Control of a Point Absorber Using Reinforcement Learning

Enrico Anderlini, David I. M. Forehand, Paul Stansell, Qing Xiao, and Mohammad Abusara

***Abstract*—This work presents the application of reinforcement learning for the optimal resistive control of a point absorber. The model-free Q-learning algorithm is selected in order to maximise energy absorption in each sea state. Step changes are made to the controller damping, observing the associated penalty, for excessive motions, or reward, i.e. gain in associated power. Due to the general periodicity of gravity waves, the absorbed power is averaged over a time horizon lasting several wave periods. The performance of the algorithm is assessed through the numerical simulation of a point absorber subject to motions in heave in both regular and irregular waves. The algorithm is found to converge towards the optimal controller damping in each sea state. Additionally, the model-free approach ensures the algorithm can adapt to changes to the device hydrodynamics over time and is unbiased by modelling errors.**

***Index Terms*—Wave energy converter (WEC), power take-off (PTO) system, reinforcement learning (RL), Q-learning.**

I. Introduction

WAVE power is a renewable energy source that can significantly contribute to the reduction of our dependence on fossil fuels in the future due to its enormous scale, with a potential of up to 3 TW of wave power globally [1]. However, the commercialization of wave energy converter (WEC) devices is still in its infancy, with a large number of possible designs having been proposed. A comprehensive review of the current technologies can be found in [2]. Point absorbers represent an established offshore WEC technology, with examples being the devices produced by Ocean Power Technologies [2]. These devices comprise of a small floating body subject to wave loading, whose motions are resisted by a power take-off (PTO) system of either hydraulic or electrical nature. Although point absorbers are expected to be deployed in arrays so as to exploit the advantage of economies of scale [3], in this work a single, axisymmetric device is considered for simplicity, in particular analysing only heaving motions.

Since the initial studies of WECs, different control strategies have been analysed in order to maximize energy absorption, as reviewed by [4]. A more recent review of the state-of-the-art control methods can be found in [5]. From hydrodynamic considerations, complex-conjugate control results in optimal power extraction, as it aims to obtain resonance between the system and the incident waves [4]. However, achieving optimal control in practice may result in excessive motions of, and loads on, the device in energetic sea states, and requires knowledge of the future wave excitation. Since the 1970s, alternative suboptimal control schemes have been developed, including physical constraints on the motions, forces and power rating of the device [3]. These strategies usually optimize the control variables for maximum time-averaged power extraction through an iterative process [3].

Latching and model-predictive control are examples of acausal real-time control strategies for WECs, since their performance strongly depends on having future information of the wave excitation force, typically over a short time horizon [5]. On the one hand, latching control, originally proposed by [6], tries to maximize energy absorption by controlling the duration of the time interval when the device is locked in place through a special mechanism (as opposed to being linearly damped) so as to achieve resonance conditions [7]–[9]. On the other hand, at each time instant, model predictive control applies the force that results in maximum future energy extraction over a pre-defined time horizon, whilst still respecting any constraints on the motions or loading of the device [10]–[12]. Whereas latching control is difficult to scale to array problems, model predictive control has been successfully applied to multi-body devices and even small array problems [13]–[15]. However, the greatest problem with the latter strategy is that the optimization process is not guaranteed to converge, so that alternative solutions may be required. Since this is performed in real-time, it may impose a serious computational burden on the controller. An additional real-time control strategy is the Simple but Effective control proposed by [16]. With this technique, the control force is adjusted in order to meet a prescribed force or velocity setpoint, which is obtained by modelling the current excitation force as a narrow banded function [5]. The performance of this simple method lies close to that of model predictive control and even outperforms it in long waves with a short wave height [5].

Manuscript received December 22, 2015; revised April 4, 2016; accepted May 11, 2016. Date of publication June 1, 2016; date of current version October 7, 2016. This work was supported in part by the Energy Technology Institute, in part by the RCUK Energy Program, in part by the EPSRC, in part by Wave Energy Scotland, in part by the Energy Technology Institute and the Research Council Energy Program as part of the IDCORE program under Grant EP/J500847, and in part by the Engineering and Physical Sciences Research Council under Grant EP/J500847/1. Paper no. TSTE-01065-2015.

E. Anderlini is with the Industrial Doctoral Centre in Offshore Renewable Energy, Edinburgh EH9 3JL, U.K. (e-mail: E.Anderlini@ed.ac.uk).

D. I. M. Forehand is with the Institute of Energy Systems, University of Edinburgh, Edinburgh EH9 3DW, U.K. (e-mail: D.Forehand@ed.ac.uk).

P. Stansell is with Dell SecureWorks, Edinburgh EH3 5DA, U.K. (e-mail: paulstansell@gmail.com).

Q. Xiao is with the Department of Naval Architecture, Ocean, and Marine Engineering, University of Strathclyde, Glasgow G4 0LZ, U.K. (e-mail: qing.xiao@strath.ac.uk).

M. Abusara is with the College of Engineering, Mathematics and Physical Sciences, University of Exeter, Penryn TR10 9FE, U.K. (e-mail: M.Abusara@exeter.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2016.2568754

Alternatively, suboptimal causal control schemes have also been researched extensively. Although they do not require any future wave information, they employ time-averaged sea conditions, thus requiring the assumption of stationary sea state for a specified time interval [3]. Through numerical modelling, it is possible to find the PTO linear damping (resistive or passive control) or combination of damping and stiffness (reactive or phase control) that result in maximum energy absorption for each sea state, whilst still respecting displacement and force constraints. Whereas this method results in a loss in efficiency as compared with on-line control schemes [15], resistive and reactive control are conceptually simple and have lower controller computational cost than model predictive control. Moreover, the control algorithm can be easily scaled up to arrays of WECs, as considered by [3].

The main disadvantage of the aforementioned methods is that they rely on internal models of the body dynamics to determine the optimal control variables. As a consequence, not only do modelling errors affect negatively the energy absorption of WECs, but also changes to the device over time, whether due to slow marine growth or sudden non-critical subsystems failures, cannot be taken into account. Hence, this paper proposes the application of reinforcement learning (RL) for the on-line, model-free optimal control of WECs. This is a type of unsupervised learning that has greatly contributed to the development of autonomous robots over the past two decades [17]. Additionally, [18] have recently used it for the improvement of the maximum point tracking algorithm for the control of wind energy turbines.

As a first application, this paper focuses on the development of RL-based passive control for a point absorber. The performance of the novel control algorithm is assessed through the numerical simulation of a single-degree-of-freedom point absorber. Realistic force constraints are applied to the generator and the efficiency of the PTO system is taken into account. Initially, single sea states are considered for regular and irregular waves. Afterwards, the device is tested in irregular waves with varying sea state conditions.

II. Optimum Passive Control of a Point Absorber

A. System Description

A diagram of the point absorber analysed in this work can be seen in Fig. 1. The mechanical energy derived from the motions of the float due to the wave excitation is converted into hydraulic and then electrical energy by the PTO system. A hydraulic PTO unit, whose design is taken from [19]–[21], is selected due to its robustness, capacity for energy storage and speed control [21]. The motion of the float drives a two-way ram that pumps high-pressure (HP) oil into the circuit. A rectifying valve ensures the hydraulic motor is driven only in one direction. Additionally, the motor rotational speed, ω_{m}, is smoothed out through a gas accumulator system, made of HP and low-pressure cylinders, the latter designed to prevent cavitation [21]. The motor is connected to an induction generator. The produced electrical power, with current I, at voltage V, phase shift ϕ, is fed into the electrical network after stepping up the voltage through a transformer. No expensive, fully-rated power converters are required

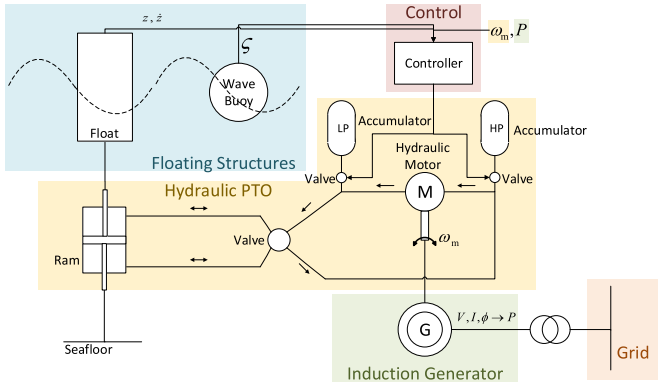

Fig. 1. Diagram of the grid-connected point absorber with its hydraulic PTO.

because the hydraulic PTO unit enables the controllability of the output current [21]. The controller can increase or decrease the flow in the hydraulic circuit by opening or closing the valves connected to the accumulators based on the feedback value of ω_{m}. Furthermore, in order to maximize the output power, the controller relies on knowledge of the vertical displacement, z, and velocity of the float, $\dot{z}$, obtained through an accelerometer, the wave elevation, ζ, fed-in by an external neighbouring wave buoy, and the generated real power, $P = \sqrt{3} IV \cos\phi$.

