Zheng, Yuemin; Tao, Jin; Sun, Qinglin; Sun, Hao; Sun, Mingwei; Chen, Zengqiang

# An intelligent course keeping active disturbance rejection controller based on double deep Q-network for towing system of unpowered cylindrical drilling platform

*Published in:*
International Journal of Robust and Nonlinear Control

*DOI:*
10.1002/rnc.5740

Published: 25/11/2021

*Document Version*
Publisher's PDF, also known as Version of record

Received: 27 February 2021 | Accepted: 2 August 2021

DOI: 10.1002/rnc.5740

RESEARCH ARTICLE

WILEY

# An intelligent course keeping active disturbance rejection controller based on double deep Q-network for towing system of unpowered cylindrical drilling platform

**Yuemin Zheng**[1] | **Jin Tao**[1,2] | **Qinglin Sun**[1] | **Hao Sun**[1] | **Mingwei Sun**[1] | **Zengqiang Chen**[1,3]

[1]College of Artificial Intelligence, Nankai University, Tianjin, China

[2]Department of Electrical Engineering and Automation, Aalto University, Espoo, China

[3]Key Laboratory of Intelligent Robotics of Tianjin, Tianjin, China

**Correspondence**
Jin Tao, College of Artificial Intelligence, Nankai University, Tianjin, China.
Email: taoj@nankai.edu.cn

**Funding information**
Academy of Finland, Grant/Award Number: 315660; China Postdoctoral Science Foundation, Grant/Award Number: 2020M670633; Key Technologies Research and Development Program of Tianjin, Grant/Award Number: 19JCZDJC32800; National Key Research and Development Project, Grant/Award Number: 2019YFC1510900; National Natural Science Foundation of China, Grant/Award Numbers: 61973172, 61973175, 62003175, 62003177

**Abstract**

Towing is a widely used mode of transportation in offshore engineering, and towing of unpowered platforms is of particular significance. However, the addition of unpowered facilities has increased the difficulty of ship maneuvering. Moreover, the marine environment is complex and changeable, and sudden winds or waves can have unpredictable effects on the towing process. Therefore, it is of great significance to overcome the influence of the harsh marine environment while navigating the towing system following a planned course to a target sea area. To tackle the time-varying disturbances, a course control method for a towing system of unpowered cylindrical drilling platform is designed based on double deep Q-network (DQN) optimized linear active disturbance rejection control (LADRC). To be specific, to tackle the difficulty of LADRC tuning, double DQN is applied to select the best parameters of the LADRC at any time according to the states of the system, without relying on the specific information of the model and the controller. The course control performance of the towing system is evaluated in a simulation environment under various disturbances. Moreover, the Monte Carlo experiment is used to test the robustness of the controller when the ship's mass changes and the robustness of the proposed method is verified by testing with various heading angles. The results show that the LADRC with adaptive parameters optimized by double DQN performs well under external interference and inherent uncertainty, and compared with the traditional LADRC, the proposed method has better course control effects.

**KEYWORDS**

double deep Q-network, linear active disturbance rejection control, reinforcement learning, towing system of unpowered cylindrical drilling platform

## 1 | INTRODUCTION

In recent years, with the development of the shipping industry and offshore engineering, the demand for towing businesses has been increasing. The towing system mainly consists of the tugboat, the towing line, and the towed vessel, and

its main application include two aspects. One is to tow non-self-propelled ships in a maritime accident to a safe area to prevent other ships from being blocked. The other one is to tow large structures such as offshore drilling platforms and floating docks to help the realization of marine resource extraction tasks.[1] In other words, the unpowered facility requires the assistance of tugboats to move to the target area. For a long time, towing operations have mainly relied on the experiences of the captain and pilot, which is challenging to set and control the towing course accurately and may increase the risk factor of towing navigation. For example, ships are affected by special weather such as sea wind, waves, or fog, and human experiences have severe limitations. Therefore, studying the course control of the towing operation system is of great significance for ensuring the safety of towing operations, reducing maritime towing transportation accidents, and preventing marine pollution.[2]

Currently, the research on ship maneuverability mainly focuses on single ships, while the study on the maneuverability of towing systems is relatively few. Moreover, the main difficulties faced by the maneuverability control of the towing system are the dynamic model is very complex; the navigation environment is challenging to predict and overcome; the actuators like the rudder are saturated, to name a few. As the dynamic model of the towing system is highly complex, the current related research mainly focused on its simplified model,[3] where the "separate" modeling method proposed by the manipulative modeling group (MMG)[4-6] was used commonly. For example, Bo Woo Nam[7] utilized a three degrees-of-freedom (DOF) maneuvering mathematical model to describe the nonlinear dynamics of the towed vessel in a calm sea, which involved the surge, swaying, and yaw motion of the ship, and the simulation results are directly compared with the test data of the model. Marco et al. additionally considered the rolling motion and established a four DOF towing system model.[8] Hongbo Sun et al. established the six DOF motion equations of tugs and barges and proposed a towline-barge coupled motion model.[9] This article mainly considers the motion control of the towing system in a plane. Therefore, a three DOF dynamic model is enough.

