Renewable Energy 246 (2025) 122887

Contents lists available at ScienceDirect

Renewable Energy

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

Adaptive optimization of wave energy conversion in oscillatory wave surge converters via SPH simulation and deep reinforcement learning

Mai Ye [a], Chi Zhang [b], Yaru Ren [c], Ziyuan Liu [b], Oskar J. Haidn [a], Xiangyu Hu [a,*]

[a] TUM School of Engineering and Design, Technical University of Munich, Boltzmannstraße 15, 85748 Garching bei München, Germany
[b] Huawei Technologies Munich Research Center, Riesstraße 25, 80992, Munich, Germany
[c] State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Section of Chengdu No.24 Southern Yihuan, 610065, Chengdu, China

ARTICLE INFO

Dataset link: https://github.com/Xiangyu-Hu/SPHinXsys.git

Keywords:
Smoothed particle hydrodynamics (SPH)
Oscillating wave surge converter (OWSC)
Wave–structure interactions
Deep reinforcement learning (DRL)
Damping coefficient

ABSTRACT

The nonlinear damping characteristics of the oscillating wave surge converter (OWSC) significantly impact the performance of the power take-off system. This study presents a framework by integrating deep reinforcement learning (DRL) with numerical simulations of OWSC to identify optimal adaptive damping policy under varying wave conditions, thereby enhancing wave energy harvesting efficiency. The open-source multiphysics library SPHinXsys establishes the numerical environment for wave interaction with OWSCs. Subsequently, a comparative analysis of three DRL algorithms is conducted using the two-dimensional (2D) numerical study of OWSC interacting with regular waves. The results reveal that artificial neural networks capture the nonlinear characteristics of wave–structure interactions and provide efficient PTO policies. Notably, the soft actor–critic algorithm demonstrates exceptional robustness and accuracy, achieving a 10.61% improvement in wave energy harvesting. Furthermore, policies trained in a 2D environment are successfully applied to the three-dimensional study, with an improvement of 22.54% in energy harvesting. The optimization effect becomes more significant with longer wave periods under regular waves with consistent wave height. Additionally, the study shows that energy harvesting is improved by 6.42% for complex irregular waves. However, for the complex dual OWSC system, optimizing the damping characteristics alone is insufficient to enhance energy harvesting.

1. Introduction

Considering the significant environmental issues caused by the extensive use of fossil fuels, including pollution, greenhouse gas emissions, and ecological destruction, there has been a marked increase in the study of clean and renewable energy sources. Among these, wave energy stands out due to its substantial potential (with a minimum estimated capacity of around 0.2 TW), high energy density, and the advantage of not occupying land resources [1]. These features have attracted significant research and development investments, making wave energy a crucial component in transitioning to a sustainable energy future. Historically, most research has focused on extracting energy from the heave motion of deep-water systems, mainly due to the common belief that nearshore wave resources are significantly lower than those in deeper waters [2]. However, since the beginning of the 21st century, the concept of exploitable wave energy resources has become more realistic [3]. In many nearshore locations, the exploitable resource is typically only 10%–20% lower than that offshore [4]. As a result, the extraction of wave energy from the surge motion of waves in nearshore waters has gained increasing attention.

Typical wave energy converters (WECs) can be classified into three main categories based on their working principles: oscillating water column (OWC) devices, which use the oscillating water column to compress air and drive a turbine [5], over-topping devices, which utilize the potential energy of waves as they spill over a barrier [6], and wave-activated bodies, which exploit the heave, surge, roll, or pitch motions depending on their construction [7]. Oscillating wave surge converters (OWSCs) are typical wave-activated bodies used in nearshore waters, usually employing bottom-hinged flap mechanisms. Notable examples include the products of WaveRoller, and Oyster [8]. The primary distinction between these two lies in the positioning of their flaps: WaveRoller's flap is wholly submerged in seawater, whereas Oyster's flap has an upper edge that protrudes above the water surface [9]. Research conducted at Queen's University Belfast suggests that while partial submersion enhances the impact pressure exerted by the rotating flap, the flap inherently decouples from the wave as the oscillation amplitude increases [10]. This decoupling effect ensures that the wave-induced loads remain manageable, thereby safeguarding the structural integrity of the Oyster, even under extreme

* Corresponding author.
E-mail address: xiangyu.hu@tum.de (X. Hu).

https://doi.org/10.1016/j.renene.2025.122887
Received 3 October 2024; Received in revised form 11 March 2025; Accepted 12 March 2025
Available online 20 March 2025

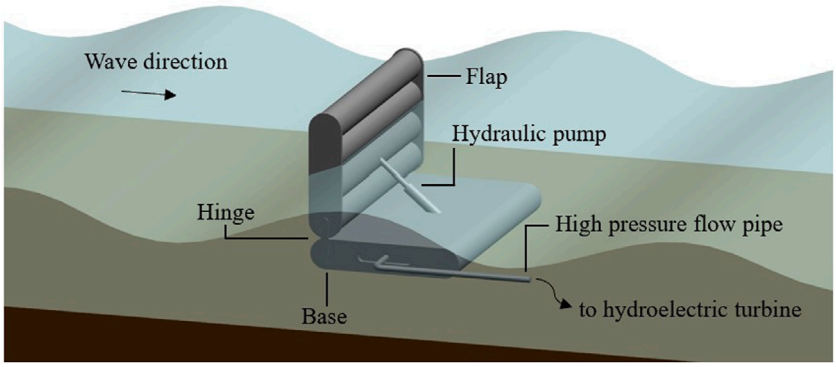


Fig. 1. Schematic of Oyster® (a type of OWSC) under waves.

sea conditions. The structure of the Oyster is shown in Fig. 1. The flap is connected to the base via a hinge and oscillates back and forth in response to the incident waves. This oscillatory motion drives the power take-off (PTO) system, which utilizes a hydraulic pump to channel high-pressure water through a pipeline to a hydroelectric turbine, generating electricity [11].

According to the experimental study of the wave interacting with OWSCs by Henry et al. [12], the flap of an OWSC actively impacts the trough of incoming waves, generating breaking waves. Chow et al. [13] discovered that a system of two tandem OWSCs could improve overall wave energy conversion efficiency and determined the optimal separation distance between them based on Bragg reflection. Brito et al. [14] demonstrated through scale model experiments that the capture width ratio (CWR) and response amplitude operator of an OWSC exhibit weak correlation under both regular and irregular wave conditions.

The numerical study is attracting more and more attention as it can be applied to explore the detailed mechanisms of wave–structure interactions (WSI), optimize the structure of OWSCs, and improve the energy harvesting efficiency of PTO systems. Cheng et al. [9,15] developed a two-dimensional (2D) and three-dimensional (3D) higher-order boundary element method (HOBEM) model to estimate the performance of an OWSC. Renzi and Dias [16] proposed a semi-analytical model for the 3D computation of OWSCs. Although these methods offer high computational efficiency, they lack the accuracy of predicting nonlinear phenomena such as slamming and overtopping. Two approaches for this problem are based on the Navier–Stokes equations with either mesh or particle-based methods. For mesh-based methods, Schmitt et al. [17] explore the nonlinear relationship between wave height and the optimal damping of the OWSC. The numerical simulations by Wei et al. [18] also agreed with experimental results in predicting the wave height and the pressure distribution on the flap. Jiang et al. [19] discussed the impact of different damping strategies on power generation efficiency in PTO systems. However, these methods come with significant computational costs due to the inclusion of additional Volume of Fluid (VOF) equations and the necessity of using dynamic mesh techniques to solve for the motion of the flap [20].

Particle-based methods, i.e., smoothed particle hydrodynamics (SPH), are particularly well-suited for addressing challenges involving large deformations and complex free surface flows [21,22], making them a compromising alternative to study the hydrodynamic interactions between waves and WECs. Specifically, for OWSC calculations, Henry et al. [12] and Rafiee et al. [23] used a modified SPH method to perform 2D and 3D numerical simulations. Their results indicated that 3D simulations can more accurately predict the pressure distribution on the flap. Brito et al. [24] presented a numerical model combining DualSPHysics for wave computation and Chrono for nonlinear mechanical constraint systems of OWSCs. Later, they updated the numerical model to consider most constraints, such as the PTO system, revolute joints, and frictional contacts [25]. Liu et al. [26] quantitatively analyzed the effects of parameters such as load, flap mass, thickness, hinge height, and damping of the PTO system on motion resonance and wave absorption of OWSC. Zhang et al. [27] used a Riemann-based weakly compressible method based on SPHinXsys and Simbody to compute incident waves and WSI. They demonstrated that the solver could accurately predict wave height and the pressure distribution on the flap while significantly reducing computation time, showing great potential for practical applications.

Currently, the performance optimization of OWSCs primarily relies on theoretical analysis [28] and numerical simulations [29] under specific wave conditions to conduct parametric studies on the structure and position of flaps or damping of the PTO system. There is still a lack of research on effective control strategies to enhance the performance of OWSCs. In comparison, control strategy methods have already been applied to optimize the wave energy absorption performance of other WEC devices [30]. However, the interaction between waves, especially irregular waves, and OWSCs is complex, highly stochastic, and nonlinear. Models such as latching and model predictive control (MPC) exhibit poor robustness and are highly dependent on the accuracy of the predictive model. Inaccurate models can significantly affect performance [31]. In the past decade, with the rapid advancement of artificial intelligence (AI), deep reinforcement learning (DRL) has demonstrated significant capabilities in the field of active flow control [32]. DRL combines artificial neural networks (ANN) and reinforcement learning (RL). Since ANN uses nonlinear activation functions and can effectively fit any function, DRL, compared to traditional RL, enhances the exploration and capture of high-dimensional state spaces, making it suitable for WSI problems [33]. Research on optimizing WECs using DRL employs potential wave theory as the environment. Studies have shown that DRL can enhance the energy harvesting efficiency of WECs [34,35]. Additionally, compared to directly using computational fluid dynamics (CFD) as the environment, this approach significantly reduces the time required to train the ANN. However, potential wave theory is less effective at capturing the coupling effects between multiple physical fields. Only Liang et al. [36] have combined 2D CFD with DRL to optimize the wave energy conversion of horizontal floating cylinders under irregular waves.

In this paper, we will first establish a new platform that combines CFD with DRL. Based on our previous work, we will use SPHinXsys, a multi-physics library based on the SPH method, and Simbody as the numerical computation platform for the bottom-hinged OWSC [37]. Given that mainstream DRL platforms like Tianshou employ neural networks such as PyTorch, which are based on the Python environment, we will use Pybind11 to package the relevant OWSC code into a dynamic link library for invocation in the OpenAI Gymnasium environment [38]. This standardized environment will facilitate the direct application of various DRL algorithms available on the Tianshou platform, enabling us to explore the impact of these algorithms on performance improvement. The remainder of this paper is organized as follows: Section 2 introduces the Riemann-based SPH method for FSI modeling

in SPHinXsys and Simbody. Section 3 discusses the details of our CFD-DRL framework, the DRL training environment, and the mainstream RL algorithms. Section 4 analyzes the performance differences of various DRL algorithms, explores the applicability of 3D numerical computations and different wave period to the policies, verifies that the adaptive damping coefficient policy can be applied to random irregular waves, and investigates the feasibility of wave energy optimization under the dual OWSC systems.

2. Numerical modeling

2.1. Governing equations

In the Lagrangian framework, the mass and momentum conservation equations for incompressible and viscous fluid can be written as

$$\begin{cases} \dfrac{d\rho}{dt} = -\rho\nabla\cdot\mathbf{v}, \\ \dfrac{d\mathbf{v}}{dt} = -\dfrac{1}{\rho}\nabla p + \nu\nabla^2\mathbf{v} + \mathbf{g}. \end{cases} \tag{1}$$

Here, ρ is the density of the fluid, $\mathbf{v}$ the velocity, p the pressure, ν the kinematic viscosity, and $\mathbf{g}$ is the gravity. An artificial isothermal equation of state (EoS) is used to close the system of Eq. (1)

$$p = c^2(\rho - \rho^0), \tag{2}$$

where ρ^0 is the initial density, $c = 10v_{max}$ the artificial speed of sound [39]. $v_{max} = 2\sqrt{gh}$ is the maximum anticipated particle velocity in the flow, $g = |\mathbf{g}|$, h the water depth.

Eq. (1) can be discretized as

$$\begin{cases} \dfrac{d\rho_i}{dt} = 2\rho_i\sum\limits_j V_j(U^* - \mathbf{v}_i\cdot\mathbf{e}_{ij})\dfrac{\partial W_{ij}}{\partial r_{ij}}, \\ \dfrac{d\mathbf{v}_i}{dt} = -m_i\sum\limits_j \dfrac{2P^*}{\rho_i\rho_j}\nabla W_{ij} + m_i\sum\limits_j \dfrac{2\mu}{\rho_i\rho_j}\dfrac{\mathbf{v}_{ij}}{r_{ij}}\dfrac{\partial W_{ij}}{\partial r_{ij}} + \mathbf{g}_i. \end{cases} \tag{3}$$

Here, m_i and ρ_i are the mass and density of particle i, V_j the particle volume, $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$ particle relative velocity, and μ is the dynamic viscosity. Also, $\nabla W_{ij} = \mathbf{e}_{ij}(\partial W(r_{ij}, h)/\partial r_{ij})$ with $\mathbf{e}_{ij} = \mathbf{r}_{ij}/r_{ij}$ and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, and W_{ij} represents the Kernel gradient [40]. Besides, U^* and P^* are the solutions of the one-dimensional Riemann problem constructed along the line pointing from particle i to j. The left and right initial states of the Riemann problem can be reconstructed as

$$\begin{cases} (\rho_L, U_L, P_L, c_L) = (\rho_i, -\mathbf{v}_i\cdot\mathbf{e}_{ij}, p_i, c_i), \\ (\rho_R, U_R, P_R, c_R) = (\rho_j, -\mathbf{v}_j\cdot\mathbf{e}_{ij}, p_j, c_j). \end{cases} \tag{4}$$

The linearized Riemann solver coupled with the weighted kernel gradient correction (WKGC) [41] is adopted to solve this Riemann problem

$$\begin{cases} U^* = \dfrac{\rho_L c_L U_L + \rho_R c_R U_R + P_L - P_R}{\rho_L c_L + \rho_R c_R}, \\ P^* = \dfrac{\rho_L c_L P_R\widetilde{\mathbf{B}}_i + \rho_R c_R P_L\widetilde{\mathbf{B}_j} + \rho_L c_L\rho_R c_R\beta(U_L - U_R)}{\rho_L c_L + \rho_R c_R}, \end{cases} \tag{5}$$

where the low dissipation limiter $\beta = min(3max(U_L - U_R, 0)/\overline{c}, 1)$, $\overline{c} = (\rho_L c_L + \rho_R c_R)/(\rho_L + \rho_R)$, the correction matrix $\widetilde{\mathbf{B}}_i = \omega_1\mathbf{B}_i + (1-\omega_1)\mathbf{I}$, $\mathbf{B}_i = (-\sum_j \mathbf{r}_{ij}\otimes\nabla W_{ij}V_j)^{-1} = (\mathbf{A}_i)^{-1}$, $\mathbf{I}$ the identity matrix, $\omega_1 = |\mathbf{A}_i|/(0.3 + |\mathbf{A}_i|)$.