B. Hydrodynamic Modelling

For simplicity, the point absorber is constrained to oscillations in heave, which is indicated by the index 3. Assuming linear wave theory and small body motions, the response of the device can be obtained from the combination of the inertial, hydrostatic, radiation and excitation forces in addition to the force exerted by the controller [22]. Hence, using Cummin's formulation for the radiation force [23], the equation of motion of the device can be expressed in the time domain as [21]:

$$\left(M + A_{3,3}(\infty)\right)\ddot{z}(t) + \int_0^t K_{3,3}(t-\tau)\dot{z}(\tau)d\tau + C_{3,3}z(t) = F_3(t) + F_{\mathrm{PTO}}(t), \quad (1)$$

where M is the displaced mass of the device, $C_{3,3}$ the hydrostatic restoring stiffness coefficient, $A_{3,3}(\infty)$ the added mass at infinite wave frequency, and $K_{3,3}(t)$ the radiation impulse response function. These variables can be calculated using the commercial program WAMIT. Furthermore, in (1), F_3 represents the excitation force, which is calculated from the convolution of the diffraction coefficients, calculated by WAMIT, and the wave elevation as described in [24], and F_{PTO} the control force.

Eq. (1) is represented by the block diagram in Fig. 2, where the radiation convolution integral is approximated by a state-space formulation due to its lower associated computational cost. Frequency-domain system identification is employed in order to obtain state-space matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ according to the procedure described by [21].

C. Optimum Passive Control

In passive or resistive control, the controller action is modelled as a damping term [3], as shown in Fig. 2, where the control

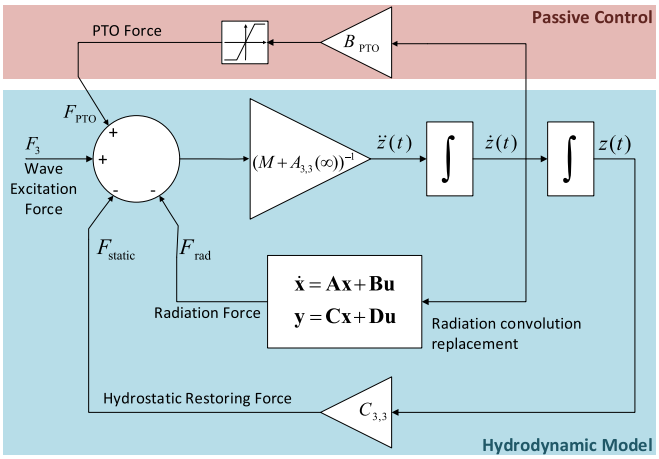

Fig. 2. Block diagram used for the calculation of the motion of the float.

force is given by:

$$F_{\mathrm{PTO}}(t) = -B_{\mathrm{PTO}}\dot{z}(t). \tag{2}$$

The PTO damping coefficient can be modified by changing the pressure within the hydraulic circuit. Hence, this work focuses on the control of B_{PTO} directly, without developing a detailed wave-to-wire model. In practice, there is a limit F_{Max} on the force that can be exerted due to the rating of the motor. Hence, the magnitude of the PTO force is bounded within $\pm F_{\mathrm{Max}}$ in the simulation through the saturation block shown in Fig. 2.

In the real device, power losses occur in the actuator, the hydraulic system, and the electrical generator [3]. These are modelled with an efficiency measure for the PTO system, η. In this work, a value of 75% has been employed due to the low-energy sea states analysed based on [3]. Hence, it is possible to compute the generated real electrical power, which corresponds to $P = \sqrt{3}IV\cos\phi$, as:

$$P = -\eta F_{\mathrm{PTO}}\dot{z}. \tag{3}$$

If there are no force constraints, the optimal PTO damping coefficient for maximum power absorption, $B_{\mathrm{PTO\,opt}}$, is a function of the wave period, T, in regular waves [25], whilst it depends on the mean zero-crossing period in irregular waves. When the force clip is modelled, such a relationship does not exist, since the significant wave height is important to determine when the limit is applied. Hence, in order to find $B_{\mathrm{PTO\,opt}}$ it is necessary to run an optimization in each sea state, e.g. with the Nelder–Mead simplex algorithm [3], using multiple wave traces in irregular waves.

The optimal damping coefficient is stored for each sea state in a table. During the actual operation of the device, the controller tries to achieve the value corresponding to the current sea state by changing the pressure in the hydraulic system. Nevertheless, this approach can be heavily biased by the modelling errors and it cannot take into account modifications to the hydrodynamics of the device over time, e.g. due to marine growth.

In regular waves, the performance of passive control can be assessed against the theoretical maximum limit on the power

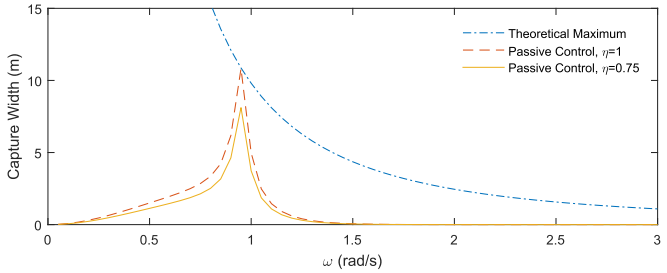

Fig. 3. Performance of a point absorber using passive control with two PTO efficiency values as compared with the case of theoretical maximum power absorption.

extraction by using the concept of capture width, which is defined to be the ratio at each frequency of the mean absorbed power by a WEC to the mean wave power per unit width [26]. In deep water, the mean wave power per unit width is given by [26]:

$$P_{\mathrm{w}} = \frac{1}{4}\frac{\rho g^2 A^2}{\omega}, \tag{4}$$

where $\rho = 1000$ kg/m^3 is the water density, $g = 9.80665$ m/s^2 the gravitational acceleration, A the wave amplitude and ω the circular wave frequency. For an axisymmetric buoy moving in heave, different authors have shown that the theoretical maximum capture width in deep water is given by [26]:

$$L_{\mathrm{opt}} = \frac{\lambda}{2\pi} = \frac{g}{\omega^2}. \tag{5}$$

For a cylindrical point absorber with a diameter of 10 m, and a draught of 8 m (later used in Section IV), the capture width of the device with passive control is shown in Fig. 3 in regular waves of unit amplitude and with the circular wave frequency ranging from 0 to 3 rad/s in steps of 0.005 rad/s. Two values for the efficiency of the PTO system are used (100% and 75%). The absorbed power has been calculated using the optimal PTO damping coefficient for each wave frequency. This value has been divided by the wave power per unit width for each wave frequency as given by Eq. (4). The curves are compared against the optimal capture width [see Eq. (5)], whose values are very high for low wave frequencies. As it can be seen from Fig. 3, with passive control the best performance is achieved at the natural frequency of the device.