We all know that a controller needs to overcome the adverse effects of the environment, such as wind or waves during the navigation of the ship. As far as we know, there are not many studies on the motion control of the towing system, where the PID method is still widely used today. For example, Liang et al. used a PD controller to control the tug rudder, and their research showed that changing the PD parameters can effectively reduce the tug's heading angle oscillation amplitude.[10] Pang et al. applied fuzzy-PID controller to the towing system, thus achieving the planned path. However, PID can only produce control actions after the disturbance has an impact on the system.[11] Therefore, Han[12,13] proposed the active disturbance rejection control (ADRC), which has the characteristic of estimating unknown disturbance and eliminating it. However, ADRC has many parameters that are needed to be adjusted, so it is not conducive to engineering applications. Later, Dr. Gao proposed linear ADRC (LADRC) on the basis of ADRC, which significantly simplifies the design of the controller.[14,15] The LADRC does not depend on the model's accurate information, which uses the linear extended state observer (LESO) to estimate all unknown disturbances of the system and uses the PD control combination to obtain the control input so as to suppress the disturbances' influences acting on the system. And compared with other disturbance attenuation control methods, such as $H_\infty$ control and stochastic control, ADRC has more potential for industrial applications with the advantages of model-free and the design process is rather simple. In terms of motion control of the towing system, Tao et al.[16] used LADRC to achieve linear path tracking control, and the simulation results proved that the control performance of LADRC is better than PID. Besides, LADRC has also shown empirical results in other fields, such as hydraulic system,[17] unmanned aerial vehicle (UAV),[18] and power system.[19]

Parameter adjustment has always been a non-negligible part of the controller design process. Compared with ADRC, LADRC has been simplified significantly, but parameter adjustment is still challenging. Generally speaking, parameter adjustment methods are mainly divided into two categories. One is to use heuristic algorithms, such as genetic algorithm (GA),[20] particle swarm optimization (PSO),[21] and whale optimization algorithm (WOA)[22] to optimize a set of fixed parameters. The other is to adjust the controller parameters in real time, such as using fuzzy adaptive control.[23] However, the former can only get a set of parameters under a certain specific situation. When the environment suddenly changes, this set of parameters may not be able to obtain a good control effect. Moreover, the second type of method relies on model information or human experience. In order to make up for the deficiencies of these methods, this article uses reinforcement learning (RL) algorithms for controller parameter tuning.

RL is an algorithm that solves sequential decision problems, which can select the optimal action according to the reward value obtained from the interaction between the environment and the agent, thereby completing the final task. Due to this feature, RL has shown good results in robot path planning,[24] multiagents,[25] and computer games.[26] However, the application of RL in the traditional control field is still in the research stage. As a matter of fact, the adjustment of controller parameters can be regarded as a process of finding the optimal strategy. Therefore, it is theoretically feasible for the reinforcement learning algorithm to optimize the controller parameters. At present, scholars have used Q learning

of RL to optimize parameters of the LADRC. For example, Chen et al.[27] used LADRC to control the heading angle of the ship without considering dynamics and used Q learning to optimize the controller parameters. Zheng et al.[28] applied the Q-learning-optimized LADRC to the power system and got a good control effect. However, in practice, in the optimization process of Q learning, the system state must be discretized, which may generate a large state space and is inconvenient for the storage and calculation of the Q table. Based on Q learning, deep Q-network (DQN) is developed, which is no longer limited to state discretization. DQN combines Q learning with deep learning and uses a deep neural network to map the state-action value function of RL,[29] where the replay buffer is added to train the deep neural network, and a target network is designed to calculate the loss function. However, DQN would overestimate the action-state value function during the training process, which affects the final decision and fails to obtain the optimal strategy. To solve this problem, many scholars put forward improved algorithms such as double DQN,[30] dueling DQN,[31] and rainbow DQN.[32] As far as we know, the double DQN plays a role in tuning optimization of the parameters for LADRC in this article for the first time.

Aiming for the course control of a towing system with an unpowered cylindrical drilling platform, we first built a three DOF motion model, then apply the double DQN algorithm to adjust the LADRC parameters in real time for course control, finally carry out the towing system course control simulations. The contributions of this article can be summarized as:

(1) A three DOF motion equation for a towing system of an unpowered cylindrical drilling platform is established. The LADRC controller is designed to realize the course control despite the nonlinear characteristics and large inertia within the system as well as the disturbances in the marine environment.

(2) Double DQN is applied to optimize adaptive parameters of LADRC, in which a multilayer fully connected neural network is designed, and the error-based reward function is defined according to the system model.

(3) The robustness of the LADRC and double DQN has been verified.

The remainder of this article is organized as follows. Section 2 describes the mathematical model of the towing system with unpowered cylindrical drilling platform under environmental disturbances. Section 3 designs the double DQN optimized LADRC based course controller of the towing system. Simulation results are reported and discussed in Section 4. And Section 5 concludes the article.

## 2 | TOWING SYSTEM MODELING

The towing system studied in this article consists of a tugboat and a cylindrical drilling platform without self-propelled capability. The system modeling is based on the 3-DOF MMG model, and the motion between the tugboat and the drilling platform is only coupled by the streamer, which satisfies the catenary model. Considering the influence of the ship's fluid power, propeller thrust, rudder force, towing force, and their corresponding moments, we model the towing system as follows.

