Journal of
Marine Science and Engineering

MDPI

Article

Comparison of Advanced Control Strategies Applied to a Multiple-Degrees-of-Freedom Wave Energy Converter: Nonlinear Model Predictive Controller versus Reinforcement Learning

Ali S. Haider [1,*], Kush Bubbar [1] and Alan McCall [2]

[1] System-Model Development Engineering Lab, University of New Brunswick, Fredericton, NB E3B 5A3, Canada; kush.bubbar@unb.ca

[2] Dehlsen Associates, LLC, Santa Barbara, CA 93101, USA; amccall@ecomerittech.com

[*] Correspondence: ali.haider@unb.ca

Abstract: Achieving energy maximizing control of a Wave Energy Converter (WEC) not only needs a comprehensive dynamic model of the system—including nonlinear hydrodynamic effects and nonlinear characteristics of Power Take-Off (PTO)—but to treat the entire system using an integrated approach, i.e., as a cyber–physical system considering the WEC dynamics, control strategy, and communication interface. The resulting energy-maximizing optimization formulation leads to a non-quadratic and nonstandard cost function. This article compares the (1) Nonlinear Model Predictive Controller (NMPC) and (2) Reinforcement Learning (RL) techniques as applied to a class of multiple-degrees-of-freedom nonlinear WEC–PTO systems subjected to linear as well as nonlinear hydrodynamic conditions in simulation, using the WEC-Sim™ toolbox. The results show that with an optimal choice of RL agent and hyperparameters, as well as suitable training conditions, the RL algorithm is more robust under more stringent operating requirements, for which the NMPC algorithm fails to converge. Further, RL agents are computationally efficient on real-time target machines with a significantly reduced Task Execution Time (TET).

Keywords: energy maximizing control; nonlinear model predictive control; cyber–physical modeling; wave energy converter; reinforcement learning; nonlinear viscous drag; non-ideal power take-off

check for updates

Citation: Haider, A.S.; Bubbar, K.; McCall, A. Comparison of Advanced Control Strategies Applied to a Multiple-Degrees-of-Freedom Wave Energy Converter: Nonlinear Model Predictive Controller versus Reinforcement Learning. *J. Mar. Sci. Eng.* **2023**, *11*, 2120. https://doi.org/10.3390/jmse11112120

Academic Editors: Giuseppe Giorgi and Mauro Bonfanti

Received: 2 October 2023
Revised: 30 October 2023
Accepted: 1 November 2023
Published: 6 November 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Renewable energy technologies offer a feasible, sustainable, and green solution to increasing global energy needs, and the ocean offers an immense, untapped resource of energy with the potential to become an integral part of the world's energy mix [1,2]. The prospect of ocean wave energy has triggered researchers to explore techniques to maximize energy capture [3] for wave energy converters under operating conditions deviating from ideality, to include practical PTO system constraints [4] and the nonlinear hydrodynamics effects of ocean waves. Energy maximization for a WEC system is in practice a multi-objective optimization problem, requiring considerations of the physical geometry of the WEC, the PTO system design, the mooring system design, the ocean conditions of the deployment site, the communication interface, and the control methodology.

On the control front, Model Predictive Control (MPC) yields superior overall system performance for wave energy converters because it optimizes energy capture while enforcing the electro-mechanical operating limits of the system [5]. MPC is a constrained online optimal control strategy that forecasts future trajectories of the system dynamics to solve an optimization program over a receding horizon window and determine the best instantaneous control action to maximize the output power of the WEC. The MPC algorithm uses an internal model of the plant to predict the system's future states. However, WEC systems are increasingly growing in complexity [6], and there is a need for the control algorithm to handle the resulting non-ideal operating conditions. MPC algorithms suffer convergence issues under stringent non-ideal operating conditions, due to the limitations of the complex online optimization algorithm. These convergence issues become more prominent as the complexity of the optimization problem increases, due to the inclusion of multiple-DoF PTO mechanisms. The performance of the MPC algorithm is also vulnerable when incorporating nonlinearities such as viscous drag effects and nonlinear hydrodynamic forces.

Moreover, the MPC controller typically does not consider cyber-related issues, such as communication latency and packet loss between the real-time target machine that implements the controller and the WEC hardware. These factors contribute not only to the degree of optimality of the MPC solution but also to the degree of convergence of the optimal control problem. There is always an intrinsic limitation of the mathematical model of a WEC when simulating a real-world system, and if the internal plant prediction is too simple, the MPC optimization algorithm generates a poor solution under non-ideal conditions and may even become unstable.

Reinforcement Learning (RL) is a data-driven, goal-oriented, computational technique. In the RL approach, a computer interacts with a given unknown dynamic system through the RL inputs (i.e., observations and reward) and RL outputs (i.e., actions). During these interactions, the RL approach trains an agent to perform a task based on a reward from the environment [7]. Given a suitable training environment and RL agent structure, the agent can be trained for any practical environment. For a given energy-maximizing problem for a WEC, if the training environment includes effects such as non-ideal PTO behavior, communications interface latencies, and nonlinear hydrodynamic responses, then the trained RL agent learns to maximize the reward (i.e., optimization objective) in the presence of these effects. The adaptability of the RL approach to a given environment has led to an uptake in the usage of this technique for energy-maximizing problems for WEC systems. For example, an RL approach based on the Q-learning approach is presented to maximize the energy extraction in regular and irregular sea states for a point-absorber-type wave energy converter in [8], where the controller damping and stiffness are adjusted based on a reward function. Resistive control of a realistic WEC model using an RL approach based on a least-squares policy iteration is presented in [9]. A nonlinear reactive control strategy for a two-body point-absorber wave energy converter using the Q actor–critic learning method is presented for a two-body 1-DoF point absorber in [10]. A deep-RL-agent-based real-time control is presented in [11] for a 1-DoF heaving point absorber under a linear environment and is compared with a linear MPC.

This work presents the energy-maximizing control of a 2-DoF WEC array device for the digital twin of Dehlsen Associates' three-pod CENTIPOD™ device [12]. The optimizing objective is to maximize the energy harnessed by PTO machines in heave and pitch axes, subject to the electro-mechanical constraints of the system. The objective function is a nonlinear and non-quadratic function of PTO current, heave velocity, and pitch velocity, considering the practical electric machine loss characteristics of the PTOs. Moreover, the wave energy converter model includes nonlinear hydrodynamic effects due to the quadratic drag of fluid, yielding a WEC model with nonlinear dynamics. To enact the energy-maximizing control of the WEC plant, we designed two controllers: (1) a Nonlinear MPC (NMPC) and (2) an RL-agent-based controller. For NMPC design, we extended the approach in [13] to two degrees of freedom, exploiting the technique of pseudo-quadratization using the ACADO Toolkit [14]. The WEC plant is modeled in surge–heave–pitch degrees of freedom using Cummin's equation, where the radiation force convolution terms are approximated by state-space models [15]. For array devices, more thorough energy-based modeling approaches are possible, such as the port-Hamiltonian approach [16]; however, for this study, array effects and body-to-body interactions are neglected. On the RL side, we trained a Deep Deterministic Policy Gradient (DDPG) RL agent for the heave and pitch degrees of freedom. The simulation results of (1) NMPC and (2) RL are compared under the operation of the device in linear sea conditions as well as with the nonlinear hydrodynamics effects enabled in WECSim™ [17]. The WEC digital twin is simulated on an emulator machine and interfaced with the controller/training machine over EtherCAT and Universal Datagram Port (UDP) buses.

2. Developing Time-Domain Equations of the WEC

This work is related to the investigation of the power capture performance of advanced controllers for three-pod Centipod devices made by Dehlsen Associates, LLC (the multi-pod CENTIPOD) [12]. Figure 1 shows a 35th-Froude-scaled model of the WEC. However, for this work, a full-scale WECSIM [17] based digital twin of the Centipod device is considered, as shown in Figure 2; this is an array of three floating bodies (pods) that are free to heave and pitch against reaction bodies (spars) attached to a single submerged backbone structure, which is moored with three taut lines. The backbone structure is the main contributor of reaction damping to the PTO, as well as providing a stable common junction point for multiple pod–spars mechanisms. The backbone is taut-moored to the seabed, as shown in Figure 2. The pods have linear direct-drive permanent magnetic AC generator PTO machines in the heave axis and rotatory direct-drive permanent magnetic AC generator PTO machines in the pitch axes. For this study, body-to-body radiation coupling between pods is ignored; it will be evaluated in future research.

