Noname manuscript No.
(will be inserted by the editor)

A noval guidance control of a thruster-assisted position mooring system with model-based reinforcement learning

Xu Jiang1 · Lei Wang · Shangyu Yu · Huacheng He · Te Yu

Received: date / Accepted: date

Abstract *Thruster-assisted position mooring (PM) systems use both mooring lines and thrusters for station keeping of marine structures in ocean environments. In order to operate in an energy-efficient manner in moderate sea conditions, setpoints need to be appropriately chosen for the setpoint controller, so that the mooring system counteracts main environmental loads, while the thrusters reduce oscillatory motions of the marine structure. Reinforcement learning (RL) is a powerful and effective approach to designing decision making agents for setpoint selection. We propose a model-based reinforcement learning method which interplays direct and indirect learning to update Q-function. The reward function of the world model is approximated by support vector regression. Extensive simulation results indicate that the proposed RL agents can successfully identify the optimal setpoints through continuous planning, acting and learning in an unknown and stochastic environment, and the model-based approach accelerates the learning speed of the decision making agent.*

Keywords thruster-assisted position mooring · optimal setpoint · reinforcement learning · Model-based Q-learning · neural network

1 INTRODUCTION

For offshore exploitation and exploration in intermediate water depths, mooring vessels with thruster assistence (position mooring-PM) are often the most cost-effective and feasible solution to stationkeeping operation compared with fixed platforms and dynamic positioned (DP) vessels[1]. The reason is that a PM system is uaually regarded as the combination of the mooring and DP system. And the mooring system provides effective passive control to compensate the mean environment loads at the intermediate water depths, while the DP systems can perform active control to reduce structure offset and keep the mooring line tensions within a safety limit in order to prevent line breakage. And PM systems have been commercially available since the 1980s and generally becomes a flexible choice for drilling and oil exploitation on the marginal fields[2].

There have been extensive studies on control strategies for DP and PM systems. [3] proposed the nonlinear feedback linearization and back-stepping control for DP. In the work of [4], a passive nonlinear observer with adaptive wave filtering is proposed, which reduces the number of tuning parameters. Further, [5] presented to use nonlinear sliding mode control for DP and couducted corresponding experiments. In terms of PM systems, [6] conducted model tests for a PM FPSO in a wide range of water depths. A PID controller was implemented in the tests to keep the center of the vessel turret at a reference position, and the heading into the waves. [7] presented a mathematical model of a PM vessel, and defined four control modes of operation for surge, sway and yaw, including manual, damping, setpoint and tracking control. Simulation results showed that heading setpoint control was crucial for turret-anchored ships, and damping control of surge and sway reduced oscillatory motions, which are induced by slowly varying environmental loads. [8] used the finite element method to model the dynamics of the mooring lines, and proposed the design of a dynamic line tensioning controller to minimize the energy consumption of thrusters by changing the length of mooring lines. [9] designed a nonlinear passive observer for PM ship-

Lei Wang
E-mail: wanglei@sjtu.edu.cn

State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai, China 200240
Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai, China 200240
School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China 200240

s, and presented the results of a full-scale test with a turret-anchored FPSO. [10] performed numerical simulations of a DP assisted turret moored FPSO to show that a light mooring system with a DP system is equivalent to withstand the same survival condition as a heavy passive mooring system. [11] conducted a fully coupled hull/mooring/riser dynamic analysis of a thruster-assisted turret-moored FPSO. The simulation results showed that the vessel's horizontal-plane response and the tensions of mooring lines and the riser were greatly reduced by the PM system.

As the DP technology became more mature, guidance systems are taken into consideration for satisfying the integration of operational requirements and the performance of vessel control systems. [12] represented the concept of optimal setpoint chasing for deep-water drilling. Further, [13] proposed to use structural reliability criteria of the drilling risers for the setpoint chasing. Considering the automatic switching between several controllers for a PM system when the environmental condition changes, [1] proposed to integrate the controllers into a supervisory control system, and employed both mean disturbance loads and wave peak frequency as the parameters to switch between controllers. By integrating an appropriate bank of controllers and models at the pant control level into a hybrid DP system, such supervisory-switched methodology is able to operate in varying environmental conditions. [14] developed a fault-tolerant control strategy for a PM vessel based on supervisory control method as well. A bank of passive observers were implemented to detect mooring line breakages and generate monitoring signals for the supervisor to select the most suitable controller.

When a PM system is in setpoint control mode, appropriate selection of setpoints is crucial for both the safety and performance of the PM system. [15] proposed a control strategy using mooring line tensioning and heading control by thrusters to reduce riser end angle deviations. The optimal setpoint was determined using the finite element model of the riser. Simulations and experiments verified that vessel offsets and riser end angle could be reduced by applying thruster actions and actively changing the mooring line length. [16] proposed a setpoint chasing approach for a PM system, which calculated the optimal setpoint based on a structural reliability criterion in order to keep all mooring lines at a specified reliability level. [17] proposed a setpoint chasing algorithm for PM systems. The low-frequency position of the vessel was chosen as the setpoint for thruster control in moderate sea conditions where the mooring lines compensate the mean environmental loads while thruster actions provided additional damping and restoring forces. In extreme conditions, the simulations and experiments suggested a position closer to the field zero point than the equilibrium point as the setpoint in order to reduce the mooring line tensions and vessel offset. [18] tested the setpoint chasing algorithm on a thruster-assisted turret-moored drillship in model tests. The experiments showed that the control algorithm reduced thruster usage and maximized the utilization of mooring system. [19] undertook both numerical and experimental studies on the influence of setpoint positions on the performance of a PM system. The results illustrated that the equilibrium point of the mooring system is the optimal setpoint in terms of positioning accuracy and power consumption of the thrusters.

In recent years, with the heated discussion of machine learning techniques especially reinforcement learning,many researchers have been searching the potential of using these techniques for industrial assets. Reinforcement learning is a maturing field in artificial intelligence, where a significant portion of the research is concerned with approaches in virtual environments [20]. During the training process, the algorithm learns how the transition function changes and estimates a state-value function V or state-action value Q that represent the value of being in the current state.[21] adopted a model-free Q-learning method, a powerful reinforcement learning technique, to predict the wave energy distribution for energy collection.[22] proposed a model-based reinforcement learning algorithm which combined a medium-size neural network model with model predictive control. The simulation results demonstrated that this model-based approach can quickly become competent for a given task. However, this method might be inapplicable with high-dimensional actions spaces. [23] developed a biomimetic underwater vehicle controller with Q-learning. The control policy was first trained by a simulator, then the updated neural network parameters were transferred to a real vehicle for validation. Recently, [24] designed a path following algorithm for unmanned surface vehicle (USV) with a deep deterministic policy gradient (DDPG) approach. Model tests demonstrated that by interacting with the environment, the USV learned the optimal policy successfully.

The objective of this paper is to develop an automated intelligent support agent which provides optimal setpoints for low-level setpoint controllers to follow. The novel decision-making strategy is based on both model-free and model-based reinforcement learning, goal-oriented machine learning tools with which the decision maker learns a policy from scratch to maximize long-term rewards through interactions with the unknown environment. By implementing the intelligent setpoint control strategies, the utilization of the mooring system is maximized to compensate the mean environmental loads, while the thruster actions reduce the oscillatory motion of the platform around the optimal setpoint.

The rest of the paper is organized as follows: the mathematical dynamics of a PM floating structure is introduced in section 2. In section 3, the plant control design of the PM system is addressed. Time-domain simulations are performed to demonstrate the control performances of the PM

system in both damping and setpoint control modes. The automated intelligent support agent for generating optimal setpoints is then presented in section 4. The decision-making strategy based on reinforcement learning is adopted, and numerical experiments are conducted for validation and discussion. Finally, the main conclusions are presented in section 5.