### 2.1 | Motion mathematical model of towing system

Since the heave motion, pitch motion, and roll motion of the ship have a relatively small influence on the course, and they are usually ignored when designing the ship course controller. The plane motion coordinate system of the towing system is shown in Figure 1. As we can see, $OXY$ is the inertial coordinate system, $O_1X_1Y_1$ is the towing coordinate system, and $O_2X_2Y_2$ is the towed coordinate system. And the movement of the tugboat and the unpowered platform is described by forward speed $u_1, u_2$, lateral drift speed $v_1, v_2$, and yaw angular velocity $r_1, r_2$ in the body-fixed frame.

Assuming the ship is a rigid body, the position coordinates and heading angle of the ship are expressed as:

$$\begin{cases} \dot{x}_i = u_i \cos\varphi_i - v_i \sin\varphi_i \\ \dot{y}_i = u_i \sin\varphi_i + v_i \cos\varphi_i \\ \dot{\varphi}_i = r_i \end{cases}, \tag{1}$$

where $\varphi_i$ is the heading angle of the ship. It can be seen that the premise of obtaining the position and heading angle of the tug and the drilling platform is to know their respective speeds $u_i, v_i$ and steering angle speeds $r_i$.

According to the 3-DOF MMG model, the motion equations of the tugboat and the cylindrical drilling platform in the towing system can be expressed as:

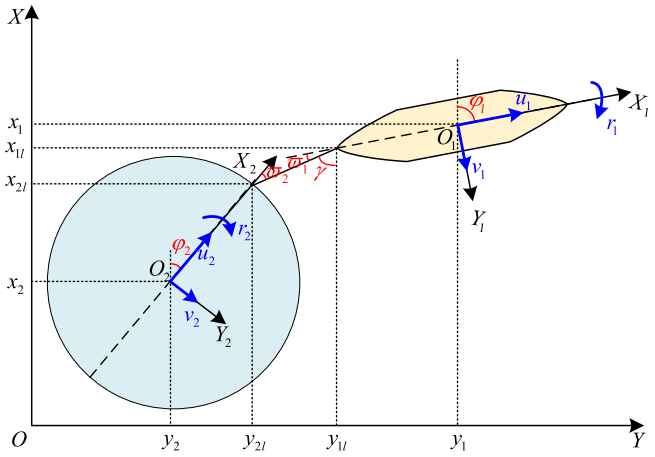



**FIGURE 1** Schematic diagram of towing system coordinates

$$\begin{cases} \left(m_i + m_{xi}\right)\dot{u}_i - \left(m_i + m_{yi}\right)v_i r_i = X_{Hi} + X_{Pi} + X_{Ri} + X_{Ti} + X_{Ei} \\ \left(m_i + m_{yi}\right)\dot{v}_i + \left(m_i + m_{xi}\right)u_i r_i = Y_{Hi} + Y_{Pi} + Y_{Ri} + Y_{Ti} + Y_{Ei} \\ \left(I_{zzi} + J_{zzi}\right)\dot{r}_i = N_{Hi} + N_{Pi} + N_{Ri} + N_{Ti} + N_{Ei} \end{cases}, \tag{2}$$

where $m_i$ is the mass of the corresponding ship. The expression of ship's additional mass $m_{xi}$, $m_{yi}$, and additional moment of inertia are shown in Equation (4). In addition, $X$, $Y$ on the right side of the equation represent the forces acting on the $x$ and $y$ axes, respectively, and $N$ represents the moment acting on the ship. The subscripts $H$, $P$, $R$, $T$, and $E$ denote hull hydrodynamic force, propeller force, rudder force, rope pulling force, and external environmental forces, respectively. It should be noted that the towed platform is only driven by the pulling force of the towline, thus,

$$\begin{cases} X_{P2} = Y_{P2} = N_{P2} = 0 \\ X_{R2} = Y_{R2} = N_{R2} = 0 \end{cases}. \tag{3}$$

The additional inertial mass and the additional moment of inertia are expressed as:

$$\begin{cases} m_{xi} = 0.01 m_i \begin{bmatrix} 0.398 + 11.97 C_{bi}\left(1 + 3.73 d_i/B_i\right) - 2.89 C_{bi} L_i/B_i \left(1 + 1.13 d_i/B_i\right) + \\ 0.175 C_{bi} (L_i/B_i)^2 \left(1 + 0.541 d_i/B_i\right) - 1.107 \left(L_i/B_i\right)\left(d_i/B_i\right) \end{bmatrix} \\ m_{yi} = m_i \begin{bmatrix} 0.882 - 0.54 C_{bi}\left(1 - 1.6 d_i/B_i\right) - 0.156 \left(1 - 0.673 C_{bi}\right) L_i/B_i + \\ 0.826 \left(d_i/B_i\right)\left(L_i/B_i\right)\left(1 - 0.678 d_i/B_i\right) - \\ 0.638 C_{bi} \left(d_i/B_i\right)\left(L_i/B_i\right)\left(1 - 0.669 d_i/B_i\right) \end{bmatrix} \\ J_{zzi} = 0.01 L_i^2 \left[33 - 76.85 C_{bi}\left(1 - 0.874 C_{bi}\right) + 3.43 L_i/B_i \left(1 - 0.63 C_{bi}\right)\right] \end{cases}, \tag{4}$$

where $C_{bi}$ and $d$ are the square coefficient and average draft of the ship, respectively. $B$ and $L$ represent the width and length of the ship.