III. RL Control

A. Background

In RL [27], an agent, which is in a particular state s_n, interacts with the surrounding environment by taking an action a_n, where n defines the time step of the RL algorithm. The agent then moves to a new state, s_{n+1}, and the action is followed by a reward, r_{n+1}, depending on its outcome. The action selection process is modelled as a Markov decision process based on the value function, which expresses the estimate of the future reward. The agent is expected to learn an optimal behaviour, known as policy, over time for the maximization of the total reward.

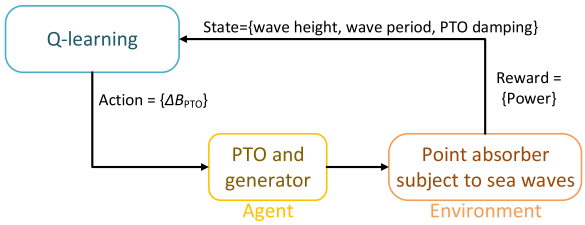

Fig. 4. Block diagram of the RL control of the point absorber.

If the agent selects an action based purely on the aim of maximising the reward function (i.e. exploiting the environment), it will never visit states other than the usual ones, and these other states may in fact result in higher rewards. This is known as the issue of exploration versus exploitation. Hence, it is still beneficial to adopt an approach that ensures some exploration at the expense of exploitation, particularly for the initial stages. Once the simulation has been initialized, the balance may be shifted towards exploitation.

RL methods can be divided into three main categories: dynamic programming, temporal difference and Monte-Carlo methods [27]. Of these, temporal difference strategies seem most appropriate, since they present a real-time implementation. Additionally, in order to limit modelling errors and to pick up changes in the device behaviour over time, model-free techniques are of interest, which use the action-value function $\boldsymbol{Q}(s,a)$. Of these methods, Q-learning has been selected, which is extensively used in the robotics industry [17].

The one-step update of the algorithm is:

$$\boldsymbol{Q}_{n+1}\left(s_n,a_n\right) = \boldsymbol{Q}_n\left(s_n,a_n\right) + \alpha_n\left[r_{n+1} + \gamma \max_{a'\in A}\boldsymbol{Q}_n \times \left(s_{n+1},a'\right) - \boldsymbol{Q}_n(s_n,a_n)\right], \quad (6)$$

where α_n is known as the learning rate, which regulates how much previous learning is retained in the update of the action-value table, and γ is the discount factor, which determines whether preference should be given to immediate or future rewards. As the optimal action-value function is estimated independently of the current policy, Q-learning is classified as an off-policy scheme.

B. Application to the Passive Control of WECs

As shown in Fig. 4, RL can be used to learn the optimal PTO damping coefficient in each sea state by relying purely on observations of the environment, i.e. the device interacting with the waves, rather than internal models. At each step, the controller selects a change in B_{PTO}, the action, which is implemented by the hydraulic PTO unit, the agent. This results in a reward that is a function of the generated power and in a change of state, where each state is represented by one value for the significant wave height, H_{s}, the mean zero-crossing period, T_{z}, and the PTO damping coefficient.

Due to the oscillatory nature of gravity waves, the generated power in the reward function needs to be averaged over at least one wave cycle. The averaging is performed over a horizon, H, during which the state s_n and action a_{n-1} are constant, so that all time steps $n-1$, n, etc. now have length H. Then, a new action a_n is selected, which results in an immediate change of state to s_{n+1} and a new averaging process.

The state and action spaces, reward function, learning and exploration rates, and discount factor of the WEC control RL formulation are described in detail in the following sections.

1) State Space: As mentioned before, the state variables are taken to be the significant wave height, mean zero-crossing period and PTO damping coefficient so that the adopted RL state space is:

$$S = \left\{ s|s_{j,k,l} = (H_{\mathrm{s},j}, T_{\mathrm{z},k}, B_{\mathrm{PTO},l}), \quad \begin{matrix} j=1:J, \\ k=1:K, \\ l=1:L \end{matrix} \right\}. \quad (7)$$

A compromise needs to be found in the selection of J, K, and L, since a large number of states may result in excessively slow convergence, while small values may strongly affect the learning accuracy [18]. The values of J and K are usually given by the wave data at the site of deployment. Ranges of $H_{\mathrm{s}} = [0:9]$ m and $T_{\mathrm{z}} = [5:15]$ s are typical, in steps of either 0.5 or 1 m or s respectively [28]. With a hydraulic PTO system, the value of L will be set by the number of accumulators. Indeed, as shown in [19], the time series of the PTO force is characterized by a number of discrete values.

2) Action Space: Considering the selected state space, for passive control the action space is thus:

$$A = \{a|(-\Delta B_{\mathrm{PTO}}, 0, +\Delta B_{\mathrm{PTO}})\}, \quad (8)$$

where $\Delta B_{\mathrm{PTO}} = B_{\mathrm{PTO},k+1} - B_{\mathrm{PTO},k}$. The states corresponding to the minimum or maximum damping coefficient, i.e. $B_{\mathrm{PTO},1}$ and $B_{\mathrm{PTO},L}$, have a limited (from 3 to 2) number of actions in order to prevent the controller from exceeding the state space boundary. For instance, for $B_{\mathrm{PTO},1}$, the action $-\Delta B_{\mathrm{PTO}}$ is precluded in the current state.

3) Reward: The reward function represents the goal that the controller is expected to maximise. Hence, for the passive control of WECs, the reward function needs to be a function of the absorbed power. However, the mean generated power, P_{avg}, is more influenced by changes in the significant wave height than variations in the PTO damping coefficient. This can be dealt with by using $P_{\mathrm{avg}}/{H_{\mathrm{s}}}^2$ as a reward, since the absorbed power is proportional to the square of the significant wave height [28]. In addition, due to the coarse discretization of the state variables and the stochastic nature of irregular waves, not only should the generated power in (3) be averaged over a horizon H to produce P_{avg}, but the reward function needs to be built on the mean of a number M of these values for each state. This can be achieved by storing the M most recent $P_{\mathrm{avg}}/{H_{\mathrm{s}}}^2$ values for each state in a matrix, $\boldsymbol{R}$, whose size is at most $n_s \times M$, with $n_s = J \times K \times L$ being the number of states. It is then possible to obtain the mean value in each state and express it with the vector $\boldsymbol{m} = \langle \boldsymbol{R}(s,m)\rangle_{m=1:(M \vee \mathrm{end})}$ of size n_s. It is important to notice that in $\boldsymbol{m}$ the states are arranged with a vectorized version of Eq. (7), so that discrete values of B_{PTO} represent the

inner-most loop, the discrete values of H_s the middle loop and the discrete values of T_z the outermost loop.

Depending on the magnitude of ΔB_{PTO}, for $B_{\mathrm{PTO}} > 0$ there can be very little difference between the mean of neighbouring PTO damping coefficient values. This could cause serious issues for the convergence of the Q-learning algorithm, where the benefit of picking the optimal damping coefficient in each sea state should be evident. This problem can be addressed by raising the values within $\boldsymbol{m}$ to a power. In order to avoid rewards that require excessive memory, it is advantageous from a mathematical perspective to first normalize the value of the vector for each state with the maximum value for each sea state. Hence, for the state s_n, the maximum value needs to be searched between the indices $o = \text{floor}\left(\frac{s_n - 1}{L}\right) L + 1$ and $p = \text{floor}\left(\frac{s_n - 1}{L}\right) L + L$ of the vector $\boldsymbol{m}$. The power value, u, needs to be an odd number in order to keep the sign of the generated power values. The finer the discretization of the PTO damping coefficient, the greater u should be in order to speed up the learning process. However, a very large value may cause convergence problems in irregular waves due to the possible noise in the mean power values, so care should be taken in the selection of u.