2.2. Dual-criteria time stepping

In order to improve computational efficiency, dual-criteria time-stepping method is adopted. Specifically, the update frequency of the particle configuration is controlled by the advection criterion, while the integration of pressure relaxation is determined by a smaller time step size based on the acoustic criterion. Following Zhang et al. [42], the time step size of the advection criterion Δt_{ad} and the acoustic criterion Δt_{ac} are

$$\begin{cases} \Delta t_{ad} = CFL_{ad}\min(\dfrac{h}{|\mathbf{v}|_{max}}, \dfrac{h^2}{\nu}), \\ \Delta t_{ac} = CFL_{ac}(\dfrac{h}{c + |\mathbf{v}|_{max}}), \end{cases} \tag{6}$$

where $CFL_{ad} = 0.25$ and $CFL_{ac} = 0.6$.

2.3. SPHinXsys and Simbody coupling

The fluid density ρ_i will be initialized firstly at the beginning of the advection time step Δt_{ad} as

$$\rho_i = \max(\rho^*, \rho^0\frac{\sum W_{ij}}{\sum W_{ij}^0}), \tag{7}$$

where ρ^* denotes the density before re-initialization and ρ^0 represents the initial reference value. The viscous force $\mathbf{f}_{av}$ is also computed at this stage. Subsequently, pressure relaxation is carried out over the next several acoustic time steps Δt_{ac} using the position-based Verlet scheme proposed by Zhang et al. [43]

$$\begin{cases} \rho_i^{n+\frac{1}{2}} = \rho_i^n + \dfrac{1}{2}\Delta t_{ac}(\dfrac{d\rho_i}{dt})^{n+\frac{1}{2}}, \\ \mathbf{r}_i^{n+\frac{1}{2}} = \mathbf{r}_i^n + \dfrac{1}{2}\Delta t_{ac}\mathbf{v}_i^n. \end{cases} \tag{8}$$

The particle's velocity $\mathbf{v}_i$, density ρ_i, and position $\mathbf{r}_i$ are updated to the mid-points as

$$\begin{cases} \mathbf{v}_i^{n+1} = \mathbf{v}_i^n + \Delta t_{ac}(\dfrac{d\mathbf{v}_i}{dt})^{n+1}, \\ \rho_i^{n+1} = \rho_i^{n+\frac{1}{2}} + \dfrac{1}{2}\Delta t_{ac}(\dfrac{d\rho_i}{dt})^{n+\frac{1}{2}}, \\ \mathbf{r}_i^{n+1} = \mathbf{r}_i^{n+\frac{1}{2}} + \dfrac{1}{2}\Delta t_{ac}\mathbf{v}_i^{n+1}. \end{cases} \tag{9}$$

Then, the forces acting on the flap of the OWSC will be computed, which are composed of two main components, the pressure force $\mathbf{f}_{ap}$ and the viscous force $\mathbf{f}_{av}$, as

$$\begin{cases} \mathbf{f}_{ap} = -2\sum\limits_i V_i V_a\dfrac{p_i\rho_a^d + p_a^d\rho_i}{\rho_i + \rho_a^d}\nabla_a W_{ai}, \\ \mathbf{f}_{av} = 2\sum\limits_i \nu V_i V_a\dfrac{\mathbf{v_i} - \mathbf{v_a^d}}{r_{ai}}\dfrac{\partial W_{ai}}{\partial r_{ai}}. \end{cases} \tag{10}$$

Here, the subscript a denotes the solid particle index. The imaginary pressure p_a^d and velocity $\mathbf{v}_a^d$ read as follows

$$\begin{cases} p_a^d = p_i + \rho_i\max(0, (\mathbf{g} - \dfrac{\mathbf{dv}_a}{dt})\cdot\mathbf{n})(\mathbf{r}_{ai}\cdot\mathbf{n}), \\ \mathbf{v}_a^d = 2\mathbf{v}_i - \mathbf{v}_a, \end{cases} \tag{11}$$

where $\mathbf{n}$ represents the normal direction of the solid body to fluid. The total force and total torque τ acting on the flap can be written as

$$\begin{cases} \mathbf{F} = \sum\limits_{a\in N}\mathbf{f}_a = \sum\limits_{a\in N}(\mathbf{f}_{ap} + \mathbf{f}_{av}), \\ \tau = \sum\limits_{a\in N}(\mathbf{r}_a - \mathbf{r}_G)\times\mathbf{f}_a, \end{cases} \tag{12}$$

where N denotes the total number of solid particles and $\mathbf{r}_G$ is the position vector of the flap mass center.

At the end of each time step, the total force and total torque calculated from SPHinXsys are transmitted to Simbody for solving the Newton–Euler equations

$$\begin{cases} \mathbf{F} = m\mathbf{I}_0\dfrac{d\mathbf{v}}{dt}, \\ \tau = \mathbf{J}_0\dfrac{d\Omega}{dt} - k_d\Omega. \end{cases} \tag{13}$$

where m is the mass of the flap, $\mathbf{I}_0$ the identity matrix, $\mathbf{J}_0$ the moment of the inertia about the center of mass, ω the angular velocity and k_d the

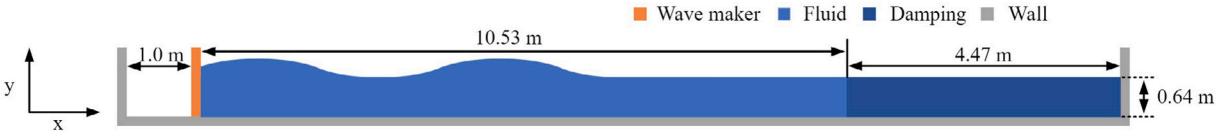


Fig. 2. The 2D water tank geometry for wave generation verification.

damping coefficient. After updating the position, velocity, and normal direction, the new kinematic state will be imported back to SPHinXsys, and the loop will continue.

It is worth noting that during the computation of Eq. (13), the k_d is variable within a specific range, and it can be rapidly implemented through the damping update definition in Simbody.

2.4. Wave generation

A piston-type wave maker is used to generate the regular and irregular waves [44]. For fluid particles around the wall region, the interaction is determined by solving a one-sided Riemann problem along the wall-normal direction [42] where the left state is defined

$$(\rho_L, U_L, P_L) = (\rho_i, -\mathbf{n}_w \cdot \mathbf{v}_i, p_i), \tag{14}$$

where $\mathbf{n}_w$ is the normal direction of the wall, and i represents the fluid particles. The right state of the velocity U_R and pressure P_R are assumed as

$$\begin{cases} U_R = -U_L + 2\mathbf{u}_w, \\ P_R = P_L + \rho_i \mathbf{g} \cdot \mathbf{r}_{iw}, \end{cases} \tag{15}$$

where $\mathbf{u}_w$ is wall velocity, $\mathbf{r}_{iw} = \mathbf{r}_w - \mathbf{r}_i$, ρ_i is computed from Eq. (2).

The displacement function of the wave maker r_a for regular waves relies on

$$\begin{cases} r_a = E\sin(2\pi f t + \psi), \\ E = \dfrac{H(\sinh(2kh) + 2kh)}{\sinh(2kh)\tanh(2kh)}. \end{cases} \tag{16}$$

Here, E is the wave stroke, f the wave frequency, ψ the wave phase, H the wave height, and h the water depth. The wave number k followed by the dispersion relation [45]

$$\omega^2 = gk\tanh(kh), \tag{17}$$

where $\omega = 2\pi f$ is the wave angular frequency.

For irregular waves, the JONSWAP spectrum exhibits a more concentrated wave energy than the Pierson–Moskowitz spectrum, making it more effective in describing the energy distribution of waves in the developmental stage [5]. Considering that OWSCs are predominantly placed in nearshore areas, where waves are typically in this developmental stage, it is appropriate to use the JONSWAP spectrum to generate irregular waves [46]

$$\begin{cases} S(f) = \beta_J H_p^2 T_p^{-4} f^{-5} \exp[-1.25(T_p f)^{-4}] \gamma_J^{\exp[-\frac{(T_p f - 1)^2}{(2\delta_J^2)}]}, \\ \beta_J = \dfrac{0.0624(1.094 - 0.01915\ln\gamma_J)}{0.230 + 0.0336\gamma_J - 0.185(1.9 + \gamma_J)^{-1}}, \end{cases} \tag{18}$$

where H_p is the main wave height, T_p the peak wave period, $f = \omega/2\pi$ the wave frequency, and $\gamma_J = 3.3$ the peak enhancement factor, δ_J dependent on the peak frequency $f_p = 1/T_p$ [47]. Our study employed a combination of N random regular waves to simulate nonlinear waves. The wave number k_n for each regular wave was consistent with Eq. (17). The displacement equation r_N can be written as

$$r_N = \sum_{n=1}^{N} E(f_n)\cos(2\pi f_n t + \psi_n). \tag{19}$$

Here, $E(f_n) = \sqrt{2S_\omega(f_n)\Delta f}$, $f_1 = 0$ Hz, $f_N = 3f_p$, $\Delta f = 1/T_{total}$ with T_{total} the total time of simulation, ψ_n represents the random phases, and $S_\omega(f_n)$ is defined as

$$S_\omega(f_n) = S(f_n)\left(\frac{4\sinh^2(kh)}{2kh + \sinh(2kh)}\right)^{-2}. \tag{20}$$

In addition, to prevent numerical divergence caused by excessive movement of the wave maker during the initial computation, we introduced a relaxation time $t_{rex} = 1$ s, and the final wave maker displacement function is as follows

$$r_N = \begin{cases} \sin(\dfrac{\pi t}{2}) r_N, & t \le t_{rex} \\ r_N, & t > t_{rex}. \end{cases} \tag{21}$$

Also, to mitigate the impact of wave reflection off the wall on the motion of the OWSC, the wave–particle velocity $\mathbf{v}$ in the damping zone is given by

$$\mathbf{v} = \mathbf{v}_0(1.0 - \alpha\Delta t(\frac{\mathbf{r} - \mathbf{r}_0}{\mathbf{r}_1 - \mathbf{r}_0})). \tag{22}$$

$\mathbf{v}_0$ the fluid particle velocity at the entrance of the damping zone, the reduction coefficient α is set as 5.0, $\mathbf{r}_0$ and $\mathbf{r}_1$ are the initial and final position vectors of the damping zone.

2.5. Numerical model validation

In this section, we verify the accuracy of generating regular and irregular waves and validate the numerical simulations of the 2D and 3D OWSC.

The geometry of the 2D numerical simulation is shown in Fig. 2. The total length of the tank L is 15 m, the water depth h 0.691 m, and the damping zone is 4.47 m long. The parameters of the second-order Stokes wave are $H = 0.2$ m and wave period $T = 2$ s [45]. As shown in Fig. 3, our numerical result of wave frequency and amplitude is very close to the analytical solution [45], which indicates that the wave is generated correctly.

Furthermore, we set three different particle resolutions from coarse to fine. The resolution of $dp = 0.015$ m not only improves the accuracy of the calculation results but also reduces the computation time. Therefore, this resolution will be used for all subsequent 2D simulations.

In Fig. 4, the typical parameters for irregular waves are $H_p = 0.2$ m and $T_p = 2.0$ s. Note that T_{total} is set for 40 s and 100 s for comparison. The frequency spectrum is obtained by applying the Fast Fourier Transform (FFT) to the free surface height, as shown in Fig. 5. Notably, our numerical results demonstrate a high degree of agreement with the analytical results derived from the JONSWAP spectrum, and for RL training, $T_{total} = 40$ s is set for one training episode.

The present method for modeling wave interaction with the OWSC is validated in both 2D and 3D. The water tank and OWSC geometries are based on the experiment conducted at Queen's University Belfast [18], as shown in Fig. 6. The length, width, and height of the wave tank are 18.4 m, 4.58 m, and 1.0 m. The flap shape of the OWSC is simplified as a box-type with the dimensions of 0.12 m × 1.04 m × 0.48 m. It is set at 7.92 m far from the wave maker in the x direction and the middle in the z direction. The water depth h is 0.691 m, and the hinge height is 0.16 m. The mass and inertia of the flap are 33 kg and 1.84 kg m^2. The damping coefficient k_d is set to 20. In addition, we set three free surface height sensors in the x direction, which are WP04 (3.99 m), WP05 (7.02 m), and WP12 (8.82 m).

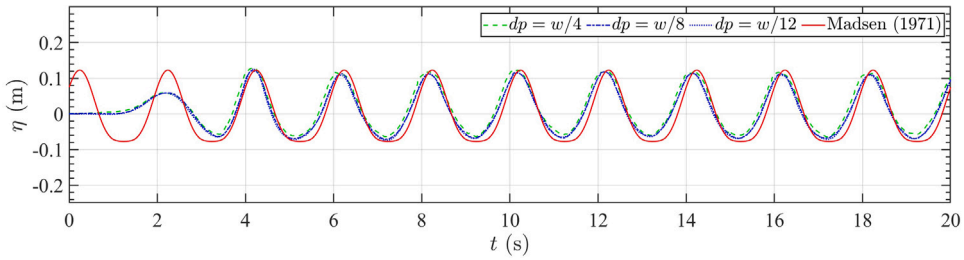


Fig. 3. Comparison of free surface heights under different resolutions at $x = 4.0$ m with theoretical results.

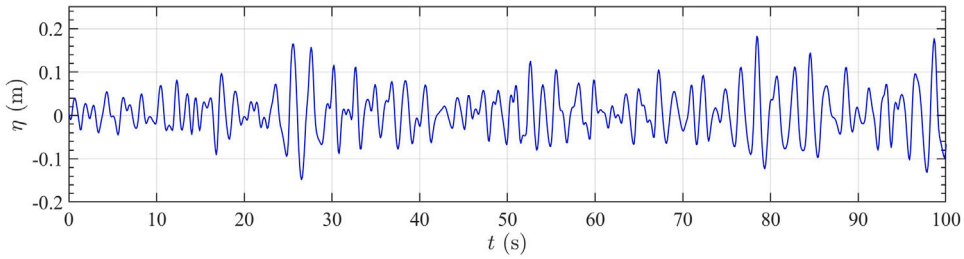


Fig. 4. Time series of the free surface elevation at $x = 0.2$ m for the irregular wave.