2 MATHEMATICAL MODELING OF A PM FLOATING STRUCTURE

As shown in Fig. 1, A global coordinate system is defined as (X_G, Y_G, Z_G) with origin O_G. For a PM system, the origin of the global coordinate system is usually the field zero point of the mooring system. The 6 degree-of-freedom dynamics of a semi-submersible platform in the time domain can be described using the Cummins equations [25], which are defined in the local coordinate system (X, Y, Z) fixed with respect to the mean position of the semi-submersible platform:

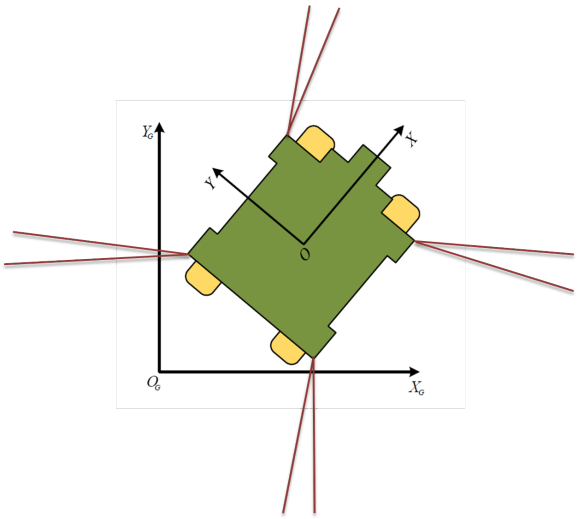

Fig. 1 The global and local coordinate systems.

$$(M_{RB} + A(\infty))\ddot{p} + \int_0^t K(t-\tau)\dot{p}(\tau)\mathrm{d}\tau + Cp = F(t) \tag{1}$$

where $M_{RB}, A(\infty), K(t), C \in \mathbb{R}^{6\times 6}$ are respectively the mass matrix of the platform, the constant infinite-frequency added mass matrix, the matrix of retardation function, and the matrix of hydrostatic restoring forces, and p represents the displacements in the six modes of response. Given the frequency-dependent added mass $A(\omega)$ and damping matrices $B(\omega)$, we have:

$$K(t) = \frac{2}{\pi}\int_0^\infty B(\omega)\cos(\omega t)\mathrm{d}\omega \tag{2}$$

$$A(\infty) = A(\omega) + \frac{1}{\omega}\int_0^\infty K(\tau)\sin(\omega\tau)\mathrm{d}\tau \tag{3}$$

The term $F(t)$ takes into account other time-varying loads, such as wave loads, viscous damping forces, wind and current loads, as well as mooring and thruster forces. The wave load effect on the platform includes both first-order wave excitation forces and second-order wave drift forces, which can be obtained using the force response amplitude operators (RAOs) and quadratic transfer functions (QTFs), respectively.

Due to the environmental loads, both the platform motion and mooring line tension have a mean static component plus wave frequency and low frequency dynamic components. The responses can be estimated in the frequency domain and time domain [26]. There are mainly two dominating treatments of the mooring line dynamics in the time domain: the slender rod model and the lumped mass model. In the lumped mass method, mooring line is spatial discretized into lumped mass nodes which are connected with massless springs. Discretization in this way simplifies the mathematical model and provides numerical efficiency [27–29]. Furthermore, each item in the lumped mass model has a clear physical meaning, which facilitates the understanding and programming.

3 PLANT CONTROL DESIGN OF PM SYSTEMS

The design of the PM control system is divided into two levels: low-level plant control of the semi-submersible platform, and high-level decision making of the optimal setpoint for the low-level controller. In this section, the designs of the low-level feedback controllers are presented.

The PM system usually has several automatic control modes for operation, such as damping and setpoint control. The position and heading angle of the semi-submersible platform are measured in the global coordinate system, which contains high-frequency motions induced by the first-order wave excitation forces. Therefore, an observer is necessary to estimate low-frequency position and velocity of the platform for the feedback controllers. A thrust allocation algorithm is also introduced to allocate commanded forces among the azimuth thrusters of the platform.

3.1 Controller Design

When designing the low-level controllers for the PM system, it is advantageous to present the platform dynamics in a

polar coordinate, so that weathervaning control can be conveniently achieved in several control modes. Follow [30], the Cartesian coordinates (x,y) is related to the polar coordinates by:

$$x = \rho\cos\gamma, \quad y = \rho\sin\gamma \tag{4}$$

where $\rho \in \mathbb{R}^+$ is the radius, and $\gamma \in S$ is the polar angle. Define the state vector $x = [\rho,\gamma,\psi]^T \in \mathbb{R}\times S^2$, we have:

$$\dot{\eta} = R(\gamma)H(\rho)\dot{x} \tag{5}$$

where $p = [x,y,\psi]^T \in \mathbb{R}^2 \times S$ represent the platform's position and orientation with respect to the global coordinate system, and the rotation matrix $R(\gamma) \in \mathrm{SO}(3)$, and $H(\rho)$ are given by:

$$R(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad H(\rho) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

The control plant model is usually defined with respect to the body-fixed frame of reference, where the platform velocity is represented by $\nu = [u,v,r]^T \in \mathbb{R}^3$. The kinematic relationship between the Earth-fixed and platform velocities is given by: $\dot{\eta} = R(\psi)\nu$. Thus we have:

$$\dot{x} = T(x)\nu \tag{7}$$

where the transformation matrix is defined by:

$$T(x) = H^{-1}(\rho)R^T(\gamma)R(\psi) \tag{8}$$

The low-frequency motion of the semi-submersible platform can be described by the following model [31]:

$$M\dot{\nu} + C(\nu)\nu + D(\nu)\nu + G\eta = \tau + w \tag{9}$$

where $M, C(\nu), D(\nu), G \in \mathbb{R}^{3\times3}$ respectively represent inertia including added mass, Coriolis-centripetal forces, damping forces, and linearized restoring forces due to the mooring system. $\tau \in \mathbb{R}^3$ is the control force. $w \in \mathbb{R}^3$ denotes the unknown external forces due to wind, wave and current, which are assumed to be slowly varying. Apply the transformation given by Eq. 7, the dynamic model can be expressed in polar coordinate as:

$$M_x(x)\ddot{x} + C_x(\nu,x)\dot{x} + D_x(\nu,x)\dot{x} + G_x(x) = T^{-T}\tau + T^{-T}w \tag{10}$$

where

$$M_x(x) = T^{-T}(x)MT^{-1}(x) \tag{11}$$

$$C_x(\nu) = T^{-T}(x)\left(C(\nu) - MT^{-1}(x)\dot{T}(x)\right)T^{-1}(x) \tag{12}$$

$$D_x(\nu,x) = T^{-T}(x)D(\nu)T^{-1}(x) \tag{13}$$

$$G_x(x) = T^{-T}(x)G\eta(x)T^{-1}(x) \tag{14}$$

Denote the reference trajectory for the control system as $x_d = [\rho_d,\gamma_d,\psi_d]^T \in C^3$, a virtual reference trajectory can be defined as:

$$\dot{x}_r = \dot{x}_d - \Lambda z_1 \tag{15}$$

where $z_1 = x - x_d$ is the tracking error expressed in the global coordinate system and $\Lambda > 0$ is a diagonal matrix. Moreover, define a new variable z_2 as:

$$z_2 = \dot{x} - \dot{x}_r = \dot{z}_1 + \Lambda z_1 \tag{16}$$

we obtain the following expressions:

$$\dot{x} = z_2 + \dot{x}_r, \qquad \ddot{x} = \dot{z}_2 + \ddot{x}_r \tag{17}$$

Substitute them into Eq. 10, the dynamic model can be rewritten as:

$$\begin{aligned} M_x(x)\dot{z}_2 + C_x(\nu,x)z_2 + D_x(\nu,x)z_2 + G_x(x) &= T^{-T}\tau \\ -M_x(x)\ddot{x}_r - C_x(\nu,x)\dot{x}_r - D_x(\nu,x)\dot{x}_r + T^{-T}w & \end{aligned} \tag{18}$$

By choosing a nonlinear PD control law as:

$$\tau = T^T\left[M_x(x)\ddot{x}_r + C_x(\nu,x)\dot{x}_r + D_x(\nu,x)\dot{x}_r - K_pz_1 - K_dz_2\right] \tag{19}$$

where $K_p, K_d \in \mathbb{R}^{3\times3}$ are positive definite diagonal matrices, the resulting dynamical system is given by:

$$\begin{aligned} \begin{bmatrix} K_p & O_{3\times3} \\ O_{3\times3} & M_x(x) \end{bmatrix}\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = &- \begin{bmatrix} K_p\Lambda & O_{3\times3} \\ O_{3\times3} & C_x(\nu,x) + D_x(\nu,x) + K_d \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\ &+ \begin{bmatrix} O_{3\times3} & K_p \\ -K_p & O_{3\times3} \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\ &+ \begin{bmatrix} O_{3\times1} \\ -G_x(x) + T^{-T}w \end{bmatrix} \end{aligned} \tag{20}$$

which is a non-autonomous system. Note that $G_x(x)$ is not directly balanced by the control actions, and instead treated as an unknown disturbance. Based on different control objectives, several control strategies can be proposed.