## 2.2 | Force and moment of the towing system

### 2.2.1 | Hull hydrodynamics

In this article, the Kishima model[33] is used to estimate the viscous fluid dynamics and moments, which can be expressed by:

$$\begin{cases} X_{Hi} = X_i(u) + X_{vvi}v_i^2 + X_{vri}v_i r_i + X_{rri}r_i^2 + X_{vvvvi}v_i^4 \\ Y_{Hi} = Y_{vi}v_i + Y_{ri}r_i + Y_{|v|vi}|v_i|v_i + Y_{|r|ri}|r_i|r_i + Y_{vvri}v_i^2 r_i + Y_{vrri}v_i r_i^2 \\ N_{Hi} = N_{vi}v_i + N_{ri}r_i + N_{|v|vi}|v_i|v_i + N_{|r|ri}|r_i|r_i + N_{vvri}v_i^2 r_i + N_{vrri}v_i r_i^2 \end{cases}, \tag{5}$$

where $X_i(u)$ is the hull resistance of the ship during direct voyage. On the right side of the equation, $X_{vvi}, X_{vri}, X_{rri}, X_{vvvvi}$ are the longitudinal nonlinear hydrodynamic derivatives, $Y_{vi}, Y_{ri}, Y_{|v|vi}, Y_{|r|ri}, Y_{vvri}, Y_{vrri}$ are the lateral linear and nonlinear hydrodynamic derivatives of the ship, respectively, and $N_{vi}, N_{ri}, N_{|v|vi}, N_{|r|ri}, N_{vvri}, N_{vrri}$ are rotational linear and nonlinear hydrodynamic derivatives.

### 2.2.2 | Hydrodynamic estimation of tugboat propeller and rudder

In the MMG model,[34] the longitudinal thrust $X_{P1}$ generated by the propeller needed to be calculated:

$$\begin{cases} X_{P1} = \left(1 - t_p\right)\rho n^2 D_P^4 K_T(J_P) \\ Y_{P1} = 0 \\ N_{P1} = 0 \end{cases}, \tag{6}$$

where $t_p$ refers to the number of the propeller thrust deratings. $\rho$, $n$, and $D_p$ are the sea water density, the speed, and diameter of the propeller. $K_T$ and $J_P$ represent the propeller thrust coefficient and advance speed coefficient, respectively.

The calculation of the tugboat rudder force can be estimated as[34]:

$$\begin{cases} X_{R1} = -(1 - t_R)F_N \sin\delta \\ Y_{R1} = -(1 + a_H)F_N \cos\delta \\ N_{R1} = -(x_R + a_H x_H)F_N \cos\delta \end{cases}, \tag{7}$$

where $\delta$ is the actual rudder angle of the tugboat. $t_R$ represents the derating of rudder resistance. $a_H$ denotes the coefficient of the influence of the steering on the lateral force of the hull. $x_R$ and $x_H$ denote the longitudinal coordinate of the rudder center and the distance from the center of the lateral force of the hull to the center of gravity of the ship, respectively. And $F_N$ is the rudder normal force.

### 2.2.3 | Modeling of the towline

In this article, the catenary model is used to establish the towline model, and the towline force is decomposed into the motion coordinate system of the tugboat and the towed cylindrical drilling platform:

$$\begin{cases} X_{T1} = -(T_H + R_t)\cos\varpi_1 \\ Y_{T1} = -(T_H + R_t)\sin\varpi_1 \\ N_{T1} = x_{p1}Y_{T1} \\ X_{T2} = T_H\cos\varpi_2 \\ Y_{T2} = T_H\sin\varpi_2 \\ N_{T2} = x_{p2}Y_{T2} \end{cases}, \tag{8}$$

where $T_H$ and $R_t$ are horizontal towing cable tension and horizontal resistance of towing cable at the towing point, respectively. $x_{pi}$ denotes the longitudinal distance from the towing point to the center of gravity of the respective ship. And $\varpi_i$ can be derived from Figure 1 .

### 2.2.4 | Disturbance dynamic model

The marine environment dramatically influences the maneuverability of ships, and many shipwrecks occur in harsh seas. Under the circumstances, how to overcome the impact of wind and waves is of great significance. Since the data required for accurate winds or wave models are challenging to obtain, disturbances are generally established based on wind direction angle (wave direction) or wind speed (wave speed). In order to verify the effective anti-interference performance of the proposed method through simulation, this article models winds and waves as follows.

*Wave:* Assuming the relative flow speed is $U_c$, and the relative flow direction angle is $\alpha_c$, then the relative speed of ship motion is:

$$\begin{cases} u_{ri} = u_i + U_c \cos(\alpha_c - \varphi_i) \\ v_{ri} = v_i + U_c \sin(\alpha_c - \varphi_i) \end{cases}. \tag{9}$$

