IJE TRANSACTIONS C: Aspects Vol. 37 No. 06, (June 2024) 1085-1097

International Journal of Engineering

Journal Homepage: www.ije.ir

# Deep Reinforcement Learning with Immersion- and Invariance-based State Observer Control of Wave Energy Converters

H. Mohammadian KhalafAnsara, J. Keighobadi*

*Faculty of Mechanical Engineering, University of Tabriz, East Azerbaijan, Iran*

*P A P E R I N F O*

*Paper history:*
Received 10 November 2017
Received in revised form 23 December 2017
Accepted 04 Januray 2018

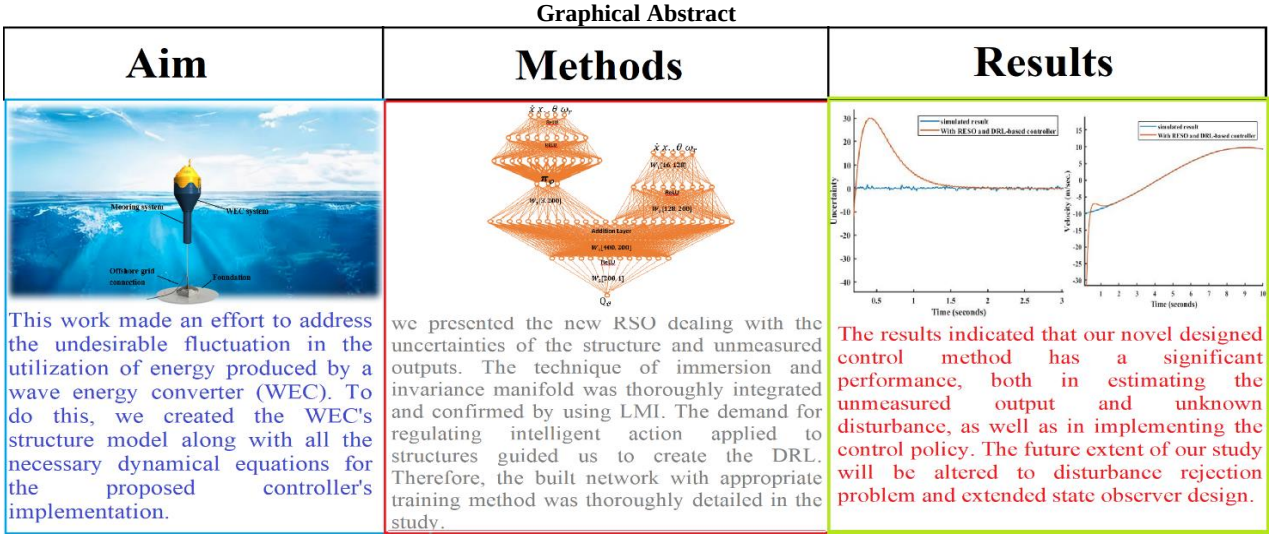
*Keywords*:
*Wave Energy Converter*
*Extended State Observer*
*Immersion- and Invariance-based Control*
*Deep Reinforcement Learning*
*Uniform Energy*

*A B S T R A C T*

Composable life under the extensive global warming of the Earth encourages the progress of renewable energy devices and the adoption of new technologies, such as artificial intelligence. Regarding enormous potential of wave energy and its consistency, wave energy converter (WEC) plays vital role in uniform energy harvesting field. In this paper, the significant environmental changes in the ocean prompt us to propose an intelligent feedback control system to mitigate the impact of disturbances and variable wind effects on the efficacy of WECs. Deep reinforcement learning (DRL), as a powerful machine intelligence technique, is capable of identifying WECs as black-box models. Therefore, based on the DRL model, the disturbance and unmeasured state variables are simultaneously estimated in the extended state observer section. Leakage in identification data and real-time application requirements of limited number of layers in the deep neural networks are compensated by implementation of immersion and invariance-based extended state observer which improves coping with the unwanted exogenous noises as well. In the overall intelligent control system, the estimated parameters are inputted into the DRL as the actor-critic networks. The initial actor network is responsible for predicting the control action, while the subsequent critic network determines the decision criterion for evaluating the accuracy of the actor's estimated amount. Next, the output value of the critic stage is backpropagated through the layers to update the network weights. The simulation test results in MATLAB indicate the convergence of unmeasured parameters/states to the corresponding true values and the significance of newly designed intelligent DRL method.

***doi**: 10.5829/ije.2024.37.06c.05*

**Graphical Abstract**




*Corresponding Author Email: keighobadi@tabrizu.ac.ir (J. Keighobadi)

Please cite this article as: Mohammadian KhalafAnsara H, Keighobadi J. Deep Reinforcement Learning with Immersion- and Invariance-based State Observer Control of Wave Energy Converters. International Journal of Engineering, Transactions C: Aspects. 2024;37(06):1085-97.

**NOMENCLATURE**

| | | | |
|---|---|---|---|
| WEC | Wave energy converter | ESO | Extended state observer |
| DRL | Deep reinforcement learning | RSO | reduced state observer |
| BEM | Boundary element method | $m_s$ | mass of the buoy(kg) |
| MPC | Model predictive control | $z_w$ | water level(m) |
| DOF | Degree of freedom | $f_s$ | lateral restoring force(N) |
| ADRC | Adaptive disturbance rejection control | $\rho$ | water's density(kg/m$^3$) |
| $g$ | gravity acceleration(m/s$^2$) | $f_r$ | radiation force(N) |
| $A$ | cross-section of the float(m$^2$) | $m_\infty$ | extra mass(kg) |
| $z_v$ | heave motion of the float(m) | $h_r$ | core of the radiation force |
| $w$ | unknown disturbance | PTO | power take-off |

## 1. INTRODUCTION

Global warming, a major consequence of the increasing use of fossil fuels, has led to a growing tendency among people to seek benefits from renewable energy sources. Accordingly, offshore energy converter devices such as wind turbines and WECs are developed to generate consistent energy for everyday needs. In this paper, we propose an artificial intelligence algorithm in the form of a closed-loop control system for the WEC plant to mitigate obstacles to achieving uniform clean energy production. As a means of reducing test platform costs, it is recommended to employ dynamic modeling of the WEC and simulate it along with the proposed control algorithms. The boundary element method (BEM) employs a linear potential theory, assuming small displacements of the structure relative to the wavelength in the presence of non-rotational flows. Hence, the performance of BEM solutions in stormy seas is restricted (1). The theoretical maximum energy has been determined to be roughly 3.1013 kWh/year, which is equivalent to almost 20% of the total-energy consumption in 2019. But the usable resource is roughly ten times smaller due to technical and budgetary restrictions. WECs convert wave energy into electrical current. While there have been efforts to generalize wave power as a regularly used source since 1890, it is not currently being widely utilized. Around 16 MW of operational wave power were globally generated in 2020, which is approximately five times less than about 2 TW, required to fully reach of the world's wave energy potential. One important aspect is the production costs per kWh, which in 2020 were approximately 10 times higher than those of offshore wind projects.

In the last three decades, various ways for conversion of wave power have been created, resulting in hundreds of patents in recent years. Diversity of wave energy ideas are now being explored by industry and academic research organizations all across the globe. Despite the fact that various operational designs have been built and validated, only a few concepts have incorporated modeling and wave tank testing into the sea.