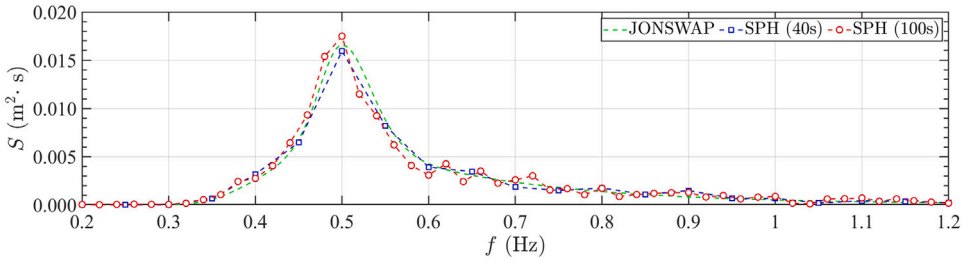


Fig. 5. Comparison between the analytical JONSWAP spectrum of Eq. (18) and the SPH spectrum corresponding to different time series.

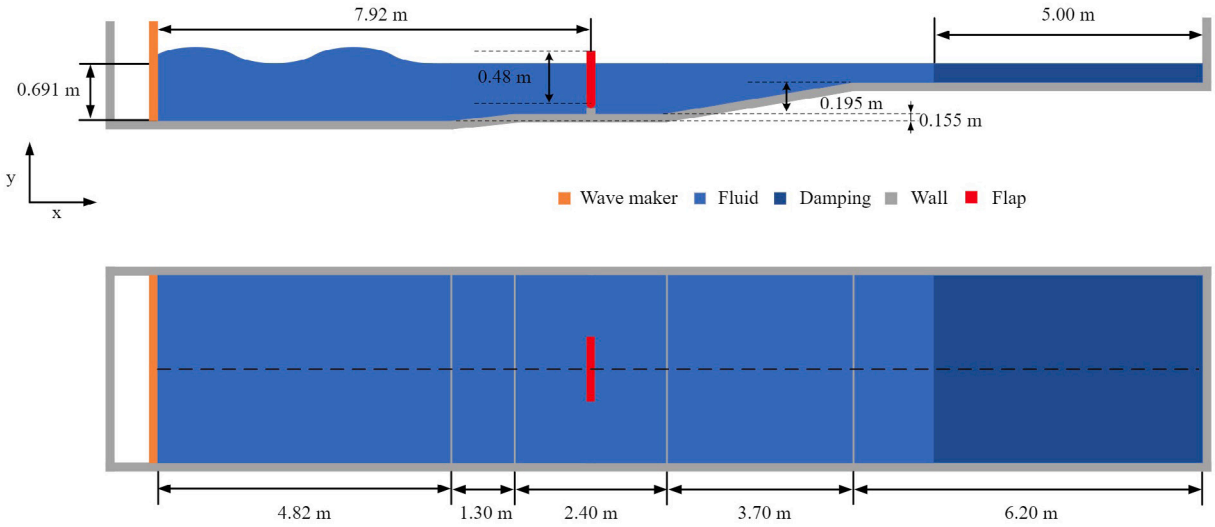
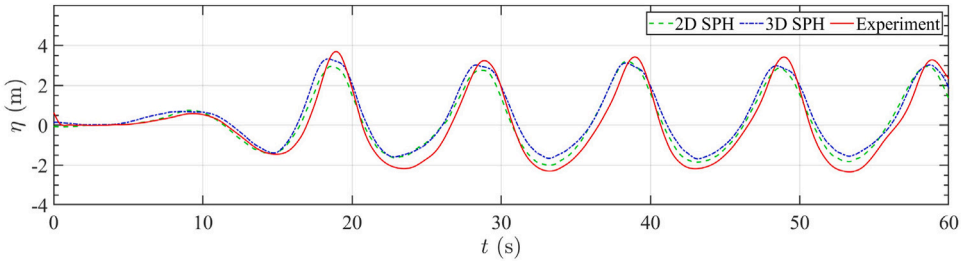
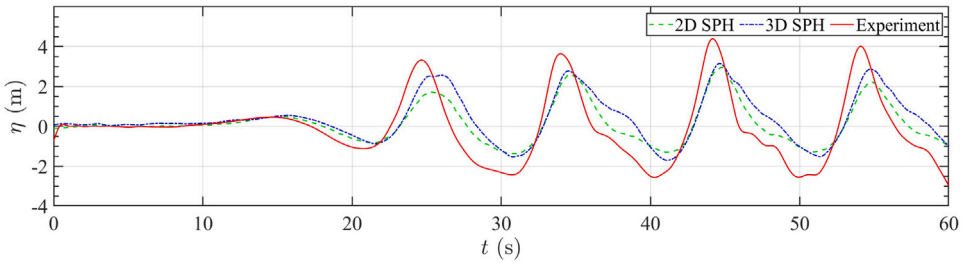


Fig. 6. Schematic of the wave tank and the OWSC mode.

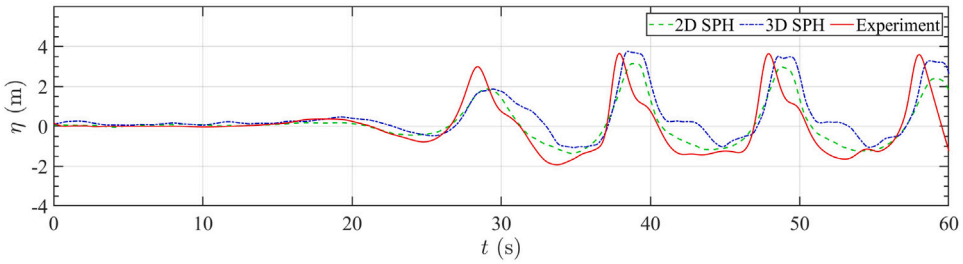
The wave maker creates a simple harmonic wave with the wave height $H = 0.2$ m and wave period $T = 2$ s. The simulation parameter is 1 : 25 in scale, and the results presented herein have been converted to full scale as in Ref. [27]. The 2D simulation and training is computed on a Mac OS system, with an Apple M1 Max core and 32 GB RAM, while the 3D simulation is carried on a Windows system, with an AMD Ryzen 9 7950X core and 48 GB RAM. The total fluid particle numbers are 40 027 and 1.542 million, for 2D and 3D simulations, respectively. From Fig. 7, we can see that WP04 is far away from the flap, so the interaction of the wave and flap has little influence on the free surface height, and our results show that the wave maker can generate an accurate sine wave with minor errors. WP05 and WP12 are set near the flap, and we can find that our simulations can capture the influence of interaction on free surface height in both places. Furthermore, by



(a) WP04



(b) WP05



(c) WP12

Fig. 7. Comparison of free surface elevations for wave height $H = 5.0$ m and wave period $T = 10$ s at different wave probes. (a) $x = 3.99$ m for WP04, (b) $x = 7.02$ m for WP05, and (c) $x = 8.82$ m for WP12.

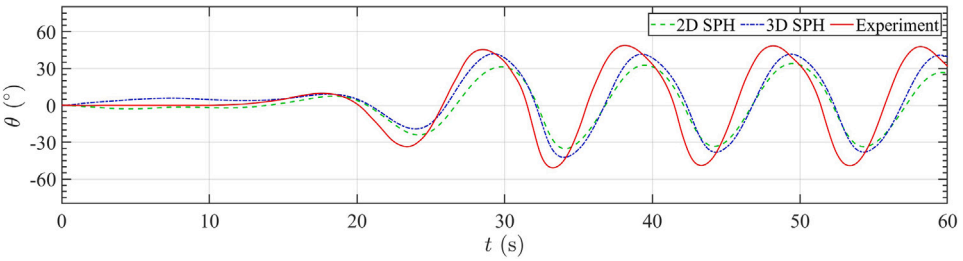


Fig. 8. Comparison of the flap rotation.

comparing the free surface heights at WP05 and WP12, it is evident that the free surface height significantly decreases after passing through the OWSC, indicating an energy reduction. This reduction indirectly demonstrates that a portion of the energy has been converted into wave energy. Fig. 8 shows that the flap rotation simulation results are consistent with the experiment. The 3D results are much better than 2D simulations, where antisymmetric diffracted waves traveling in all directions, including tangentially to the flap, can induce near-resonant phenomena that enhance the exciting torque on the converter [48].

Notably, a 2D simulation requires only 1.2 min of computing time for 12 s of physical time calculation, whereas a 3D simulation demands 4.5 h. Given that each DRL training session necessitates at least 200 episodes, we employ 2D simulations for training and utilize 3D simulations for policy validation.

3. Direct deep reinforcement learning

Fig. 9 provides an overview of the platform. The DRL training process consists of two key components: the environment and the agent. The sampled state vector from SPHinXsys is normalized and passed to the agent. The reward is calculated based on the change in state between two consecutive time steps and the variation in reward parameters during the action application. We show a typical framework of the DRL algorithm: soft actor critics (SAC), which takes actor–critic architecture [49]. The policy network (actor) outputs actions fed into Simbody and critic networks in real-time. The critic networks evaluate the quality of these actions, and the smaller Q_i is chosen to update the policy network with gradient ascent. The critic networks will be updated with target networks using gradient descent.

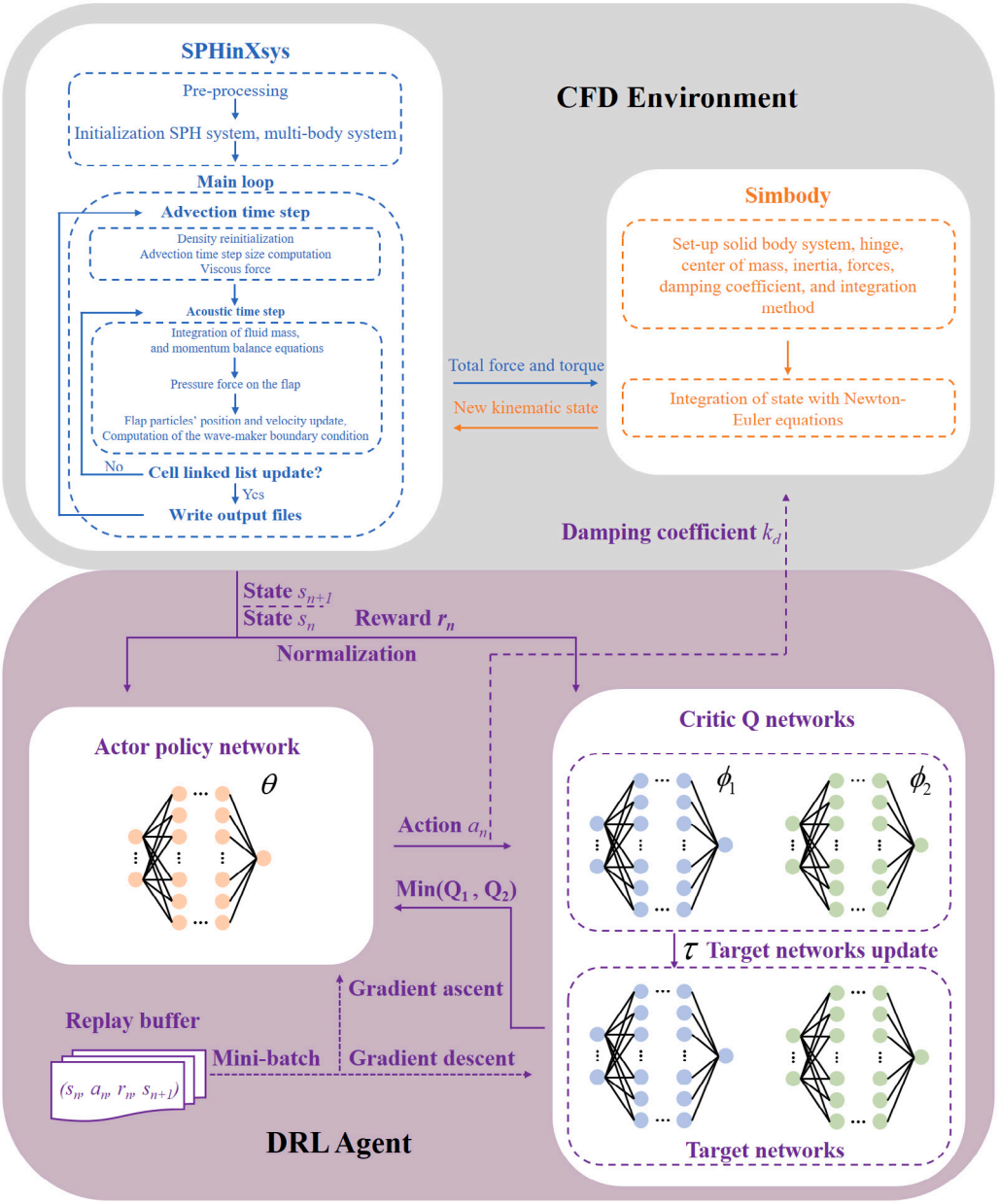


Fig. 9. The current CFD-DRL framework encompasses an integrated interaction process between two primary components: the CFD Environment and the DRL Agent.

3.1. DRL environments

The numerical simulation environment, observation probes, and action transition functions are created using SPHinXsys. The standard DRL environment is built based on OpenAI's Gym library in Tianshou [50], which includes two essential functions: reset and step. In the reset function, the numerical environment is initialized, and the initial observations are collected using probes. The step function receives action values from the DRL algorithm, passes them to SPHinXsys for numerical calculations under the current action, collects new observations, and gets the reward.

Rabault et al. [51] show that more observations will give the agent more information to update its network and improve the training result. Therefore, 38 observations are set as the state vector s_n from the environment to capture the structure of the regular wave and its impact on the flap. The components of the observations are based on two parts, as shown in Fig. 10. The first part is the wave properties, including velocity and free surface height at five positions, starting from $x = 3.0$ m to $x = 7.6$ m, which is close to one wavelength. Another 5 points from $y = 0.3$ m to $y = 0.7$ m on the front panel of the OWSC device are also set to get the wave velocity. The second part is the characteristics of the flap, including flap rotation and angular velocity, and damping coefficient k_d of the PTO system. Besides, it is ideal that the observations from 2D and 3D simulations can be close. Thus, the observation positions are set on the middle plane in the z-axis as shown in Fig. 6, and only x-axis and y-axis properties are considered.