In this case, Equations (1) and (2) can be estimated as:

$$\begin{cases} \dot{x}_i = u_i \cos\varphi_i - v_i \sin\varphi_i + U_c \cos\alpha_c \\ \dot{y}_i = u_i \sin\varphi_i + v_i \cos\varphi_i + U_c \sin\alpha_c \\ \dot{\varphi}_i = r_i \\ (m_i + m_{xi})\dot{u}_i - (m_i + m_{yi}) v_i r_i = X_{Hi} + X_{Pi} + X_{Ri} + X_{Ti} + (m_i + m_{xi}) U_c r_i \sin(\alpha_c - \varphi_i) \\ (m_i + m_{yi})\dot{v}_i + (m_i + m_{xi}) u_i r_i = Y_{Hi} + Y_{Pi} + Y_{Ri} + Y_{Ti} - (m_i + m_{yi}) U_c r_i \cos(\alpha_c - \varphi_i) \\ (I_{zzi} + J_{zzi})\dot{r}_i = N_{Hi} + N_{Pi} + N_{Ri} + N_{Ti} + 0 \end{cases}. \tag{10}$$

*Wind:* For wind with a constant direction and speed, the model is generally established as

$$\begin{cases} X_{\text{windi}} = -\frac{1}{2}\rho_a A_f \left(u_{rwi}^2 + v_{rwi}^2\right) C_{wx}(\alpha_w) \\ Y_{\text{windi}} = \frac{1}{2}\rho_a A_s \left(u_{rwi}^2 + v_{rwi}^2\right) C_{wy}(\alpha_w) \\ N_{\text{windi}} = \frac{1}{2}\rho_a A_s L \left(u_{rwi}^2 + v_{rwi}^2\right) C_{wn}(\alpha_w) \end{cases}, \tag{11}$$

where $\rho_a$ is the air density. $A_f$ and $A_s$ are the orthographic area and the side projection area above the ship's waterline. $u_{rwi}$ and $v_{rwi}$ denote the relative speed. $C_{wx}$, $C_{wy}$, and $C_{wn}$ are the wind coefficient on the $x$-axis, the $y$-axis, and the wind moment coefficient around the $z$-axis, respectively.

It is worth noting that there are mismatched disturbances and uncertainties in the towing system, such as unmodeled dynamics, external winds and waves, and parameter perturbations. Therefore, it is very important to eliminate or suppress the influence of these disturbances and uncertainties on the system. Moreover, since the drilling platform itself has no driving force, this also increases the challenge of controlling the towing system.

## 3 | DOUBLE DQN OPTIMIZED LADRC

The LADRC has significant advantages in suppressing the influence of disturbances on the system. They can also estimate and compensate for disturbances without knowing the specific model of the system. Therefore, this article applies the LADRC to the course control for the cylindrical drilling platform's towing system. The purpose of the controller is to obtain a suitable $\delta_1$, so that the heading angles of the towing system $\varphi_1$ and $\varphi_2$ could follow the planned course.

### 3.1 | LADRC controller design

Because only the tugboat involves the force of the rudder in the towing system, the design of the controller is only for the tugboat. According to Equations (1) and (2), the following equation can be derived:

$$\ddot{\varphi}_1 = \frac{1}{I_{zz1} + J_{zz1}} \left(N_{H1} + N_{P1} + N_{R1}\left(\delta_1\right) + N_{T1} + N_{E1}\right). \tag{12}$$

Deformation of the above formula can be written as:

$$\ddot{\varphi}_1 = b_0\delta_1 + f, \tag{13}$$

where $f$ can be regarded as the total disturbance containing both the model dynamics and external disturbances of the system and $b_0$ is the nominal input gain. A core idea in LADRC is to add a state to estimate the disturbance, which is defined as one of the following states:

$$x_1 = \varphi_1, \quad x_2 = \dot{\varphi}_1, \quad x_3 = f. \tag{14}$$

Thus, the following state space equation can be obtained:

$$\begin{cases} \dot{x} = Ax + B\delta_1 + Eh \\ y = Cx \end{cases}, \tag{15}$$

where $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ b_0 \\ 0 \end{bmatrix}$, $C = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T$, $E = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $h = \dot{f}$.

Then the extended state observer equation can be constructed as:

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + B\delta_1 + L\left(y - \hat{y}\right) \\ \hat{y} = C\hat{x} \end{cases}, \tag{16}$$

where $L = [\beta_1, \beta_2, \beta_3]^T$ is the observer gain. The value of $L$ has a great influence on the accuracy of states estimation. Gao[35] converted the adjustment of $L$ into the tuning of observer pole:

$$\begin{aligned} |sI - (A - LC)| &= s^3 + \beta_1 s^2 + \beta_2 s + \beta_3 \\ &= (s + \omega_o)^3. \end{aligned} \tag{17}$$

That is, take $\beta_1 = 3\omega_o, \beta_2 = 3\omega_o^2, \beta_3 = \omega_o^3$, where $-\omega_o$ is the observer pole.

Another important structure in LADRC is the PD control combination, which plays the role of eliminating the influence of the estimated disturbance $\hat{x}_3$.

Take $\delta_1$ of Equation (13) as:

$$\delta_1 = u = \frac{-\hat{x}_3 + u_0}{b_0}. \tag{18}$$

Under the premise of $f \approx \hat{x}_3$, substituting Equation (18) into Equation (13) can get:

$$\ddot{\varphi}_1 \approx u_0. \tag{19}$$

Let

$$u_0 = k_p\left(\varphi_d - \hat{x}_1\right) + k_d\left(\dot{\varphi}_d - \hat{x}_2\right), \tag{20}$$

where $\varphi_d$ is the planned value of the heading angle $\varphi_1$. And $K = \left[k_p, k_d, 1\right]$ is the state feedback control gain. Similarly, the pole configuration method can be used to obtain:

$$\begin{aligned} |sI - BK| &= s\left(s^2 + k_d s + k_p\right) \\ &= s(s + \omega_c)^2, \end{aligned} \tag{21}$$

where $k_p = \omega_c^2, k_d = 2\omega_c$ are represented by the state feedback control pole $-\omega_c$.