Wave energy is still a crucial source of clean and renewable energy, even though it won't likely ever correspond with the global power output of wind and solar energy. Wave energy has the ability to deliver the electricity in hard-to-reach locations like coastal communities and remote islands that today depend on expensive, carbon-intensive diesel imports as it is more predictable and stable than solar / wind energy. Also, military tasks that need access to deeper seas, offshore fishing, and marine research may all be powered by wave energy devices. Waves in the US provide nearly 80% of the energy required by the country. The industry may access part of that energy, albeit not all of it, to make it simpler for the country to move to 100% clean energy. WECs, also known as point absorbers, are buoy-style devices that collect wave energy from all directions. They are positioned at or near the surface of the ocean. Wave energy is captured by a vertically submerged buoy and then converted into power by a piston or linear generatorAmong numerous research studies in the field of control for WECs, Nielsen et al. proposed a model predictive control (MPC) approach to demonstrate its superiority in terms of collecting up to 25% more energy compared to conventional reactive controllers. These reactive controllers take into account the power losses during the conversion from mechanical to electrical energy (2). Also, the authors showed the hydrodynamic and economic performance of an oscillating wave surge energy converter is significantly influenced by the installation depth and height of the incident wave (3). In (4), the authors introduced a novel configuration for a linear permanent magnet Vernier machine, specifically tailored for harnessing wave energy and enhancing the operational efficiency of the existing Vernier machine prototype. Bayani et al. presented an overview of an offshore point-absorber WEC developed by the Hydrodynamics, Acoustics and Marine Propulsion Group at Babol Noshirvani University of Technology (5). Alizadeh Kharkeshi et al. assessed the effectiveness of dimensionless coefficients in multi-reservoirs within a hydrodynamic oscillating column converter for sea water waves. The experiment, which took place in Mazandaran, involved extracting

parameters that influence the system's behavior in order to conduct a hydrodynamic analysis of the oscillating water column (6). Son and Yeung conducted experimental tests to validate the nonlinear MPC using a solid-state electrical relay with pulse-width modulation to simulate analog current flow (7). Korde explored the sub-optimal time-domain responsive control of WECs and contrasted the increased reactive energy consumption for the submerged buoy (8). Anderlini et al. gave the wave altitude, wave energy span, and the power take-off damping as an input data to the artificial neural network to assess the lack of confidence in linear model (9). Latching control for heave motion of wave converter was developed to extremize the motion's measurement irrespective of the wave's frequency (10). Wu et al. devised the simultaneous latching control for the single Duck WEC in unbalanced waves to increase the act of structure in sea conditions distant from the normal frequency (11). Using MPC, Faedo et al. examined several methods of implementing a latching control system in the context of WEC (12). Upper limit value for wave excitation force was determined using Kalman filter to monitor energy absorption with 90% of the ideal constrained optimal collected energy (13). Using limited MPC, O'Sullivan and Lightbody produced the large quantity of absorbed electrical power from the point absorber WEC working in heave mode independently, linked to a linear permanent magnet generator (14). Also, Li et al. developed the multilayer perceptron using the deep neural network's techniques to anticipate the temporary wave forces and transferred the output of network via MPC strategy to conduct online latching control action to a point absorber WEC (15). For maximum power extraction, a robust MPC using Laguerre polynomials was developed to alleviate the computational burden (16). A heave-pitch-surge 3-DOF WEC, with the pitch-surge optimization separated from heave, increased the gathered energy by more than three times compared to the captured energy from the lurch-only WEC (17). Investigation of blocked optimum control on a small asymmetrical float indicated that, despite the improvement in power capture achievable with the current control, the actuation forces are significant (18). Moreover, Burgaç and Yavuz implemented the discrete Fourier transform technique to estimate the dominant wave frequency in the fuzzy controller for defining the power take-off control settings (19). Zhan et al. (20) proposed a categorized adaptive optimum control framework for WECs to enhance energy conversion efficiency and reduce the modeling effort required for control design.

The newly built controller in this study was not adequately evaluated in the field of WEC and requires further investigation. The following summarizes the most crucial related works to our newly proposed controller. In (21), Han J introduced an active disturbance rejection control that combines an error-driven control rule, state observer feature, and the power of a nonlinear controller. Zhejiang Gao proved that by unifying the controller and addressing the issue of disturbance rejection, the performance of traditional controllers when faced with disturbances is improved. The proposal of the paper was to improve the developed controller for the official model, with the aim of rejecting disturbances in the plant model (22). Zhao and Guo introduced the novel ADRC, which has the capacity to track reference signals, reject disturbances, and maintain closed-loop stability for a group of single-input single-output systems (23). Also, Feng and Guo have researched the output feedback stability for indefinite structures described by partial differential equations (24). Hosseini and Keighobadi developed an extended state observer-based robust active control to approximate both the speed and perturbation trajectories of the gyro's dynamics using the location signs (25). In (26), Guo and Zhao sought to demonstrate the stability issue of an extended state observer by utilizing the error equation. They aimed to decrease the influence of the disturbance by implementing a high-gain approach. Li et al. (27) presented the necessary conditions for the convergence of the quantized nonlinear extensible state observer using linear matrix inequality. Zhao and Guo developed a temporally varying gain ESO to reduce the peaking effect near the main time caused by the constant high gain technique (28).

In this research work, to obtain a nearly compatible model with real plant, the development of the overall dynamics of a WEC with detailed consideration of acting forces on the structure is explained. Next, the paper proposes a new estimation technique for the disturbance and unmeasured parameters of WEC, specifically the immersion and invariance-based extended state observer. Through mapping of considered observer manifold, the unmeasured variables are estimated and are fed as input to the intelligent controller presumed DRL, which consists of two networks named actor-critic. The output of the actor network is the input of the main plant. The Q_value, which represents the output of the critic network, is used to update the weights and biases of the networks using the backpropagation method. To prove the convergence of the control system, the direct stability method of the Lyapunov is assessed. The results of the controller are evaluated for the reduced state observer (RSO) as well. The RSO focuses on estimating unmeasured states and accounting for exogenous inputs. The simulation is conducted to design a linear observer for these unmeasured variables. The wide range simulation results are assessed with the noisy data to guarantee the strength and robustness of the suggested controller.

## 2. MATERIAL AND METHODS

In this paper, the software simulation of WEC is conducted using the point absorber model to demonstrate the effectiveness of the proposed control system. To achieve this, we assume the nominal second-order model of the WEC, considering all acting forces, in order to obtain an accurate model. The model contains an immersed object with a cylinder on the surface of the sea. Wave energy is gathered through the use of a power take-off system.

**2. 1. State Space Model of WEC** The state space calculations of the proposed system include forces, where $\boldsymbol{z_w}$ represents the water level and $\boldsymbol{z_v}$ represents the heave movement of the middle point of the float. The power generated by the generator is proportional to the force $\boldsymbol{f_u}$ applied to the piston. Therefore, the obtained power will be equal to $\boldsymbol{P = f_u v}$, with the value of $\boldsymbol{v = \dot{z}_v}$. By applying Newton's second law, the dynamics of the system can be determined:

$$m_s \ddot{z}_v = -f_s - f_r + f_e + f_u \tag{1}$$

where $m_s$ represents the mass of the buoy. The following describes the definition of the forces present in Equation 1.

$f_s$ is the lateral restoring force.

$$f_s = \rho g A z_v \tag{2}$$

In the given equation, $\rho$ stands for the density of water, $g$ represents the acceleration due to gravity, $A$ shows the cross-section of the float, and $z_v$ is the heave motion of the float.

The radiation force $f_r$ is obtained as follows.

$$f_r = m_\infty \ddot{z}_v + \int_{-\infty}^{\infty} h_r(\tau)\dot{z}_v(t-\tau)d\tau \tag{3}$$

where $m_\infty$ indicates the additional mass and $h_r$ represents the core of the radiation force. Using a finite-dimensional state-space model, the convolutional term in the force value is estimated as follows.

$$\begin{aligned} \dot{x}_r &= A_r x_r + B_r \dot{z}_r \\ f_r &= C_r x_r \approx \int_{-\infty}^{t} h_r(\tau)\dot{z}_v(t-\tau)d\tau \end{aligned} \tag{4}$$

Now, considering the equivalent state-space representation $D_r \sim (A_r, B_r, C_r, 0)$ leads to the $f_r \approx D_r(s)\dot{z}_v$.

Exterior perturbation force acting on the float is determined as $f_e = D_e(s)z_w$, where $D_e$ is the equivalent representation of the following realization:

$$\begin{aligned} \dot{x}_e &= A_e x_e + B_e z_w \\ f_e &= C_e x_e \approx \int_{-\infty}^{t} h_e(\tau) z_w(t-\tau)d\tau \end{aligned} \tag{5}$$

By applying the force amounts, the state-space model of WEC is derived.

$$\begin{aligned} \dot{x} &= A_c x + B_{uc} u + B_{wc} w \\ z &= C_z x \end{aligned} \tag{6}$$

where $w = z_w$, $z = z_v$. $y = \dot{z}_v$, $x = [z_v, \dot{z}_v]^T$ and

$$A_c = \begin{bmatrix} 0 & 1 \\ -\frac{k_s}{m} & -\frac{D_r(\delta_r)}{m} \end{bmatrix}, B_{wc} = \begin{bmatrix} 0 \\ \frac{D_e(\delta_e)}{m} \end{bmatrix}, B_{uc} = \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}, C_z = [1\ 0] \tag{7}$$

The value of $m$ is $m_a + m_\infty$. Additionally, the dynamics of the radiation and excitation forces include structure uncertainties, represented by $\delta_r$ and $\delta_e$, respectively. The state space form of a dynamic model is commonly referred to as:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{k_s}{m}x_1 + \frac{1}{m}u - \frac{D_r}{m}x_2 + \frac{D_e}{m}w \end{aligned} \tag{8}$$

Therefore, we will estimate the amount of $x_2$ as unmeasured output and the quantity of $w$ as an unknown disturbance value. The viscosity force, assuming the negligible velocity of WEC compared to the wave velocity, only includes constant terms that refer to $D_e(\delta_e)$ as an uncertainty term. The characteristics of the sea required for the training of DRL were discussed in Section 3. Moreover, Table 1 includes a detailed table about this process and some training data.