3.1.1 Damping Control

When $K_p = O_{3\times3}$ and $\Lambda = O_{3\times3}$, Eq. 20 can be simplified as:

$$\dot{z}_1 = z_2 \tag{21}$$

$$M_x(x)\dot{z}_2 = -\left[C_x(\nu,x) + D_x(\nu,x) + K_d\right]z_2 - G_x(x) + T^{-T}w \tag{22}$$

If the slowly varying environmental disturbance w can be totally compensated by the mooring system, then the equilibrium point $z_2 = O$ is exponentially stable, and instead of the tracking objective x_d, x converges to the equilibrium position x_e of the mooring system under the effect of w, which is given by $G_x(x_e) = T^{-T}w$. Damping control of the PM system introduces additional damping to the platform, which reduces the oscillatory motions around the equilibrium position x_e.

3.1.2 Setpoint Control

When $K_p \neq O_{3\times3}$ and $\Lambda \neq O_{3\times3}$, the equilibrium point of the dynamical system can be obtained by solving the following equations:

$$z_2 = \Lambda z_1 \tag{23}$$

$$\left[C_x(\nu,x) + D_x(\nu,x) + K_d\right]z_2 + K_p z_1 = -G_x(x) + T^{-T}w \tag{24}$$

in which $x = z_1 + x_d$. Depending on the choice of the setpoint for the PD controller, the closed-loop system converges to different positions.

If the tracking objective x_d follows the equilibrium position exactly so that $x_d = x_e$, then $z_1 = z_2 = O_{3\times1}$ is the equilibrium point of the dynamical system given by Eq. 20. Generally speaking, the equilibrium position x_e of the mooring system is unfortunately unknown. Denote $P = [C_x(\nu,x) + D_x(\nu,x) + K_d]\Lambda + K_p$, we have:

$$Px + G_x(x) = Px_d + T^{-T}w \tag{25}$$

It is shown that if $x_d \neq x_e$, the equilibrium point x^* of the dynamical system falls somewhere between x_d and x_e with respect to the field zero point of the mooring system, i.e., the origin of the global coordinate system. Define new variables $y_1 = x - x^*$ and $y_2 = z_2 - \Lambda e$, where $e = x^* - x_d$, Eq. 20 can be transformed into:

$$\begin{bmatrix} K_p + G_L & O_{3\times3} \\ O_{3\times3} & M_x \end{bmatrix}\begin{bmatrix}\dot{y}_1 \\ \dot{y}_2\end{bmatrix} = -\begin{bmatrix} (K_p+G_L)\Lambda & O_{3\times3} \\ O_{3\times3} & C_x + D_x + K_d \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} O_{3\times3} & K_p + G_L \\ -K_p - G_L & O_{3\times3} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{26}$$

where G_L is the linearized restoring matrix of the mooring system around the equilibrium point x^*. Define $y = [y_1^T, y_2^T]^T \in \mathbb{R}^6$, the equations can be compactly written as:

$$\mathscr{M}(x)\dot{y} = -\mathscr{H}(\nu,x)y + \mathscr{S}y \tag{27}$$

where the matrices are defined as:

$$\mathscr{M}(x) = \mathscr{M}^T(x) = \begin{bmatrix} K_p + G_L & O_{3\times3} \\ O_{3\times3} & M_x \end{bmatrix} \tag{28}$$

$$\mathscr{H}(\nu,x) = \begin{bmatrix} (K_p+G_L)\Lambda & O_{3\times3} \\ O_{3\times3} & C_x + D_x + K_d \end{bmatrix} > 0 \tag{29}$$

$$S = -S^T = \begin{bmatrix} O_{3\times3} & K_p + G_L \\ -K_p - G_L & O_{3\times3} \end{bmatrix} \tag{30}$$

It can be seen that the origin $y = O_{6\times1}$ is exponentially stable according to Lyapunov stability theory. Therefore, we have $x \to x^*$ and $\dot{x} \to \dot{x}_d$ as $t \to +\infty$. The selection of the tracking objective x_d in setpoint control is crucial for the performance of the PM system. It is always advantageous to chase the exact equilibrium position x_e of the mooring system in order to maximize the utilization of the mooring lines in moderate sea conditions. Since it is impossible to determine x_e accurately, choosing a setpoint closer to the field zero point than the equilibrium position is generally better than further to the field zero point, in terms of preventing the thrusters from fighting against the mooring system.

3.2 Observer Design

Following the methods proposed by [4], a nonlinear passive observer can be derived as:

$$\dot{\hat{\xi}} = A_w\hat{\xi} + K_1\tilde{y} \tag{31}$$

$$\dot{\hat{\eta}} = R(\hat{\psi})\hat{\nu} + K_2\tilde{y} \tag{32}$$

$$\dot{\hat{b}} = -T^{-1}\hat{b} + K_3\tilde{y} \tag{33}$$

$$M\dot{\hat{\nu}} = -D\hat{\nu} - G\hat{\eta} + R^T(\hat{\psi})\hat{b} + \tau + K_4R^T(\hat{\psi})\tilde{y} \tag{34}$$

$$y = \hat{\eta} + C_w\hat{\xi} \tag{35}$$

where $K_1 \in \mathbb{R}^{6\times3}, K_2, K_3, K_4 \in \mathbb{R}^{3\times3}$ are the observer gain matrices. $\tilde{y} = y - \hat{y}$ is the estimation error. $\hat{\eta}$ is the estimated low-frequency motion in the global coordinate system. $\hat{b}$ represents the estimated slowly varying environmental disturbances and unmodeled dynamics. $C_w\hat{\xi}$ gives the estimated wave-frequency motion of the semi-submersible platform. The details of the observer design can be found in [4]. Although an additional term of the mooring system stiffness $G\hat{\eta}$ is introduced to the observer compared with the original design for a DP system, it does not affect the result of the stability analysis.

3.3 Thrust Allocation

The control forces and moments calculated by the feedback controllers are distributed among the azimuth thrusters of the semi-submersible platform. The thrust allocation at each time-step can be formulated as a quadratic programming problem:

$$\min \quad J(u,\Delta\alpha,s) = u^T W u + s^T Q s + \Delta\alpha^T \Omega \Delta\alpha \tag{36}$$

$$\text{s.t.} \quad s + B(\alpha_0)u + \frac{\partial}{\partial\alpha}\left(B(\alpha)u\right)|_{\substack{\alpha=\alpha_0\\u=u_0}}\Delta\alpha = \tau \tag{37}$$

$$u_{\min} \le u \le u_{\max} \tag{38}$$

$$\Delta\alpha_{\min} \le \Delta\alpha \le \Delta\alpha_{\max} \tag{39}$$

where $W \in \mathbb{R}^{n\times n}, Q \in \mathbb{R}^{3\times 3}, \Omega \in \mathbb{R}^{n\times n}$ are the weight matrices, $u \in \mathbb{R}^n$ represents the thrusts of n available azimuth thruster, $\alpha \in \mathbb{R}^n$ denotes their azimuth angles, $\Delta\alpha \in \mathbb{R}^n$ is the change of azimuth angle, $s \in \mathbb{R}^3$ is the slack variable, $B(\alpha) \in \mathbb{R}^{3\times n}$ represents the configuration matrix of the thrusters, $\tau \in \mathbb{R}^3$ is the commanded thrust given by the feedback controllers, and $[u_{\min}, u_{\max}], [\Delta\alpha_{\min}, \Delta\alpha_{\max}]$ are respectively the constraints on $u, \Delta\alpha$ at each time-step.