In the present work, the action a_n in one action time step $t_a = 0.1$ s of variation in the damping coefficient $|\Delta k_d| \leq 25$ N m s/rad. However, directly changing the damping coefficient has the potential risk to the computation divergence. Therefore, the current numerical damping coefficient k_d^n and the subsequent numerical damping coefficient k_d^{n+1} have a simple linearly increasing change of $\Delta k_d / M$ during the time step of t_a / M. Considering that the $\Delta t_{ac} = 0.00021$ s, $M = 10$ is sufficient to ensure accuracy in the 2D simulation. Note that a large damping coefficient will make the flap's rotation challenging. A small damping coefficient will result in larger deflection angles, which may

Table 1
The variations of the total wave energy conversion E_t in terms of damping coefficients.

k_d (N m s/rad)	10	20	30	40	45	50	60	70	80	90	100
E_t (J)	195.19	281.38	309.56	321.26	321.96	316.52	310.86	300.92	291.02	279.37	267.47

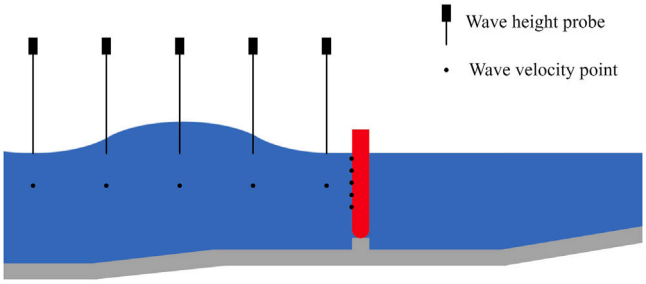


Fig. 10. Two main components of the observation vector.

cause the OWSC to break in real applications. In the present work, $0 \le k_d$ (N m s/rad) ≤ 100 is applied. As observed in Fig. 8, we can see that even if the damping coefficient is set to 0, the most extensive rotation of the flap is 50°, which is smaller than the critical value. Although the entire OWSC can still run under extreme boundary conditions, a penalty term is given in the reward as $P^* = -1 \cdot \left[(k_d + \Delta k_d < 0) \vee (k_d + \Delta k_d > 100)\right]$.

Based on Senol and Raessi's work [52], the instance energy harvesting in one action time step can be defined as

$$P_{take-off} = \sum_{n=0}^{M-1} k_d^n \left(\frac{\Omega_{n+1} + \Omega_n}{2}\right)^2. \tag{23}$$

Here, Ω_n is the flap angular velocity. It is evident that larger $P_{take-off}$ means more instantaneous wave energy harvesting by the OWSC. The study of Zhang et al. [27] shows that the average wave energy harvesting factor (CF) of 3D OWSCs can reach a peak when k_d^* is around 40 N m s/rad. Combined with Table 1, the verification results show that in the case of 2D OWSC, the total wave energy conversion reaches a maximum value at $k_d^* = 45$ N m s/rad. Then, the instance energy harvesting in one action time step with k_d^* is recorded as the baseline $P_{baseline}$ in the reward to help the agent distinguish good from harmful actions. The final reward can be calculated as

$$r_n = P_{take-off} - P_{baseline} + P^*. \tag{24}$$

3.2. DRL algorithms

RL algorithms can be broadly categorized into two types: on-policy and off-policy. On-policy algorithm updates its policy network after one episode and uses the newly updated policy to collect data in the next episode. The typical algorithm is proximal policy optimization (PPO) [53]. On the other hand, the policy for updating and collecting is different for the off-policy algorithm. The typical methods are twin delayed deep deterministic policy gradient (TD3) [54] and SAC.

More specifically, the action value function $Q_{\pi_\theta}(s_n, a_n)$ and the state value function $V_{\pi_\theta}(s_n)$ can be defined as:

$$\begin{cases} Q_{\pi_\theta}(s_n, a_n) = \mathbf{E}_{\pi_\theta}[\sum_{t=n}^{\infty}(\gamma_t r_t | s_t, a_t)], \\ V_{\pi_\theta}(s_n) = \mathbf{E}_{a_n \sim \pi_\theta}[\sum_{t=n}^{\infty}(\gamma_t r_t | s_t)], \end{cases} \tag{25}$$

where π_θ is the policy network with parameter θ, $\gamma_t \in [0, 1]$ the discount factor for the future reward, and $\sum_{t=n}^{\infty} \gamma_t r_t$ usually defined as return G_n. The policy network plays an important role in the optimization process, as its input is the current state s_n, and its output is the action a_n or the probability density function of the action $\pi_\theta(\cdot|s_n)$. The action value function $Q_{\pi_\theta}(s_n, a_n)$ represents the expected return obtained by taking action a_n in the current state s_n and then following the policy π_θ. The state value function $V_{\pi_\theta}(s_n)$ represents the expected return obtained by following the policy π_θ and the current state s_n.

3.2.1. The PPO algorithm

The core of the PPO algorithm is to build an objective function J to characterize the return G_n under the parameter θ of the current policy network. The best policy and return are obtained by updating θ to maximize the objective function. Based on the Policy Gradient Theorem (PGT), the objective function $J(\theta)$ can be written as [53]

$$J(\theta) = \mathbf{E}_{s_n \sim D}[\mathbf{E}_{a_n \sim \pi_{\theta_k}}[\frac{\pi_\theta(a_n|s_n)}{\pi_{\theta_k}(a_n|s_n)} \cdot A^{\pi_{\theta_k}}(s_n, a_n)]], \tag{26}$$

where D is the replay buffer, θ_k the old parameter of the policy network. Also, $A^{\pi_{\theta_k}}(s_n, a_n)$ is the advantage function which can be defined as

$$A^{\pi_{\theta_k}}(s_n, a_n) = r_n + \gamma V_\phi^{\pi_{\theta_k}}(s_{n+1}) - V_\phi^{\pi_{\theta_k}}(s_n), \tag{27}$$

where $V_\phi^{\pi_{\theta_k}}(s_n)$ is the estimated value of state s_n by the critic network with parameters ϕ, the superscript π_{θ_k} indicates that the data is collected using the policy employed during the kth iteration, and γ is the discount factor. The advantage function will not affect the expectation but improve the policy's performance [55]. A key feature of PPO is the clipped surrogate objective, which is designed to prevent huge policy updates. The objective function of $J(\theta)$ is then rewritten as follows

$$J(\theta) = \mathbf{E}_{s_n \sim D}\left[\mathbf{E}_{a_n \sim \pi_{\theta_k}}\left[\min\left(r_\theta(s_n, a_n), \right.\right.\right. \\ \left.\left.\left. \text{clip}(r_\theta(s_n, a_n), 1-\sigma, 1+\sigma)\right) \cdot A^{\pi_{\theta_k}}(s_n, a_n)\right]\right], \tag{28}$$

where $r_\theta(s_n, a_n) = \pi_\theta(a_n|s_n)/\pi_{\theta_k}(a_n|s_n)$, and $\sigma = 0.2$.

The critic network V_ϕ in PPO is primarily used to represent the state value function and its loss function is defined as

$$L(\phi) = \mathbf{E}_{s_n \sim D}\left[\left(V_\phi(s_n) - (r_n + \gamma V_{\phi_k}(s_{n+1}))\right)^2\right] \tag{29}$$

3.2.2. The TD3 algorithm

The TD3 algorithm is based on policy gradient methods, where the policy network π_θ outputs actions a_n directly. The value networks Q_{ϕ_i}, called critics, provide value estimates based on s_n and a_n suggested by the policy network. TD3 employs two value networks to mitigate overestimation issues and obtains more reliable value estimates using the minimum value predicted by the two networks [54]. Each network (the policy and the two value networks) is paired with a corresponding target network. Therefore, a total of six networks are utilized during the training process.

The objective function of a TD3 policy network is defined as

$$J(\theta) = \mathbf{E}_{s_n \sim D}\left[Q_{\phi_1}(s_n, \pi_\theta(s_n))\right]. \tag{30}$$

Also the loss for the critic networks is given by temporal difference (TD)

$$L(\phi_i) = \mathbf{E}_{s_n \sim D}\left[\left(Q_{\phi_i}(s_n, a_n) - y_n\right)^2\right], i = 1, 2. \tag{31}$$

Here, the target value $y_n = r_n + \gamma \min_{i=1,2} Q_{\phi_i'}(s_{n+1}, \pi_{\theta'}(s_{n+1}) + \epsilon)$, where $Q_{\phi_i'}$ means the target critic network, $\pi_{\theta'}$ the target policy network and ϵ is the truncated Gaussian noise.

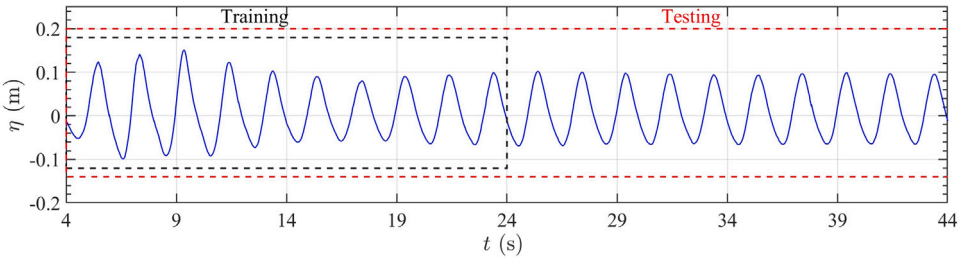


Fig. 11. Free surface height in front of the flap for training and testing.

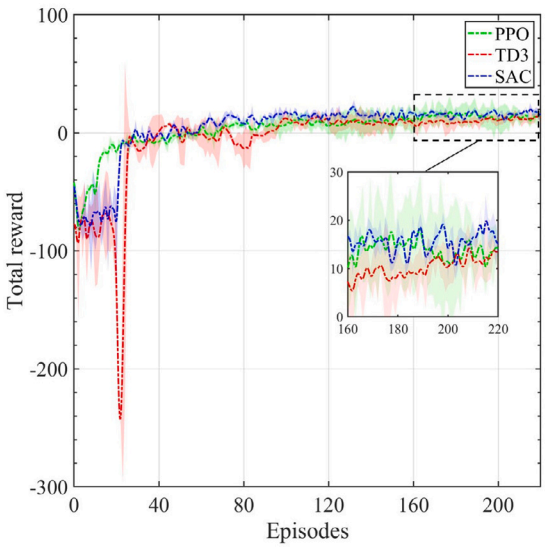


Fig. 12. Total reward curves in the training process with three agents.

3.2.3. The SAC algorithm

SAC consists of a policy network that outputs an action probability density function, two value networks, and two target networks corresponding to the value networks.

In the PPO state space exploration primarily relies on sampling from the action probability distribution output by the policy network, whereas TD3 achieves exploration by artificially adding noise to the output of the action. Compared with the PPO and TD3 algorithms, the SAC algorithm incorporates the entropy of the policy into the state-value function, encouraging exploration by maximizing the return regularized by entropy

$$V_{\pi_\theta}(s_n) = \mathbf{E}_{a_n\sim\pi_\theta}[\sum_{t=n}^{\infty}(\gamma_t r_t|s_t) + \beta\mathcal{H}(\pi_\theta(s_n))], \tag{32}$$

where β is the regularization coefficient.

The objective function of the policy network is

$$J(\theta) = \mathbf{E}_{s_n\sim D}\left[\min_{i=1,2} Q_{\phi_i}(s_n, \tilde{a}_n) - \beta\log\pi_\theta(\tilde{a}_n|s_n)\right], \tag{33}$$

where $\tilde{a}_n$ is the sample from $\pi_\theta(\cdot|s_n)$.

Also, the loss for the critic networks is also calculated by TD, while with a different definition of y_n

$$\begin{cases} L(\phi_i) = \mathbf{E}_{s_n\sim D}\left[\left(Q_{\phi_i}(s_n, a_n) - y_n\right)^2\right], i = 1, 2, \\ y_n = r_n + \gamma\left(\min_{i=1,2} Q_{\phi_i'}(s_{n+1}, \tilde{a}_{n+1}) - \beta\log\pi_\theta(\tilde{a}_{n+1}|s_{n+1})\right), \end{cases} \tag{34}$$

here where $\tilde{a}_{n+1}$ is the sample from $\pi_\theta(\cdot|s_{n+1})$.

Table 2
Basic hyperparameters of different DRL algorithms.

Algorithm	PPO	TD3	SAC
Activation function	tanh	tanh	tanh
Learning rate (α)	3e−4	3e−4	1e−3
Steps per epoch	2048	2048	2048
Batch size	256	256	256
Discount factor (γ)	0.99	0.99	0.99
Soft update (τ)	–	0.005	0.005

4. Result

4.1. Study of DRL algorithms

Initially, three typical DRL algorithms are trained:i.e., PPO, TD3, and SAC. The parameters of the policy and critic networks under the three algorithms are consistent, with two hidden layers and 512 neurons in each layer. Other settings of the neural network and algorithm hyperparameters are shown in Table 2.

Fig. 11 illustrates the free surface height at $x = 7.6$ m in front of the flap. It is observed that around the 4-s mark, the wave reaches and interacts with the flap. Consequently, the training phase is initiated at the 4-s mark and continues until the 24-s mark, encompassing 20 s and 200 actions. The testing phase extends over 40 s, incorporating 400 actions.

The overall reward curves with standard deviation shadows are depicted in Fig. 12. The training consisted of 220 episodes, with TD3 and SAC requiring the pre-collection of data for the first 20 episodes, a process of gathering initial data before the actual training. Pre-collection resulted in oscillations around −80 in the reward curves during this phase. However, both algorithms quickly identified effective energy enhancement strategies within the first ten episodes after training commenced. Considering TD3 incorporates noise artificially to enrich exploration, it displayed instability and only began to stabilize after 90 episodes.

On the other hand, SAC explored and converged to optimal strategies by approximately 60 episodes with entropy regularization. PPO, not requiring initial data collection, showed a slower but steady improvement from the beginning, converging to effective strategies around 80 episodes. From the reward trends observed between episodes 160 and 220, SAC demonstrates its best performance with the slightest standard deviation, indicating superior stability over other two algorithms. In addition, since the incident wave is regular, the dynamic curve of the damping coefficient is also periodic, as shown in Fig. 13. It can be observed that the overall fluctuation of the PPO algorithm is substantial, whereas the TD3 algorithm converges to a locally optimal solution due to insufficient exploration. Therefore, SAC will be applied for subsequent agent training.

Fig. 14 presents the velocity field in the x-direction and the motion state of the flap under both fixed and dynamic damping coefficients, and Fig. 15 focuses on quantitatively analyzing the flap's rotation and angular velocity. It can be observed that the adaptive change of the damping coefficient does not alter the flow field structure or the

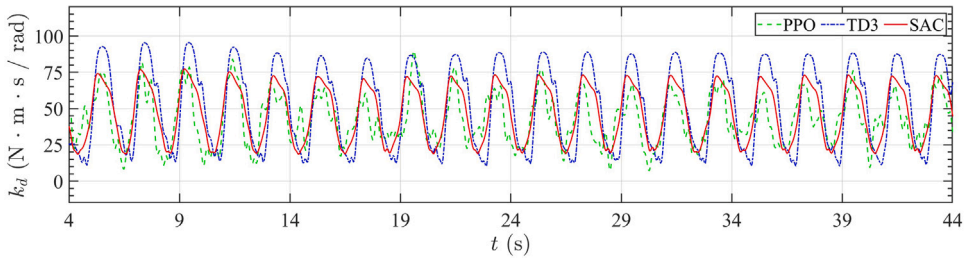


Fig. 13. Adaptive damping coefficients of the PTO system with different agents.