**2. 2. Implementation of the Control System** Suppose an n-dimensional lower-triangular nonlinear system with definition as follows (25).

$$\dot{\xi}_i(t) = \xi_{i+1}(t) + f_i(\xi_1(t), \ldots, \xi_i(t), u(t)), \quad i = 1:n-1, \tag{9}$$

where, $\xi_i \in \mathbb{R}$ denotes the state variables belonging to a compact set $\Omega$, $u \in C^1(\mathbb{R}_{\geq 0}, \mathbb{R})$ shows the input with an upper bound namely $u_0$, $w \in C^1(\mathbb{R}_{\geq 0}, \mathbb{R})$ demonstrates the unknown disturbance with an upper bound of $w_1$, $f_i \in C^0(\mathbb{R}^{i+1}, \mathbb{R})$ are defining parts of the system that are locally Lipschitz, and $g \in C^0(\mathbb{R}^{n+1}, \mathbb{R})$ stands for the uncertainty section.

**TABLE 1.** Target sea characteristic data

| Wave height | Wave period | Wavelength |
|---|---|---|
| 0.024 | 0.870 | 1.186 |
| 0.030 | 1.008 | 1.581 |
| 0.036 | 1.178 | 2.109 |
| 0.032 | 1.217 | 2.231 |
| 0.030 | 1.260 | 2.367 |
| 0.022 | 1.385 | 2.761 |
| 0.018 | 1.510 | 3.152 |

We define the measurable output as $y$, and the other trajectories and disturbances go into the aggregated variable vector $x$. Therefore, the following representation of overall system is expressed.

$$\begin{aligned} &y = \xi_i \\ &x_i = \xi_{i+1},\ \ i = 1:n-1 \\ &x_n = g\left(y, x_1, \ldots, x_{n-1}, w\right) \end{aligned} \tag{10}$$

By introducing the following definition as a new state variables vector:

$$\begin{aligned} &\bar{x}_i = \left[x_1, \ldots, x_i\right]^T \in \mathrm{R}^i,\ i = 1:n \\ &x = \bar{x}_n \in \mathrm{R}^n \end{aligned} \tag{11}$$

Equation 9 is transferred as:

$$\begin{aligned} &\dot{y}(t) = Cx(t) + f_1\left(y(t), u(t)\right) \\ &\dot{x}(t) = Ax(t) + \sum_{i=1}^{n-1} B_i f_{i+1}\left(y(t), \bar{x}_i(t), u(t)\right) \\ &+ B_n h\left(y(t), x(t), u(t), w(t), \dot{w}(t)\right) \end{aligned} \tag{12}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times n},\ B_i = [\delta_{1i},\ \ldots, \delta_{in}]^T \in \mathbb{R}^{n \times 1},\ C = [1\ 0\ \ldots\ 0] \in \mathbb{R}^{1 \times n},\ i = 1:n$$

$$\begin{aligned} &h\left(y, x, u, w, \dot{w}\right) = \frac{\partial g\left(y, \bar{x}_{n-1}, w\right)}{\partial y}\left(x_1 + f_1\left(y, u\right)\right) + \\ &\sum_{i=1}^{n-1} \frac{\partial g\left(y, \bar{x}_{n-1}, w\right)}{\partial x_i}\left(x_{i+1} + f_{i+1}\left(y(t), \bar{x}_i(t), u(t)\right)\right) \\ &+ \frac{\partial g\left(y, \bar{x}_{n-1}, w\right)}{\partial w}\dot{w} \end{aligned} \tag{13}$$

Along with the measurable output as first input to the estimator, the produced control action by the DRL is the second input. Accordingly, the proposed estimator gives the unmeasurable part of the output vector, the matched uncertainty and disturbance values as well. In this regard, the dynamical system is defined as:

$$\dot{\xi} = \alpha(\xi, y, u, t) \tag{14}$$

where $\xi$ belongs to the n-dimensional space and $\alpha$ is a continuous function from the (n+3)-dimensional space to the n-dimensional kind.

If there exists the left-invertible mapping $\phi_1$ then the system of Equation 14 is called a reduced order observer with the following manifold:

$$\mathrm{M}_1 = \{\left(y, x, \xi, u, t\right) \in \mathrm{R}^{2n+3} \mid \phi_1\left(x, y, u, t\right) = \beta_1\left(\xi, y, u, t\right)\} \tag{15}$$

Therefore, the reduced order ESO is paraphrased as the dynamical system of Equation 14 along with manifold introduced by Equation 15. Now, imagine the mapping $\phi_1$ is of the form:

$$\varphi_1\left(x, y, u, t\right) = x + \psi_1\left(y, u, t\right) \tag{16}$$

Considering this case, the defined manifold is replaced by:

$$\mathrm{M}_1 = \left\{\left(y, x, \xi, u, t\right) \in \mathrm{R}^{2n+3} \mid x + \psi_1\left(y, u, t\right) = \beta_1\left(\xi, y, u, t\right)\right\} \tag{17}$$

In the equality of the set in Equation 17, by replacing the function as estimation of $x$:

$$\eta_1\left(\xi, y, u, t\right) = \beta_1\left(\xi, y, u, t\right) - \psi_1\left(y, u, t\right) \tag{18}$$

By assuming a manifold coordinate of Equation 7 to describe the distance of the routes of systems in Equations 13 and 14:

$$z = x - \eta_1\left(\xi, y, u, t\right) \tag{19}$$

Considering the $\eta_1(\xi, y, u, t)$ as $\hat{x}$, the difference between $x$ and $\hat{x}$ is obtained as error estimation of the state variable. The derivative of the Equation 19 and the substitution of corresponding parameters in Equation 13 lead to:

$$\begin{aligned} &\dot{z}(t) = \left(A - \frac{\partial \eta_1}{\partial y}C\right)z(t) + \left(A - \right. \\ &\left.\frac{\partial \eta_1}{\partial y}C\right)\eta_1(\xi(t), y(t), u(t), t) \\ &+ \sum_{i=1}^{n-1} B_i f_{i+1}\left(y(t), \bar{x}_i(t), u(t)\right) + \\ &B_n h\left(y(t), x(t), u(t), w(t), \dot{w}(t)\right) \\ &- \frac{\partial \eta_1}{\partial y} f_1\left(y(t), u(t)\right) - \frac{\partial \eta_1}{\partial u}\dot{u}(t) - \frac{\partial \eta_1}{\partial t} - \\ &\frac{\partial \eta_1}{\partial \xi}\alpha(\xi(t), y(t), u(t), t) \end{aligned} \tag{20}$$

If $\frac{\partial \eta_1}{\partial \xi}$ doesn't have the zero determinant, we introduce the expression for $\alpha$ to simplify Equation 20:

$$\begin{aligned} &\alpha(\xi, y, u, t) = \left(\frac{\partial \eta_1}{\partial \xi}\right)^{-1}\left(\left(A - \right.\right. \\ &\left.\left.\frac{\partial \eta_1}{\partial y}C\right)\eta_1(\xi, y, u, t) + \sum_{i=1}^{n-1} B_i f_{i+1}(y, \hat{\bar{x}}_i, u)\right) + \\ &\left(\frac{\partial \eta_1}{\partial \xi}\right)^{-1}\left(-\frac{\partial \eta_1}{\partial y}f_1(y, u) - \frac{\partial \eta_1}{\partial u}\dot{u} - \frac{\partial \eta_1}{\partial t}\right) \end{aligned} \tag{21}$$

where $\hat{\bar{x}}$ is defined as component-wise based on $\hat{x}$. Consequently, the system of Equation 20 is reduced as 25:

$$\begin{aligned} &\dot{z}(t) = \left(A - \frac{\partial \eta_1}{\partial y}C\right)z(t) + \\ &\sum_{i=1}^{n-1} B_i\left(f_{i+1}\left(y(t), \bar{x}_i(t), u(t)\right) - \right. \\ &\left. f_{i+1}\left(y(t), \hat{\bar{x}}_i(t), u(t)\right)\right) + \\ &B_n h\left(y(t), x(t), u(t), w(t), \dot{w}(t)\right) \end{aligned} \tag{22}$$

The matrix inequality of Equation 22 to be stable is as follows:

$$P_1\left(A - \frac{\partial \eta_1}{\partial y}C\right) + \left(A - \frac{\partial \eta_1}{\partial y}C\right)^T P_1 + \frac{2}{\lambda_1} P_1 \leq 0 \tag{23}$$

**Theorem 1.** Considering the prolonged system of Equation 13, the system of Equation 14 with given dynamics in Equation 21 is a reduced order ESO and the corresponding state estimation is obtained as $\hat{x} = \eta_1(\xi, y, u, t)$.