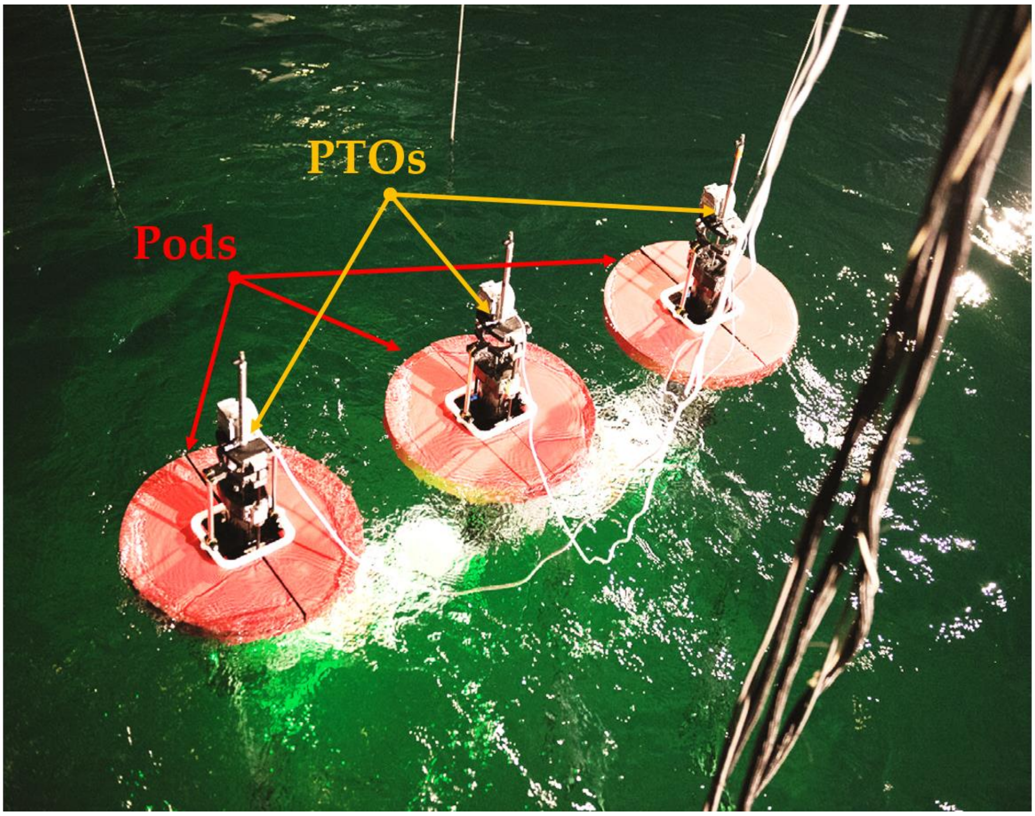
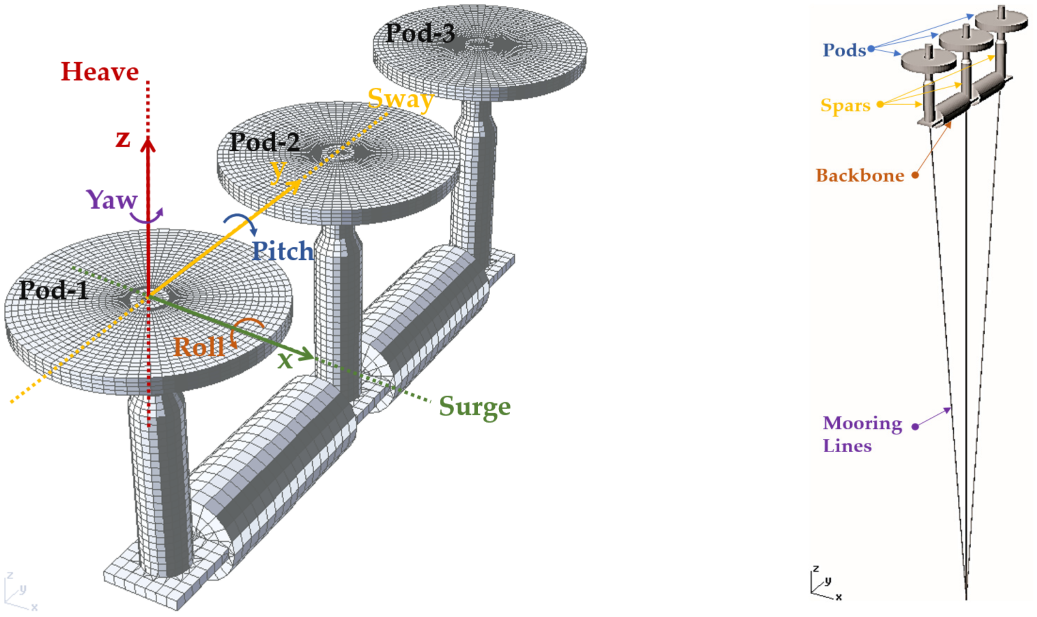


Figure 1. 35th-scale Centipod wave energy converter by Dehlsen Associates, LLC.

As per the multi-body dynamics convention for the floating pods, subscripts “1”, “3”, and “5” denote surge, heave, and pitch axes, respectively. Table 1 lists the variables and their descriptions, which are used in WEC dynamics.



(**a**) (**b**)

Figure 2. Degrees of freedom for dynamic modeling of Centipod WEC: (**a**) baseline configuration and (**b**) model with mooring lines.

Table 1. Nomenclature for WEC dynamics.

Symbol	Unit	Description
$v_i(t)$	m/s, rad/s	Generalized velocity
$x_i(t)$	m, rad	Generalized displacement
$\xi_i(t)$	–	Auxiliary state variable for radiation force dynamics
$F_{r,pq}(t)$	N, Nm	Force from wave radiation in p axis due to velocity in q axis
$F_{hs,i}(t)$	N, Nm	Buoyancy restoring force
$F_{v,i}(t)$	N, Nm	Fluid damping force
$F_{e,i}(t)$	N, Nm	Force due to wave excitation
$F_{p,i}(t)$	N, Nm	Power take-off actuation force
m	Kg	Pod mass
$A_{pq}(\infty)$	Kg, Kg m, Kg m^2	Infinite frequency generalized radiation added mass in p axis due to acceleration in q axis.
C_i	N/m, Nm/rad	Buoyancy restoring constant
$C_{vd,i}$	N/(m/s)^2, Nm/(rad/s)^2	Fluid quadratic damping constant
$A_{qp}(\omega)$	Kg, Kg m, Kg m^2	Added mass due to wave radiation in p axis due to acceleration in q axis
$B_{qp}(\omega)$	N/m/s, Nm/rad/s	Radiation damping due to wave radiation in p axis due to velocity in q axis
$K_{pq}(t)$	N/m, Nm/rad	Impulse response function for wave radiation
$Z_{qp}(\omega)$	N/m/s, Nm/rad/s	Mechanical impedance
g	m/s^2	Gravity constant
ρ	Kg/m^3	Water density

2.1. Dynamic Model of WEC in Surge, Heave, and Pitch Axes

The orientation of the Centipod device in Figure 2 with respect to incoming waves in the surge direction results in negligible roll, sway, and yaw displacements of the pods; hence, it is adequate to consider the surge–pitch–heave model of each pod for energy capture considerations. The floating pods in Figure 2 are modeled as point-absorber bodies. Heave motion is very weakly coupled to surge and pitch; hence, this coupling effect can be ignored. In the local frame of reference, the Cummins equations for the three axes of freedom (surge–pitch–heave) are

$$M_{11}\dot{v}_1 + A_{15}(\infty)\dot{v}_5 = -F_{r,11}(t) - F_{r,15}(t) - F_{v,1}(t) + F_{e,1}(t), \tag{1}$$

$$M_{33}\dot{v}_3(t) = -F_{r,33}(t) - F_{hs,3}(t) - F_{v,3}(t) - F_{p,3}(t) + F_{e,3}(t), \tag{2}$$

$$M_{55}\dot{v}_5 + A_{51}(\infty)\dot{v}_1 = -F_{r,55}(t) - F_{r,51}(t) - F_{v,5}(t) - F_{hs,5}(t) - F_{p,5}(t) + F_{e,5}(t) \tag{3}$$

Here, $M_{ii} = (m + A_{ii}(\infty))$. The radiation force, buoyancy restoring force, and fluid quadratic damping terms in (1) through (3), respectively, are given by

$$F_{r,ij}(t) = \int_{-\infty}^{t} K_{ij}(t-\tau)v_j d\tau, \tag{4}$$

$$F_{hs,i}(t) = C_i x_i, \tag{5}$$

$$F_{v,i}(t) = C_{d,i} v_i |v_i|. \tag{6}$$

The time-domain convolution integral term in (4) can be transformed into the frequency-domain expression $Z_{pq}(j\omega)V_q(j\omega)$ through the application of the Fourier transform. The frequency-domain hydrodynamic parameters for the Centipod hull geometry without the mooring system are determined using the WAMIT™ (Version 7.201-x64) software package [18]. A single-body WEC intrinsic impedance $Z_{pq}(j\omega)$ [19] is calculated using these hydrodynamic parameters plotted in Figures 3–5. A minimal-order transfer function in Laplace space is approximated for $Z_{pq}(j\omega)$ using system identification techniques, and an equivalent state-space representation can be formulated as in [15,20]. After performing algebraic manipulations, the final state-space model of the plant developed in [20] is given by

$$\dot{\mathrm{X}} = \mathrm{AX} + \mathrm{B_p F_p} + \mathrm{B_v F_v} + \mathrm{B_e F_e}, \tag{7}$$

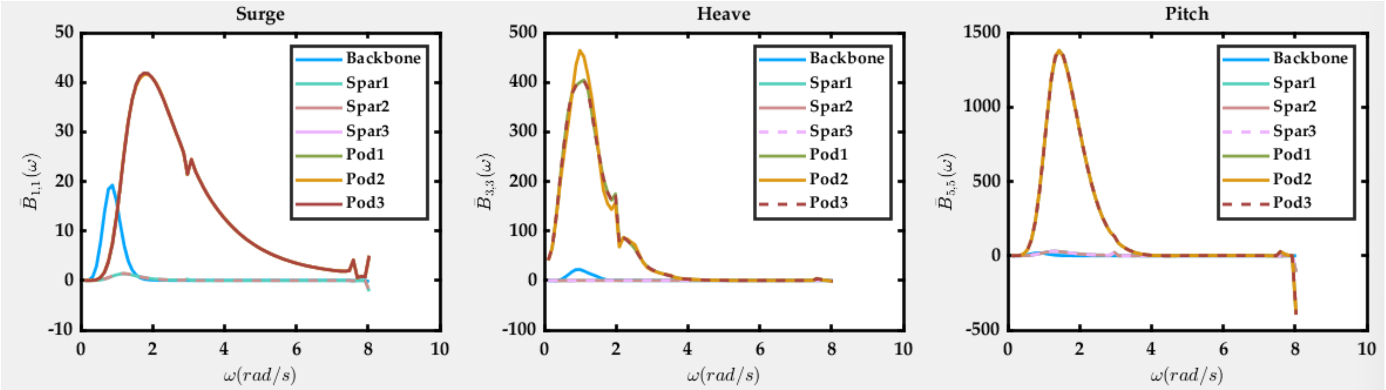
where X is the state vector and

$$\mathrm{F_p} = \begin{bmatrix} F_{p,5} & F_{p,3} \end{bmatrix}^T \tag{8}$$

$$\mathrm{F_v} = \begin{bmatrix} F_{v,1} & F_{v,5} & F_{v,3} \end{bmatrix}^T \tag{9}$$

