. Li, T. A. Haskew, R. P. Swatloski, and W. Gathings, "Optimal and direct-current vector control of direct-driven PMSG wind turbines," *IEEE Trans. Power Electron.*, vol. 27, no. 5, pp. 2335–2337, May 2012.
[28] R. Song and F. L. Lewis, "Robust optimal control for a class of nonlinear systems with unknown disturbances based on disturbance observer and policy iteration," *Neurocomputing*, vol. 390, pp. 185–195, Jan. 2020.
[29] D. Wang and F. Ding, "Least squares based and gradient based iterative identification for Wiener nonlinear systems," *Signal Process.*, vol. 91, no. 5, pp. 1182–1189, May 2011.
[30] R. Kumar, S. Srivastava, J. R. P. Gupta, and A. Mohindru, "Diagonal recurrent neural network based identification of nonlinear dynamical systems with Lyapunov stability based adaptive learning rates," *Neurocomputing*, vol. 287, pp. 102–117, Apr. 2018.

NGA THI-THUY VU received the Ph.D. degree in engineering from Donguk University, South Korea, in 1996, and the B.Eng. and M.Sc. degrees in electrical engineering from the Electrical Engineering School, Hanoi University of Science and Technology, Vietnam, in 2005 and 2009, respectively. She is currently an Associate Professor at the Hanoi University of Science and Technology, where she is also lecturing undergraduates and postgraduates in control theory and applications at the School of Electrical and Electronics. Her research interests include driven control, power electronics control, mechatronics system control, and renewable energy systems.

HA DUC NGUYEN is currently a Student at the School of Electrical and Electronics, Hanoi University of Science and Technology, Vietnam. His research interests include advanced control, intelligent control, and reinforcement learning control.

ANH TUAN NGUYEN received the B.S. and M.S. degrees in electrical engineering from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2010 and 2012, respectively, the M.S. degree in electrical engineering from the Grenoble Institute of Technology, Joseph Fourier University, Grenoble, France, in 2015, and the Ph.D. degree from the Division of Electronics and Electrical Engineering, Dongguk University, Seoul, South Korea, in 2021.

Since October 2021, he has been a Lecturer and a Researcher at Phenikaa University, Hanoi. His research interests include electric power systems, control of power converters, and DSP-based electric machine drives.