*Proof.* We present the following Lyapunov function as:

$$V_1(z) = z^T P_1 z \tag{24}$$

Taking the time derivative of Equation 24 and using the equivalent amounts of Equations 22 and 23 gives:

$$\begin{aligned}&\dot{V}_1(z(t)) \le -\frac{2}{\epsilon_1} V_1(z(t)) + \\ &2z^T(t)\textstyle\sum_{i=1}^{n-1} P_1 B_i \Big(f_{i+1}(y(t), \bar{x}_i(t), u(t)) - \\ &f_{i+1}\Big(y(t), \hat{\bar{x}}_i(t), u(t)\Big)\Big) + \\ &2z^T(t) P_1 B_n h\big(y(t), x(t), u(t), w(t), \dot{w}(t)\big)\end{aligned} \tag{25}$$

Assuming $h(.)$ is restricted to an upper bound as:

$$|h(y, x, u, w, \dot{w})| \le h_0 \tag{26}$$

Locally Lipschitz assumption of $f$ with $l_{fi}$ multiplier gives 21:

$$f_i(y, \bar{x}_{i-1}, u) - f_i(y, \hat{\bar{x}}_{i-1}, u) \le l_{fi} z, \quad i = 2{:}n \tag{27}$$

Now, the inequality (25) is rewritten as:

$$\begin{aligned}&\dot{V}_1(z(t)) \le -\Big(\frac{2}{\epsilon_1} - 2l_f \frac{\lambda_{max}(P_1)}{\lambda_{min}(P_1)}\Big) V_1(z(t)) + \\ &2h_0 \frac{\lambda_{max}(P_1)}{\sqrt{\lambda_{min}(P_1)}} \sqrt{V_1(z(t))}\end{aligned} \tag{28}$$

To complete the proof, the $\Lambda_r$-attractivity of the manifold $\mathcal{M}_1$ is explained as follows.

$\Lambda_r$-attractivity: For each $t_0 \in \mathbb{R}_{\ge 0}$, $r_0$ is available such that if $z(t_0) \in \Lambda_{r_0}$, then $z(t)$ is bounded for all $t \ge t_0$ and moreover, the positive $r_1$ and its finite reaching time are as $t_r = t_r(r_0, r_1)$, in this sense $z(t) \in \Lambda_{r_1}$ for all $t \ge t_0 + t_r$.

Now, assuming:

$$\epsilon_1 < \epsilon_1' = \frac{\lambda_{min}(P_1)}{l_f \lambda_{max}(P_1)} \tag{29}$$

Application of the comparison lemma on 28 yields in:

$$\begin{aligned}&z(t) \le \sqrt{\frac{\lambda_{max}(P_1)}{\lambda_{min}(P_1)}} \left( z(t_0) - \sqrt{\frac{\lambda_{max}(P_1)}{\lambda_{min}(P_1)}} \lambda(\epsilon_1) h_0 \right) \\ &\times \exp\left( -\frac{t - t_0}{\lambda(\epsilon_1)} \right) + \frac{\lambda_{max}(P_1)}{\lambda_{min}(P_1)} \lambda(\epsilon_1) h_0\end{aligned} \tag{30}$$

$$\lambda(\epsilon_1) = \frac{\epsilon_1}{1 - \frac{\epsilon_1}{\epsilon_1'}} \tag{31}$$

The inequality 30 implies that the trajectories starting in $\Lambda_{r_0}$ stay bounded. Furthermore, the reaching time is of the form:

$$t_r = \lambda(\epsilon_1) \ln\left( \frac{1}{r_2} \sqrt{\frac{\lambda_{max}(P_1)}{\lambda_{min}(P_1)}} \left| z(t_0) - \sqrt{\frac{\lambda_{max}(P_1)}{\lambda_{min}(P_1)}} \lambda(\epsilon_1) h_0 \right| \right) \tag{32}$$

With the positive scalars $r_2 < r_1 < r_0$ gives:

$$\epsilon_1 < \epsilon''_1 = \frac{r_1 - r_2}{h_0 \frac{\lambda_{max}(P_1)}{\lambda_{min}(P_1)} + \frac{r_1 - r_2}{\epsilon'_1}} \tag{33}$$

Noting that $\epsilon''_1 < \epsilon'_1$, and Equation 33 holds for the selected parameters, then the manifold $\mathcal{M}_1$ is $\Lambda_r$-attractive for any $r_1 < r_0$. So, the system of Equation 14 with defined dynamics of Equation 21 is a reduced order ESO.■

**Corollary 1.** If Theorem 1 satisfied, so $\lim_{\substack{(t\to\infty)\\(\epsilon_1\to 0)}} z(t) = 0$.

*Proof.* Taking the supremum of Equation 30 leads to:

$$\lim_{t\to\infty} z(t) \le \frac{\lambda_{max}(P_1)}{\lambda_{min}(P_1)} \lambda(\epsilon_1) h_0 \tag{34}$$

Since $\lim_{\epsilon_1 \to 0} \lambda(\epsilon_1) = O(\epsilon_1)$, the ultimate bound of $\|z(t)\|$ approaches to zero as $\epsilon_1 \to 0$.■

Convergence analysis via LMIs

Considering the $i^{th}$ component of $\frac{\partial \eta}{\partial v}$ as $a_i$:

$$\begin{aligned}&a_i(t) = a_i^0 + a_i^1 \theta_i(t), \\ &a_i^{min} \le a_i(t) \le a_i^{max}, \\ &a_i^0 = (a_i^{max} + a_i^{min})/2, \\ &a_i^1 = (a_i^{max} - a_i^{min})/2 \\ &|\theta_i(t)| \le 1\end{aligned} \tag{35}$$

Based on Equation 35, we obtain the following representation of coefficient of $z(t)$ in Equation 22,

$$A - \frac{\partial \eta}{\partial v}(t) C = A_0 + A_1 \Theta(t) A_2 \tag{36}$$

With the assumptions:

$$A_0 = \begin{bmatrix} -a_1^0 & 1\,0 \cdots 0 \\ -a_2^0 & 0\,1 \cdots 0 \\ \vdots & \vdots\,\vdots\,\ddots\,\vdots \\ -a_{n-1}^0 & 0\,0 \cdots 1 \\ -a_n^0 & 0\,0 \cdots 0 \end{bmatrix} \in \mathbb{R}^{n \times n},$$

$$A_1 = [A_{11}, \ldots, A_{1n}],\ A_{1j} = \begin{bmatrix} -a_1^1 \delta_{1j} & 0\,0 \ldots 0 \\ -a_2^1 \delta_{2j} & 0\,0 \ldots 0 \\ \vdots & \vdots\,\vdots\,\ddots\,\vdots \\ -a_{n-1}^1 \delta_{(n-1)j} & 0\ 0 \ldots 0 \\ -a_n^1 \delta_{nj} & 0\ 0 \ldots 0 \end{bmatrix} \in \mathbb{R}^{n \times n} \tag{37}$$

$$\Theta(t) = \mathrm{diag}\big(\Theta_j(t)\big)_{j=1}^{n},\ \Theta_j(t) = \theta_j(t) \in \mathbb{R}^{n \times n},$$

$$A_2 = [A_{21}, \ldots, A_{2n}]^T,$$

$$A_{2j} = \begin{bmatrix} 1\,0\,0 \cdots 0 \\ 0\,0\,0 \cdots 0 \\ \vdots\,\vdots\,\vdots\,\ddots\,\vdots \\ 0\,0\,0 \cdots 0 \\ 0\,0\,0 \cdots 0 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

**Theorem 2.** Suppose that Equation 37 is valid for the components of $\frac{\partial \eta}{\partial \nu}$ and the matrix $P$ is available as following:

$$\begin{bmatrix} PA_0 + A_0^T P + \frac{2}{\epsilon} P & PA_1 & A_2^T \\ A_1^T P & -I & 0 \\ A_2 & 0 & -I \end{bmatrix} \leq 0 \tag{38}$$

Then, Equation 23 holds with positive $\epsilon_1 = \epsilon$ and matrix $P$.