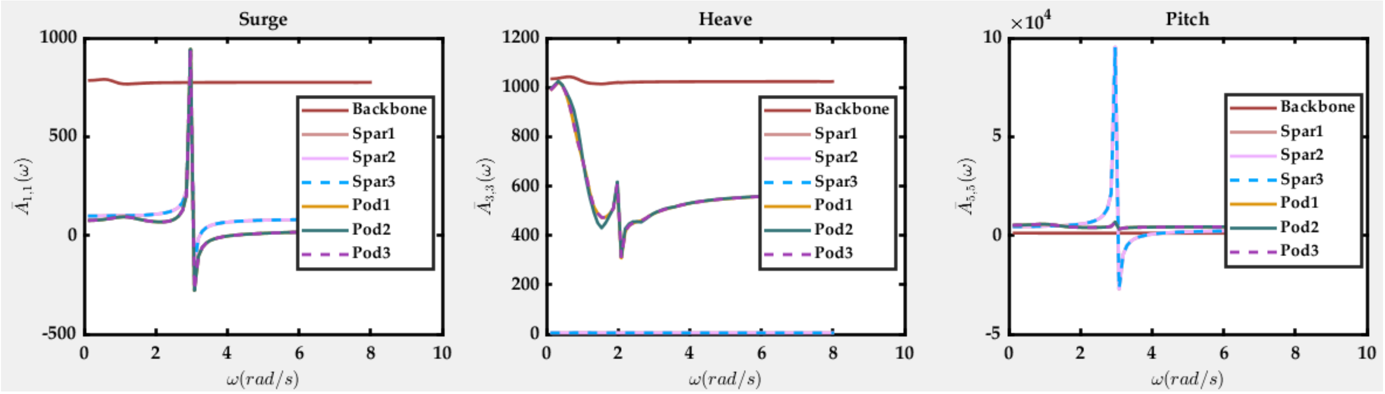
$$\mathrm{F_e} = \begin{bmatrix} F_{e,1} & F_{e,5} & F_{e,3} \end{bmatrix}^T \tag{10}$$

The state matrix and input matrices in (7) are given by (11) and (12) with appropriate systems parameter constants m_{ij}, a_i, and b_i.



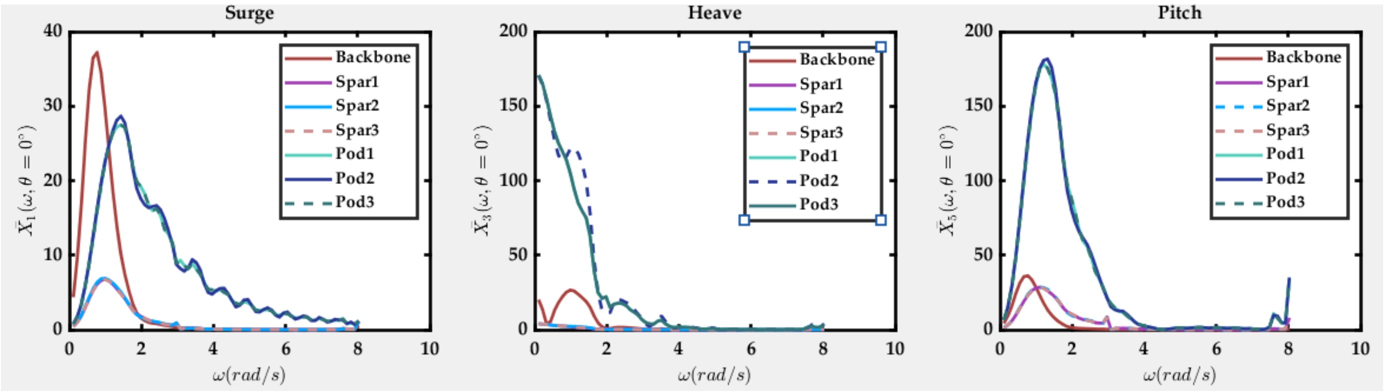
(**a**) (**b**) (**c**)

Figure 3. Normalized radiation damping $B/\omega\rho$ of the Centipod WEC: (**a**) surge axis, (**b**) heave axis, and (**c**) pitch axis.



(**a**) (**b**) (**c**)

Figure 4. Normalized added mass A/ρ of the Centipod WEC: (**a**) surge axis, (**b**) heave axis, and (**c**) pitch axis.



(**a**) (**b**) (**c**)

Figure 5. Normalized excitation amplitude $X/\rho g$ of the Centipod WEC: (**a**) surge axis, (**b**) heave axis, and (**c**) pitch axis.

2.2. Non-Ideal Power Take-Off Model

The power take-off machine for the heave axis is a Linear Universal Modular Actuator/Absorber (LUMA) machine [21]. For the pitch-axis, PTO comprises a direct-drive permanent magnet AC generator. The non-ideal power take-off model is taken from the case study in [20], where the PTO power capture is a function of PTO force and velocity with system parameter constants c_i given by (13).

$$\mathrm{A} = \begin{bmatrix} \begin{matrix} 0 & 0 & -m_{15}C_5 & -m_{11} & 0 & -m_{11} & 0 & -m_{15} & 0 & -m_{15} & 0 \\ 0 & 0 & -m_{55}C_5 & -m_{51} & 0 & -m_{51} & 0 & -m_{55} & 0 & -m_{55} & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_3 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_4 & 0 & 0 & a_3 & a_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_5 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & b_6 & 0 & 0 & 0 & a_5 & a_6 & 0 & 0 & 0 & 0 \\ 0 & b_7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & b_8 & 0 & 0 & 0 & 0 & 0 & a_7 & a_8 & 0 & 0 \\ b_9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ b_{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_9 & a_{10} \end{matrix} & 0_{11\times 4} \\ 0_{4\times 11} & \begin{matrix} 0 & \frac{-C_3}{M_{33}} & \frac{-1}{M_{33}} & 0 \\ 1 & 0 & 0 & 0 \\ b_1 & 0 & 0 & 1 \\ b_2 & 0 & a_1 & a_2 \end{matrix} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 1.4 & -3.8e^{-6} & 0 & -3.8e^{-6} & 0 & 6.7e^{-8} & 0 & 6.7e^{-8} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2.8 & 6.7e^{-8} & 0 & 6.7e^{-8} & 0 & -1.3e^{-7} & 0 & -1.3e^{-7} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.2e^{5} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -4.7e^{5} & 0 & 0 & -5.3 & -2.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5.4e^{5} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -7e^{5} & 0 & 0 & 0 & -3.3 & -1.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.4e^{6} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2.4e^{6} & 0 & 0 & 0 & 0 & 0 & -2.6 & -1.1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5.4e^{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -6.9e^{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3.3 & -1.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2.1 & -1.2e^{-6} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6.3e^{5} & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -8.9e^{5} & 0 & -1.5 & -1.5 \end{bmatrix} \tag{11}$$

$$\mathrm{B_p} = \begin{bmatrix} -m_{15} & 0 \\ -m_{55} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \frac{-1}{M_{33}} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 6.7e^{-8} & 0 \\ -1.3e^{-7} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1.2e^{-6} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathrm{B_e} = -\mathrm{B_v} = \begin{bmatrix} m_{11} & m_{15} & 0 \\ m_{51} & m_{55} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{M_{33}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 3.8e^{-6} & -6.7e^{-8} & 0 \\ -6.8e^{-8} & 1.3e^{-7} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1.2e^{-6} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{12}$$

$$P_{E,i} = c_{0,i}F_{p,i}v_i - \left(c_{1,i}F_{p,i}^{6} + c_{2,i}F_{p,i}^{5} + c_{3,i}F_{p,i}^{4} + c_{4,i}F_{p,i}^{3} + c_{5,i}F_{p,i}^{2} + c_{6,i}F_{p,i} + c_{7,i} \right), \tag{13}$$

3. Nonlinear MPC Design for WEC

A given NMPC problem optimizes a manipulated variable $\mathrm{u} \subseteq \mathrm{w}$ to maximize some cost function P of a set of system variables "w" while respecting the given system constraints. A general class of NMPC problems has been formulated in [13], in which the cost function takes on a nonlinear piecewise polynomial form. Considering the case of finite-horizon optimization, we can mathematically describe the NMPC problem of such a class as (Tables 2 and 3)

$$\underset{\mathrm{u}}{max}\mathrm{P}(\mathrm{w}) = \begin{cases} \mathrm{P}_1(\mathrm{w}) + \rho_{N,1}(\mathrm{w}), & w_k < R_1 \\ \mathrm{P}_2(\mathrm{w}) + \rho_{N,2}(\mathrm{w}), & R_1 \leq w_k \leq R_2 \\ \vdots & \vdots \\ \mathrm{P}_\mathrm{j}(\mathrm{w}) + \rho_{N,j}(\mathrm{w}), & R_{j-1} \leq w_k \leq R_j \end{cases}, \tag{14}$$

Table 2. Nomenclature for nonlinear MPC formulation.