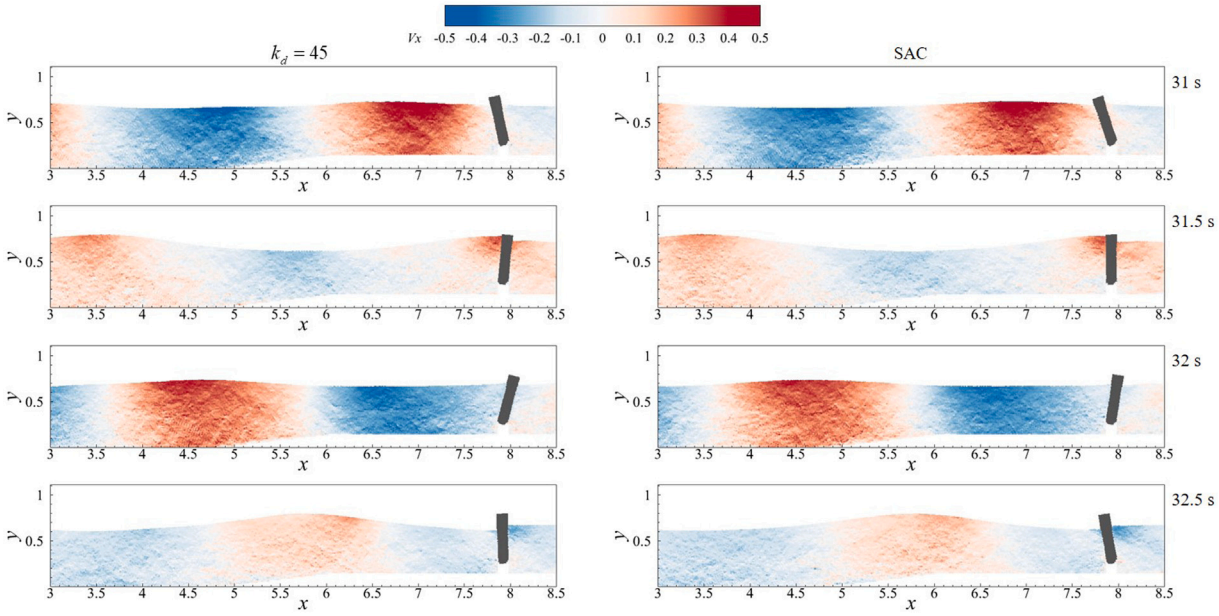


Fig. 14. Free surfaces and wave–structure interactions in one wave period. The fluid particles are colored by velocity magnitude on the x-axis.

angular amplitude of the flap, while shift the equilibrium position of the flap, from −1.81° to 4.31°. Further analysis indicates that the periodic characteristics of the free surface height at the flap are consistent with the damping coefficient of the PTO system. When the wave crest passes, the damping coefficient increases to its peak value. Given that the energy density of the wave crest is high, the angular velocity is reduced slightly, leading to an overall improvement in the PTO system, as shown in Fig. 15(c). In addition, the wave energy density of the trough itself is lower than that of the peak, and some energy has already been absorbed during the crest phase. Maintaining a high damping coefficient during this phase would rapidly decrease angular velocity. Although reducing the damping coefficient can increase the flap's angular velocity, the PTO system's power output still decreases compared to a constant damping coefficient. Overall, as shown in Fig. 16(b), during a complete wave period, the average energy harvesting by the dynamic damping system is 27.3 J, compared to 24.4 J captured under the optimal constant damping coefficient, resulting in a 10.61% improvement in wave energy harvesting.

The energy harvesting efficiency of the OWSC can be quantified by CWR [25,52]

$$\mathrm{CWR} = \frac{P_{out}}{P_0}. \tag{35}$$

Here, P_{out} is the capture of instantaneous energy within a wave period and P_0 is the mean incident power of unidirectional regular waves based on the linear theory

$$\begin{cases} P_{out} = \dfrac{1}{T_{period}} \displaystyle\int_0^{T_{period}} k_d \Omega^2 \, dt \\ P_0 = \dfrac{\rho g H^2 B \omega}{16k} \left(1 + \dfrac{2kh}{\sinh(2kh)}\right), \end{cases} \tag{36}$$

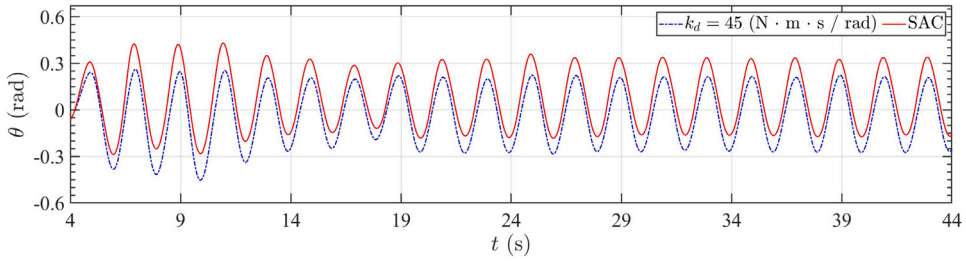
with B denoting the width of the flap. Currently, CWR under optimal constant damping coefficient is 13.12%, and 14.68% for the adaptive damping coefficient in 2D simulations, as shown in Table 3.

Table 3
Comparison of constant and adaptive damping coefficients for different wave types.

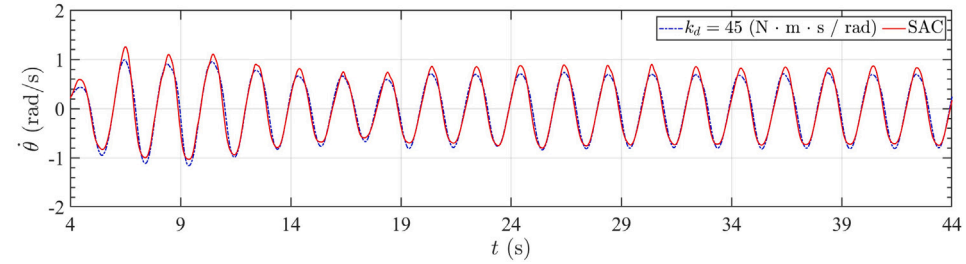
	Fixed k_d (%)	DRL (%)
2D regular wave	13.12	14.68
3D regular wave	34.15	41.86
2D irregular wave	86.79	92.38

4.2. Effects of 3D simulations

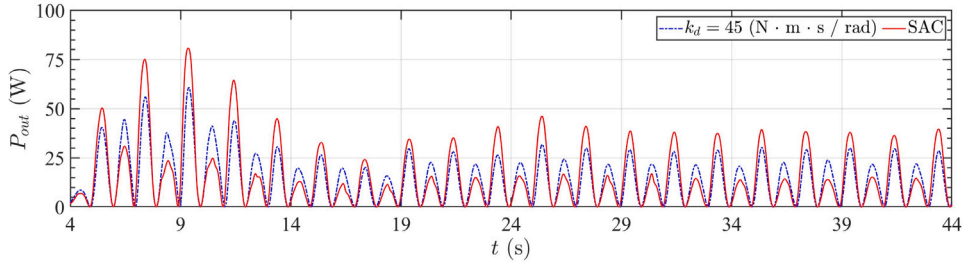
Previous studies have shown that fixed flaps in 2D simulations simplify the diffraction waves, which are theoretically equal in size to the incident waves and opposite in direction. This results in standing waves forming on the windward side of the flap [48]. In 3D simulations, diffraction waves propagate in all directions, and antisymmetric shear waves along the flap can trigger near-resonance, enhancing the torque



(a) Rotation of the flap

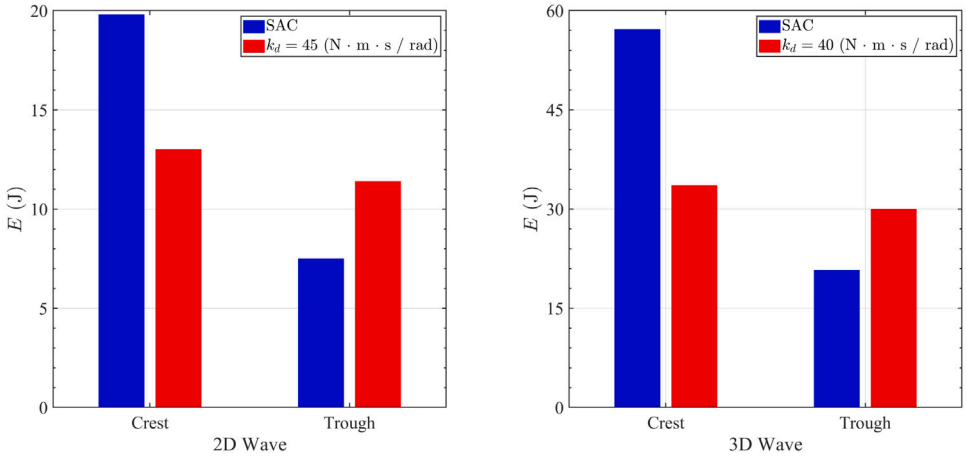


(b) Angular velocity of the flap



(c) Instantaneous power capture

Fig. 15. Influence of the damping coefficient on the (a) rotation of the flap, (b) angular velocity of the flap, and (c) instantaneous power capture.



(a) 2D wave period (b) 3D wave period

Fig. 16. Comparison of peak and trough energy in a complete cycle under the optimal constant damping coefficient and SAC strategy, (a) 2D and (b) 3D.

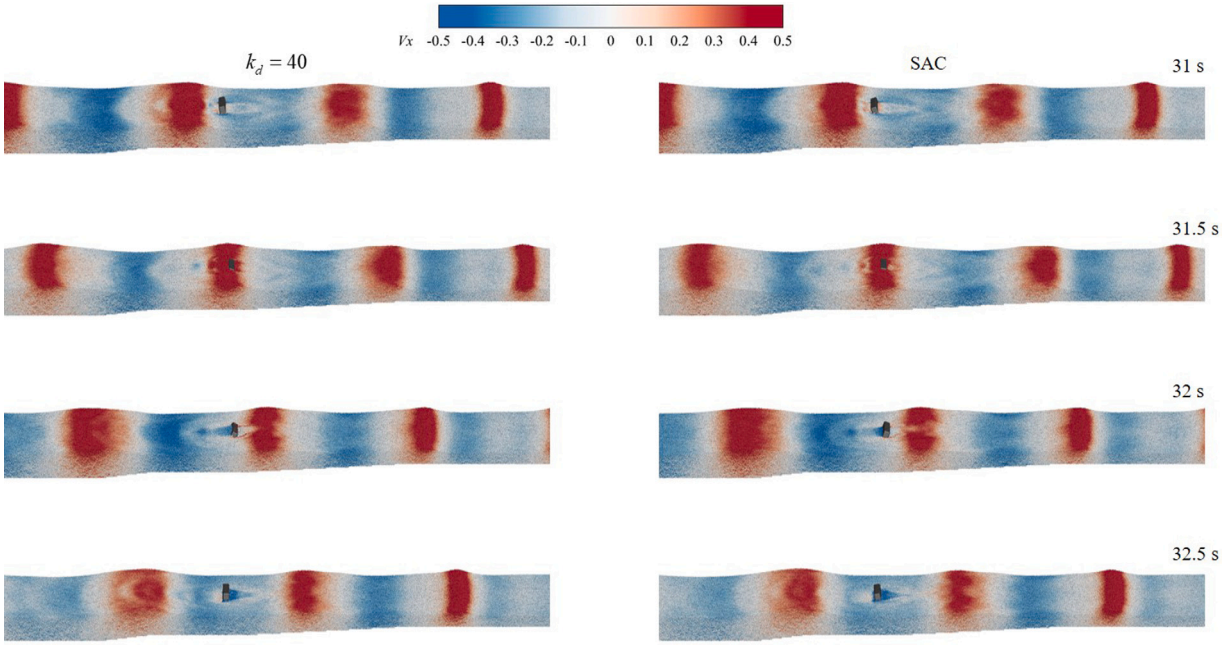


Fig. 17. 3D simulations of OWSC: Free surfaces and the flap motion in a wave period. The left side is the optimal constant damping coefficient, and the right side is the damping coefficient controlled with SAC.

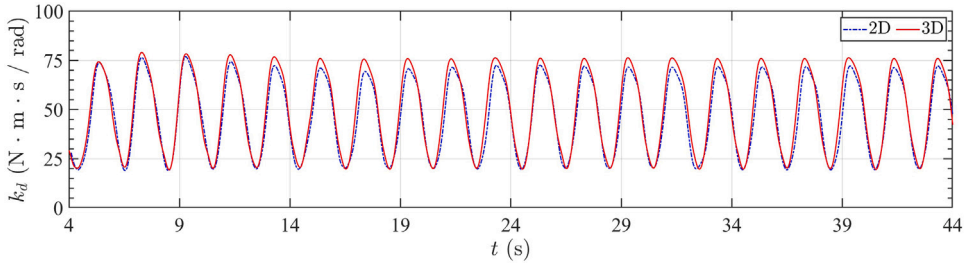


Fig. 18. Comparison of the damping coefficient with the same policy under 2D and 3D simulations.

acting on the converter. Therefore, 3D simulations can more accurately capture the motion characteristics of the OWSC device in actual operation. In this section, we conduct experiments in a 3D environment with the policy obtained from a 2D training environment. Notably, Zhang et al. [27] have proved that in 3D simulations, $k_d = 40$ N m s/rad is the optimal constant damping coefficient.

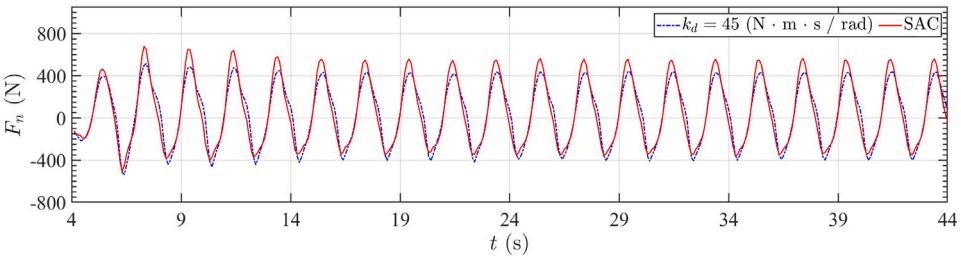
As shown in Fig. 17, the dynamic changes in the damping coefficient do not alter the structure of the flow field in the 3D simulations. Also, as shown in Fig. 18, the output of the damping coefficient in both 2D and 3D simulations is essentially consistent. This consistency indicates that the 2D assumption can accurately represent the coupling effect between the waves and the OWSC, and it also demonstrates the robustness of the trained policy network, which can be applied in real-world scenarios.