To sum up, LADRC uses the LESO to estimate the total disturbance $f$, then suppressing the disturbance by the PD control combination. The parameters that need to be adjusted in the entire control system are: $\omega_o, \omega_c$, and $b_0$.

## 3.2 | Double DQN

### 3.2.1 | Main principle of double DQN optimization

RL is an intelligent algorithm that experts and scholars have favored in recent years. It has the characteristic that the optimal strategy can be obtained through the constant interaction between the agent and the environment without knowing the specific structure of the environment or the controlled system. Q learning of RL has a good optimization effect for systems with discrete actions and discrete states. However, in practice, the state of the system is often continuous, or the number of the states is very large. In this case, Q learning shows certain limitations: the dimension of the Q table is too large, resulting in a "dimensional disaster" or the Q table is difficult to converge. DQN is an algorithm developed based on the combination of Q learning and deep learning, where the Q table is expressed by a deep neural network so that the system state is no longer required to be discrete. Nevertheless, DQN is prone to over-fitting, so this article uses the double DQN algorithm to optimize the LADRC controller parameters.

The reward value $r$ is the most essential data in the process of interaction, through which the value of taking action $a$ in a certain state $s$ can be evaluated, and a numerical basis for the choice of actions are provided. Usually, we use cumulative rewards $R_c$ considering future reward values as reference data for final parameter selection. That is:

$$Q_\pi(s,a) = E_\pi\left[R_c \left| S_t = s, A_t = a\right.\right], \tag{22}$$

where $R_c = \sum_{t=0}^{\infty} \gamma^t r_{t+1}$. And $\gamma$ is the discount factor, which can reflect the importance of estimates of future reward values.

The larger the $Q$ value, the closer the corresponding $a$ is to the optimal value. In Q learning, each state–action pair corresponds to a $Q$ value, but in double DQN, the $Q$ value is replaced by a deep neural network shown in Figure 2. Assume that the system has $n$ state values to form a group of states and $m$ actions to form a group of state values at a certain time. There are a total of $j$ groups of actions. Then in the Q learning algorithm, the action-state value $Q\left(s_i, a_j\right)$ can be obtained through the $i$th group of states and the $j$th group of actions. Moreover, in double DQN, the $i$th group of the state is input to the deep neural network, and $Q$ values with $j$ quantity can be output, and each $Q$ value corresponds to a set of action values. It can be expressed as:

$$F\left(s_i, a_j; \theta\right) \approx Q\left(s_i, a_j\right), \tag{23}$$

where $\theta$ stands for the weights of neural network.

In other words, when the network is well trained, the network can output the $Q$ values corresponding to all the action values in a certain state, then the action value corresponding to the maximum $Q$ value is the optimal action-value that needs to be selected in this state. The schematic diagram of optimizing the parameters of LADRC through double DQN is shown in Figure 3.

### 3.2.2 | Working process of double DQN

As mentioned above, the ultimate goal of double DQN is to train a deep neural network to fit $Q$ values so that the optimal solution can be obtained through the network output. The process diagram of double DQN is shown in Figure 4. There are two deep neural networks with the same structure and different weights, where $Q$-network is used to estimate the $Q$ value and $\hat{Q}$-network is to get the $Q$ value of the next moment. And it is worth noting that $\hat{Q}$-network does not train neural network weights, which are assigned by $Q$-network every $T_n$ steps. That is to say, only $Q$-network needs to update the weights $\theta$ through training, and the training of the network involves the loss function like Equation. (24) and replay memory. The replay memory, also called the experience pool or replay buffer, is used to store the neural network data.