Symbol	Description
w	Set of system variables
N	Prediction horizon
$\mathrm{X} \subseteq \mathrm{w}$	State vector of WEC dynamics
$\mathrm{u} \subseteq \mathrm{w}$	Manipulated variable vector, PTO force/torque $\mathrm{F_p}(N)$
$\rho_{N,i}$	Finite-horizon terminal cost penalty
P_i	Polynomial of system variables
Ψ_i	Constant weighting matrices
$\mathrm{B_i}$	Constant column vectors
Υ_i	Column vectors of nonlinear functions of state variables
q	Column vectors of nonlinear functions of state variables
d	Excitation force disturbance vector, $\mathrm{F_e}(N)$
R_i	Some real number

Table 3. WEC system parameters.

Parameter	Value
m	2.36×10^5 Kg
$A_{11}(\infty)$	2.88×10^4 Kg
$A_{33}(\infty)$	5.71×10^5 Kg
$A_{55}(\infty)$	4.4×10^6 Kg m^2
$A_{15}(\infty)$	1.33×10^5 Kg m
$A_{51}(\infty)$	1.33×10^5 Kg m
C_3	1.69×10^6 N/m
C_5	2.12×10^7 Nm/rad
$C_{d,1}$	1.48×10^5 N/(m/s)^2
$C_{d,3}$	1.73×10^5 N/(m/s)^2
$C_{d,5}$	1.29×10^7 Nm/(rad/s)^2
M_{11}	2.65×10^5 Kg
M_{33}	8.07×10^5 Kg
M_{55}	7.53×10^6 Kg m^2
Water depth	212 m
Pod volume	400 m^3
Pod immersed volume	359 m^3
Mooring line length	192.32 m
Mooring line type	Chain
Mooring no of lines	3
Mooring line diameter	0.175 m
Mooring mass density in air	18.375 kg/m
Mooring damping ratio	0.8
Mooring stiffness	1.11×10^4 MN
Mooring transverse drag coefficient	1.6
Mooring transverse added mass coefficient	1
Mooring tangential drag coefficient	0.05
Mooring tangential added mass coefficient	0

4. RL Agent Design for 2-DoF Heave–Pitch PTOs

Similar to Section 3, we propose a method to optimize the overall electrical power captured by the 2-DoF PTO, in this case through designing an appropriate RL agent for our problem. For continuous action and observation spaces, the typical options for the candidate agents are

1. Deep Deterministic Policy Gradient (DDPG).
2. Twin-Delayed Deep Deterministic policy gradient (TD3).
3. Proximal Policy Optimization (PPO).
4. Soft Actor-Critic (SAC).

Regarding increasing complexity, the DDPG is the simplest compatible agent, followed by TD3, PPO, and SAC. TD3 is an improved, more complex version of DDPG, while PPO has more stable updates but requires more training [22]. On the other hand, SAC is an improved and more complex version of DDPG that generates stochastic policies. We utilize the DDPG for our problem, typically the first choice for problems with continuous action and observation spaces [7].

4.1. RL DDPG Agent Reward Function and Properties

DDPG-based control aims to maximize the PTO power capture while respecting the PTO velocity limits. The observation consists of the pod velocity, and the action is the PTO force. Training is performed offline for this study. To specify a reward function to train our RL DDPG, we propose using a modified version of (14) below, which includes a penalty term for the agent for exceeding the velocity limits of the PTO mechanisms:

$$Reward_i = \mathrm{k}_p \frac{1}{2} \mathbf{h}_\mathbf{i}^\mathbf{T} (2\mathbf{W}_\mathbf{i}) \mathbf{h}_\mathbf{i} - \mathrm{k}_{v_i}(|v_i| > v_{i,max}) \tag{15}$$

Here, k_p and k_{v_i} are some appropriate scaling factors. We have designed two separate RL agents for the pitch and heave control because these DoFs are decoupled. The DDPG agent options for both DoFs are given in Table 4.

Table 4. RL DDPG agent properties for heave and pitch control.

RL Agent Options	Value
Sample time	0.1
Target smooth factor	1×10^{-6}
Discount factor	0.95
Mini batch size	512
Length of experience buffer	1×10^{6}
Noise variance	0.3
Variance decay rate	1×10^{-5}
Target update frequency	1

4.2. Design of Actor and Critic Deep Networks for RL Training

The DDPG-based RL algorithm requires the critic and actor neural networks to implement the optimal policy by generating actions in response to the given observations. A critic neural network predicts the discounted value of the cumulative long-term reward by looking at the observations and actions, and an agent neural network implements the RL policy to produce actions to maximize the predicted discounted cumulative long-term reward [22]. An experience-based design choice of the deep network structures for RL actor and critic is shown in Figure 6. Based on extensive training trials, an RL-Q-value representation for the critic network is finally selected for both heave and pitch, and the other hyperparameter choice for the critic network is given in Table 5.

An RL-deterministic representation is chosen for the actor network for both heave and pitch, based on extensive training trials; the other hyperparameters chosen for the actor network are provided in Table 6.

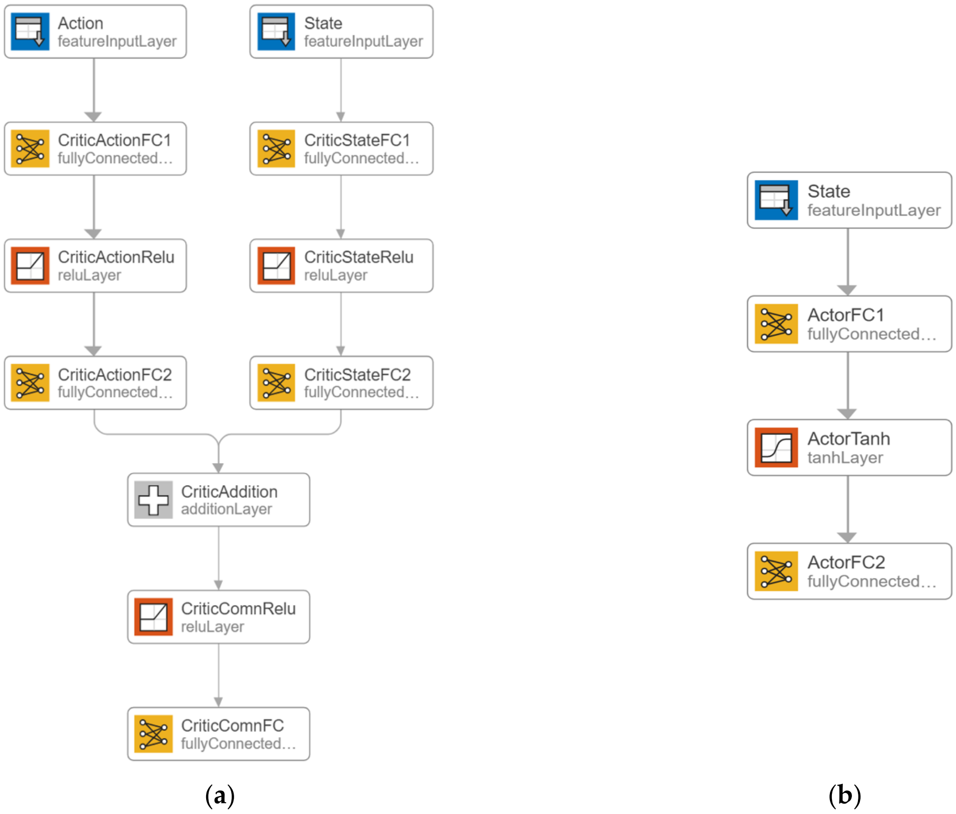


Figure 6. Deep network design for RL actor and critic in MATLAB for the heave and pitch PTOs: (**a**) deep network for the critic and (**b**) deep network for the actor.

Table 5. RL critic properties for heave and pitch control.

RL Critic Options	Value
Representation	RL Q-value
Learn rate	0.1
Gradient threshold	inf
Action feature input layer size	1
State feature input layer size	1
Action and critic fully connected layer-1 size (for unconstrained F_{pto})	64
Action and critic fully connected layer-2 size (for constrained F_{pto})	50
Critic common fully connected layer-2 (FC1) size	1

Table 6. RL actor properties for heave and pitch control.

RL Critic Options	Value
Representation	RL deterministic
Learn rate	0.1
Gradient threshold	inf
Optimizer momentum	0.95
State feature input layer size	1
Actor fully connected layer-1 size (unconstrained F_p)	32 (heave), 16 (pitch)
Actor fully connected layer-1 size (constrained F_p)	25 (heave), 16 (pitch)
Actor fully connected layer-2 size	1
Optimizer for pitch (constrained or unconstrained F_p)	Root mean square propagation (RMS-Prop)
Optimizer for heave (for constrained F_p)	Root mean square propagation (RMS-Prop)
Optimizer for heave (for unconstrained F_p)	Stochastic gradient descent with momentum (SGDM)

4.3. RL Agent Training

To train the RL agent, we generate training data by simulating a full-scale version of the 2-DoF version of Dehlsen's three-pod CENTIPOD WEC of Figure 1 in an emulator machine, by using its WEC-Sim model, as shown in Figure 7. The mean of a cluster of sea states used to execute the WEC-Sim simulation are described in Table 7, which were selected based on the geographic location and wave resources at the deployment site at PacWave [23,24]; the corresponding wave spectrum is shown in Figure 8.

Table 7. Mean of the cluster of sea states for Centipod digital twin simulation in WECSim.