3.4 Simulations of PM Control Modes

A numerical simulation program is developed using C++. Simulations are carried out with a semi-submersible platform in order to demonstrate the behaviors of a PM system in the following control modes:

- CM1: damping control for ρ, γ, ψ;
- CM2: setpoint control for ρ, and damping control for γ, ψ.

In CM1, the thruster actions reduce the oscillatory motion of the platform. The semi-submersible platform is free to move in both radial and tangential directions with respect to the global polar coordinate. If the environmental loads are stationary, the platform will converge to the equilibrium position (ρ_e, γ_e) of the mooring system. In CM2, the semi-submersible is free to move along the circle arc of radius ρ^* which is between the desired radius ρ_d and equilibrium position ρ_e. Accordingly, the PM system is capable of weather-vaning in this control mode. Meanwhile, the setpoint control of the radial offset ρ requires the decision-making of the desired radius ρ_d, which will be addressed in the next section.

Table 1 presents the main dimensions of the semi-submersible platform, and Table 2 gives the specifications of eight corresponding azimuth thrusters. The mooring system consists of eight cables connected to the platform. Each line is composed of three segments. The parameters of each segment are presented in Table 3, and the water depth is 1500 m. The semi-submersible platform is subjected to environmental disturbances due to wind, wave and current. The environmental condition used in the simulations is given in Table 4. The headings are defined in the global coordinate system.

Table 1 Dimensions of the semi-submersible platform

Parameter	Value	Units
Displacement	52509	t
Draft	19	m
VCG (from BL)	25.84	m
Roll radius of gyration	32.8	m
Pitch radius of gyration	33.2	m
Yaw radius of gyration	37.8	m

Table 2 Thruster specifications

Thruster	X (m)	Y (m)	Max. Thrust (kN)
No.1	15.70	35.50	800
No.2	47.02	24.58	800
No.3	47.02	-24.58	800
No.4	15.70	-35.50	800
No.5	-15.70	-35.50	800
No.6	-47.02	-24.58	800
No.7	-47.02	24.58	800
No.8	-15.70	35.50	800

Table 3 Properties of mooring lines in each section

Parameters	Fairlead	Middle	Ground
Length (m)	150	2650	1450
Diameter (mm)	84	160	84
Wet weight (kg/m)	139.2	4.2	141.36

Table 4 Environmental condition

Type	Condition	Heading
Wind	$U_w = 27.0$ m/s	60°
Current	$U_c = 0.75$ m/s	60°
Wave	$H_s = 5.27$ m, $T_p = 10.4$ s (Jonswap, $\gamma = 3.3$)	60°

The simulation time is chosen as three hours in the following studies. First, the control system is switched off, and the semi-submersible platform is positioned by the mooring system alone without any thruster assistance. In the second run, CM1 is activated instead so that the thruster actions provide damping effects to reduce the oscillatory motion of the platform. The results of these two simulations are illustrated in Fig. 2, where only the results of the last two

hours are displayed. It shows that the moored platform oscillates around the equilibrium position of the mooring system where $\rho_e \approx 25$ m and $\gamma_e \approx 60°$, without thruster actions. When operating in the damping control mode, the equilibrium position is still the center of the path, while the oscillatory motion is greatly subdued due to the thruster assistance, as expected.

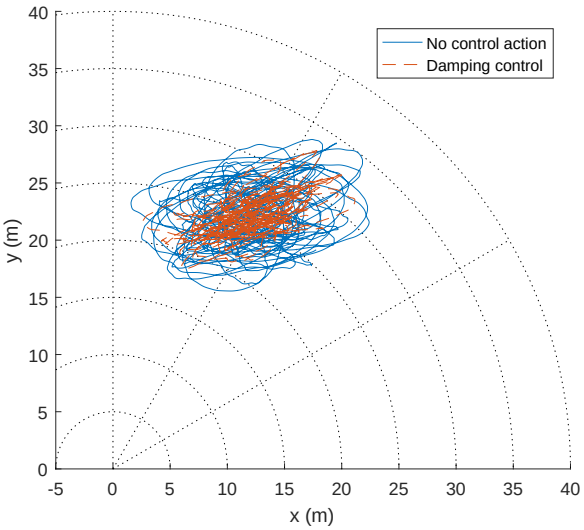

Fig. 2 Platform paths under no control action and damping control mode, respectively.

Fig. 3 depicts the performance of the platform motion when being positioned in CM2 with different fixed values of ρ_d. In each simulation where the same control parameters Λ, K_p, K_d are used, the platform mainly moves along the circle arc of radius ρ_d, as the motions along the radial direction are constrained by the mooring lines and the setpoint control actions. When the setpoint $\rho_d > \rho_e$, the thrusters fight against the mooring system and push the platform to the radial offset ρ_d. As the setpoint ρ_d moves closer to the field zero point when $\rho_d < \rho_e$, an increasing portion of the mean environmental loads is compensated by the thruster actions.

Fig. 4 shows how the selection of the radius setpoint ρ_d affects the power consumption level of the thrusters. The power consumed by a thruster in a DP system can be calculated by its torque multiplied by its speed of revolution [32]. Accordingly, the power consumption level at each radius setpoint is approximated by the mean value of $\Sigma u_i(t)^{(3/2)}$, where u_i denotes the thrust force exerted by the ith azimuth thruster. As expected, the minima of the power consumption lies at $\rho_d = \rho_e$, where the mean environmental disturbances are mainly counteracted by the mooring system. The equilibrium position of the mooring system is considered to be the optimal setpoint for the PM system in terms of the power consumption of the thrusters. In the next section, a novel

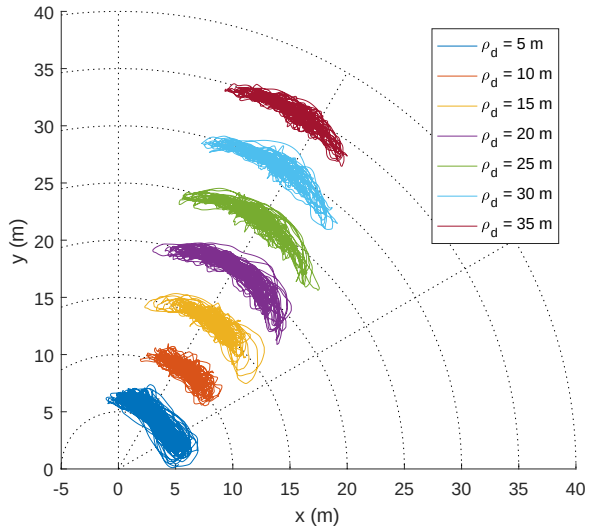

Fig. 3 Platform paths with different radius setpoint ρ_d.

method to determine the optimal setpoint is presented for a PM system operating in an unknown environment.

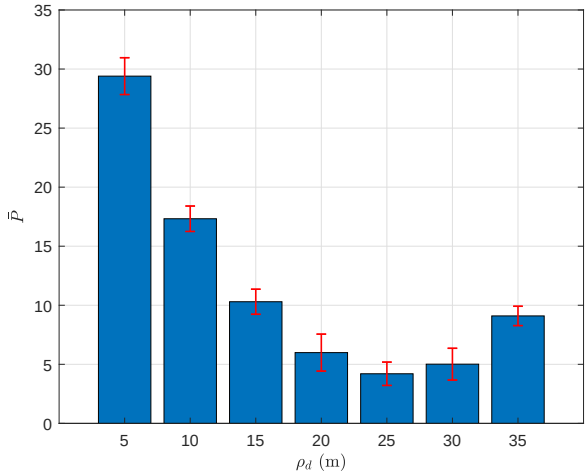

Fig. 4 The mean value and standard error of the average power consumption level $\Sigma u_i(t)^{(3/2)}$ at each radius setpoint ρ_d over a 5-min interval. Note that the displayed values of the vertical axis have been divided by 1000.