Similar to the 2D simulation, Fig. 19 shows that under adaptive damping control, the equilibrium position of the flap shifts to the left by 11.46°, approaching nearly vertical position to the base. Considering the fluid incompressibility, this shift increases the force perpendicular to the flap, as illustrated in Fig. 19(a). A control period is assumed to begin when the flap rotates to the far left, with the wave crest reaching the flap. During the first quarter of the period, the damping coefficient continues to rise. Due to the increased thrust on the flap, the angular velocity remains almost unchanged compared to the constant damping coefficient, enhancing energy harvesting. As the wave trough passes, the force on the flap decreases, and the reduction in the damping coefficient helps maintain the flap's angular velocity. Since the energy of the wave crest is inherently higher than that of the trough, the overall energy harvesting improves due to the difference in energy levels, as shown in Fig. 16(b). Over a complete wave period, the average energy harvesting by the dynamic damping system is 77.88 J, compared to 63.55 J under the optimal constant damping coefficient, resulting in a 22.54% improvement. Also, in 3D simulation, CWR under optimal constant damping coefficient is 34.15% and 41.86% with the adaptive damping coefficient.

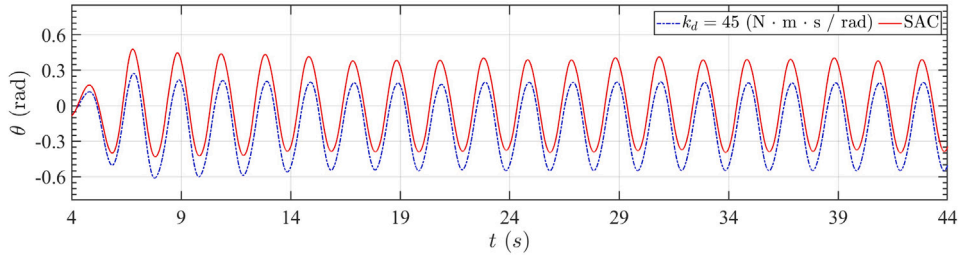
4.3. Optimization under different wave periods

It is well-established that fixed damping coefficients in vibration absorbers are suboptimal under off-resonance conditions. In this section, we first apply the optimized policy, trained under baseline wave conditions ($T = 2$ s, $H = 0.2$ m), to scenarios where either the wave period is altered ($T = 3$ s) or the wave amplitude is increased ($H = 0.4$ m). The results, shown in Fig. 20, indicate that when the wave period changes, the policy's adaptability decreases, yielding only a limited improvement in energy harvesting (3.73%). In contrast, when the wave amplitude varies, the policy remains highly adaptive, achieving a notable enhancement in energy capture (14.24%).

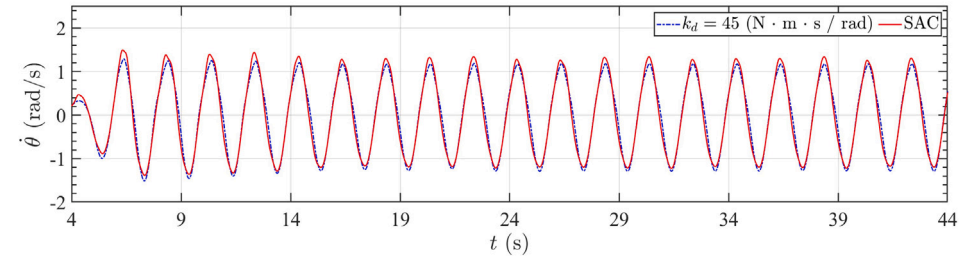
We train the policy using waves with periods of $T = 1$ s and $T = 3$ s, respectively, and perform a comparative analysis against the results obtained from training with $T = 2$ s, as shown in Fig. 21. The results indicate that the wave energy density increases as the wave



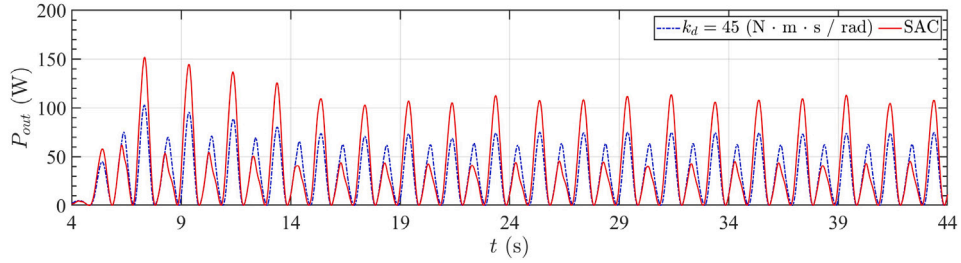
(a) Vertical force on the flap



(b) Rotation of the flap



(c) Angular velocity of the flap



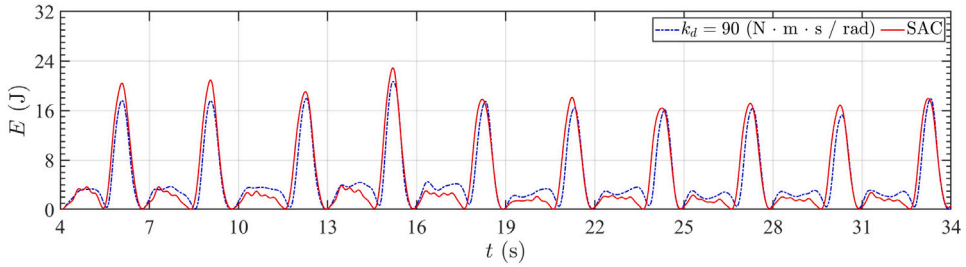
(d) Instantaneous power capture

Fig. 19. Effects of different policies on (a) vertical force on the flap, (b) rotation of the flap, (c) angular velocity of the flap, and (d) instantaneous power capture.

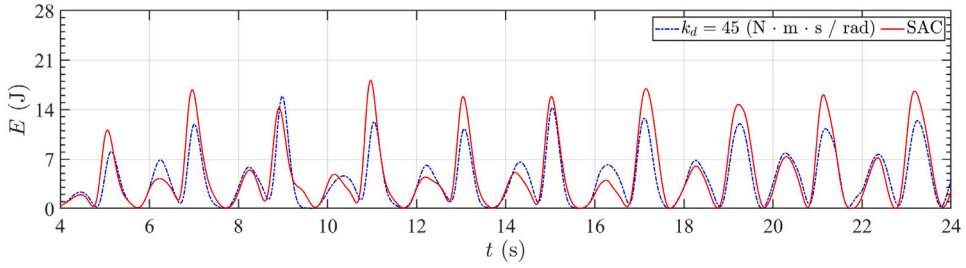
period increases. However, the energy difference between wave crests and troughs becomes more pronounced. Since the strategy primarily enhances energy capture efficiency by leveraging the energy in the crest region, the optimization effect becomes more significant with longer wave periods. Overall, energy harvesting improved by 1.6%, 8.8%, and 13.75% under the three wave periods, compared to the optimal fixed damping coefficient. Furthermore, for $T = 3$ s, the variation in damping exhibits the same periodic characteristics as the wave itself, as shown in Fig. 22. This suggests that for regular waves, the damping variation period is identical to the wave period, although the magnitude of the damping variation differs. This observation is expected, as an increase in wave period while maintaining a constant wave height results in higher overall wave energy. Consequently, the optimal fixed damping coefficient increases, leading to corresponding adjustments in damping variation.

4.4. Optimization under irregular waves

In this section, we investigate optimization problems under irregular wave conditions, the most common scenarios encountered in practical engineering applications. Considering the strong nonlinearity of the induced motion of the OWSC under irregular waves, the number of observation points increases. The initial observation position of the free surface height is set at $x = 3.5$ m, with one point placed every 0.264 m for total 17 probes. The locations of observation points for wave speed are also increased accordingly, and the value of the observation vector increased to 74 at last. In addition, to verify that the policy network can resist the random characteristics of irregular waves, two sets of random seeds are used to characterize the wave phases in Eq. (19) during the training and testing stages while H_p and T_p remained unchanged. The relation between the total energy conversion and linear damping

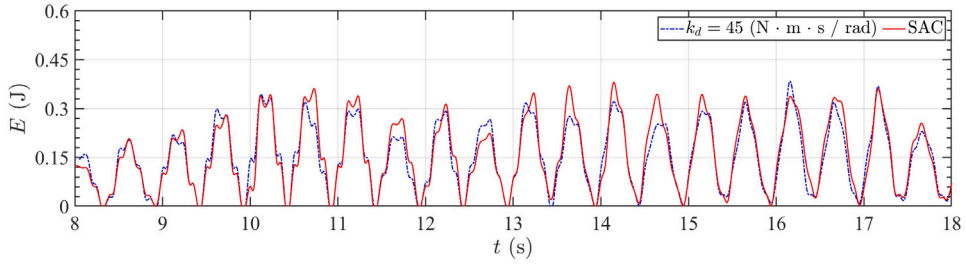


(a) $T = 3$ s, $H = 0.2$ m

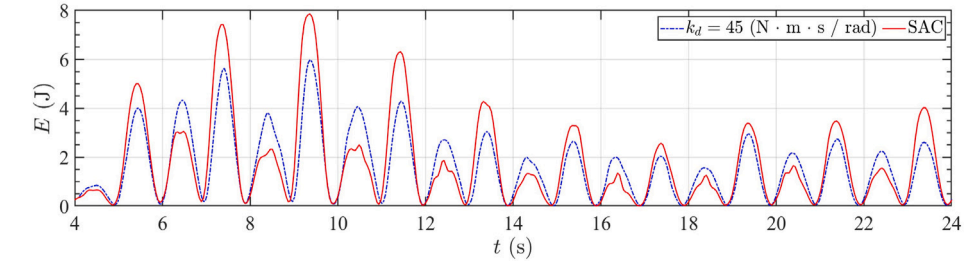


(b) $T = 2$ s, $H = 0.4$ m

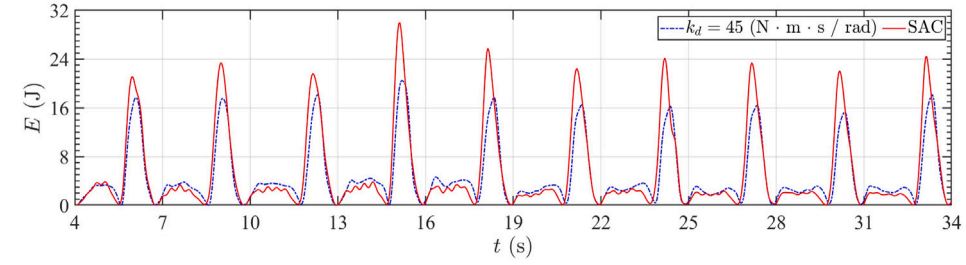
Fig. 20. Comparison of peak and trough energy in 10 complete cycles under the optimal constant damping coefficient and SAC strategy, (a) $T = 3$ s, $H = 0.2$ m and (b) $T = 2$ s, $H = 0.4$ m.



(a) $T = 1$ s



(b) $T = 2$ s



(c) $T = 3$ s

Fig. 21. Comparison of peak and trough energy in 10 complete cycles under the optimal constant damping coefficient and SAC strategy, $H = 0.2$ m, (a) $T = 1$ s, (b) $T = 2$ s, and (c) $T = 3$ s.

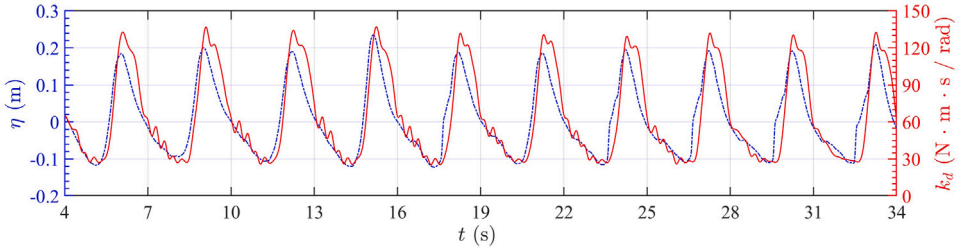
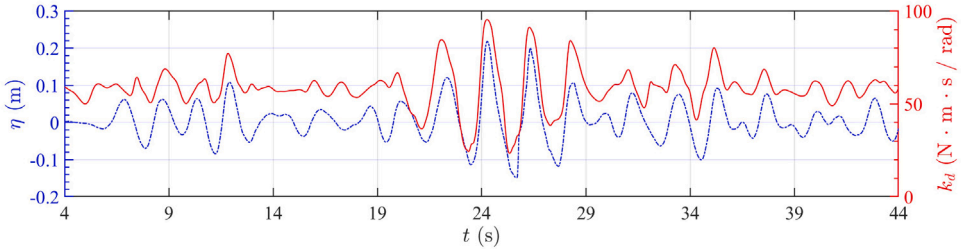
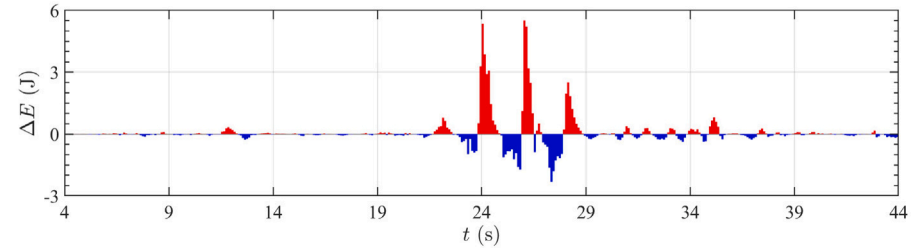


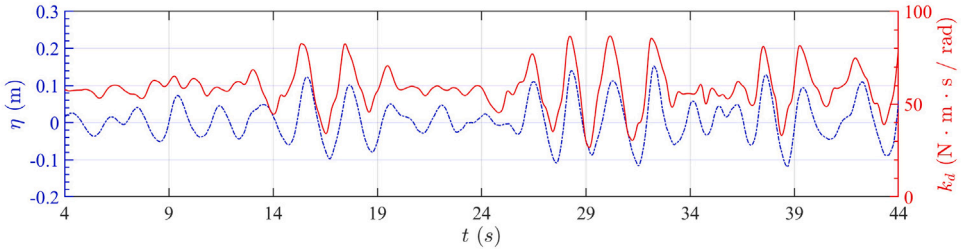
Fig. 22. The free surface height in front of the first flap ($x = 7.6$ m) and the corresponding damping coefficient.