*Proof.* Considering Equation 36, the inequality 23 is written as:

$$P\left(A_0 + A_1\Theta(t)A_2\right) + \left(A_0 + A_1\Theta(t)A_2\right)^T P + \frac{2}{\epsilon} P \leq 0 \tag{39}$$

According to Young's inequality together with constraint $\Theta^T(t)\Theta(t) \leq I$,

$$PA_1\Theta(t)A_2 + A_2^T\Theta^T(t)A_1^T P \leq \frac{1}{\mu} PA_1A_1^T P + \mu A_2^T A_2 \tag{40}$$

where $\mu$ is a positive constant. The inequality 40 satisfies if the following is true.

$$PA_0 + A_0^T P + \frac{2}{\epsilon} P + \frac{1}{\mu} PA_1A_1^T P + \mu A_2^T A_2 \leq 0 \tag{41}$$

Therefore, inequality 41 is equivalent to the viability of LMI 38 for a non-negative $P$. ■

Assuming fixed values of $\frac{\partial \eta}{\partial \nu}$ as $k_i$:

$$\frac{\partial \eta^{(i)}}{\partial \nu} = k_i, \quad i = 1,2,..,n \tag{42}$$

where $\eta^{(i)}$ stands for the $i^{th}$ component of $\eta$. Therefore, $\eta$ is written as:

$$\eta^{(i)}(\xi,\nu) = k_i\nu + \varpi_i(\xi) \tag{43}$$

where the functions $\varpi_i(\xi)$ satisfy the condition of non-zero determinant of $\frac{\partial \eta^{(i)}}{\partial \nu}$. The fixed $\frac{\partial \eta^{(i)}}{\partial \nu}$ reduces LMI 38 to:

$$PA_0 + A_0^T P + \frac{2}{\epsilon} P \leq 0 \tag{44}$$

Now, using the generalized eigenvalue solvers, the solution of $P$ is obtained.

## 3. NUMERICAL RESULTS

The Simulink diagram of the designed control model in MATLAB is illustrated in Figure 1. The figure shows that the first step is to simulate the system in order to calculate the outputs. Then, the measured output along with the assigned value for input are fed to the observer block to estimate the unmeasured output and the unknown disturbances. Besides, our main block including the DRL controller will suggest the action $u$ and it tries to reach the optimal input action along with correct estimation of the mentioned parameters with use of the immersion- and invariance-based extended state observer. The implemented DRL has the structure as shown in Figure 2. In this figure, the actor and critic networks are stacked in a row. The lagged version of two main networks is not shown but they are necessary to avoid the divergence during the updating process step. The pseudo code of the DRL as clarification is brought in Table 2. The output function for getting the optimal control action is described as:

$$y_j = \begin{cases} r_j & \text{terminates at step j+1} \\ r_j + \gamma \max_{a'} \boldsymbol{Q}'_{\vartheta'}\left(s_{j+1}, a; \vartheta'\right) & \text{otherwise} \end{cases} \tag{45}$$

where, the $\boldsymbol{Q}'_{\vartheta'}$ stands for the lagged version of the main critic network, $\gamma$ is the coefficient and $r_j$ indicates the cumulative reward during processing time. DRL, in every sea state based on the environment's observed conditions, learns the required optimal force on the piston of PTO to capture the uniform energy, i.e., the structure faces the waves instead of internal models. At each processing time, the control system determines a change in the action, which is carried out by the hydraulic PTO unit as the agent. Through accumulating rewards as the function of the produced electrical power and in an alteration of environment's state, where the state is indicated by the substantial wave altitude, $H_s$, the average zero-crossing span, T, and the PTO damping factor. For the oscillatory essence of sobriety waves, it is essential to average the captured power in the reward function over a horizon, $H$, in one wave cycle during which the state $s_n$ and action $a_{n-1}$ are fixed. Next, a new action $a_n$ is designated in an instant modification of state to $s_{n+1}$ and a new mean progression. The process of DRL is described in detail as follows.

*State Space:* As noted before, the circumstances variables are supposed to be the effective wave height, average zero-crossing span, and PTO damping multiplier so that the assumed RL state space is:

$$S = \left\{ s \mid s_{j,k,l} = H_{s,j} + T_{z,k} + B_{PTO,l} \begin{matrix} j=1:J, \\ k=1:K, \\ l=1:L \end{matrix} \right\} \tag{46}$$

*Action Space:* The action series consists of three amounts based on the selected state space as following:

$$A = \{a \mid (-B_{PTO}, 0, +B_{PTO})\} \tag{47}$$

where $\Delta B_{PTO} = B_{PTO,k+1} - B_{PTO,k}$. Equivalent states to

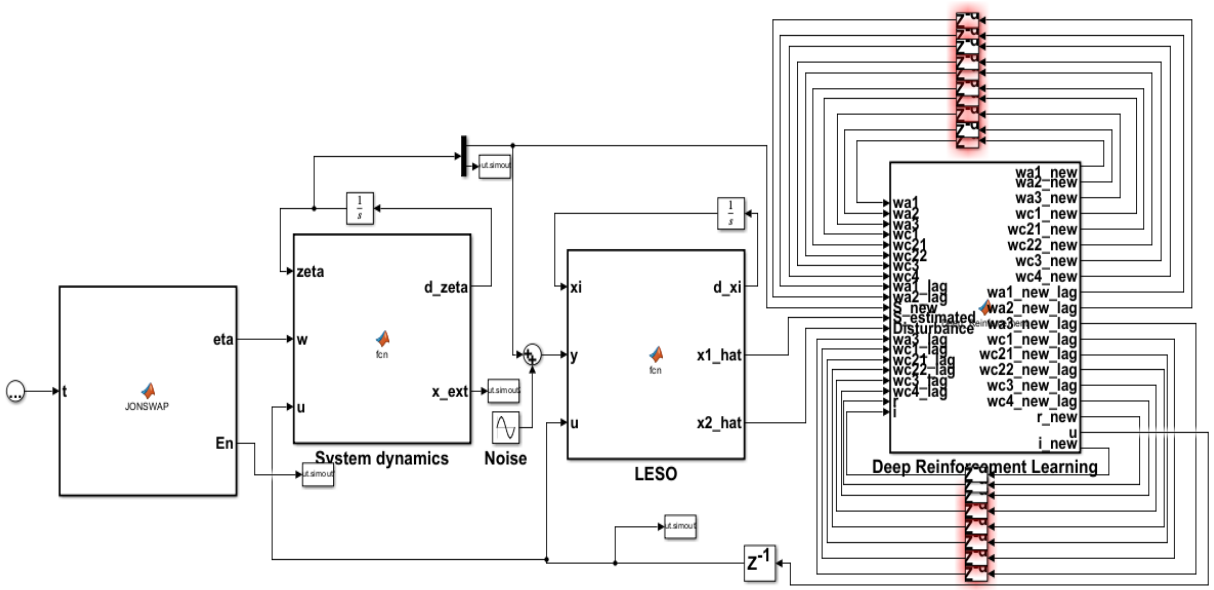



**Figure 1.** Block diagram of the entire observer and controller in MATLAB

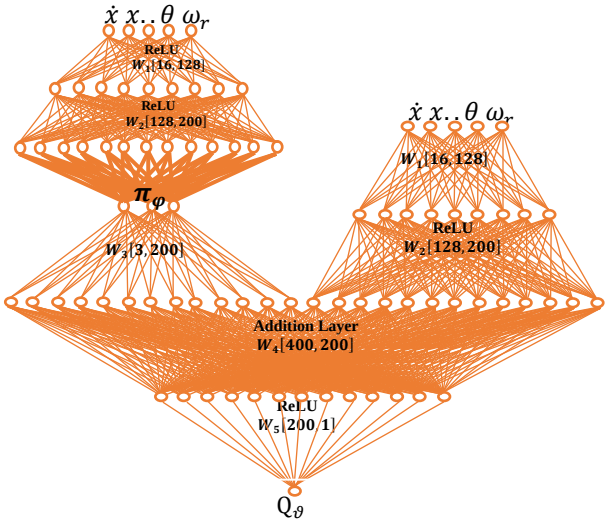



**Figure 2.** Stacked critic-actor network in DRL (29)

the extreme damping constants, i.e., $B_{PTO,1}$ and $B_{PTO,L}$, have some actions so as to preclude the controller from surpassing the boundary of state space. For example, for $B_{PTO,1}$ case, the candidate $+\Delta B_{PTO}$ is left out.