4 DECISION-MAKING APPROACH

The high-level decision support for the PM control system is responsible for automatically providing the appropriate setpoints for the low-level controllers in CM2, rather than letting the operators manually decide. Maximizing the utilization of the mooring system to compensate the mean environmental loads is the primary objective of the decision-making strategy for PM systems operating in moderate sea conditions. In this section, an automated decision support a-

gent is presented to determine the optimal setpoints, which is based on the method of reinforcement learning (RL).

4.1 Reinforcement Learning

A RL agent learns a state-action policy $\pi = P(a|s)$ through a Markov decision process (MDP) with observations (s), actions (a) and rewards (r) while interacting with the environment. The final goal of RL is to select actions which maximize cumulative future reward, which can be represented by the optimal action-value function:

$$Q^*(s,a) = \max_{\pi} \mathbb{E}\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots | s_t = s, a_t = a, \pi\right] \tag{40}$$

where r_t is the reward at time-step t, and γ is the discount factor. In practice, it is common to use a linear or nonlinear function approximator to estimate the Q-function, $Q(s,a;\theta) \approx Q^*(s,a)$ [20]. In this study, the method proposed by [33] is adopted, which uses a feedforward neural network with weights θ as the Q-function approximator. The architecture of the neural network is illustrated in Fig. 5. The state representation s is the input to the neural network, and the outputs $Q(s,a)$ correspond to the approximate Q-values of the individual actions for the input state.

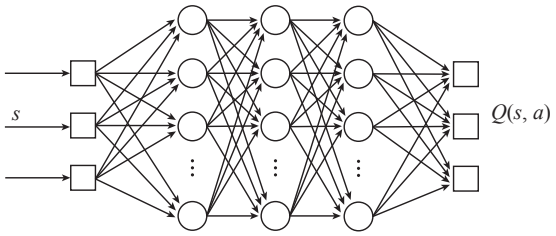

Fig. 5 Architecture of the neural network as the Q function approximator.

The training algorithm is based on Q-learning method. After taking some action a_i based on an ε-greedy policy and observing (s_i, a_i, r_i, s_i') at iteration i, the Q-network can be trained by adjusting the parameters θ_i to reduce the mean-squared error:

$$L(\theta_i) = \frac{1}{2}||Q(s_i,a_i;\theta_i) - y_i||^2 \tag{41}$$

where $y_i = r_i + \gamma \max_{a'} Q(s',a';\theta_i^-)$ is the approximate target value with parameters θ_i^- from some previous iteration. The techniques of experience replay and target network are also used in the training algorithm to provide diverse and decorrelated training samples and to prevent the action selection policy from oscillating or even diverging. The samples of the RL agent's experience (s_t, a_t, r_t, s_{t+1}) at each time-step are stored in a replay memory, and a random mini-batch is drawn from the stored samples when performing a gradient descent step on $L(\theta)$. Second, a separate network $\hat{Q}$ is used for generating the approximate targets y_i in the Q-learning update. The parameters of the target network $\hat{Q}$ are replaced by the network Q every N updates.

The selections of optimal setpoints can be formulated as an infinite MDP. The state is the radius of the platform in the polar coordinate system. Based on the observations that the platform in CM2 always oscillates within a small range around the setpoint along the radial direction, the states can be defined by the discrete setpoints for the PM system. The actions consist of increasing or decreasing the radius setpoint by $\Delta\rho$, as well as keeping the setpoint unchanged. The set of actions can be denoted by:

$$\mathscr{A} = \{0, -\Delta\rho, \Delta\rho\} \tag{42}$$

Accordingly, the states can be defined as:

$$\mathscr{S} = \{\cdots, \rho_0 - 2\Delta\rho, \rho_0 - \Delta\rho, \rho_0, \rho_0 + \Delta\rho, \rho_0 + 2\Delta\rho, \cdots\} \tag{43}$$

where ρ_0 is an initial selection of the radius setpoint. As a result, if the agent takes action a in state s, we know exactly what the next state s' is. Besides, the states must be within a certain range $(0, \rho_s]$, in which the upper bound ρ_s is chosen to ensure the safety of the riser and mooring system. A performance measure of the PM system can be selected as the reward. Fig. 4 shows that the performance at a radius setpoint can be characterized by the power consumption level $\bar{P} = \sum \bar{u}_i(t)^{(3/2)}$ of the thrusters. Accordingly, the reward function can be defined as a mapping $r : \bar{P} \to \mathbb{R}^-$, e.g., $r(s,a) = -\bar{P}$, so that a higher power consumption level applies a heavier penalty to the given state-action pair. $\bar{P}$ is computed over the interval Δt between two time-steps of the decision support agent. Since the environmental loads are constantly varying, the reward $r(s,a)$ is non-stationary. A trade-off must be made when choosing the time-step size Δt of the decision support agent. On the one hand, it should be long enough for the PM system to converge to the desired radius setpoint in order to obtain a reliable reward. On the other hand, it should be as small as possible so that enough samples can be generated in a relatively short period of time to train the Q-network.

At each time-step, the model-free RL agent selects an action from $\mathscr{A}$ based on an ε-greedy policy that follows the greedy policy $a = \text{argmax}_{a'} Q(s,a';\theta)$ with probability $1-\varepsilon$ and selects a random action with probability ε. Then, the new radius setpoint is sent to a first-order reference model so as to provide smooth transfer between two setpoints. A

upper limit is also imposed on the rate of change of the radius setpoint, in order to ensure the stability of the low-level setpoint controller.

4.2 Model-based Acceleration

In Model-free RL, the RL agent depends on sampling and observation heavily, and the inner working of the system remains unclear. On the contrary, a model-based RL has a strong advantage of being sample efficient. The reason is that a model-based RL agent can learn a world model of how the environment operates from its observations, then use the world model to train its state-action policy. In other words, the model which stores knowledge about the transitions and reward dynamics will produce a estimated next state and reward for learning a improved policy. For instance, if the agent is currently in state s_i, takes action a_i, and then observes the environment transition to state s_{i+1} with reward r_{i+1}, that information can be used to improve its estimate of $T(s_{i+1}|s_i,a_i)$ and $R(s_i,a_i)$, which can be performed using supervised learning approaches. Once the agent has adequately modeled the environment, it can use a planning algorithm with the learned model to learn a policy[34].

In order to build a model based RL agent, the state transition function and the reward function should be modeled at the beginning. The state-transition function can be defined as:

$$s_{t+1} \sim T_\eta(s_{i+1}|s_i,a_i) \tag{44}$$

where η is the parameter of the model. In the problem of optimal setpoint learning, the state of next time step s_{t+1} is actually known and uniquely determined by (s_t,a_t), since the setpoint controller of the PM system is assumed to be able to position the floating structure accurately. On the other hand, the modeling of reward function is considerably complex, which can be written as:

$$r_t = R_\eta(r_t|s_t,a_t) \tag{45}$$

Since the reward is calculated based on the average power consumption $\bar{P}$ in the real environment, in the planning model the power consumption needs to be accurately estimated by developing a regression model.

The construction of power consumption model is a typical supervised learning problem. The data set $\{(\rho_d^{(i)},\bar{P}^{(i)})\}$ collected from the real world is used to from training process to train a regression model. In this work, we use the real experience to train the regression model to estimate the average power consumption and calculate the corresponding reward value. The support vector machine (SVM) is used to establish the relation between average power consumption and radius setpoint, the radial basis function is chosen to be the kernel function of the SVM.