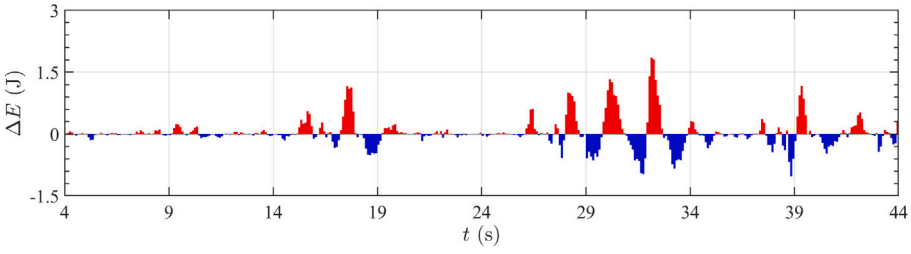
(a) Training



(b) Training



(c) Testing



(d) Testing

Fig. 23. The free surface height in front of the flap and the corresponding damping coefficient under the (a) training wave (c) testing wave, and the difference of instantaneous energy harvesting $\Delta E = E_t - E_{60}$ under the (a) training wave (c) testing wave.

coefficient for irregular waves is shown in Table 4. It is clear that with $k_d = 60$ N m s/rad, the energy harvesting factor is the highest.

From Fig. 23(a), it can be observed that, compared to regular waves, less than one-third of the free surface heights of irregular waves exceed 0.2 m throughout the entire period, which is also the region where the wave energy is primarily concentrated. The dynamic response of the damping coefficient is related to the free surface height. When the peak period occurs, the damped vibration also increases accordingly.

Table 4
The variations of the total energy conversion in terms of damping coefficients.

k_d (N m s/rad)	10	20	30	40	50	60	70	80	90	100
E_{train} (J)	195.21	290.79	340.34	365.88	378.75	383.36	382.02	379.29	377.94	368.95
E_{test} (J)	206.12	300.77	350.17	378.48	383.89	390.49	381.86	376.41	368.35	363.64

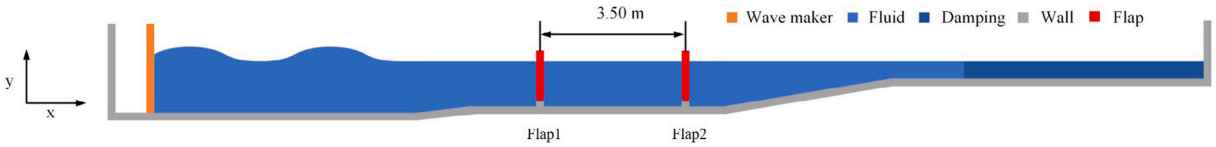


Fig. 24. The geometry of the dual OWSC system. The overall structure has stayed the same. Only the length of the plane where the base is located has been increased.

Table 5
The influence of the spacing on the total energy conversion of dual OWSCs.

Δx (m)	2.0	2.25	2.5	2.75	3.0	3.25	3.5	3.75	4.0
E_t (J)	449.01	366.32	422.91	526.69	608.52	661.68	675.47	650.44	541.84

This relationship can also be observed in the test section, indicating that the agent can accurately capture the wave characteristics under the specific spectrum.

Further combined with Fig. 23(b), we can see that compared with the constant damping coefficient, the difference in energy harvesting is mainly concentrated in the peak period, which is essentially consistent with the improvement of energy harvesting by regular waves. The near-simple harmonic damping motion will improve Capture energy in the crest section and reduce energy harvesting in the trough section. For secondary period waves, since the instantaneous energy they carry is small, the response of the damping coefficient will not cause significant changes in flap motion and energy harvesting. This part of the energy cannot be effectively improved. Therefore, over the entire 40-s period, the wave energy harvesting increased by 24.67 J, an increase of 6.42%, compared to the energy harvesting under a constant damping coefficient. In addition, for the average incident energy of the irregular wave, Eq. (36) is modified as

$$P_0 = \sum_{i=0}^{N} \frac{\rho g H_i^2 B \omega_i}{16 k_i} \left(1 + \frac{2k_i h}{\sinh(2k_i h)}\right). \tag{37}$$

CWR under optimal constant damping coefficient is 86.79% and 92.38% for the adaptive damping coefficient.

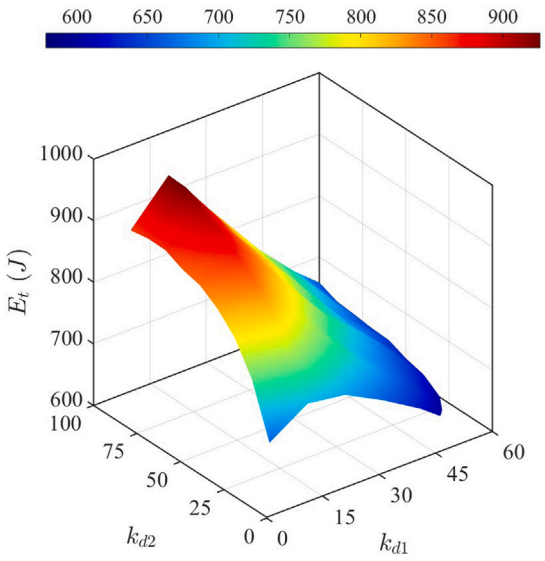


Fig. 25. The variations of the total energy conversion in terms of the damping coefficients in dual OWSCs.

4.5. Study of dual OWSC system

The dual OWSC system is illustrated in Fig. 24. Previous research has indicated that for a dual OWSC system, the maximum total energy conversion is achieved when the spacing between the two OWSCs is seven-eighths of the wavelength [13]. In this section, we initially set the damping coefficients of both flaps to 50 N m s/rad to investigate the impact of different spacings on total energy conversion. As shown in Table 5, when the spacing is 3.5 m, approximately three-quarters of the wavelength, the total energy conversion reaches its maximum.

Subsequently, the effect of varying damping coefficient combinations on the total energy harvesting at the identified optimal spacing is investigated, as depicted in Fig. 25. The analysis reveals a linear relationship between the damping coefficient of the OWSC-2 and the total energy conversion, with higher damping coefficients leading to a gradual increase in energy harvesting. Conversely, for the OWSC-1, the total energy conversion initially increases with the damping coefficient, reaching a peak before declining. Notably, the peak value occurs around k_{d1} = 20 N m s/rad, and this peak remains unaffected by variations in k_{d2}. Therefore, for the subsequent RL training, the condition with k_{d1} = 20 N m s/rad and k_{d2} = 80 N m s/rad is established as the baseline. Given the strong nonlinear characteristics of the dual OWSC system and the observed lower wave energy harvesting by the OWSC-2, the damping coefficient of the OWSC-2 is held constant during training. This approach allows us to focus on optimizing the damping coefficient of the OWSC-1. The RL training commences at the 24th second, a point in time when the wave–structure interactions have stabilized, meaning the system has reached a steady state in terms of energy harvesting and conversion.

The training results are illustrated in Fig. 26. The damping coefficient variation is consistent with the trend of free surface height changes before the flap, demonstrating apparent periodicity that aligns with previous research findings. Further analysis based on Fig. 27 reveals that, after 39 s, the system's state is stabilized. An increase in the damping coefficient during the wave peak phase significantly reduces the angular velocity, resulting in only a limited increase in energy harvesting during the peak. Conversely, reducing the damping during the trough phase does not enhance the angular velocity, which remains lower than under constant damping conditions. This reduction in energy harvesting during the trough phase leads to a notable decrease in overall energy acquisition over the entire wave period, thereby failing to improve energy harvesting efficiency. During the 60-s test period, the energy harvesting for the OWSC-1 using the adaptive damping coefficient was 1560.37 J, compared to 1626.52 J with constant damping, representing a reduction of 4.07%.

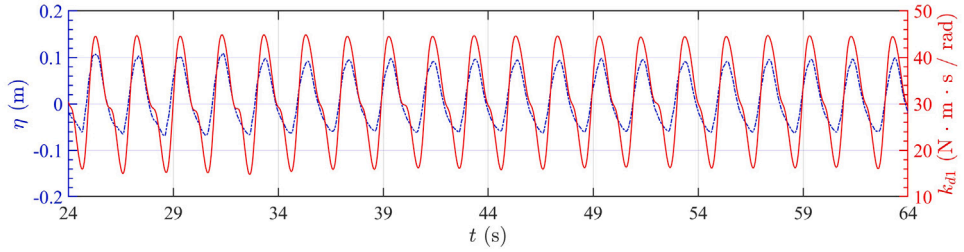
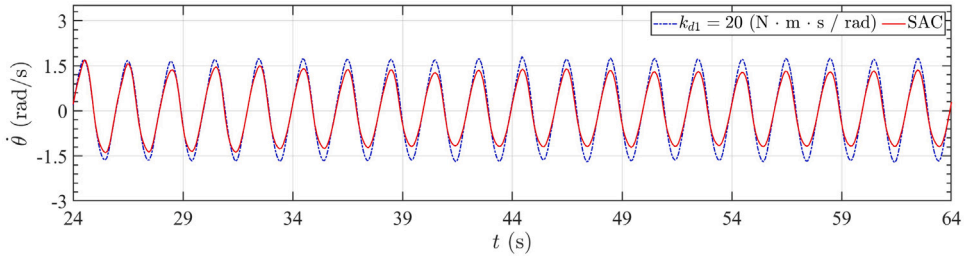
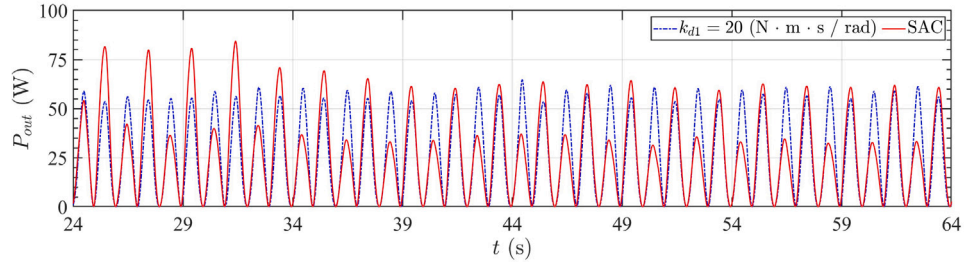


Fig. 26. The free surface height in front of the first flap ($x = 7.6$ m) and the corresponding damping coefficient.



(a) Angular velocity of the flap



(b) Instantaneous power capture

Fig. 27. The influence of the damping coefficient on the (a) angular velocity of the first flap, and (b) instantaneous power capture.

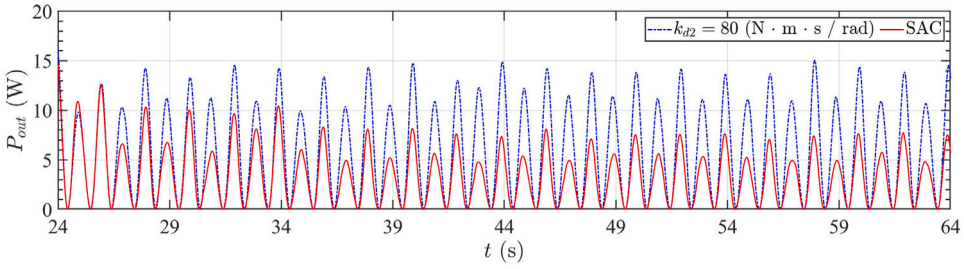


Fig. 28. The influence of the damping coefficient on the instantaneous power capture of the OWSC-2.

Additionally, for the OWSC-2, where the damping coefficient remained unchanged, the energy harvesting efficiency dropped significantly from 362.94 J to 197.91 J, a decrease of 45.47%, as shown in Fig. 28. This indicates that after the wave passes through the OWSC-1 with adaptive damping, the energy loss is more significant than with constant damping. Combined with Fig. 29, it is evident that under constant damping, significant harmonics are generated between the two OWSCs, which is beneficial for enhancing energy harvesting.

Therefore, in the dual OWSC system, the nonlinear characteristics are pronounced, and single damping control is insufficient to improve overall wave energy acquisition efficiency. Moreover, the 2D simulations constrain the design and optimization of the OWSC layout, necessitating further analysis and discussion in subsequent work.

5. Conclusion

This paper establishes a framework coupling a CFD environment based on an open-source SPH-based library with DRL, aimed at optimizing the adaptive damping coefficient of the PTO system in OWSCs for wave energy conversion. Initially, the wave-making model and the numerical model of WSI were validated. Subsequently, the performance of various RL algorithms for the optimization process was investigated. The results indicated that SAC considers policy entropy, balanced exploration, and exploitation well and provides effective policies to enhance wave energy conversion.

For regular waves, the strategy primarily utilizes the difference in energy density between wave crests and troughs. Increasing energy

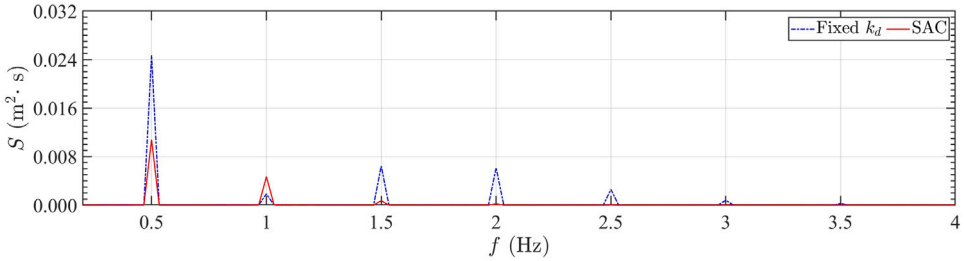


Fig. 29. The spectrum diagram at $x = 8.2$ m.

harvesting during the crest phase and reducing it during the trough phase achieves a positive net energy harvesting over each wave period. The policy trained in 2D simulations can be effectively applied in 3D simulations. Although the 2D simulations simplify wave diffraction and result in a slight decrease in calculated CWR, they accurately capture the coupling characteristics between waves and OWSCs, allowing DRL to learn practical policies that are robust and transferable. This provides a theoretical foundation for validation in experiments. Moreover, the training strategy optimized for a specific wave period remains effective under variations in wave height. However, its performance deteriorates when the wave period changes. Notably, the optimization effect becomes more pronounced with longer wave periods.

The DRL algorithm could still learn effective energy conversion optimization policies for irregular waves, primarily targeting regular waves with high energy density in the main period. The optimization principle is similar to that for regular waves, with limited enhancement in energy harvesting from the dynamic damping response for waves in the secondary period due to their lower energy density.

Finally, this paper explores the optimization of wave energy absorption in a dual OWSC system. The interaction between incident waves and OWSCs in the dual system generates harmonics with strong nonlinearity. Optimization focused on the primary OWSC showed that using similar optimization policies cannot enhance energy harvesting, as it significantly reduces the wave energy density between the OWSCs, leading to a substantial decrease in energy absorption by the secondary OWSC and an overall reduction in the system's energy harvesting. Therefore, considering the dynamic response of a single OWSC's damping coefficient is insufficient to optimize the energy conversion in complex dual OWSC systems.

Future work will introduce multi-agent reinforcement learning to directly learn corresponding energy optimization strategies for various OWSC layouts in 3D simulations.