Sea-State Parameter in WECSim	Value
Significant wave height [m]	2.5
Peak wave period [s]	7.35
Spectrum type of ocean waves	Pierson Moskowitz (PM)
Class of waves	Irregular

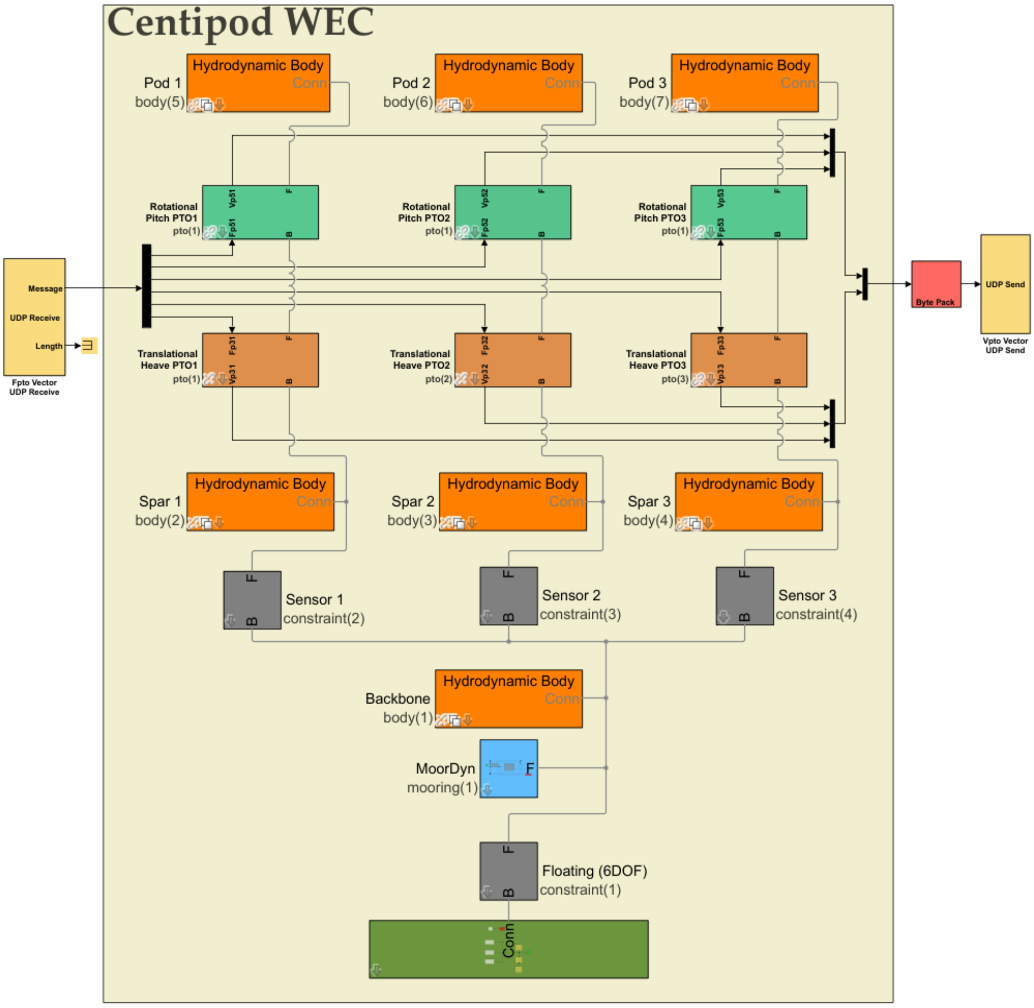


Figure 7. WECSim digital twin of 3-pod Centipod WEC with PTOs in heave and pitch axes.

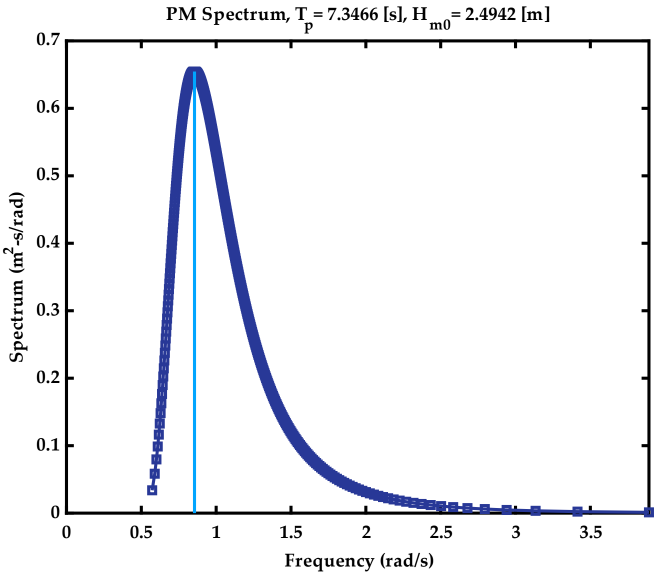


Figure 8. Wave spectrum of the mean of a cluster of sea states for WEC-Sim simulation.

The strategy to train the RL agent is shown schematically in Figure 9. An emulator machine simulates the real-time digital twin of the plant in Simulink/WEC-Sim., which is connected to another real-time controller target machine via an Ethernet/Universal Datagram Packet (UDP) link which runs the RL training algorithm in MATLAB/Simulink. A MATLAB script to train the RL algorithm in the controller machine establishes the connection between the two machines for each training episode and trains the agent for the environment marked in Figure 9.

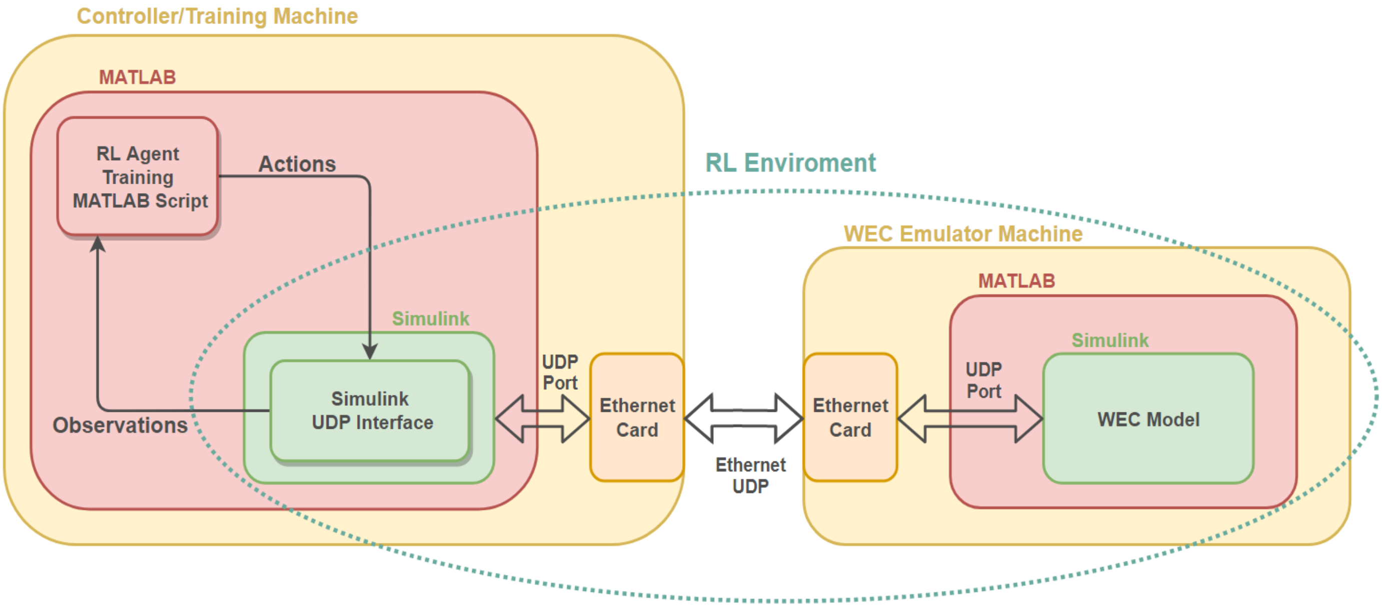


Figure 9. RL agent training through custom Simulink environment interfaced to WEC Emulator via Ethernet UDP.

The Simulink model used to implement the RL agent is shown in Figure 10. A separate RL agent is implemented for heave and pitch DoFs for each of the three pods of Centipod WEC in Figure 2. For this study, body-to-body interactions and array effects are neglected, and all pods are assumed to be identical; therefore, the DDPG RL agent is trained for the heave and pitch axes for one pod, and the trained policy is deployed for each PTO in Figure 10. The training policy is implemented in a Speedgoat Performance real-time target machine (Intel Core 3.1 GHz, 4-core, 8 GB). The velocity data (observations) of each pod are collected over the UDP link, parsed, and observed by respective policies to generate control actions, which are packed and transmitted to the WEC emulator machine over the UDP channel, as shown in Figure 10.

Two agents are trained for each DoF, one for the unconstrained PTO force and the other with a 40 kN upper bound on the PTO force magnitude. The three pods in Figure 10 are identical, so agents are trained for a single pod and the resulting agents are duplicated for the other two pods. The training stats for the heave agents for Pod 1 with constrained and unconstrained PTO forces are shown in Figures 11a and 11b, respectively. In either case, RL training converges in about 40 episodes with a 100 s simulation time for each episode. The training stats for the pitch agents for Pod 1 with constrained and unconstrained PTO forces are shown in Figures 12a and 12b, respectively. The constrained force pitch agent converges in 25 episodes, and the uncontained force pitch agent converges in 60 episodes.