In addition, in extreme seas the selected optimal damping coefficient may result in excessive motions [12], e.g. complete submergence or emergence of the machine. This may cause severe structural damage if not complete failure. In order to prevent this, a penalty, $p < 0$, is returned when the magnitude of the maximum displacement over the averaging horizon H exceeds a set value, z_{Max}. Using a penalty $p = -2$, the resulting reward function is thus:

$$r_{n+1} = \begin{cases} \left[\dfrac{\langle \boldsymbol{m}(s_n)\rangle}{\max_{i=o:p}\langle \boldsymbol{m}(i)\rangle}\right]^u, & \text{if } |\max(z)| \le z_{\mathrm{Max}} \\ -2, & \text{if } |\max(z)| > z_{\mathrm{Max}}. \end{cases} \tag{9}$$

4) Exploration Strategy, Learning Rate and Discount Factor: In order to ensure exploration, an ϵ-greedy strategy has been adopted [27]. This means that at each step of the Q-learning algorithm, the action is selected as:

$$a_n = \begin{cases} \arg\max_{a' \in A} \boldsymbol{Q}_n(s_n, a'), & \text{with probability } 1 - \epsilon_n \\ \text{random action}, & \text{with probability } \epsilon_n, \end{cases} \tag{10}$$

with ϵ_n being the exploration rate. During the initial stages of a RL run, it is desirable to explore as many of the state-action pairs as possible and then slowly shift the focus towards exploitation as the learning progresses. Hence, the exploration rate can be expressed as:

$$\epsilon_n = \begin{cases} \epsilon_0, & \text{if } N \le 0 \\ \dfrac{\epsilon_0}{\sqrt{N}}, & \text{if } N > 0, \end{cases} \tag{11}$$

where $N = \sum_{i=1:n_a} \boldsymbol{N}_n(s_n, a_i) - N_{\min\epsilon}$, with $\boldsymbol{N}_n(s_n, a_n)$ indicating the total number of visits to the current state-action pair $s_n - a_n$ (n_a is the number of actions) and $N_{\min\epsilon} = 25$ the minimum number of visits for an initial random exploration. The initial exploration rate is set to $\epsilon_0 = 0.5$.

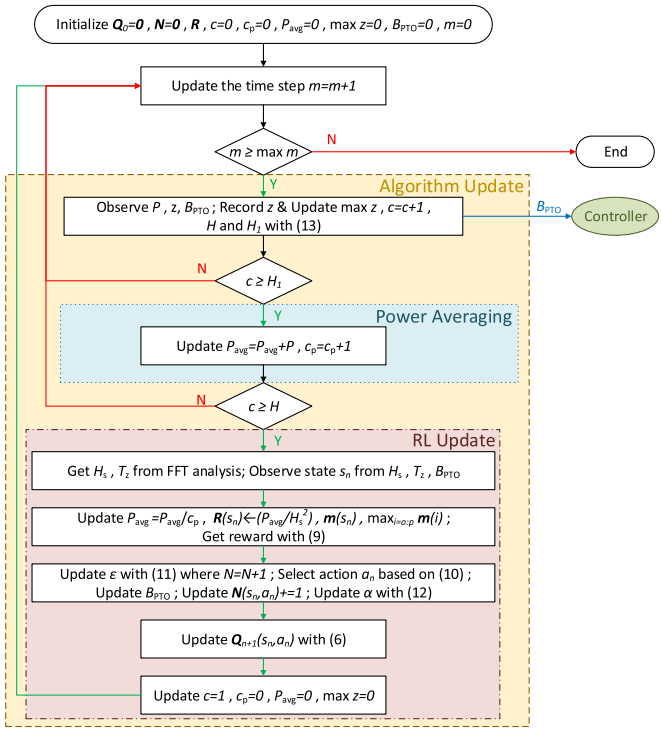

Fig. 5. Flowchart of the Q-learning algorithm for the passive control of WECs.

Similarly, a high initial learning rate is selected which slowly decays in order to ensure convergence of the Q-values:

$$\alpha_n = \begin{cases} \alpha_0, & \text{if } \boldsymbol{N}_n(s_n, a_n) \le N_{\min\alpha} \\ \dfrac{\alpha_0}{\boldsymbol{N}_n(s_n, a_n)}, & \text{if } \boldsymbol{N}_n(s_n, a_n) > N_{\min\alpha}, \end{cases} \tag{12}$$

where an initial learning rate $\alpha_0 = 0.4$ and $N_{\min\alpha} = 5$ are used throughout this work.

From a comparison of (11) and (12), it is clear that a slower decay is sought for the learning rate, so that sufficient exploration is ensured even as the learning process goes on. In order to ensure that changes to the device are taken into account, e.g. due to marine growth or even subsystems failures, the learning and exploration rates should be reset on a predefined, regular basis.

In Q-learning, the controller seeks to maximise the total discounted future rewards, so that it is necessary to specify a discount factor [17]. A value of $\gamma = 0.75$ has been used throughout this work as in [18].

C. Algorithm

The Q-learning algorithm used in this work can be seen in Fig. 5. The first stage of the algorithm is the initialization of all required variables. $\boldsymbol{Q}$ and $\boldsymbol{N}$ are matrices of dimensions $n_s \times n_a$, where the number of actions is $n_a = 3$. The value of L for the specification of the size of the matrix $\boldsymbol{R}$ has been set to 10 in regular waves and 25 in irregular waves. In order to speed up convergence, the entries of the $\boldsymbol{R}$ matrix are precomputed in a run in a similar wave trace, whilst taking random actions. Simulations can also be used to initialize the $\boldsymbol{R}$ matrix for the

full-scale device, since its entries will slowly be replaced from those of the actual environment.

After the initialization phase, the algorithm is run indefinitely until maintenance is due. At every time step m, with time step length Δt, the desired damping coefficient value is stored by the controller so that it can be implemented through changes in the hydraulic pressure in the PTO unit. Additionally, the generated power and vertical buoy displacement are sampled in order to obtain, respectively, the mean absorbed power, and maximum displacement at the end of each horizon after H time steps. Furthermore, at each update of the Q-learning algorithm, an external program returns the values of H_{s} and T_{z}, which are calculated using spectral analysis and fast Fourier transforms (FFT) from the record of the wave elevation, ζ, within the horizon as described in [28], based on a unidirectional wave spectrum for simplicity.

As can be seen from Fig. 5, the generated power in each state is averaged over the horizon H only after an interval $H_1 \approx 5T_{\mathrm{z}}$, over which transient effects due to the change in PTO damping coefficient are dominant. In selecting the horizon length H a compromise needs to be found between a small value for quicker response and a large value for a more stable algorithm. Indeed, although a sea state can be stationary for a period ranging from 15 to 30 minutes [28], individual neighbouring waves within this time can present very different characteristics. Continuous changes in the sea state from a step of the RL to the next prevent the algorithm from converging, since by taking an action a_n in state s_n the agent may land in a different state every time depending on the sea conditions. Hence, a value of $H = 30T_{\mathrm{z}}$ is selected in irregular waves, whereas $H = 10T$ may be used in regular waves to speed up convergence. Furthermore, the time discretization of the algorithm requires the horizon to be expressed in time steps rather than seconds, so that:

$$H_1 = \text{round}\left(\frac{5T_{\mathrm{z}}}{\Delta t}\right) \text{ and } H = \text{round}\left(\frac{30T_{\mathrm{z}}}{\Delta t}\right). \qquad (13)$$

IV. Simulation Results

A. Simulation System

Numerical simulations have been run for the same device used in [12], i.e. a floating vertical cylinder of radius 5 m and draught 8 m in deep water as shown in Fig. 1. The time domain solution for this problem is standard, with this specific example being treated also by [22]. As in [12], a fifth-order state-space system has been used to approximate the radiation convolution. The hydrodynamic model in Fig. 2 has been expressed in state-space format and discretized with a bilinear transform [29], where the sampling time has been set to $\Delta t = 0.1$ s. The maximum PTO force has been assumed to be 1 MN, while the float displacement has been limited to ± 5 m.