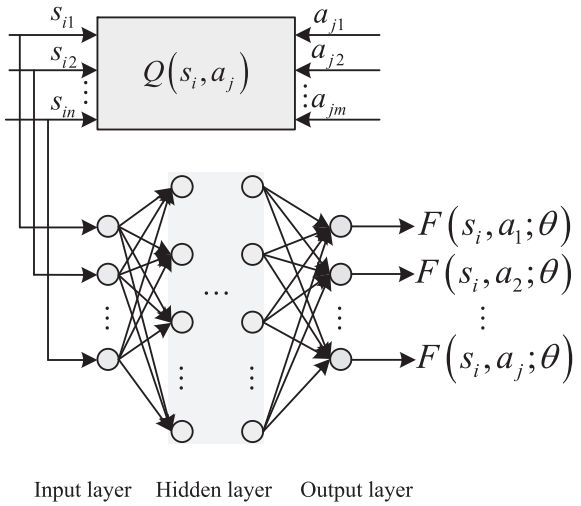



**FIGURE 2** Structure diagram of Q learning and double DQN

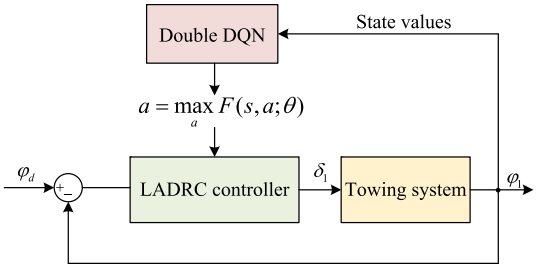



**FIGURE 3** Main structure of the control system

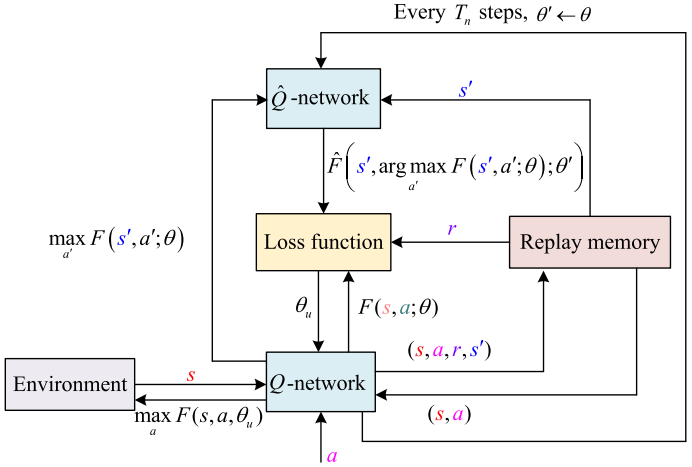



**FIGURE 4** Internal relationship structure diagram of double DQN

$$\text{Loss function} = \left( r + \gamma \hat{F}\left( s', \underset{a'}{argmax} F\left( s', a'; \theta \right); \theta' \right) - F\left( s, a; \theta \right) \right)^2, \tag{24}$$

where $r$ is the instant reward value obtained by performing an action in RL, and the design of $r$ in this article is shown in Equation (28).

According to the loss function, the updated weight value can be obtained by using the gradient descent method:

$$\theta_u = \theta + \alpha \left[ r + \gamma \hat{F}\left( s', \underset{a'}{argmax} F\left( s', a'; \theta \right); \theta' \right) - F\left( s, a; \theta \right) \right] \nabla F\left( s, a; \theta \right) \tag{25}$$

where $\alpha$ denotes the learning rate.

During the interaction between the agent and the environment, the selection of action $a$ in the current state $s$ is based on the $\varepsilon$-greedy policy:

$$\pi\left( a \left| s \right. \right) \leftarrow \begin{cases} 1 - \varepsilon + \frac{\varepsilon}{|A(s)|}, & \text{if } a = \underset{a}{argmax} Q\left( s, a \right) \\ \frac{\varepsilon}{|A(s)|}, & \text{otherwise} \end{cases}. \tag{26}$$

Generally speaking, the meaning of the above expression is: choose the action that maximizes the $Q$ value with the probability of $1 - \varepsilon$, otherwise, choose an action randomly. Similarly, it can be seen from Equation (24) that when the state is $s'$ at the next moment, the corresponding action $a'$ is the action value that maximizes $F\left( s', a'; \theta \right)$.

## 3.3 | Double DQN based LADRC parameters optimization

In this article, double DQN is used to solve the problem of adaptive parameter tuning of LADRC in the presence of uncertainties. Therefore, the towing system, including the LADRC, is regarded as the environment. Then the agent, after interacting with the environment, the agent can then obtain the state values and reward values so that the deep neural network of double DQN can be trained. Before that, we need to preprocess the states and actions for the environment.

### 3.3.1 | Define of states and discretization of actions

The state is the direct characteristic expression of the towing system, and it should be able to reflect the movement trend of the system. In this article, the state at time $k$ is defined by the error and the derivative of the error:

$$\begin{cases} s_1\left( k \right) = \varphi_d - \varphi_1\left( k \right) \\ s_2\left( k \right) = s_1\left( k \right) - s_1\left( k - 1 \right) \end{cases}. \tag{27}$$

As can be seen from Equation (27), at each sampling moment, a set of state vectors $s = [s_1, s_2]$ will be generated. Then the number of neurons in the input layer of the neural network in Figure 2 can be determined as two. Based on the state, the reward function can be designed. The reward function should encourage the agent to adopt the optimal parameters, which are reflected in the system's state. The closer the system is to the target state, the greater the reward value should be. Regarding the state, we can get the following law:

(1) The smaller the $s_1$, the closer the system is to the planned course;

(2) When $s_1 \cdot s_2 \le 0$, the towing system is approaching to the target heading angle; otherwise, it is far away.

Therefore, the instant reward function is designed as:

$$r\left( k \right) = -|s_1| + 2 * |\text{sign}\left( s_1 s_2 | \right) - 10 * |\text{sign}\left( s_1 s_2 \right)|. \tag{28}$$

The above formula shows that the closer the heading angle error is to 0, the greater the reward function is. Moreover, to avoid excessive overshoot in the system, we use the sign function to generate a certain reward and punishment signal.