*Reward:* In DRL field, the aim to propose a reward function is to optimize the performance of the agent through negative reward for incorrect action or positive for correct one, and the reward function is maximized at the end of processing time. Hence, for the I&I RSO control system of WECs, the compensation function is considered to be pertinent to the captured power. The effect of variations in the momentous wave height on the average produced power, $P_{avg}$, is more than variations in the PTO restraining constant. Considering the harvested power proportional to the square of the substantial wave elevation, the reward function is well-defined as $P_{avg}/H_s^2$ to obtain a dimensionless function. Additionally, owing to the stiff quantization of the condition variables and the random environment of periodic waves, the reward function is alleviated by averaging a number $M$ of $P_{avg}/H_s^2$ values for each state along with the average captured power over a horizon $H$. Keeping the $M$ recent $P_{avg}/H_s^2$ quantities for each state in a matrix, $R$, whose magnitude is as a maximum $n_s \times M$, with $n_s = J \times K \times L$ as the number of states.

**TABLE 2.** Pseudo code of the training process of the Deep Deterministic Policy Gradient (DDPG) algorithm

Algorithm 1: DDPG algorithm

Randomly initialization of weights $\vartheta$ and $\varphi$ in critic network $\mathbf{Q}_{\vartheta}(s, a \mid \vartheta)$ and actor $\pi_{\varphi}(s \mid \varphi)$.

Initialization of target lagged network $\mathbf{Q}'_{\vartheta'}$ and $\pi'_{\phi'}$ with weights $\vartheta' \leftarrow \vartheta . \phi' \leftarrow \varphi$

Replay buffer $R$ Initialization

for incident $= 1. M$ ensure

Take initial state $s_1$

for t $= 1.$ T do

Choose action $a_t = \mu(s_t \mid \theta^{\mu})$ based on the current policy

Accomplish action $a_t$ to detect reward $r_t$ and new-fangled state $s_{t+1}$

Save evolution $(s_t . a_t . r_t . s_{t+1})$ in $R$

Sample a spontaneous small group of $N$ transitions $(s_i . a_i . r_i . s_{i+1})$ from $R$

Set $y_i = r_i + \gamma \mathbf{Q}'_{\vartheta'}\left(\mathbf{s}'_i . \pi'_{\phi'}(\mathbf{s}'_i)\right)$

Update critic by differentiation of the loss: $L = \frac{1}{N}\sum(\mathbf{Q}_{\vartheta}(\mathbf{s}_i . \mathbf{a}_i) - y_i(\boldsymbol{s}_i . \boldsymbol{a}_i . \boldsymbol{r}_i . \mathbf{s}'_i))^2$ with respect to weights

Revise the performer procedure employing the sampled policy slope:

$$\nabla_{\varphi} J \approx \frac{1}{N}\sum_{i} \nabla_a Q(s . a \mid \vartheta)\Big|_{s=s_i . a=\pi_{\varphi}} \nabla_{\varphi}\pi_{\varphi}(s \mid \varphi)\Big|_{s_i}$$

Update the target networks:

$$\vartheta'^{(t+1)} \leftarrow \tau\vartheta + (1-\tau)\vartheta'^{(t)}$$
$$\varphi'^{(t+1)} \leftarrow \tau\varphi + (1-\tau)\varphi'^{(t)}$$

end for

end for

Therefore, the average amount in each state is described with the vector $\boldsymbol{m} = \langle R(s,m)\rangle_{m=1:(M \vee end)}$ of size $n_s$. The states of the vector $\boldsymbol{m}$ include the vector-oriented version of Equation 46, such that the discrete quantities of $B_{PTO}$, $H_s$ and $T_z$ represent the inner-most, middle and outermost loop of vectorization, respectively.

Regarding the values of $\Delta B_{PTO}$, for $B_{PTO} > 0$, there exists a bit of difference between the averages of adjacent damping coefficients of PTO, which yields extreme problems for the confluence of the Q-learning process. Hence, the benefit of desiring the optimum balancing multiplier in apiece sea condition is obvious to detour the convergence issues. Assignment of a factor as the power of the values in vector $\boldsymbol{m}$ deals with the close amounts in the neighboring coefficients. Furthermore, with the mentioned mathematical operation, the excessive memory is required to keep the reward values. Therefore, it is essential from a mathematical point of view to primarily regularize the quantity of the vector for every individual condition concerning the highest amount for each sea condition. Regarding $s_n$, we should find the maximum quantity between the indexes $o = \text{floor}\left(\frac{s_n - 1}{L}\right)L + 1$ and $p = \text{floor}\left(\frac{s_n - 1}{L}\right)L + L$ of the vector $\boldsymbol{m}$. Also, the power factor, $k$, is considered an odd digit to maintain the sign of the produced power. The smoother the quantization of PTO balancing value, the superior $k$ is considered to expedite the learning process. On the other hand, very great values for $k$ result in convergence issues especially with noise affected power in irregular wave. Furthermore, selecting the optimum damping value by itself makes the structure destruction i.e., submergence or emergence, if not complete failure. So, the penalty amount should be considered to avoid exceeding the upper bound limitation of translational movement of the structure, $z_{max}$. The reward function is defined as following with arbitrary chosen amount the same as $p = -4$ for penalty:

$$r_{n+1} = \begin{cases} \left[\frac{m(s_n)}{\max_{i=o:pm(i)}}\right]^k & if\ |max(z)| \le z_{\max} \\ -4 & if\ |max(z)| > z_{\max} \end{cases}$$

The system is acting in the presence of wave with the following regular form:

$$\xi = \xi_a \sin(\omega t - kx) \tag{48}$$

We can describe the irregular wave by using the superposition principle consisting of $n$ regular waves with variant amplitude, frequency, and phase. Therefore, the wave elevation $\xi_n$ is given as:

$$\xi(x,t) = \xi_{An} \sin(\omega_n t - k_n x + \epsilon_n) \tag{49}$$

where the value $\epsilon_n$ stands for the stochastic phase that its value is distributed between 0 and $2\pi$. If all elevations are summed up, the following shows the total elevation of the proposed wave structure:

$$\xi(x,t) = \sum_{n=1}^{N} \xi_{An} \sin(\omega_n t - k_n x + \epsilon_n) \tag{50}$$

The sea state is typically described in terms of the wave spectrum $S(\omega)$, which explains information about the wave energy for various frequencies. $E_n$, the energy in each component is defined as follows.

$$s(\omega_n)\Delta\omega = \frac{E_n}{\rho g} = \frac{1}{2}\xi_{An}^2 \tag{51}$$

Therefore, the value of the amplitude is found:

$$\xi_n = \sqrt{2S(\omega_n)\Delta\omega}\cos(\omega_n t - k_n x + \epsilon_n) \tag{52}$$

A standardized wave spectrum to model the sea states is JONSWAP, which is developed for the North Sea. The formula of this spectrum is:

$$S(\omega)=\frac{\alpha g^2}{\omega^5}e^{-\frac{5}{4}\left(\frac{\omega_p}{\omega}\right)^4}\gamma^{e^{-\frac{1}{2}\left(\frac{\omega-\omega_p}{\sigma\omega_p}\right)^2}} \quad (53)$$

where $\gamma$ is peak number, $\sigma$ decides the form of the spectrum in the high frequency part and with assuming $T_p$ as peak frequency. The value of the parameter $\omega_p$ is:

$$\omega_p=\frac{2\pi}{T_p} \quad (54)$$

The simulated disturbance and uncertainty assuming JONSWAP spectrum is shown in Figure 3 in which the randomness of the proposed wave spectrum is vivid. The measured output of the system is illustrated as Figure 4 showing the stability of the desired state. With application of the above assumptions, the simulated results in MATLAB are as follows. Through adding the measurement noise, we tried to show the efficacy of the proposed controller. The results of Figure 5 illustrate that only one second takes to achieve the exact value of the unmeasured output and no effect of fluctuation due to the noise is seen.