The program used for the simulation of the point absorber is summarized in Fig. 6 for clarity. A wave model is required in order to determine the wave elevation time series, whereas in reality buoy measurements will be used as in Fig. 1. For irregular waves, it is necessary to specify the amplitude wave spectrum $S(\omega)$ for a n_ω number of circular wave frequencies.

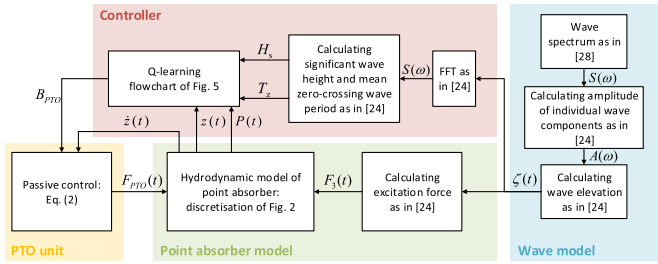

Fig. 6. Workflow diagram of the program used to simulate the point absorber.

The wave elevation is then computed from the superposition of the n_ω individual wave components, each with a wave amplitude $A(\omega) = \sqrt{2S(\omega)\Delta\omega}$, where $\Delta\omega$ is the frequency step [24]. Not only is the wave elevation used to determine H_{s} and T_{z} in each state, but also to compute the excitation force through the diffraction convolution integral [24].

The PTO system of the device has been assumed to be composed of 4 accumulators, with a maximum PTO damping coefficient of 800 kN·s/m for the sea states under study. Hence, nine RL states are used when a single sea state is considered, with the linear damping coefficient ranging from 0 to 800 kN·s/m in steps of $\Delta B_{\mathrm{PTO}} = 100$ kN·s/m. With this discretization, a value of $u = 21$ has been selected in order to decrease the learning time, while avoiding possible problems with noise in the reward function in irregular waves. When the control is tested in multiple sea states in random seas, only five damping coefficients values are employed, with the same range but $\Delta B_{\mathrm{PTO}} = 200$ kN·s/m, in order to limit the overall number of states and thus speed up convergence. However, a wider range and finer resolution are likely to be required for a more realistic implementation.

B. Results in Regular Waves

Regular waves have been analysed first in order to assess the convergence properties of the proposed RL control under deterministic conditions. A single sea state ($J = K = 1$) with unit wave amplitude and a wave period of 8 s has been considered, with the time series lasting 4 hours.

Fig. 7(a) shows the convergence of the RL algorithm towards the optimal PTO damping for this sea state, where the optimal value has been calculated through a Nelder–Mead optimization in a 20-minute wave trace. The difference in the mean absorbed power obtained using RL and that obtained using the optimal PTO damping coefficient can be seen in Fig. 7(b), with $P_{\mathrm{avg,opt}} = 70.210$ kW.

Due to the low wave height selected in all simulations, the PTO force never reaches its limit, with the maximum force being 237.910 kN for the optimal B_{PTO} in Fig. 7. In order to analyse the effects of the force clip, or saturation, on the optimal PTO damping coefficient and the learning process, the force limit has been reduced to $F_{\mathrm{Max}} = 237.910$ kN. Then, the wave amplitude has been slightly increased to 1.1 m. This is analogous to the device reaching the original saturation limit in more extreme waves, whilst the validity of the assumption of linear wave theory in the hydrodynamic model is ensured.

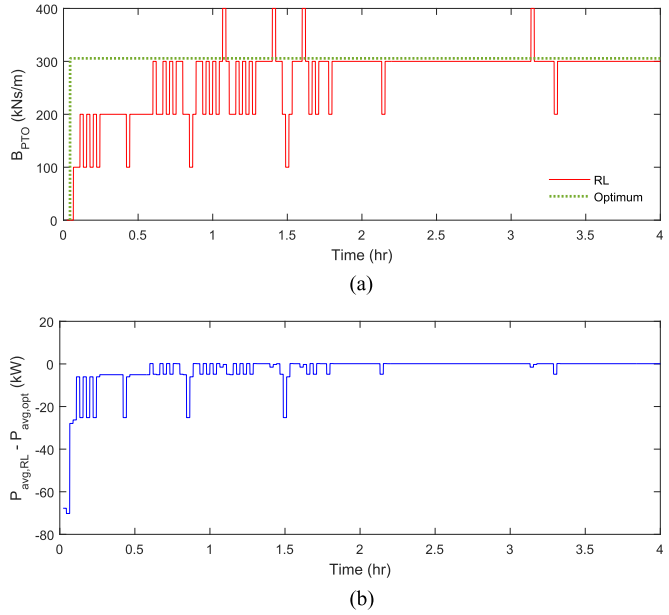

Fig. 7. RL-control-selected and optimal PTO damping coefficient (a) and difference in the corresponding mean absorbed power (b) in regular waves of unit amplitude and a wave period of 8 s.

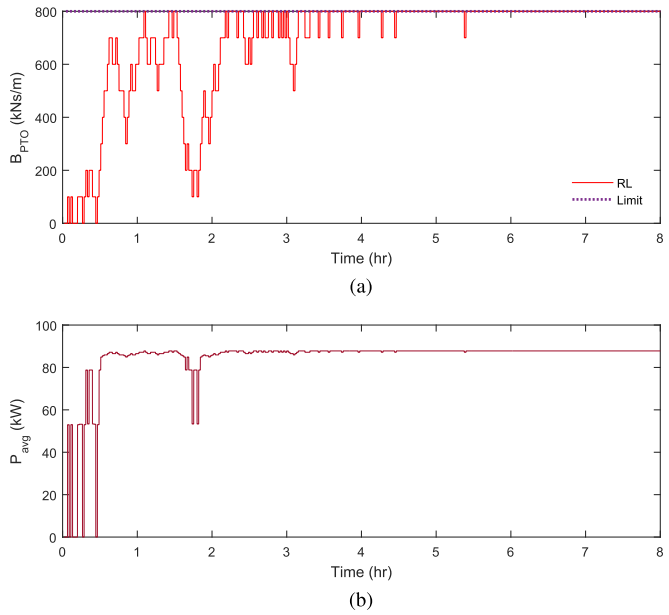

Fig. 8. (a) RL-control-selected PTO damping coefficient and (b) corresponding mean generated power in regular waves with $H_{\mathrm{s}} = 2.2$ m and $T_{\mathrm{z}} = 8$ s, when $F_{\mathrm{Max}} = 237.910$ kN.

The convergence of the RL algorithm towards a new PTO damping coefficient and the corresponding mean absorbed power can be seen in Fig. 8(a) and (b) respectively. Note that the optimal B_{PTO} value would be far beyond the state space we have defined, so that it is saturated at 800 kN·s/m. The reason for this behaviour can be understood by looking at Fig. 9, which shows the variation of the PTO velocity and force over time with the two different PTO damping coefficients, 300 and 800 kN·s/m, in regular waves of unit amplitude and a wave period of 8 s. With the lower saturation limit $F_{\mathrm{Max}} = 237.910$ kN, the controller

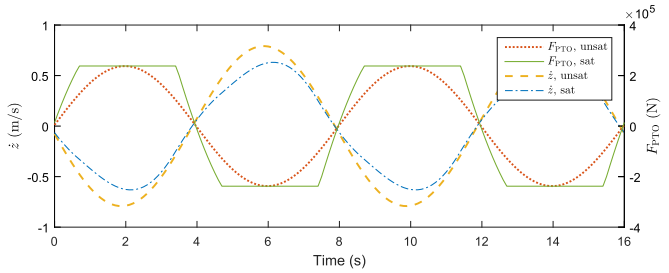

Fig. 9. PTO velocity and force over two wave periods in regular waves with $H_{\mathrm{s}} = 2$ m and $T_{\mathrm{z}} = 8$ s for the cases of unsaturated ($B_{\mathrm{PTO}} = 300$ kN·s/m) and saturated ($B_{\mathrm{PTO}} = 800$ kN·s/m) PTO force, when $F_{\mathrm{Max}} = 237.910$ kN.