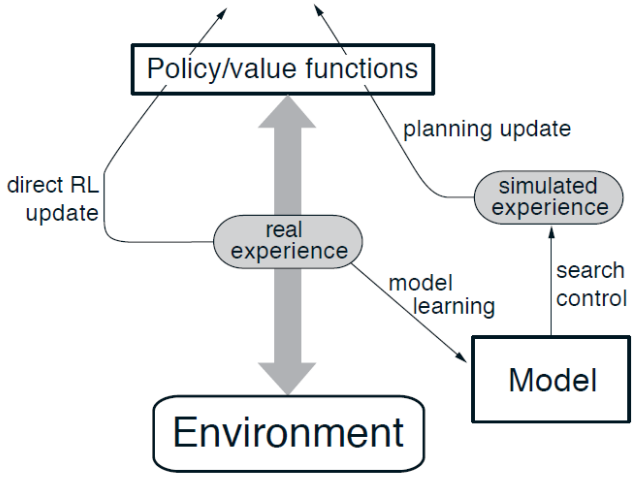

Fig. 6 The general model-based Architecture, which takes advantage of both real experience and simulated experience.(adapted from[20])

Fig. 6 shows the general architecture of the model-based RL method, which is proposed by [20] . In this architecture, the training data set consists of real experience and simulated experience, and the direct model-free RL process run together with the indirect model-based RL process. At each decision time step, the model-based agent uses the trajectory $\{(\rho_d^{(i)},\bar{P}^{(i)})\}$ of the past 3 hours to train the regression model, and it then creates a collection of simulated experience along the setpoints the agent chose in the past three hours. Also, based on the learned world model, the agent uses these data to train the Q-network together with the real experience. The additional training experience would efficiently accelerate the learning of an optimal policy.

4.3 Simulation Results and Discussion

The overall scheme of the intelligent PM system with model-based acceleration is given in Fig. 7, where the model-based decision level represents the high-level for generating the sequence of radius setpoints. Simulations are conducted with both the high-level decision support agent and the low-level setpoint controller involved. In this study, both model-free and model-based RL agents are implemented using the Google TensorFlow framework. The Q-network has one input, one hidden layer with 20 nodes, and three outputs. ReLU activation functions are applied to the nodes in the hidden layer. We use the same network architecture and hyperparameter values of the learning algorithm (see Table 5) throughout the paper. The learning rate in the table

is used by the RMSProp algorithm in the gradient descent updates.

In the simulations, the time-step size Δt of the RL agent is chosen as 5 mins, and the action step $\Delta\rho$ is 1.5 m. The initial radius setpoint of the platform is 10 m, and the radius setpoints must be smaller than 40 m. The reward function is defined as $r(s,a) = k_0(\bar{P}-k_1)(\bar{P}-k_2)$, in which k_0, k_1, k_2 are carefully chosen constants.

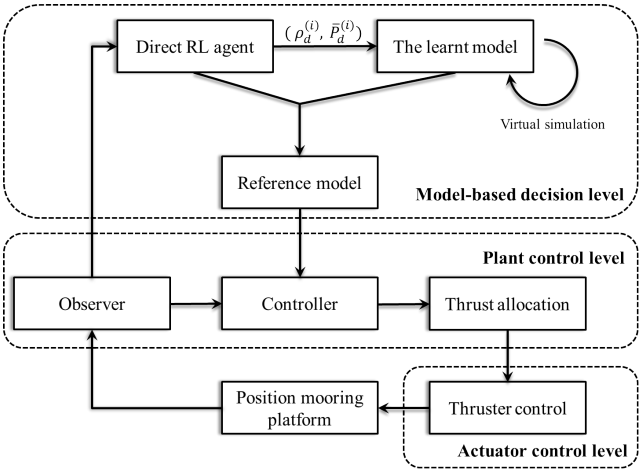

Fig. 7 The overall scheme of the intellegent PM system with model-based acceleration

Table 5 Hyperparameters of the RL agent

Hyperparameters	Value
minibatch size	32
replay memory size	200
target network update frequency	36
discount factor	0.9
learning rate	0.05
exploration rate	0.1

Table 6 Environmental condition for simulations

Type	Condition	Heading
Wind	$U_w = 27.0$ m/s	60° 45° 90°
Current	$U_c = 0.75$ m/s	60° 45° 90°
Wave	$H_s = 5.27$ m, $T_p = 10.4$ s (Jonswap, $\gamma = 3.3$)	60° 45° 90°

Fig. 8 illustrates the simulation results of both model-free and model-based RL agent with the incident angle equals to 60 degree. Distinct random seeds for the ε-greedy policy, the initial values of the Q-network parameters and selection of mini-batch are used in the 15 simulations for each RL agent. It is shown that the valley of the power consumption can be successfully found within 10 hours by the

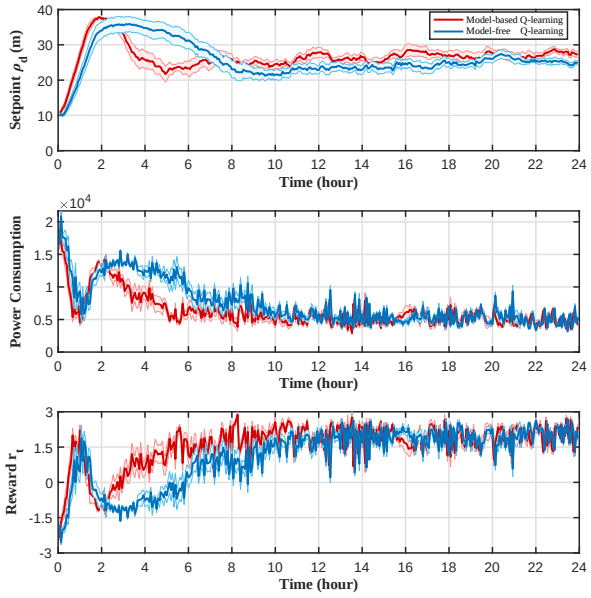

Fig. 8 The average setpoints ρ_d determined by the RL agents, the corresponding power consumption level and reward r_t during the whole simulation, with the incident angle equals to 60 degree. The shades represent the standard errors.

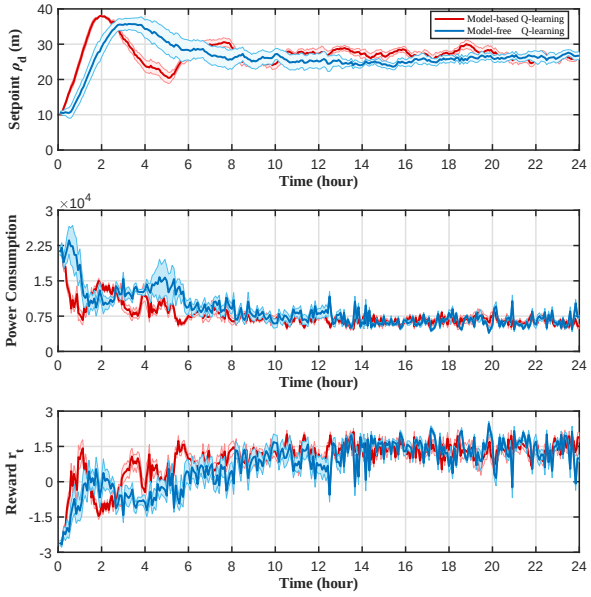

Fig. 9 The average setpoints ρ_d determined by the RL agents, the corresponding power consumption level and reward r_t during the whole simulation, with the incident angle equals to 45 degree. The shades represent the standard errors.