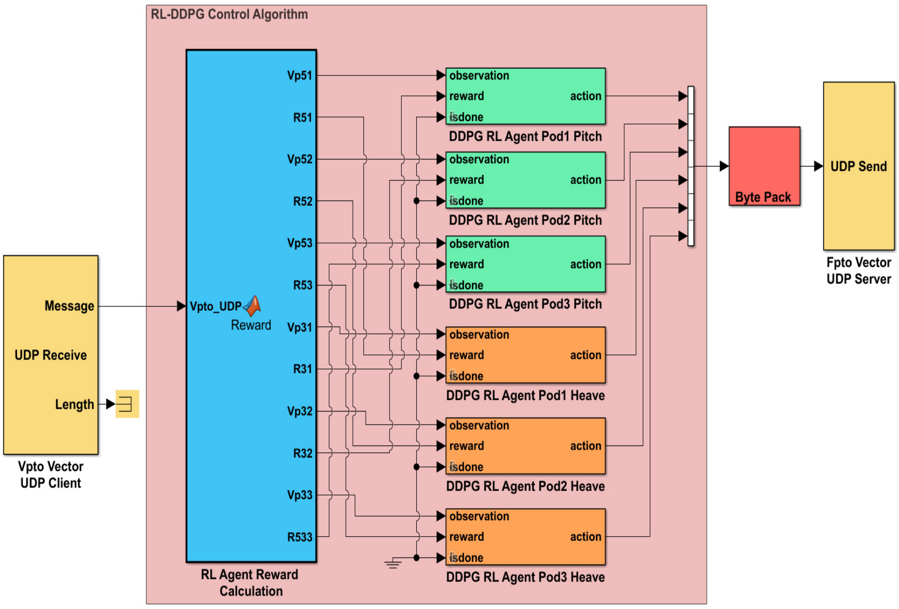
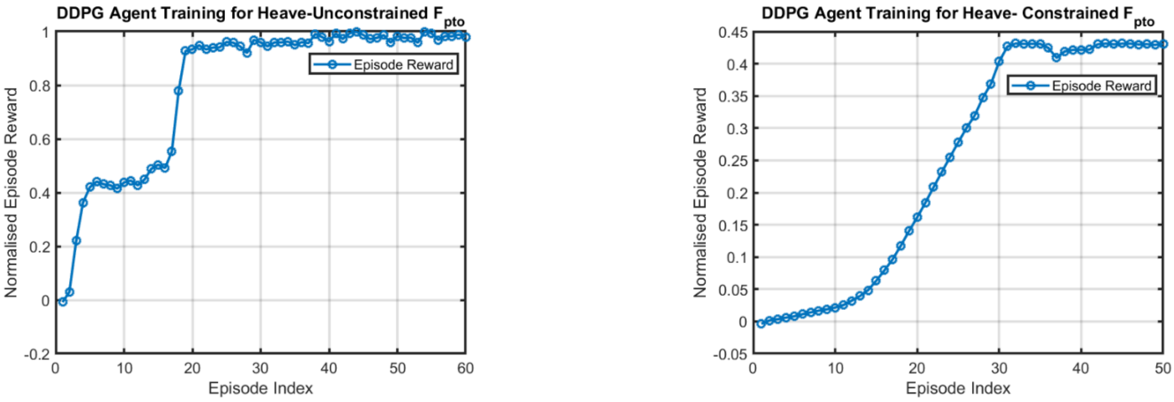


Figure 10. RL-DDPG controller for 2-DoF 3-pod CENTIPOD WEC.



(**a**) (**b**)

Figure 11. RL DDPG agent training for heave PTO: (**a**) training with unconstrained PTO force and (**b**) training with constrained PTO force, $|F_p| \leq 40$ kN.

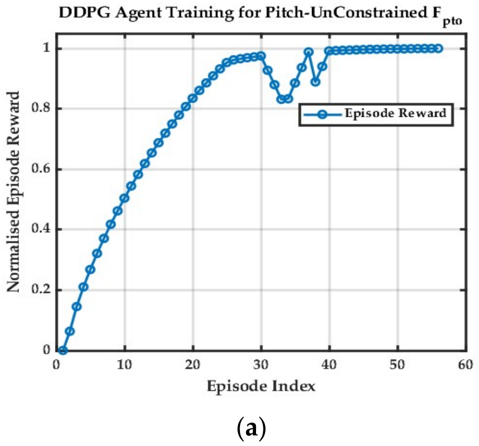


Figure 12. RL DDPG agent training for pitch PTO: (**a**) training with unconstrained PTO force and (**b**) training with constrained PTO force, $|F_p| \leq 40$ kN.

5. Results

The training setup shown in Figure 9 is also used for executing system simulations with the trained agents. When executing simulations, the agents in Figure 10 are replaced with the trained agents, and the training machine in Figure 9 plays the role of a controller machine. A controller model similar to the RL model in Figure 9 is developed to simulate the NMPC by extending the scheme in [13] to the NMPC designed in Section 4, as shown in Figure 13. The RL and NMPC controllers in Figures 10 and 13, respectively, are tested with the WEC-Sim model of Figure 10 running on the emulator machine in Figure 9, with the same sea-state parameters given in Table 7. Tests are run with constrained and unconstrained PTO force conditions. Each case is simulated with linear wave conditions as well as with nonlinear buoyancy and Froude–Krylov excitations enabled in WECSim. The mechanical velocity constraints of the power take-off machines are enforced as follows: |heave velocity| ≤ 2 m/s and |pitch velocity| ≤ 0.5 rad/s.

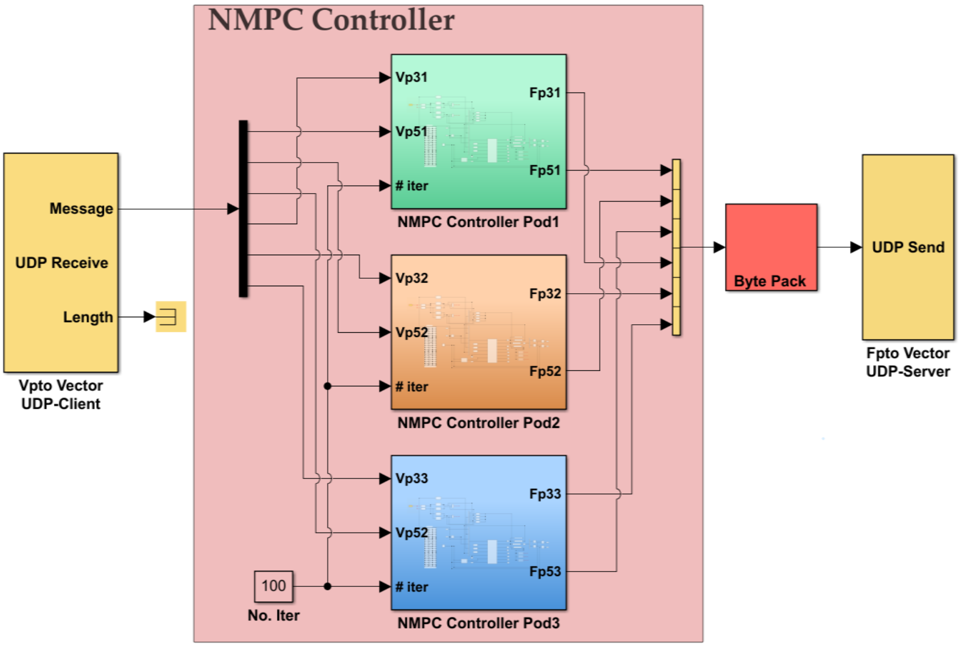
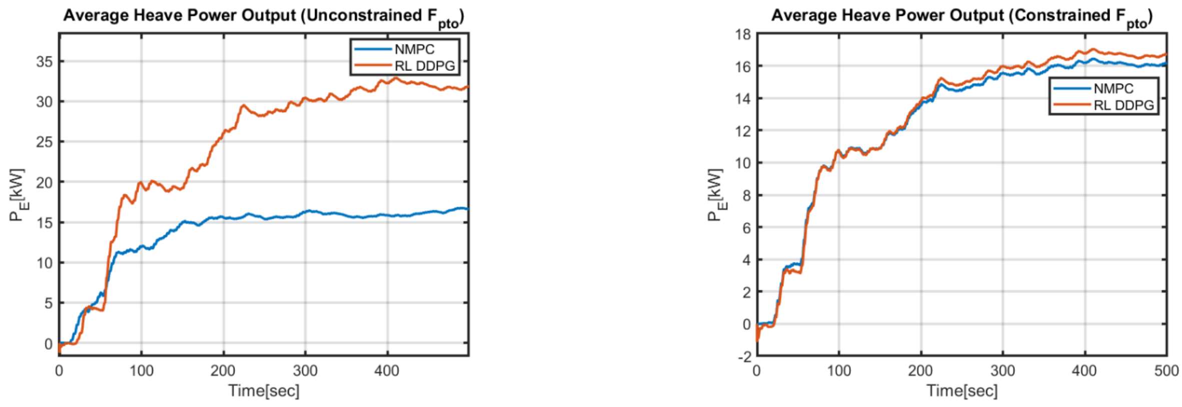


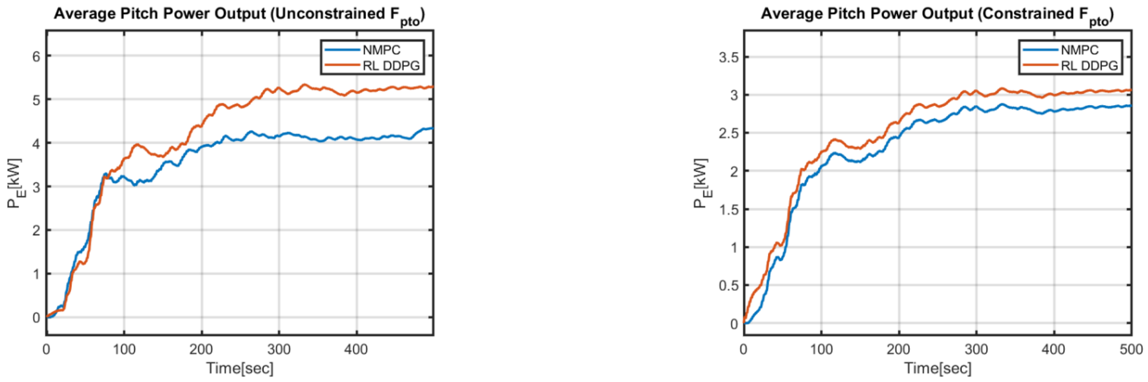
Figure 13. Nonlinear MPC controller for digital twin of three-pod Centipod device.