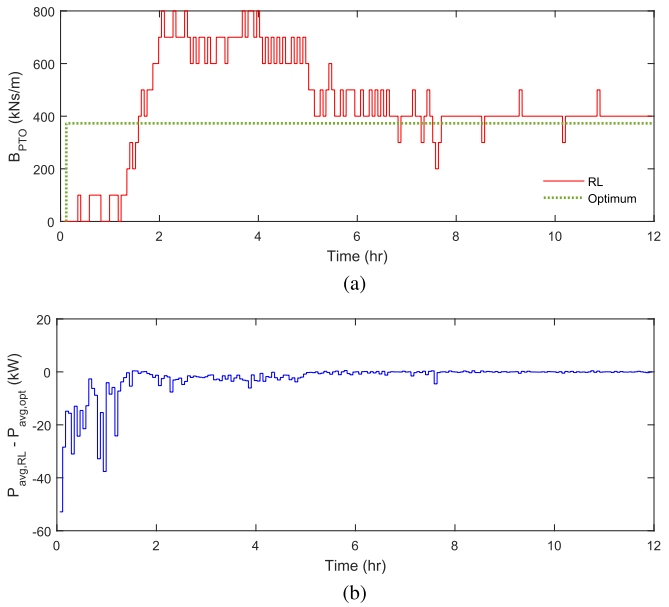

Fig. 10. (a) RL-control-selected and optimal PTO damping coefficient and (b) difference in the corresponding mean absorbed power in irregular waves with $H_{\mathrm{s}} = 2$ m and $T_z = 7$ m, generated using a JONSWAP spectrum.

tries to maximise the absorbed power by maximising the area under the curve of the PTO force through a square wave. The limit on the PTO damping coefficient prevents the realization of a fully non-linear, bang-bang type of control response.

C. Results in Irregular Waves

In irregular waves, longer wave traces are considered, each lasting 12 hours and 15 minutes. In order to ensure the motions of the model are fully developed, the RL control is run only after 15 minutes from the start of the time series. In all cases considered in this section, the force and displacement constraints are not reached.

1) Single Sea State: Firstly, a wave trace generated using a single JONSWAP spectrum [30] is considered, with a significant wave height of 2 m and a peak wave period of 9 s, corresponding to $T_{\mathrm{z}} = 7$ s from the FFT analysis. Although there are oscillations in the predicted values of H_{s} and T_{z} over neighbouring horizons, $J = K = 1$ have been used for simplicity, so that $n_s = 9$.

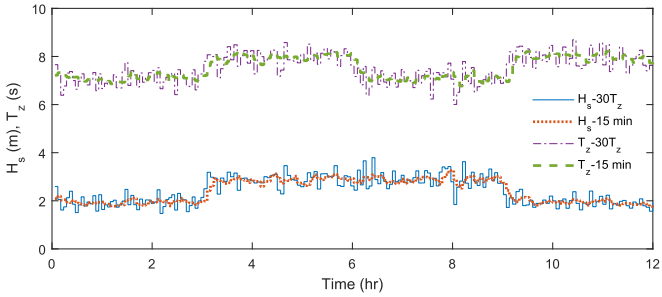

Fig. 11. Significant wave height and mean zero-crossing period calculated over each horizon (continuous lines) and over overlapping 15-minute windows every minute (dotted lines) for the multiple sea state wave trace.

Fig. 10(a) shows the convergence of the RL-selected PTO damping coefficient towards the optimum. The optimal value has been calculated by taking the average of the results obtained through Nelder–Mead optimizations in a 20-minute wave trace with a JONSWAP spectrum with $H_s = 2$ m and a peak wave period of 9 s using five different seed values. The difference in the mean absorbed power obtained using RL and the optimal PTO damping coefficient can be seen in Fig. 10(b), where the optimal mean absorbed power has an average value of 25.686 kW over the 12-hour wave trace.

2) Multiple Sea States: In ocean waves, sea states can last from a minimum of 30 minutes to a maximum of six to eight hours, with swells lasting typically between 3 and 6 hours [28]. Hence, a semi-realistic wave trace (see Fig. 11) has been generated from the concatenation of four sea states, each lasting three hours and corresponding to a JONSWAP spectrum. In order to achieve convergence, the wave trace has been repeated 4 times for a total of 48 hours.

Although only four wave spectra are employed to generate the sea state, determining the sea state over the horizon length H results in four discrete values of both the significant wave height (1–4 m, in steps of 1 m) and the mean zero-crossing wave period (6–9 s, in steps of 1 s). As a result, $J = 4$ discrete T_z are used. However, since the wave energy is too low for the generator to reach its force limit within this wave trace, the optimal damping coefficient is dependent only on the wave period. Therefore, it is sufficient to employ only one discretized value for the significant wave height, $I = 1$, in order to speed up the learning time. Thus, $n_s = 1 \times 4 \times 5 = 20$.

Fig. 12 shows the initial behaviour of the Q-learning algorithm, while Fig. 13 shows the control performance after the optimal PTO damping coefficient has been learnt in each sea state. Figs. 12(a) and 13(a) also present the optimal value for the PTO damping coefficient, calculated as described in the previous section for the four individual sea states. However, as opposed to the RL method, in this case the values of H_s and T_z are obtained from 15-minute moving windows every minute, as shown by the dotted lines in Fig. 11. In Figs. 12(b) and 13(b), it is possible to see the difference in the mean absorbed power obtained using RL and the optimal PTO damping coefficient, where the optimal mean absorbed power has an average value of 26.147, 56.429, 59.191, and 25.208 kW in each sea state respectively, over the 3-hour wave traces.

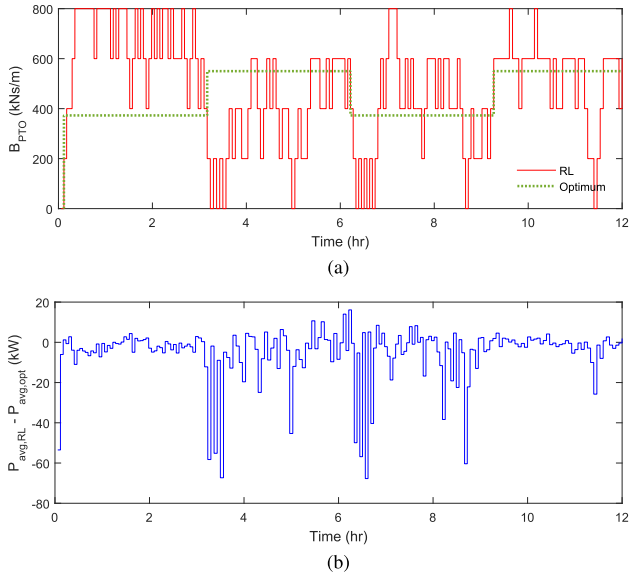

Fig. 12. (a) Optimal and RL-control-selected ($J = 1$, $K = 4$, $L = 5$) PTO damping coefficient and (b) corresponding mean absorbed power in irregular waves with four sea states, generated from the combination of $H_s = 2, 3$ m and $T_z = 7, 8$ s.