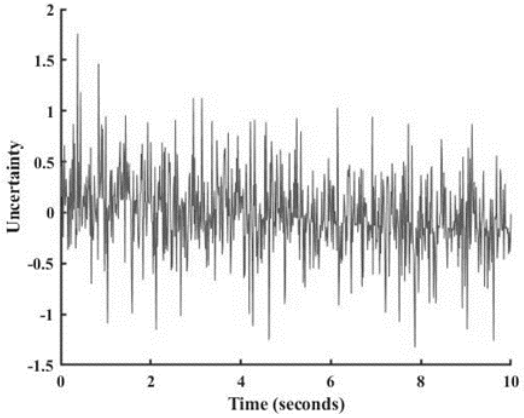



**Figure 3.** Uncertainty analysis based on the JONSWAP spectrum

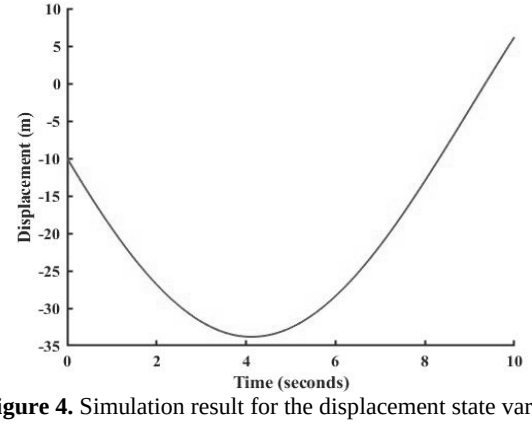



**Figure 4.** Simulation result for the displacement state variable

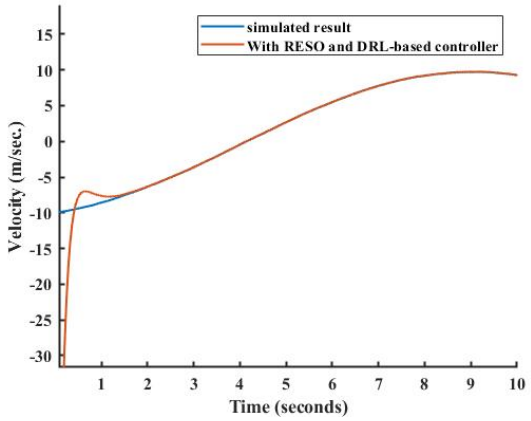



**Figure 5.** Comparison of unmeasured velocity in simulated and estimated phase

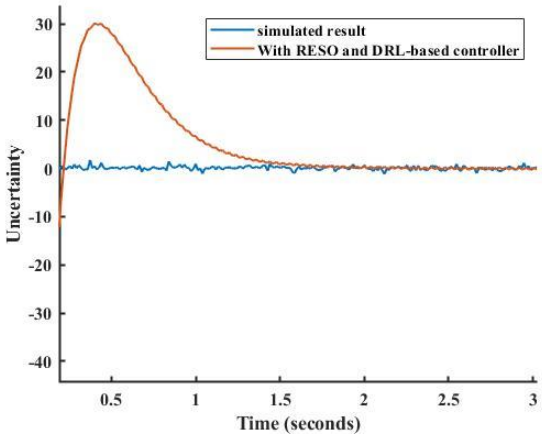



**Figure 6.** Unknown disturbance and uncertainty with simulated and estimated scenario

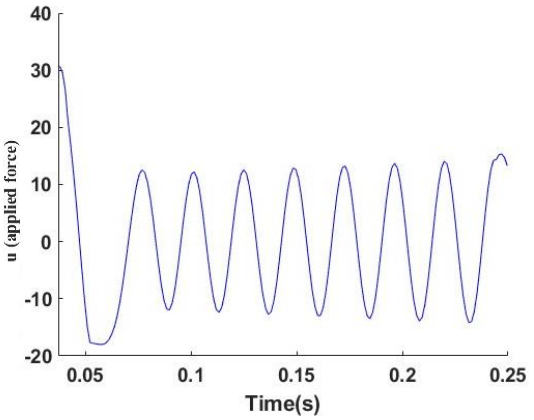



**Figure 7.** Control action applied on the PTO with DRL

As shown in Figure 6, the uncertainty of the problem is tracked quickly so that the tracking error approaches to zero after two seconds. The estimated results can be

used in the other control purposes including disturbance rejection. The significant point of the correct estimation with low settling time refers to the application of the concise control action. Figure 7 illustrates the applied force required to reach the uniform and maximum amount of energy and correct tracking. The accumulated reward during the processing time is demonstrated in Figure 8. The accumulated reward is increased so that its final value stands for the optimum control policy applied by the actor network and also, the maximum amount of energy. The captured energy during this task is demonstrated as Figure 9. Although, the fluctuation in some periods is clear, oscillation is not bothersome and is in the acceptable range for the consumer.

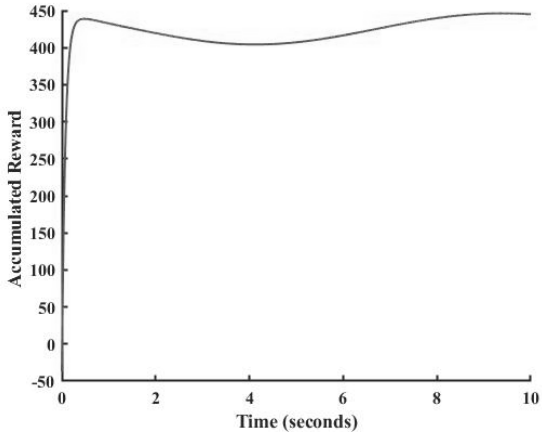



**Figure 8.** Captured energy during the processing time by DRL

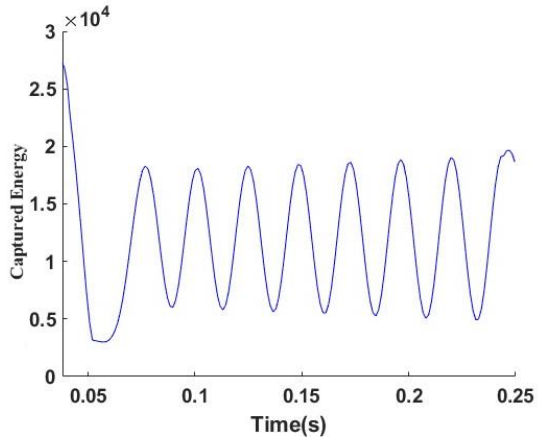



**Figure 9.** Captured energy during the processing time by DRL

## 4. CONCLUSION AND REMARKS

This work made an effort to address the undesirable fluctuation in the utilization of energy produced by a WEC. To do this, we created the WEC's structure model along with all the necessary dynamical equations for the proposed controller's implementation. Next, we presented the new RSO dealing with the uncertainties of the structure and unmeasured outputs. The technique of immersion and invariance manifold was thoroughly integrated and confirmed by using LMI. The demand for regulating intelligent action applied to structures guided us to create the DRL. Therefore, the built network with appropriate training method was thoroughly detailed in the study. Finally, the numerical results carried out using MATLAB to examine and verify the performance of the suggested control system. The results indicated that our novel designed control method has a significant performance, both in estimating the unmeasured output and unknown disturbance, as well as in implementing the control policy. The future extent of our study may be altered to disturbance rejection problem and extended state observer design.

## CREDIT AUTHOR STATEMENT

Hadi Mohammadian KhalafAnsar and Jafar Keighobadi: Conceptualization, Methodology, Data testing, Writing original draft, and editing.

## DECLARATION OF COMPETING INTEREST

The authors declare no financial interests/personal relationships which may be considered as potential competing interests.

## DATA AVAILABILITY

The data is accessible with request.

## ACKNOWLEDGMENTS

This work was supported by University of Tabriz.