The moving averaged electrical power outputs with nonlinear MPC and RL are plotted in heave and pitch for Pod 1 with unconstrained and constrained PTO force cases subject to linear wave conditions, as shown in Figures 14 and 15, respectively. It is also important to evaluate the performance of the two algorithms under nonlinear buoyancy and Froude–Krylov wave excitations in WEC-Sim. The average electrical heave and pitch power outputs for Pod 1 with unconstrained and constrained PTO force cases under nonlinear hydrodynamic conditions are shown in Figures 16 and 17, respectively.



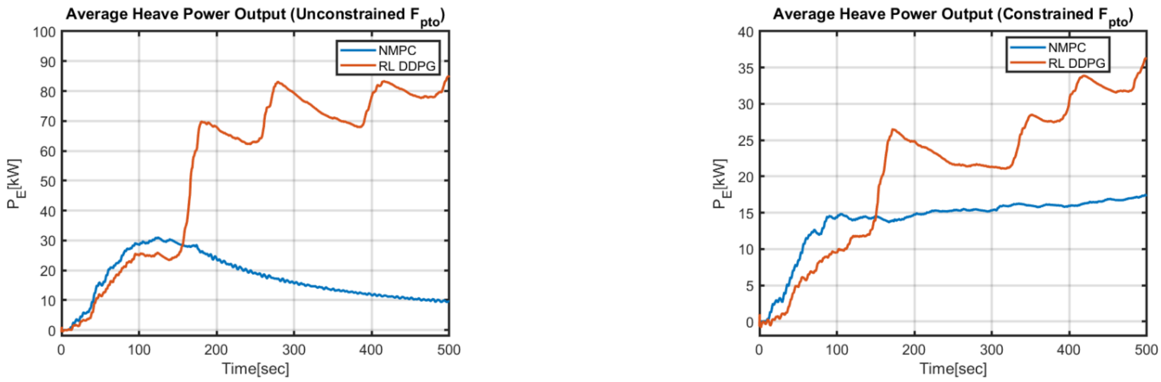
(**a**) (**b**)

Figure 14. Heave average electrical power output per pod with linear waves enabled in WECSim: (**a**) with unconstrained PTO force and (**b**) with constrained PTO force, $|F_p| \leq 40$ kN.



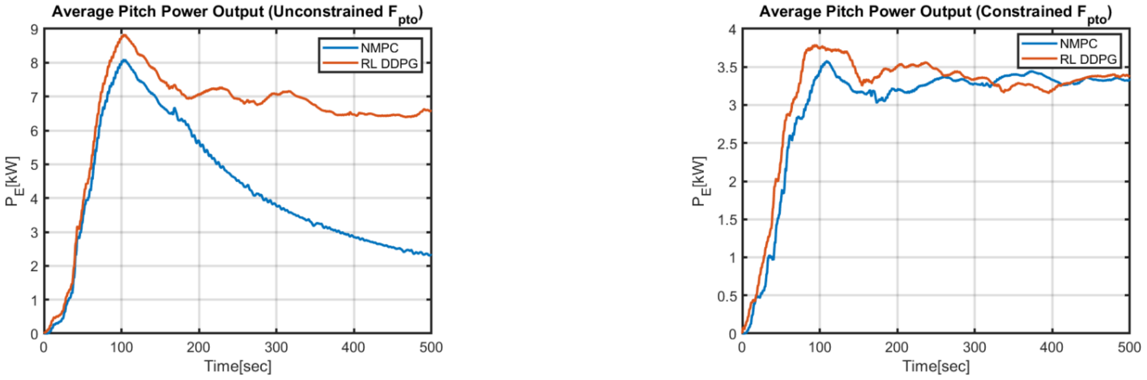
(**a**) (**b**)

Figure 15. Pitch average electrical power output per pod with linear waves enabled in WECSim: (**a**) with unconstrained PTO force and (**b**) with constrained PTO force, $|F_{pto}| \leq 40$ kN.



(**a**) (**b**)

Figure 16. Heave average electrical power output per pod with nonlinear buoyancy and Froude–Krylov excitations enabled in WECSim: (**a**) with unconstrained PTO force and (**b**) with constrained PTO force, $|F_{pto}| \leq 40$ kN.



(**a**) (**b**)

Figure 17. Pitch average electrical power output per pod with nonlinear buoyancy and Froude–Krylov excitations enabled in WECSim: (**a**) with unconstrained PTO force and (**b**) with constrained PTO force, $|F_{pto}| \leq 40$ kN.

The results for the PTO power capture performance in heave and pitch from Figure 14 through Figure 17 are summarized in Table 8. The summary of the computational performance statistics for the NMPC and RL-DDGP control algorithms is given in Table 9, where the average Task Execution Time (TET) for each algorithm and RL agent training times for the PTO machines in heave and pitch axis are listed.

Table 8. Moving mean of electrical output power [kW] per PTO in WEC-Sim with irregular waves.

Controller Algorithm	**Linear Wave Conditions**		**Nonlinear Buoyancy and Froude–Krylov Excitation**	
	F$_{pto}$ Unconstrained	**$\lvert F_{pto}\rvert \leq$40 kN**	**F$_{pto}$ Unconstrained**	**$\lvert F_{pto}\rvert \leq$40 kN**
NMPC	17	16	Unstable	17
L-DDPG	32	17	85	37
	Average electrical power [kW] for pitch			
NMPC	4.50	2.80	Unstable	3.25
RL-DDPG	5.25	3.10	6.50	3.30

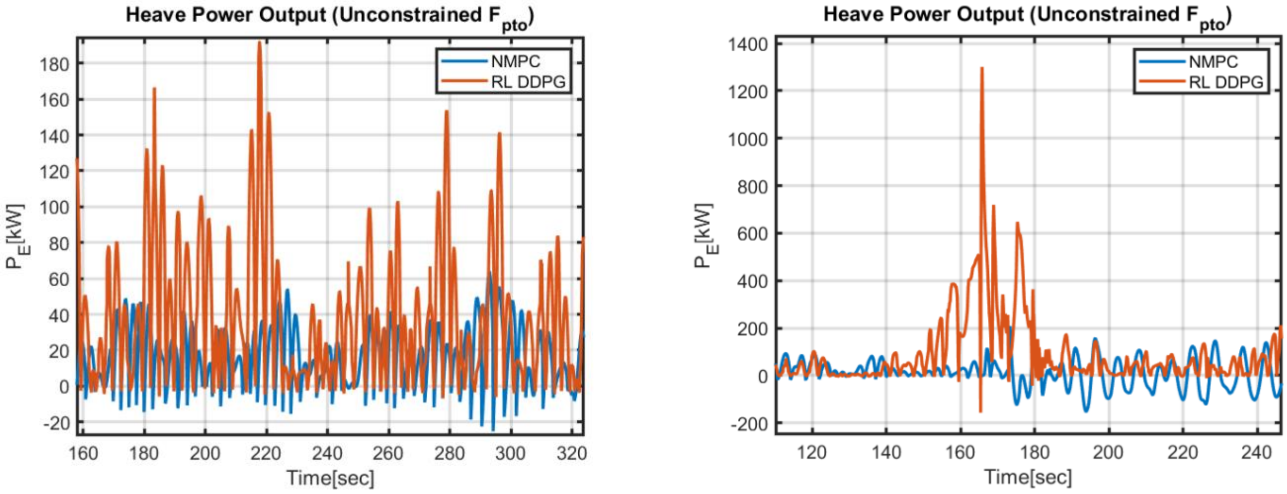
Table 9. Timings stats for NMPC and RL DDPG control.

	Training Time [h]	**Task Execution Time (TET) [s]**
NMPC heave and pitch combined	-	7.92×10^{-3}
NMPC per DoF	-	3.96×10^{-3}
RL-DDPG heave	1.12	4.32×10^{-4}
RL-DDPG pitch	1.67	7.52×10^{-4}

6. Discussion

The observations of the moving mean of electrical output power from the PTO mechanisms in Figures 14–17 reveal an improvement in the power output in the case of the RL-DPPG agent compared to the NMPC. This observation may be attributed to the fact that the operation of the NMPC is based upon the prediction model of the WEC plant, which, from the definition of the NMPC method, is an approximate representation of the actual process. On the other hand, the RL-DPPG agent was trained on the actual process in Figure 9 and observed the full process dynamics to determine how to act accordingly. In the constrained linear hydrodynamic cases in Figures 14b and 15b, the performance of the RL-DPPG very closely resembles the performance of NMPC. However, in the cases with unconstrained PTO forces (Figures 14a and 15a), nonlinear effects in the process dynamics become prominent as large PTO force magnitudes emerge and the NMPC performance degrades. This degradation can be attributed to the NMPC algorithm being susceptible to unmodelled or poorly modeled nonlinearities. The plot of the instantaneous electrical power output for the heave DoF is shown in Figure 18a, which corresponds to the average power output plot in Figure 14a. A significantly improved RL-DPPG control strategy operation is evident compared to the NMPC.

The performance degradation of the NMPC is attributed to the unmodeled process nonlinearities and becomes fully visible when observing the system's operation under nonlinear hydrodynamic wave conditions in WEC-Sim, as detailed in Figures 19 and 17a. The NMPC performs poorly in these figures, and the controller becomes unstable. The plot of instantaneous electrical power output for heave is shown in Figure 18b, corresponding to the average power output plot in Figure 19 under unconstrained PTO force conditions. It can be observed in Figure 18b that the NMPC algorithm fails to converge after around 170 s because it is unable to respond appropriately to nonlinear wave hydrodynamic conditions. On the other hand, RL-DPPG remains stable under the same nonlinear wave conditions. The heave PTO force plots corresponding to Figure 18b for NMPC and RL-DPPG are shown in Figure 19.