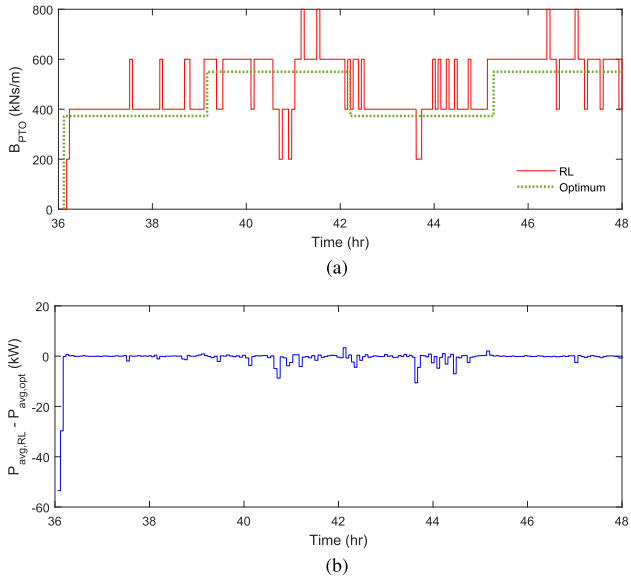

Fig. 13. (a) Optimal and RL-control-selected ($J = 1$, $K = 4$, $L = 5$) PTO damping coefficient and (b) corresponding mean absorbed power in irregular waves with four sea states, generated from the combination of $H_s = 2, 3$ m and $T_z = 7, 8$ s after learning has occurred.

V. Discussion

A. Regular Waves

As can be seen from Fig. 7, in regular waves the RL algorithm can converge towards the optimal PTO damping coefficient for passive control in less than 3 hours starting from a random initialization. This is possible because of the deterministic nature of regular waves, which also enables the use of a relatively short averaging horizon. Similarly, the use of the tabular approach for the reward function would not be necessary. It is also interesting to notice that due to the selected exploration strategy, random

actions may be taken even after the Q-table entries have fully converged.

From Fig. 8, it is clear that the application of the force clip results in the optimal PTO damping coefficient moving to the upper limit. As aforementioned, the reason for this behaviour is the fact that the control force tends to a square wave shape (see Fig. 9), which maximises the area under the curve. Conversely, due to the force saturation, the magnitude of the body velocity, which corresponds to the velocity at the PTO in this simple case, is not significantly affected by the PTO force. Since the absorbed power is proportional to the product of the PTO velocity and force, a square wave shape of the PTO force maximises the amount of generated energy. Hence, the controller is able to turn to a bang-bang type of control action when the force saturates, which can result in greater energy absorption than resistive control, as for instance shown by [31], despite a stronger generator loading. Nevertheless, as this work focuses on the application of RL to resistive control, a relatively low limit has been imposed on the PTO damping coefficient to prevent the controller behaviour from becoming strongly non-linear.

In Fig. 9, it is also interesting to notice that the saturated body velocity, like the PTO force, is no longer sinusoidal. The two curves are still in phase, but the velocity is affected by the higher order harmonics of the PTO force due to the saturation.

Similarly, although the specified limit on the body displacement is never reached in the tests considered, the RL control would be expected to return a higher PTO damping coefficient than the optimal value if this were the case. Indeed, stronger damping is associated with a smaller motion amplitude.

B. Irregular Waves

The statistical reward function is proven to be very effective in the treatment of irregular waves, as it is clear from Fig. 10. However, a longer time is required for convergence to occur as compared with regular waves. This is evident from the comparison of Figs. 12 and 13, which respectively show a random response while the controller is learning and the optimal performance once convergence is achieved. From this analysis of multiple sea states, it is possible to deduce that the controller needs to spend a minimum of 12 hours in each sea state in order to learn the optimal policy by ensuring sufficient exploration, when 5 values are used for the PTO damping coefficient (for a total of 20 states, not all of which are encountered). This time is likely to rise when a finer mesh is used for the RL state space. In particular, assuming the learning time to be proportional to the number of states, a very large number of discrete B_{PTO} values can seriously affect the convergence properties of the algorithm, since the number of states is equal to the product of L and the number of sea states.

Although a 12-hour learning time seems much longer than the 20-minute window used for the Nelder–Mead optimization, multiple iterations are required for convergence with any search technique, so that RL does in fact converge faster in an on-line application. In fact, a real-time, model-free implementation of an exhaustive search method would be impossible. Since in the real environment a wave trace is never repeated exactly, any search scheme would be unable to recognize whether a change in the cost function is due to the change in PTO damping or wave noise. Conversely, as Fig. 13 shows, the proposed RL strategy is able to start the optimization in any sea state from where it left off the last time it entered that specific sea state. Once convergence is achieved, the RL approach is reduced to a look-up-table method until the exploration rate is increased in order to check if there have been any changes in the dynamics of the device. This can be done every season, but it will result in much shorter learning times during which the performance will never be far from the optimum, since the Q-table is already initialized. Thus, as the operational life of a WEC is planned as 25 years, a relatively poor efficiency during the very first stages of operation should not affect the economic performance of the device.

From Fig. 13(a), it may look like the Q-learning algorithm has still not fully learnt the optimal policy even after 48 hours, despite a much better performance as compared with Fig. 12(a). In fact, the Q-table has by now converged towards the correct optimal PTO damping coefficient in each sea state. However, the optimal values in the $\boldsymbol{m}$ vector, used to calculate the reward function, lie closest to $B_{\mathrm{PTO}} = 200$ kN·s/m for $T_{\mathrm{z}} = 6$ s, $B_{\mathrm{PTO}} = 400$ kN·s/m for $T_{\mathrm{z}} = 7$ s, $B_{\mathrm{PTO}} = 600$ kN·s/m for $T_{\mathrm{z}} = 8$ s, and $B_{\mathrm{PTO}} = 800$ kN·s/m for $T_{\mathrm{z}} = 9$ s. Hence, the oscillations in the PTO damping coefficient selected by the Q-learning algorithm in fact correspond to changes in sea state, as it is possible to understand from a close comparison with Fig. 11. As a result, the RL method even presents higher power absorption at some points as compared with the standard resistive control in Fig. 13(b) despite the use of a very coarse RL state space at this stage.

No comparison has been made at this stage with other control strategies, such as latching or model predictive control, because RL is considered to be a method to make existing control strategies independent of the hydrodynamic model of the WEC. Hence, its performance is only as good as the control scheme itself.

VI. Conclusion

An on-line, model-free RL algorithm has been proposed in order to obtain the optimal PTO damping in each sea state for the resistive control of WECs, including penalties for large displacements. Its performance has been assessed through numerical simulations of a single-degree-of-freedom point absorber. In regular waves, the control converges quickly towards the optimal coefficient due to their deterministic nature. In irregular waves, convergence is ensured by employing a statistical reward function, which returns the average over multiple power absorption values recorded in each state. This approach is shown to be effective not only in a single sea state, but also in random waves made from the concatenation of four sea states. Since the control does not rely on internal models of the device, it can be easily implemented on a full-scale machine and it can account for changes to the unit over time, such as due to marine growth or non-critical failures. Additionally, this method can be extended to the phase control of a WEC, although the increase in complexity may result in slower learning times, during

which considerable power losses could be incurred if random actions are taken. Similarly, the technique can be generalised and applied to the control of arrays of the devices.

ACKNOWLEDGMENT

In addition, Mr. Anderlini would like to thank Wave Energy Scotland for sponsoring his Eng.D. research project.