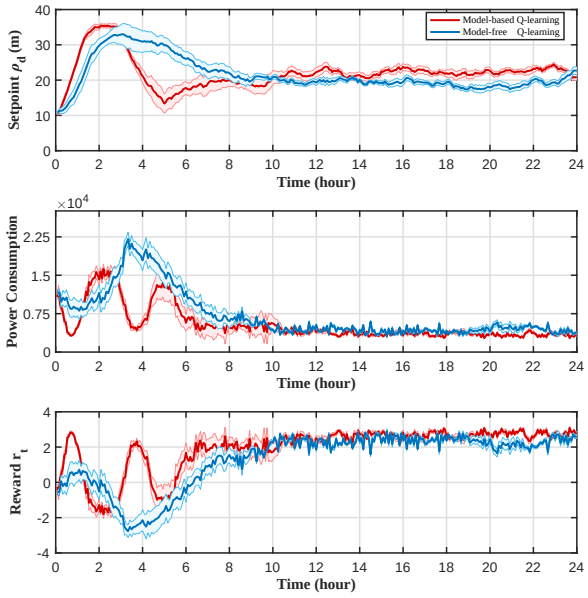

Fig. 10 The average setpoints ρ_d determined by the RL agents, the corresponding power consumption level and reward r_t during the whole simulation, with the incident angle equals to 90 degree. The shades represent the standard errors.

model-free RL agent, and 8 hours by the model-based RL agent. Most of the selected radius setpoints fall roughly into the range between 25.0 and 28.0 m. Even though the sea condition is stationary in the simulation, the wave drift force varies continuously in each time-step. Accordingly, due to the changing environmental disturbance and the dynamics of the PM system, the reward function is non-stationary, and the radius setpoint and reward keeps fluctuating in the valley of the power consumption curve of the thrusters.

It is shown clearly that both the model-free and model-based RL approach can help the PM system find the optimal setpoint, while the later approach has a greater efficiency to reach the convergence of the optimal strategy, which may thank to the construction of the simulated experience. At the beginning of the training process, the model-free agent and the model-based agent generally generated the same radius setpoint, this is because the model-based agent can only conduct simulated experience after it collected enough state-action pair for modeling a relatively precise model, otherwise the incorrect model will negatively influence the learning process. As mentioned before, the model-free RL relies solely on the real experience from the interaction with the environment to learn the state-action policy. Accordingly, after the radius setpoint of the PM system moves to the upper bound for the first time, the learning process can be extremely slow, because the model-free RL agent has no experience of moving back from the upper bound at all.As a consequence, it takes quite a long time for the model-free RL agent to figure out that it is better to leave the upper bound than hanging there, in order to lower the total power consumption.

The model-free RL, however, uses simulated experience, which includes the samples of moving back from the upper bound, to train the Q-network together with the real experience. Therefore, the key to success of the model-based RL agent is the learning of the world model. Fig. 11 to Fig. 16 show how regression model of power consumption is built over time. The red dots represent the real average power consumption data at each experienced setpoint radius for the last three hours, and the blue curve is the fitted curve by the support vector machine. It can be seen that the support vector machine with the radial basis function as the kernel function can fit the data points quite well. It can not only generate reliable simulated experience in the local areas, but also demonstrate certain generalization capabilities in predicting the power consumption in a global manner. Therefore, the model-based agent reaches the optimal position quicker than the model-free agent.

On top of that, further numerical experiments with different environment conditions were taken to validate the robustness of the proposed scheme (see Table 6). Fig. 9 and Fig. 10 gives the simulation results with different environment incident angles. Under both situations, the model-free and model-based agent successfully found the optimal radius setpoint region within 10 and 8 hours respectively. A larger fluctuation happened when the incident angle is 45°, the optimal radius setopint region under such environment condition is between 26m and 28m. On the other hand, when the incident angle is 90°, the floating structure reached a steady setpoint radius(between 20 and 22m) with a relatively stable power consumption level and a better reward after a quick exploration in first six hours.

Compared with 60° and 90° case, the simulation with a 45° heading direction experienced a larger power consumption fluctuation, this is because the external environment force is asymmetrically applied to the floating structure, resulting an uneven force sharing for thrusters and mooring lines. As mentioned before, the accuracy of the model has great impact on the model-based reinforcement learning. In 45° heading direction case, the same decided radius setpoint may get different power consumption level at different times, which might negatively influence the model-based learning process.

Overall, the simulation results demonstrate that the RL agent based on both model-based and model-free approach is fairly effective in finding the optimal radius setpoints for the PM system on the basis of the continuous decreasing trend of the power consumption level during the training process, and the model-based approach is shown to increase the learning speed of the RL agent.

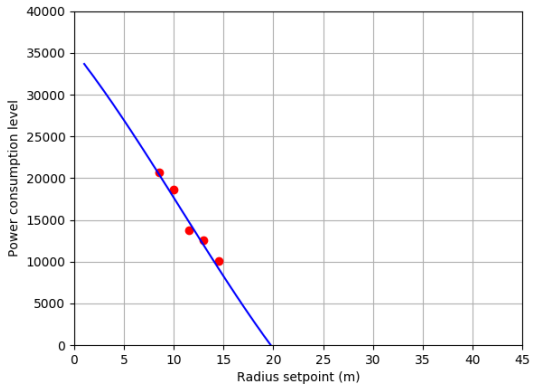

Fig. 11 The learned model of power consumption when t = 25 min.

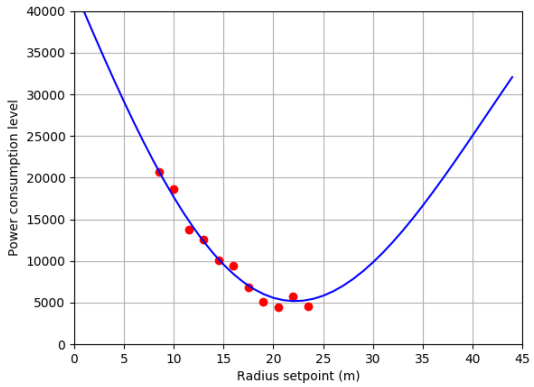

Fig. 12 The learned model of power consumption when t = 55 min.

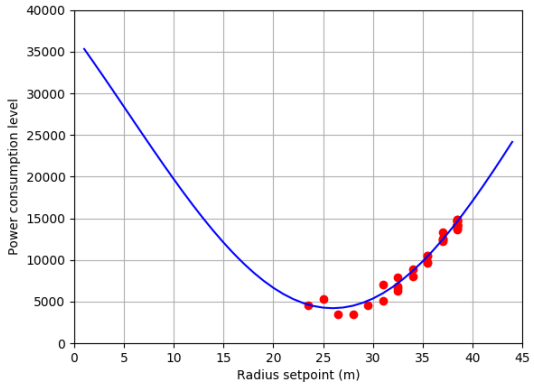

Fig. 13 The learned model of power consumption when t = 230 min.

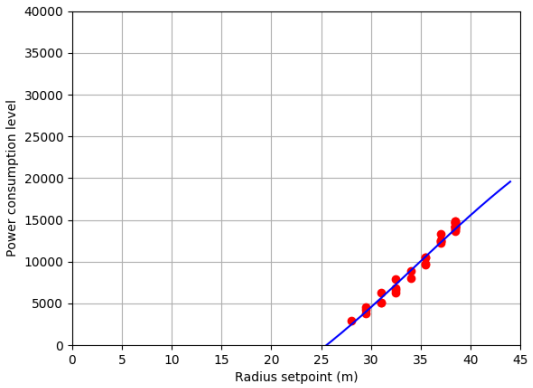

Fig. 14 The learned model of power consumption when t = 260 min.

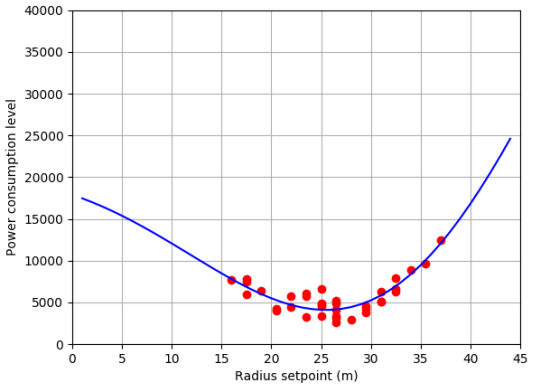

Fig. 15 The learned model of power consumption when t = 375 min.