CRediT authorship contribution statement

Mai Ye: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis. **Chi Zhang:** Writing – review & editing, Software, Investigation. **Yaru Ren:** Resources, Formal analysis, Conceptualization. **Ziyuan Liu:** Methodology, Data curation. **Oskar J. Haidn:** Validation, Supervision, Investigation. **Xiangyu Hu:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

M. Ye was supported by China Scholarship Council (No. 202006120018) when he conducted this work.

Data availability

The corresponding code of this work is available on GitHub at https://github.com/Xiangyu-Hu/SPHinXsys.git. The corresponding data of this work will be made available on reasonable request.

References

[1] E. Callaway, To catch a wave: ocean wave energy is trying to break into the renewable-energy market, but many challenges remain, Nat. 450 (7167) (2007) 156–160, http://dx.doi.org/10.1038/450156a.

[2] D. Evans, Maximum wave-power absorption under motion constraints, Appl. Ocean Res. 3 (4) (1981) 200–203, http://dx.doi.org/10.1016/0141-1187(81)90063-8.

[3] M. Folley, T. Whittaker, Analysis of the nearshore wave energy resource, Renew. Energy 34 (7) (2009) 1709–1715, http://dx.doi.org/10.1016/j.renene.2009.01.003.

[4] M. Folley, B. Elsaesser, T. Whittaker, Analysis of the wave energy resource at the European marine energy centre, in: Coasts, Marine Structures and Breakwaters: Adapting To Change: Proceedings of the 9th International Conference Organised By the Institution of Civil Engineers and Held in Edinburgh on 16 To 18 September 2009, Thomas Telford Ltd, 2010, pp. 660–669, http://dx.doi.org/10.1680/cmsb.41301.0058.

[5] I. López, R. Carballo, G. Iglesias, Site-specific wave energy conversion performance of an oscillating water column device, Energy Convers. Manage. 195 (2019) 457–465, http://dx.doi.org/10.1016/j.enconman.2019.05.030.

[6] J.P. Kofoed, P. Frigaard, E. Friis-Madsen, H.C. Sørensen, Prototype testing of the wave energy converter wave dragon, Renew. Energy 31 (2) (2006) 181–189, http://dx.doi.org/10.1016/j.renene.2005.09.005.

[7] A. Day, A. Babarit, A. Fontaine, Y.-P. He, M. Kraskowski, M. Murai, I. Penesis, F. Salvatore, H.-K. Shin, Hydrodynamic modelling of marine renewable energy devices: A state of the art review, Ocean Eng. 108 (2015) 46–69, http://dx.doi.org/10.1016/j.oceaneng.2015.05.036.

[8] T. Whittaker, M. Folley, Nearshore oscillating wave surge converters and the development of Oyster, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 370 (1959) (2012) 345–364, http://dx.doi.org/10.1098/rsta.2011.0152.

[9] Y. Cheng, C. Ji, G. Zhai, Fully nonlinear analysis incorporating viscous effects for hydrodynamics of an oscillating wave surge converter with nonlinear power take-off system, Energy 179 (2019) 1067–1081, http://dx.doi.org/10.1016/j.energy.2019.04.189.

[10] T. Whittaker, D. Collier, M. Folley, M. Osterried, A. Henry, M. Crowley, The development of Oyster—a shallow water surging wave energy converter, in: Proceedings of the 7th European Wave and Tidal Energy Conference, 2007, pp. 11–14.

[11] E. Renzi, K. Doherty, A. Henry, F. Dias, How does Oyster work? The simple interpretation of Oyster mathematics, Eur. J. Mech. B Fluids 47 (2014) 124–131, http://dx.doi.org/10.1016/j.euromechflu.2014.03.007.

[12] A. Henry, P. Schmitt, T. Whittaker, A. Rafiee, F. Dias, The characteristics of wave impacts on an oscillating wave surge converter, in: ISOPE International Ocean and Polar Engineering Conference, ISOPE, 2013, pp. ISOPE–I.

[13] Y.-C. Chow, Y.-C. Chang, C.-C. Lin, J.-H. Chen, S.-Y. Tzang, Experimental investigations on wave energy capture of two bottom-hinged-flap WECs operating in tandem, Ocean Eng. 164 (2018) 322–331, http://dx.doi.org/10.1016/j.oceaneng.2018.06.010.

[14] M. Brito, R.M. Ferreira, L. Teixeira, M.G. Neves, R.B. Canelas, Experimental investigation on the power capture of an oscillating wave surge converter in unidirectional waves, Renew. Energy 151 (2020) 975–992, http://dx.doi.org/10.1016/j.renene.2019.11.094.

[15] Y. Cheng, G. Li, C. Ji, T. Fan, G. Zhai, Fully nonlinear investigations on performance of an OWSC (oscillating wave surge converter) in 3D (three-dimensional) open water, Energy 210 (2020) 118526, http://dx.doi.org/10.1016/j.energy.2020.118526.

[16] E. Renzi, F. Dias, Resonant behaviour of an oscillating wave energy converter in a channel, J. Fluid Mech. 701 (2012) 482–510, http://dx.doi.org/10.1017/jfm.2012.194.
[17] P. Schmitt, H. Asmuth, B. Elsäßer, Optimising power take-off of an oscillating wave surge converter using high fidelity numerical simulations, Int. J. Mar. Energy 16 (2016) 196–208, http://dx.doi.org/10.1016/j.ijome.2016.07.006.
[18] Y. Wei, A. Rafiee, A. Henry, F. Dias, Wave interaction with an oscillating wave surge converter, Part I: Viscous effects, Ocean Eng. 104 (2015) 185–203, http://dx.doi.org/10.1016/j.oceaneng.2015.05.002.
[19] X. Jiang, S. Day, D. Clelland, Hydrodynamic responses and power efficiency analyses of an oscillating wave surge converter under different simulated PTO strategies, Ocean Eng. 170 (2018) 286–297, http://dx.doi.org/10.1016/j.oceaneng.2018.10.050.
[20] H.R. Mottahedi, M. Anbarsooz, M. Passandideh-Fard, Application of a fictitious domain method in numerical simulation of an oscillating wave surge converter, Renew. Energy 121 (2018) 133–145, http://dx.doi.org/10.1016/j.renene.2018.01.021.
[21] M. Luo, C. Koh, Shared-memory parallelization of consistent particle method for violent wave impact problems, Appl. Ocean Res. 69 (2017) 87–99, http://dx.doi.org/10.1016/j.apor.2017.09.013.
[22] C. Zhang, X.Y. Hu, N.A. Adams, A generalized transport-velocity formulation for smoothed particle hydrodynamics, J. Comput. Phys. 337 (2017) 216–232, http://dx.doi.org/10.1016/j.jcp.2017.02.016.
[23] A. Rafiee, B. Elsaesser, F. Dias, Numerical simulation of wave interaction with an oscillating wave surge converter, in: International Conference on Offshore Mechanics and Arctic Engineering, Vol. 55393, American Society of Mechanical Engineers, 2013, V005T06A013, http://dx.doi.org/10.1115/OMAE2013-10195.
[24] M. Brito, R. Canelas, R. Ferreira, O. García-Feal, J. Domínguez, A. Crespo, M. Neves, Coupling between DualSPHysics and Chrono-Engine: towards large scale HPC multiphysics simulations, in: 11th International SPHERIC Workshop, Munich, Germany, 2016.
[25] M. Brito, R. Canelas, O. García-Feal, J. Domínguez, A. Crespo, R.M. Ferreira, M.G. Neves, L. Teixeira, A numerical tool for modelling oscillating wave surge converter with nonlinear mechanical constraints, Renew. Energy 146 (2020) 2024–2043, http://dx.doi.org/10.1016/j.renene.2019.08.034.
[26] Z. Liu, Y. Wang, X. Hua, Numerical studies and proposal of design equations on cylindrical oscillating wave surge converters under regular waves using SPH, Energy Convers. Manage. 203 (2020) 112242, http://dx.doi.org/10.1016/j.enconman.2019.112242.
[27] C. Zhang, Y. Wei, F. Dias, X. Hu, An efficient fully Lagrangian solver for modeling wave interaction with oscillating wave surge converter, Ocean Eng. 236 (2021) 109540, http://dx.doi.org/10.1016/j.oceaneng.2021.109540.
[28] Y. Liu, N. Mizutani, Y.-H. Cho, T. Nakamura, Performance enhancement of a bottom-hinged oscillating wave surge converter via resonant adjustment, Renew. Energy 201 (2022) 624–635, http://dx.doi.org/10.1016/j.renene.2022.10.130.
[29] R.P. Gomes, M.F. Lopes, J.C. Henriques, L.M. Gato, A.F.d.O. Falcão, The dynamics and power extraction of bottom-hinged plate wave energy converters in regular and irregular waves, Ocean Eng. 96 (2015) 86–99, http://dx.doi.org/10.1016/j.oceaneng.2014.12.024.
[30] M. Shadman, G.O.G. Avalos, S.F. Estefen, On the power performance of a wave energy converter with a direct mechanical drive power take-off system controlled by latching, Renew. Energy 169 (2021) 157–177, http://dx.doi.org/10.1016/j.renene.2021.01.004.
[31] H. Liang, D. Qiao, X. Wang, G. Zhi, J. Yan, D. Ning, J. Ou, Energy capture optimization of heave oscillating buoy wave energy converter based on model predictive control, Ocean Eng. 268 (2023) 113402, http://dx.doi.org/10.1016/j.oceaneng.2022.113402.
[32] D. Fan, L. Yang, Z. Wang, M.S. Triantafyllou, G.E. Karniadakis, Reinforcement learning for bluff body active flow control in experiments and simulations, Proc. Natl. Acad. Sci. 117 (42) (2020) 26091–26098, http://dx.doi.org/10.1073/pnas.2004939117.
[33] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Netw. 2 (5) (1989) 359–366, http://dx.doi.org/10.1016/0893-6080(89)90020-8.
[34] E. Anderlini, S. Husain, G.G. Parker, M. Abusara, G. Thomas, Towards real-time reinforcement learning control of a wave energy converter, J. Mar. Sci. Eng. 8 (11) (2020) 845, http://dx.doi.org/10.3390/jmse8110845.
[35] S. Zou, X. Zhou, I. Khan, W.W. Weaver, S. Rahman, Optimization of the electricity generation of a wave energy converter using deep reinforcement learning, Ocean Eng. 244 (2022) 110363, http://dx.doi.org/10.1016/j.oceaneng.2021.110363.
[36] H. Liang, H. Qin, H. Su, Z. Wen, L. Mu, Environmental-sensing and adaptive optimization of wave energy converter based on deep reinforcement learning and computational fluid dynamics, Energy 297 (2024) 131254, http://dx.doi.org/10.1016/j.energy.2024.131254.
[37] C. Zhang, M. Rezavand, Y. Zhu, Y. Yu, D. Wu, W. Zhang, J. Wang, X. Hu, SPHinXsys: An open-source multi-physics and multi-resolution library based on smoothed particle hydrodynamics, Comput. Phys. Comm. 267 (2021) 108066, http://dx.doi.org/10.1016/j.cpc.2021.108066.
[38] J. Weng, H. Chen, D. Yan, K. You, A. Duburcq, M. Zhang, Y. Su, H. Su, J. Zhu, Tianshou: A highly modularized deep reinforcement learning library, J. Mach. Learn. Res. 23 (267) (2022) 1–6, http://jmlr.org/papers/v23/21-1127.html.
[39] J.P. Morris, P.J. Fox, Y. Zhu, Modeling low Reynolds number incompressible flows using SPH, J. Comput. Phys. 136 (1) (1997) 214–226, http://dx.doi.org/10.1006/jcph.1997.5776.
[40] H. Wendland, Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree, Adv. Comput. Math. 4 (1995) 389–396, http://dx.doi.org/10.1007/BF02123482.
[41] Y. Ren, P. Lin, C. Zhang, X. Hu, An efficient correction method in Riemann SPH for the simulation of general free surface flows, Comput. Methods Appl. Mech. Engrg. 417 (2023) 116460, http://dx.doi.org/10.1016/j.cma.2023.116460.
[42] C. Zhang, X. Hu, N.A. Adams, A weakly compressible SPH method based on a low-dissipation Riemann solver, J. Comput. Phys. 335 (2017) 605–620, http://dx.doi.org/10.1016/j.jcp.2017.01.027.
[43] C. Zhang, M. Rezavand, X. Hu, Dual-criteria time stepping for weakly compressible smoothed particle hydrodynamics, J. Comput. Phys. 404 (2020) 109135, http://dx.doi.org/10.1016/j.jcp.2019.109135.
[44] J.-F. Lee, J.-R. Kuo, C.-P. Lee, Transient wavemaker theory, J. Hydraul. Res. 27 (5) (1989) 651–663, http://dx.doi.org/10.1080/00221688909499116.
[45] O.S. Madsen, On the generation of long waves, J. Geophys. Res. 76 (36) (1971) 8672–8683, http://dx.doi.org/10.1029/JC076i036p08672.
[46] E. Renzi, F. Dias, Application of a moving particle semi-implicit numerical wave flume (MPS-NWF) to model design waves, Coast. Eng. 172 (2022) 104066, http://dx.doi.org/10.1016/j.coastaleng.2021.104066.
[47] Y. Goda, Random Seas and Design of Maritime Structures, vol. 33, World Scientific Publishing Company, 2010, http://dx.doi.org/10.1142/7425.
[48] E. Renzi, A. Abdolali, G. Bellotti, F. Dias, Wave-power absorption from a finite array of oscillating wave surge converters, Renew. Energy 63 (2014) 55–68, http://dx.doi.org/10.1016/j.renene.2013.08.046.
[49] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, in: International Conference on Machine Learning, PMLR, 2018, pp. 1861–1870, https://proceedings.mlr.press/v80/haarnoja18b.html.
[50] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, W. Zaremba, Openai gym, 2016, http://dx.doi.org/10.48550/arXiv.1606.01540, arXiv preprint arXiv:1606.01540.
[51] J. Rabault, M. Kuchta, A. Jensen, U. Réglade, N. Cerardi, Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control, J. Fluid Mech. 865 (2019) 281–302, http://dx.doi.org/10.1017/jfm.2019.62.
[52] K. Senol, M. Raessi, Enhancing power extraction in bottom-hinged flap-type wave energy converters through advanced power take-off techniques, Ocean Eng. 182 (2019) 248–258, http://dx.doi.org/10.1016/j.oceaneng.2019.04.067.
[53] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, 2017, http://dx.doi.org/10.48550/arXiv.1707.06347, arXiv preprint arXiv:1707.06347.
[54] S. Fujimoto, H. Hoof, D. Meger, Addressing function approximation error in actor-critic methods, in: International Conference on Machine Learning, PMLR, 2018, pp. 1587–1596, https://proceedings.mlr.press/v80/fujimoto18a.html.
[55] J. Schulman, P. Moritz, S. Levine, M. Jordan, P. Abbeel, High-dimensional continuous control using generalized advantage estimation, 2015, http://dx.doi.org/10.48550/arXiv.1506.02438, arXiv preprint arXiv:1506.02438.