## 5. REFERENCES

1. Keighobadi J, KhalafAnsar HM, Naseradinmousavi P. Adaptive neural dynamic surface control for uniform energy exploitation of floating wind turbine. Applied Energy. 2022;316:119132. https://doi.org/10.1016/j.apenergy.2022.119132
2. Nielsen KM, Pedersen TS, Andersen P, Ambühl S. Optimizing control of wave energy converter with losses and fatigue in power take off. IFAC-PapersOnLine. 2017;50(1):14680-5. https://doi.org/10.1016/j.ifacol.2017.08.2497
3. Sadripour G, Shafaghat R, Alizadeh Kharkeshi B, Tabassom R, Mahmoudi A. Installation Depth and Incident Wave Height Effect on Hydrodynamic Performance of a Flap Type Wave Energy Converter: Experimental Analysis. International Journal of Engineering. 2022;35(12):2283-90. https://doi.org/10.5829/ije.2022.35.12c.02

4. Arish N, Marignetti F. Evaluation of Linear Permanent Magnet Vernier Machine Topologies for Wave Energy Converters. International Journal of Engineering. 2021;34(2):403-13. https://doi.org/10.5829/ije.2021.34.02b.12
5. Bayani R, Farhadi M, Shafaghat R, Alamian R. Experimental evaluation of irwec1, a novel offshore wave energy converter. International Journal of Engineering. 2016;29(9):1292-9. https://doi.org/10.5829/idosi.ije.2016.29.09c.15
6. Alizadeh Kharkeshi B, Shafaghat R, Jahanian O, Alamian R. Experimental evaluation of the effect of dimensionless hydrodynamic coefficients on the performance of a multi-chamber oscillating water column converter in laboratory scale. Modares Mechanical Engineering. 2021;21(12):823-34. http://mme.modares.ac.ir/article-15-52993-en.html
7. Son D, Yeung RW. Real-time implementation and validation of optimal damping control for a permanent-magnet linear generator in wave energy extraction. Applied energy. 2017;208:571-9. https://doi.org/10.1016/j.apenergy.2017.09.097
8. Korde UA. Preliminary consideration of energy storage requirements for sub-optimal reactive control of axisymmetric wave energy devices. Annual Reviews in Control. 2015;40:93-101. https://doi.org/10.1016/j.arcontrol.2015.08.004
9. Anderlini E, Forehand D, Bannon E, Xiao Q, Abusara M. Reactive control of a two-body point absorber using reinforcement learning. Ocean Engineering. 2018;148:650-8. https://doi.org/10.1016/j.oceaneng.2017.08.017
10. Babarit A, Duclos G, Clément AH. Comparison of latching control strategies for a heaving wave energy device in random sea. Applied Ocean Research. 2004;26(5):227-38. https://doi.org/10.1016/j.apor.2005.05.003
11. Wu J, Yao Y, Zhou L, Göteman M. Real-time latching control strategies for the solo Duck wave energy converter in irregular waves. Applied Energy. 2018;222:717-28. https://doi.org/10.1016/j.apenergy.2018.04.033
12. Faedo N, Olaya S, Ringwood JV. Optimal control, MPC and MPC-like algorithms for wave energy systems: An overview. IFAC Journal of Systems and Control. 2017;1:37-56. https://doi.org/10.1016/j.ifacsc.2017.07.001
13. Hals J, Falnes J, Moan T. Constrained optimal control of a heaving buoy wave-energy converter. 2011. https://doi.org/10.1115/1.4001431
14. O'Sullivan AC, Lightbody G. Co-design of a wave energy converter using constrained predictive control. Renewable energy. 2017;102:142-56. https://doi.org/10.1016/j.renene.2016.10.034
15. Li L, Yuan Z, Gao Y. Maximization of energy absorption for a wave energy converter using the deep machine learning. Energy. 2018;165:340-9. https://doi.org/10.1016/j.energy.2018.09.093
16. Jama M, Wahyudie A, Noura H. Robust predictive control for heaving wave energy converters. Control Engineering Practice. 2018;77:138-49. https://doi.org/10.1016/j.conengprac.2018.05.010
17. Zou S, Abdelkhalik O, Robinett R, Korde U, Bacelli G, Wilson D, et al. Model Predictive Control of parametric excited pitch-surge modes in wave energy converters. International journal of marine energy. 2017;19:32-46. https://doi.org/10.1016/j.ijome.2017.05.002
18. Korde UA, Lyu J, Robinett III RD, Wilson DG, Bacelli G, Abdelkhalik OO. Constrained near-optimal control of a wave energy converter in three oscillation modes. Applied Ocean Research. 2017;69:126-37. https://doi.org/10.1016/j.apor.2017.10.004
19. Burgaç A, Yavuz H. Fuzzy Logic based hybrid type control implementation of a heaving wave energy converter. Energy. 2019;170:1202-14. https://doi.org/10.1016/j.energy.2018.12.090
20. Zhan S, Wang B, Na J, Li G. Adaptive optimal control of wave energy converters. IFAC-PapersOnLine. 2018;51(29):38-43. https://doi.org/10.1016/j.ifacol.2018.09.466
21. Han J. From PID to active disturbance rejection control. IEEE transactions on Industrial Electronics. 2009;56(3):900-6. https://doi.org/10.1109/TIE.2008.2011621
22. Gao Z. On the centrality of disturbance rejection in automatic control. ISA transactions. 2014;53(4):850-7. https://doi.org/10.1016/j.isatra.2013.09.012
23. Zhao Z-L, Guo B-Z. On convergence of nonlinear active disturbance rejection control for SISO nonlinear systems. Journal of Dynamical and Control Systems. 2016;22(2):385-412. https://doi.org/10.1007/s10883-015-9304-5
24. Feng H, Guo B-Z. Active disturbance rejection control: Old and new results. Annual Reviews in Control. 2017;44:238-48. https://doi.org/10.1016/j.arcontrol.2017.05.003
25. Hosseini-Pishrobat M, Keighobadi J. Extended state observer-based robust non-linear integral dynamic surface control for triaxial MEMS gyroscope. Robotica. 2019;37(3):481-501. https://doi.org/10.1017/S0263574718001133
26. Guo B-Z, Zhao Z-l. On the convergence of an extended state observer for nonlinear systems with uncertainty. Systems & Control Letters. 2011;60(6):420-30. https://doi.org/10.1016/j.sysconle.2011.03.008
27. Li J, Xia Y, Qi X, Wan H. On convergence of the discrete-time nonlinear extended state observer. Journal of the Franklin Institute. 2018;355(1):501-19. https://doi.org/10.1016/j.jfranklin.2017.11.019
28. Zhao Z-L, Guo B-Z. Extended state observer for uncertain lower triangular nonlinear systems. Systems & Control Letters. 2015;85:100-8. https://doi.org/10.1016/j.sysconle.2015.09.004

## Persian Abstract

**چکیده**

زندگی در دنیای گرمتر زمین، باعث ترغیب و پیشرفت دستگاه‌های تولید انرژی تجدیدپذیر و بهره‌گیری از فناوری‌های جدید مانند هوش مصنوعی می‌شود. با توجه به پتانسیل بالای انرژی امواج دریا و تطابق آن با محیط زیست، مبدل‌های انرژی موج نقش حیاتی در برداشت انرژی یکنواخت ایفا می‌کنند. در این مقاله، تغییرات زیست محیطی چشمگیر در اقیانوس‌ها را در نظر گرفته‌ایم تا یک سیستم کنترل فیدبک هوشمند را برای کاهش تأثیر اغتشاشات و اثرات باد متغیر بر عملکرد مبدل‌های انرژی امواج پیشنهاد دهیم. یادگیری تقویتی عمیق به عنوان یک تکنیک هوش مصنوعی قدرتمند، توانمند به تشخیص مبدل‌های امواج مانند مدل‌های جعبه سیاه است. بنابراین، با استفاده از یادگیری تقویتی عمیق، متغیرهای حالتی که نمی‌توان به طور مستقیم اندازه گیری کرد و اغتشاش‌ها به طور همزمان در بخش نظارت بر حالت تخمین زده می‌شوند. ضعف در داده‌های اندازه گیری و نیاز به کاربرد بلادرنگ تعداد محدودی از لایه‌ها در شبکه‌های عصبی عمیق را با پیاده‌سازی مشاهده‌گر حالت توسعه یافته مبتنی بر روش غوطه‌وری و تغییر ناپذیری که مقابله با نویزهای خارجی ناخواسته را بهبود می‌بخشد، جبران می‌شود. در سیستم کنترل هوشمند کلی، پارامترهای تخمینی به عنوان شبکه‌های عملگر–منتقد وارد مدل یادگیری تقویتی عمیق می‌شوند. شبکه عملگر اولیه مسئول پیش‌بینی ورودی کنترل است، در حالی که شبکه منتقد بعدی معیار تصمیم‌گیری برای ارزیابی دقت مقدار پیش‌بینی عملگر را تعیین می‌کند. سپس، مقدار خروجی مرحله منتقد از طریق لایه‌ها به عقب برگردانده می‌شود تا وزن‌های شبکه به‌روز شوند. نتایج آزمون شبیه‌سازی در نرم‌افزار متلب نشان‌دهنده هم‌گرایی پارامترها/حالت‌های اندازه‌گیری نشده با مقادیر واقعی متناظر و اهمیت روش یادگیری تقویتی عمیق هوشمندانه طراحی‌شده جدید است.