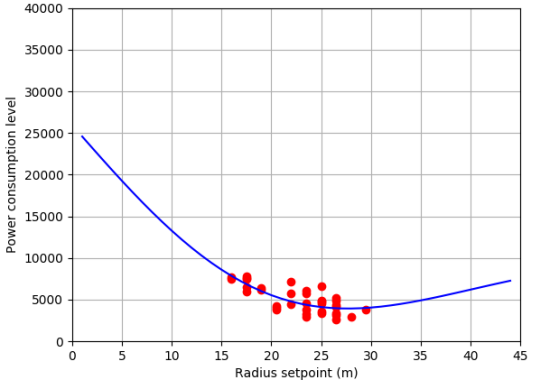

Fig. 16 The learned model of power consumption when t = 430 min.

5 CONCLUSIONS

In this paper, a novel decision-making strategy for selecting optimal setpoints is developed for PM systems operating in moderate sea conditions. Based on simulation results, it is shown that the equilibrium position of the mooring system is the optimal setpoint to follow, which can reduce the power consumption of the thrusters and maximize the utilization of the mooring lines.

The optimal setpoints are determined by the proposed intelligent decision support agent based on both model-free and model-based reinforcement learning, in which neural networks are used to approximate the action-value functions of the learning algorithm. The simulation results

demonstrate that for a given unknown stochastic sea condition,the RL agents can successfully find the desired setpoints through extensive interactions with the environment, while the model-based approach can profoundly accelerate the learning speed of the agent, showing its promising potential in future applications.

In the future, it will be important to conduct model tests for further validation. Moreover, the proposed method will be implemented for a non-collinear unsteady ocean environment, and the case of damaged mooring line will be studied as well, in order to further validate the capabilities of the intelligent decision-making agent in more realistic environments.

Acknowledgements The authors greatly acknowledge the support of the China Postdoctoral Science Foundation (Grant No. 2017M621479), the Open Foundation of State Key Laboratory of Ocean Engineering (Grant No. 1717) and the 7th Generation Ultra Deep Water Drilling Unit Innovation Project.

References

1. Nguyen DT, Sørensen AJ (2009) Switching control for thruster-assisted position mooring. Control Engineering Practice 17(9):985–994
2. Sørensen AJ (2011) A survey of dynamic positioning control systems. Annual reviews in control 35(1):123–136
3. Aarset MF, Strand JP, Fossen TI (1998) Nonlinear vectorial observer backstepping with integral action and wave filtering for ships. IFAC Proceedings Volumes 31(30):77–82
4. Fossen TI, Strand JP (1999) Passive nonlinear observer design for ships using lyapunov methods: Full-scale experiments with a supply vessel. Automatica 35(1):3 – 16
5. Tannuri E, Agostinho A, Morishita H, Moratelli Jr L (2010) Dynamic positioning systems: An experimental analysis of sliding mode control. Control Engineering Practice 18(10):1121–1132
6. Aalbers A, Merchant A (1996) The hydrodynamic model testing for closed loop dp assisted mooring. In: Offshore Technology Conference, Offshore Technology Conference
7. Strand JP, Sørensen AJ, Fossen TI (1998) Design of automatic thruster assisted mooring systems for ships. Modeling, Identification and Control 19(2):61–75
8. Aamo OM, Fossen TI (1999) Controlling line tension in thruster assisted mooring systems. In: Control Applications, 1999. Proceedings of the 1999 IEEE International Conference on, IEEE, vol 2, pp 1104–1109
9. Sorensen A, Strand JP, Fossen TI (1999) Thruster assisted position mooring system for turret-anchored fpsos. In: Control Applications, 1999. Proceedings of the 1999 IEEE International Conference on, IEEE, vol 2, pp 1110–1117
10. Wichers J, van Dijk R, et al (1999) Benefits of using assisted dp for deepwater mooring systems. In: Offshore Technology Conference, Offshore Technology Conference
11. Ryu S, Kim M (2003) Coupled dynamic analysis of thruster-assisted turret-moored fpso. In: OCEANS 2003. Proceedings, IEEE, vol 3, pp 1613–1620
12. Sørensen AJ, Leira B, Peter Strand J, Larsen CM (2001) Optimal setpoint chasing in dynamic positioning of deep-water drilling and intervention vessels. International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal 11(13):1187–1205
13. Leira B, Sørensen A, Berntsen P, Aamo O (2006) Structural reliability criteria and dynamic positioning of marine vessels. International Journal of Materials & Structural Reliability 4(2):161–174
14. Ren Z, Skjetne R, Hassani V (2015) Supervisory control of line breakage for thruster-assisted position mooring system. IFAC-PapersOnLine 48(16):235–240
15. Nguyen DH, Nguyen DT, Quek ST, Sørensen AJ (2010) Control of marine riser end angles by position mooring. Control Engineering Practice 18(9):1013–1021
16. Fang S, Leira BJ, Blanke M (2013) Position mooring control based on a structural reliability criterion. Structural Safety 41:97–106
17. Nguyen DT, Sørensen AJ (2009) Setpoint chasing for thruster-assisted position mooring. IEEE Journal of Oceanic Engineering 34(4):548–558
18. Bjørnø J, Heyn HM, Skjetne R, Dahl AR, Frederich P (2017) Modeling, parameter identification and thruster-assisted position mooring of c/s inocean cat i drillship. In: ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering, American Society of Mechanical Engineers
19. Wang L, Yang J, He H, Xu S, Su Tc, et al (2016) Numerical and experimental study on the influence of the set point on the operation of a thruster-assisted position mooring system. International Journal of Offshore and Polar Engineering 26(04):423–432
20. Sutton RS, Barto AG (1999) Reinforcement Learning: An Introduction. MIT Press
21. Anderlini E, Forehand D, Bannon E, Xiao Q, Abusara M (2018) Reactive control of a two-body point absorber using reinforcement learning. Ocean Engineering 148:650–658
22. Nagabandi A, Kahn G, Fearing RS, Levine S (2018) Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp 7559–7566

23. Magalhães J, Damas B, Lobo V (2018) Reinforcement learning: The application to autonomous biomimetic underwater vehicles control. In: IOP Conference Series: Earth and Environmental Science, IOP Publishing, vol 172, p 012019
24. Woo J, Yu C, Kim N (2019) Deep reinforcement learning-based controller for path following of an unmanned surface vehicle. Ocean Engineering 183:155–166
25. Cao Y, Tahchiev G, Zhang F, Aarsnes JV, Glomnes E-B (2010) Effects of hydrostatic nonlinearity on motions of floating structures. In: Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE, Shanghai, China, vol 4, pp 257 – 267
26. Gao Z, Moan T (2009) Mooring system analysis of multiple wave energy converters in a farm configuration. In: 8th European Wave and Tidal Energy Conference (EWTEC), pp 509–518
27. Chai Y, Varyani K, Barltrop N (2002) Three-dimensional lump-mass formulation of a catenary riser with bending, torsion and irregular seabed interaction effect. Ocean Engineering 29(12):1503–1525
28. Low Y, Langley R (2006) Time and frequency domain coupled analysis of deepwater floating production systems. Applied Ocean Research 28(6):371–385
29. Xiong L, Yang J, Zhao W (2016) Dynamics of a taut mooring line accounting for the embedded anchor chains. Ocean Engineering 121:403–413
30. Fossen TI, Strand JP (2001) Nonlinear passive weather optimal positioning control (wopc) system for ships and rigs: experimental results. Automatica 37(5):701 – 715
31. Fossen TI (2011) Handbook of marine craft hydrodynamics and motion control. John Wiley & Sons
32. Arditti F, Souza F, Martins T, Tannuri E (2015) Thrust allocation algorithm with efficiency function dependent on the azimuth angle of the actuators. Ocean Engineering 105:206 – 216
33. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller MA, Fidjeland AK, Ostrovski G, et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533
34. Silver D, Sutton RS, Müller M (2008) Sample-based learning and search with permanent and transient memories. In: Proceedings of the 25th international conference on Machine learning, ACM, pp 968–975