(**a**) (**b**)

Figure 18. Heave instantaneous electrical power output per pod with unconstrained PTO force: (**a**) with linear wave conditions in WECSim and (**b**) with nonlinear buoyancy and Froude−Krylov excitations enabled in WECSim.

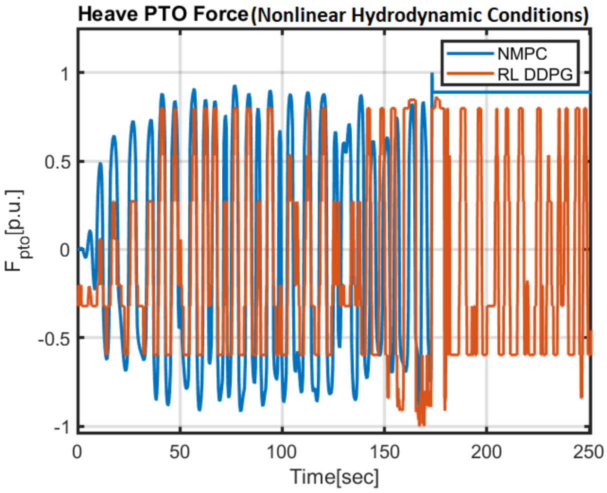


Figure 19. Heave PTO force output of NMPC and RL controllers with nonlinear buoyancy and Froude−Krylov excitations enabled in WECSim.

In Figure 19 the unstable output of the NMPC can be observed. The performance of the RL-DPPG remains stable and robust against process uncertainties in the same conditions. This might also be attributed to the fact that no underlying online optimization problem is being solved at every time step in the RL-DPPG policy deployment stage as opposed to the case of the NMPC, which also explains the significantly reduced TET in Table 9 for the RL-DDPG agent compared to the NMPC.

7. Conclusions

This article presents a comparison of two strategies, (a) NMPC and (b) RL-DPPG, for controlling the power-capture dynamics of the nonlinear WEC device by Dehlsen Associates' (the three-pod CENTIPOD WEC), with a PTO operating simultaneously in the heave axis and pitch axis. A state-space model of the WEC plant is formulated, including nonlinear quadratic viscous drag, and we consider a case study PTO model with a non-quadratic cost function. Two controllers are designed to optimize power capture from the PTO, (a) NMPC and (b) RL-DPPG, by training agents for the PTO machines in the heave and pitch axis. Both control algorithms are tested against the same simulated WEC model in WEC-Sim running on an external emulator machine. The heave and pitch PTO power output results are obtained for the linear wave conditions as well as with nonlinear buoyancy, and Froude–Krylov excitations are enabled in WECSim for cases where the PTO force is constrained or unconstrained. Results depict a significant enhancement in the performance of the proposed RL-DDPG algorithm when compared to the NMPC controller, based on various performance metrics, including a reduction in the Task Execution Time (TET), an increase in the power extraction, an improvement in the robust operation when subject to exogenous conditions, and more overall flexibility and ease of design.

Author Contributions: Conceptualization, A.S.H. and A.M.; methodology, A.S.H., K.B. and A.M.; software, A.M. and K.B.; validation, A.S.H., A.M. and K.B.; formal analysis, A.S.H. and K.B.; investigation, A.S.H., A.M. and K.B.; resources, A.M. and K.B.; data curation, A.M.; writing—original draft preparation, A.S.H. and K.B.; writing—review and editing, A.S.H., A.M. and K.B.; visualization, A.S.H., K.B. and A.M.; supervision, K.B. and A.M.; project administration, K.B. and A.M.; funding acquisition, K.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the McCain Foundation postdoctoral fellowship award in innovation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors appreciate technical support from Dehlsen Associates, LLC, California, and the provision of experimental data for the PTO machines from McCleer Power, Michigan.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Muetze, A.; Vining, J.G. Ocean Wave Energy Conversion—A Survey. In Proceedings of the Conference Record of the 2006 IEEE Industry Applications Conference Forty-First IAS Annual Meeting, Tampa, FL, USA, 8–12 October 2006; Volume 3, pp. 1410–1417.
2. Richter, M.; Magana, M.E.; Sawodny, O.; Brekken, T.K.A. Nonlinear Model Predictive Control of a Point Absorber Wave Energy Converter. *IEEE Trans. Sustain. Energy* **2013**, *4*, 118–126. [CrossRef]
3. Genest, R.; Ringwood, J.V. A Critical Comparison of Model-Predictive and Pseudospectral Control for Wave Energy Devices. *J. Ocean Eng. Mar. Energy* **2016**, *2*, 485–499. [CrossRef]
4. Falcão, A.F.O.; Henriques, J.C.C. Effect of Non-Ideal Power Take-off Efficiency on Performance of Single- and Two-Body Reactively Controlled Wave Energy Converters. *J. Ocean Eng. Mar. Energy* **2015**, *1*, 273–286. [CrossRef]
5. Brekken, T.K.A. On Model Predictive Control for a Point Absorber Wave Energy Converter. In Proceedings of the 2011 IEEE Trondheim PowerTech, Trondheim, Norway, 19–23 June 2011; pp. 1–8.
6. Bubbar, K.; Buckham, B.; Wild, P. A Method for Comparing Wave Energy Converter Conceptual Designs Based on Potential Power Capture. *Renew. Energy* **2018**, *115*, 797–807. [CrossRef]

7. What Is Reinforcement Learning?—MATLAB & Simulink—MathWorks United Kingdom. Available online: https://uk.mathworks.com/help/reinforcement-learning/ug/what-is-reinforcement-learning.html (accessed on 2 October 2023).
8. Anderlini, E.; Forehand, D.I.M.; Bannon, E.; Xiao, Q.; Abusara, M. Reactive Control of a Two-Body Point Absorber Using Reinforcement Learning. *Ocean. Eng.* **2018**, *148*, 650–658. [CrossRef]
9. Control of a Realistic Wave Energy Converter Model Using Least-Squares Policy Iteration. Available online: https://ieeexplore.ieee.org/document/7911321 (accessed on 2 October 2023).
10. Zadeh, L.G.; Glennon, D.; Brekken, T.K.A. Nonlinear Control Strategy for a Two-Body Point Absorber Wave Energy Converter Using Q Actor-Critic Learning. In Proceedings of the 2020 IEEE Conference on Technologies for Sustainability (SusTech), Santa Ana, CA, USA, 23–25 April 2020; pp. 1–5.
11. Anderlini, E.; Husain, S.; Parker, G.G.; Abusara, M.; Thomas, G. Towards Real-Time Reinforcement Learning Control of a Wave Energy Converter. *J. Mar. Sci. Eng.* **2020**, *8*, 845. [CrossRef]
12. Rij, J.; Yu, Y.-H.; McCall, A.; Coe, R.G. Extreme Load Computational Fluid Dynamics Analysis and Verification for a Multi-Body Wave Energy Converter. In Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering. American Society of Mechanical Engineers, Glasgow, UK, 9–14 June 2019.
13. Haider, A.S.; Brekken, T.K.A.; McCall, A. A State-of-the-Art Strategy to Implement Nonlinear Model Predictive Controller with Non-Quadratic Piecewise Discontinuous Cost Index for Ocean Wave Energy Systems. In Proceedings of the 2020 IEEE Energy Conversion Congress and Exposition (ECCE), Detroit, MI, USA, 11–15 October 2020; pp. 1873–1878.
14. Houska, B.; Ferreau, H.J.; Diehl, M. An Auto-Generated Real-Time Iteration Algorithm for Nonlinear MPC in the Microsecond Range. *Automatica* **2011**, *47*, 2279–2285. [CrossRef]
15. Lu, H.; Chang, S.; Chen, C.; Fan, T.; Chen, J. Replacement of Force-to-Motion Relationship with State–Space Model for Dynamic Response Analysis of Floating Offshore Structures. *Appl. Ocean. Res.* **2022**, *119*, 102977. [CrossRef]
16. Barradas-Berglind, J.J.; Muñoz-Arias, M.; Wei, Y.; Prins, W.A.; Vakis, A.I.; Jayawardhana, B. Towards Ocean Grazer's Modular Power Take-Off System Modeling: A Port-Hamiltonian Approach. *IFAC-PapersOnLine* **2017**, *50*, 15663–15669. [CrossRef]
17. WEC-Sim (Wave Energy Converter SIMulator)—WEC-Sim Documentation. Available online: https://wec-sim.github.io/WEC-Sim/ (accessed on 27 March 2021).
18. Wamit, Inc. The State of the Art in Wave Interaction Analysis. Available online: https://www.wamit.com/ (accessed on 28 March 2021).
19. Falnes, J. Wave-Energy Conversion through Relative Motion between Two Single-Mode Oscillating Bodies. *J. Offshore Mech. Arct. Eng.* **1999**, *121*, 32–38. [CrossRef]
20. Haider, A.S.; Brekken, T.K.A.; McCall, A. Real-Time Nonlinear Model Predictive Controller for Multiple Degrees of Freedom Wave Energy Converters with Non-Ideal Power Take-Off. *J. Mar. Sci. Eng.* **2021**, *9*, 890. [CrossRef]
21. Technology—Centipod. Available online: https://centipodwave.com/technology/ (accessed on 24 October 2023).
22. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous Control with Deep Reinforcement Learning. *arXiv* **2019**. [CrossRef]
23. PacWave—TESTING WAVE ENERGY FOR THE FUTURE. Available online: https://pacwaveenergy.org/ (accessed on 24 October 2023).
24. Dunkle, G.; Zou, S.; Robertson, B. Wave Resource Assessments: Spatiotemporal Impacts of WEC Size and Wave Spectra on Power Conversion. *Energies* **2022**, *15*, 1109. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.