REFERENCES

[1] G. Mørk, S. Barstow, A. Kabuth, and M. T. Pontes, "Assessing the global wave energy potential," *Proc. 29th Int. Conf. Offshore Mech. Artic Eng.*, 2010, pp. 447–454.
[2] A. F. D. O. Falcão, "Wave energy utilization: A review of the technologies," *Renewable Sustain. Energy Rev.*, vol. 14, no. 3, pp. 899–918, 2010.
[3] A. J. Nambiar, D. I. M. Forehand, M. M. Kramer, R. H. Hansen, and D. M. Ingram, "Effects of hydrodynamic interactions and control within a point absorber array on electrical output," *Int. J. Marine Energy*, vol. 9, pp. 20–40, 2015.
[4] S. H. Salter, J. R. M. Taylor, and N. J. Caldwell, "Power conversion mechanisms for wave energy," *Proc. Inst. Mech. Eng. M, J. Eng. Maritime Environ.*, vol. 216, no. 1, pp. 1–27, 2002.
[5] J. V. Ringwood and G. Bacelli, "Energy-Maximizing control of wave-energy converters: The development of control system technology to optimize their operation," *IEEE Control Syst. Mag.*, vol. 34, no. 5, pp. 30–55, Oct. 2014.
[6] K. Budal and J. Falnes, "Optimum operation of wave power converter," *Marine Sci. Commun.*, vol. 3, no. 2, pp. 133–150, 1977.
[7] A. Babarit, G. Duclos, and A. H. Clément, "Comparison of latching control strategies for a heaving wave energy device in random sea," *Appl. Ocean Res.*, vol. 26, no. 5, pp. 227–238, 2004.
[8] A. Babarit and A. H. Clément, "Optimal latching control of a wave energy device in regular and irregular waves," *Appl. Ocean Res.*, vol. 28, no. 2, pp. 77–91, 2006.
[9] A. F. D. O. Falcão, "Phase control through load control of oscillating-body wave energy converters with hydraulic PTO system," *Ocean Eng.*, vol. 35, nos. 3/4, pp. 358–366, 2008.
[10] J. Hals, J. Falnes, and T. Moan, "Constrained optimal control of a heaving buoy wave-energy converter," *J. Offshore Mech. Arctic Eng.*, vol. 133, no. 1, p. 011401, 2011.
[11] T. K. A. Brekken, "On model predictive control for a point absorber wave energy converter," *Proc. IEEE Trondheim PowerTech*, 2011, pp. 1–8.
[12] J. a. M. Cretel, G. Lightbody, G. P. Thomas, and A. W. Lewis, "Maximisation of energy capture by a wave-energy point absorber using model predictive control," *IFAC Proc. Volumes (IFAC-PapersOnline)*, vol. 18, no. PART 1, pp. 3714–3721, 2011.
[13] K. U. Amann, M. E. Magaña, S. Member, and O. Sawodny, "Model predictive control of a nonlinear 2-body point absorber wave energy converter with estimated state feedback," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 336–345, Apr. 2015.
[14] D. Oetinger, M. E. Magaña, S. Member, and O. Sawodny, "Decentralized model predictive control for wave energy converter arrays," *IEEE Trans. Sustain. Energy*, vol. 5, no. 4, pp. 1099–1107, Oct. 2014.
[15] O. Sawodny, D. Oetinger, and M. E. Magaña, "Centralised model predictive controller design for wave energy converter arrays," *IET Renewable Power Generation*, vol. 9, no. 2, pp. 142–153, 2015.
[16] F. Fusco and J. V. Ringwood, "A simple and effective real-time controller for wave energy converters," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 21–30, Jan. 2013.
[17] S. G. Khan, G. Herrmann, F. L. Lewis, T. Pipe, and C. Melhuish, "Reinforcement learning and optimal adaptive control: An overview and implementation examples," *Annu. Rev. Control*, vol. 36, no. 1, pp. 42–59, 2012.
[18] C. Wei, Z. Zhang, W. Qiao, and L. Qu, "Reinforcement learning-based intelligent maximum power point tracking control for wind energy conversion systems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6360–6370, Oct. 2015.
[19] R. Henderson, "Design, simulation, and testing of a novel hydraulic power take-off system for the Pelamis wave energy converter," *Renewable Energy*, vol. 31, no. 2, pp. 271–283, 2006.
[20] A. F. D. O. Falcão, "Modelling and control of oscillating-body wave energy converters with hydraulic power take-off and gas accumulator," *Ocean Eng.*, vol. 34, no. 14-15, pp. 2021–2032, 2007.
[21] D. Forehand, A. E. Kiprakis, A. Nambiar, and R. Wallace, "A bi-directional wave-to-wire model of an array of wave energy converters," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 118–128, Jan. 2016.
[22] J. N. Newman, *Marine Hydrodynamics*. Cambridge, MA, USA: MIT Press, 1977.
[23] W. E. Cummins, "The impulse response function and ship motions," *Schiffstechnik*, vol. 47, no. 9, pp. 101–109, 1962.
[24] J. Falnes, *Ocean Waves and Oscillating Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[25] E. Tedeschi, M. Carraro, M. Molinas, and P. Mattavelli, "Effect of control strategies and power take-off efficiency on the power capture from sea waves," *IEEE Trans. Energy Convers.*, vol. 26, no. 4, pp. 1088–1098, Dec. 2011.
[26] J. Cruz, *OceanWave Energy*. New York, NY, USA: Springer-Verlag, 2008.
[27] R. S. Sutton and A. G. Barto, *Reinforcement Learning*. Cambridge, MA, USA: MIT Press, 1998.
[28] L. H. Holthuijsen, *Waves Oceanic Coastal Waters*. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[29] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, 6th ed. London, U.K.: Pearson, 2008.
[30] Det Norske Veritas, "Environmental conditions and environmental loads," *Recommended Practice*, Oct. 2010, pp. 9–123.
[31] G. Li, G. Weiss, M. Mueller, S. Townley, and M. R. Belmont, "Wave energy converter control by wave prediction and dynamic programming," *Renewable Energy*, vol. 48, pp. 392–403, 2012.

Enrico Anderlini received the M.Eng. degree in naval architecture from the University of Southampton, Southampton, U.K., in 2013. He is currently working toward the Eng.D. degree in offshore renewable energy with the Industrial Doctoral Centre in Offshore Renewable Energy, a partnership of the University of Edinburgh, University of Exeter, and University of Strathclyde, U.K.

His research interests include marine hydrodynamics, control of wave energy converters, and machine learning.

David I. M. Forehand received the B.Sc. (Hons.) degree in mathematics in 1990, the Ph.D. degree in numerical modelling of free-surface waves from the University of Edinburgh, Edinburgh, U.K., in 1999, and the M.Sc. degree in applied mathematics from the University of Oxford, Oxford, U.K., in 1993.

He is currently a Research Associate with the Institute for Energy Systems, University of Edinburgh. His research interests include nonlinear engineering dynamics and the numerical hydrodynamic modeling of wave energy converters and floating bodies.

Paul Stansell received the B.Sc. (Hons.) degree in theoretical physics from the University of Exeter, Penryn, U.K., in 1988, and the Ph.D. degree in computational fluid mechanics from the University of Edinburgh, Edinburgh, U.K., in 1995.

He then researched in the fields of computational fluid dynamics, turbulence and mixing, surface gravity waves, and ocean wave statistics at Edinburgh, Strathclyde, and Heriot-Watt Universities. From 2005 to 2014, he held roles at Pelamis Wave Power Ltd., culminating in that of Lead Machine Learning Engineer where he applied reinforcement learning techniques to optimise wave energy captured by the Pelamis. He is currently working in the field of cybersecurity as data science principal software engineer at Dell SecureWorks.

Qing Xiao received the Ph.D. degree in mechanical engineering from the National University of Singapore, Singapore, in 2001.

She is currently a Senior Lecturer of marine hydrodynamics at the University of Strathclyde, Glasgow, U.K. After experiences at the Temasek Laboratory and the Insititute of High Performance Computing in Singapore, she is currently leading the Computational Fluid Dynamics and Computational Structural Dynamics research group at the University of Strathclyde, with a particular interest in biomimetics, renewable energy devices, and offshore fluid-structure-interaction problems.

Mohammad Abusara received the B.Eng. degree from Birzeit University, Birzeit, Palestine, in 2000, and the Ph.D. degree from the University of Southampton, Southampton, U.K., in 2004, both in electrical engineering.

He is currently a Senior Lecturer in renewable energy with the University of Exeter, Penryn, U.K. He has more than ten years of industrial experience with Bowman Power Group, Southampton, in the field of research and development of digital control of power electronics. During his years in industry, he led the development of a number of commercial products that include grid- and parallel-connected inverters, microgrid, dc/dc converters for hybrid vehicles, and sensorless drives for high-speed permanent magnet machines.