As mentioned earlier, the parameters $\omega_o, \omega_c$, and $b_0$ that the controller needs to optimize are the action values in double DQN. Studies have pointed out that when $b_0$ is within a certain range,[36] the stability and convergence of LESO can be guaranteed. Therefore, $\omega_o$ and $\omega_c$ are mainly adjusted in this article. In addition, double DQN can only handle discrete action spaces, so we perform the following discretization:

$$\begin{cases} \omega_o \in \{\omega_{o\min}, \omega_{o\min} + h_1, \ \dots \ , \omega_{o\max}\} \\ \omega_c \in \{\omega_{c\min}, \omega_{c\min} + h_2, \ \dots \ , \omega_{c\max}\} \end{cases}, \tag{29}$$

where $h_1$ and $h_2$ are sampling intervals. In other words, $\omega_o$ and $\omega_c$ are divided into action spaces with elements numbers of $n_1 = \frac{\omega_{o\max} - \omega_{o\min}}{h_1}$ and $n_2 = \frac{\omega_{c\max} - \omega_{c\min}}{h_2}$, respectively. Therefore, there are a total of $n_1 * n_2$ action vectors to form the final action space. Then the number of neurons in the output layer of the neural network in Figure 2 can be determined as $n_1 * n_2$.

### 3.3.2 | Parameter tuning

In order to help the reader understand the whole process more clearly, the flowchart is given in Figure 5.

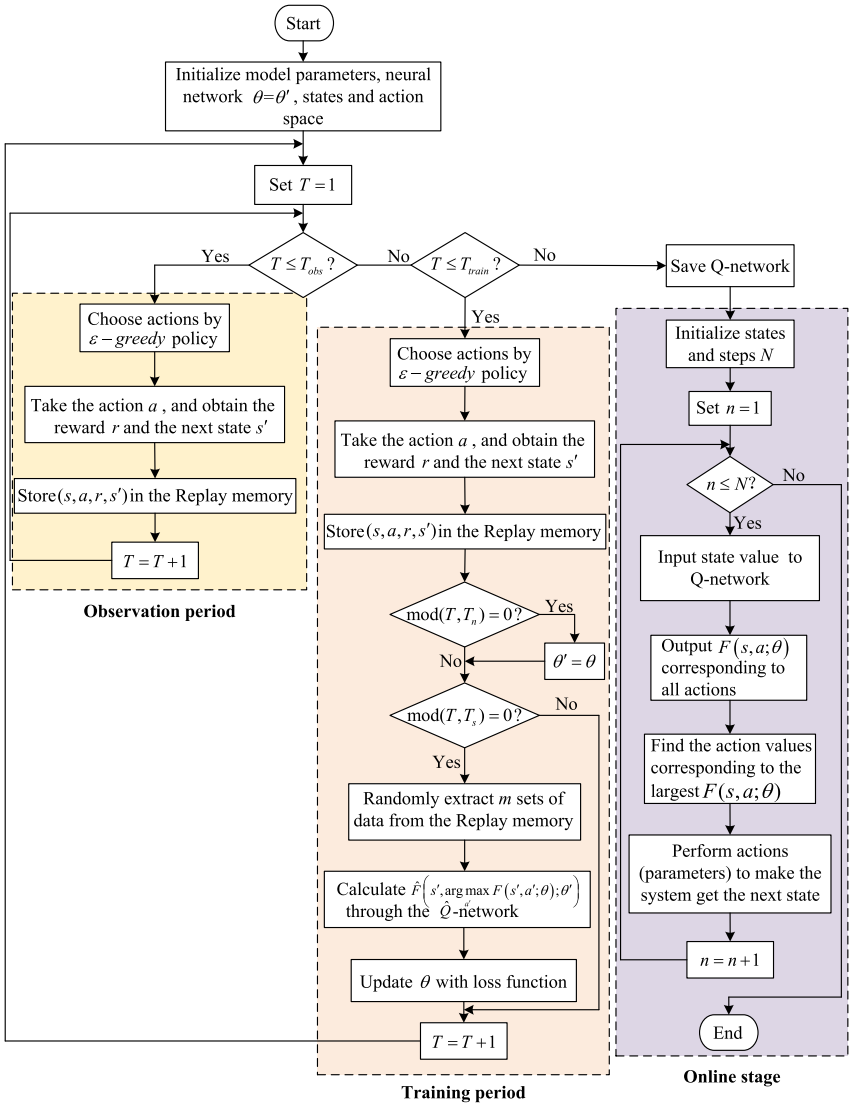



**FIGURE 5** Flowchart of double DQN optimization

**TABLE 1** Main dimensions of tugboat model

| Parameters | Displacement | Length | Width | Square factor | Draught |
|---|---|---|---|---|---|
| Values | 4522*t* | 63.6 m | 16.4 m | 0.692 | 6.22 m |
| **Parameters** | **Propeller diameter** | **Propeller pitch** | **Rudder area** | **Rudder height** | **Aspect ratio** |
| Values | 5 m | 5.2 m | 15 $m^2$ | 5 m | 1.7 |

**TABLE 2** Main dimensions of cylindrical drilling platform

| Parameters | Displacement | Diameter | Square factor | Draught |
|---|---|---|---|---|
| Values | 33,000*t* | 84 m | 0.7854 | 6.4 m |

**TABLE 3** Main dimensions of towline

| Parameters | Cable diameter | Towline length | Cable weight | Young's modulus |
|---|---|---|---|---|
| Values | 54.6 mm | 500 m | 10.04 kg/m | $9.2 \times 10^8$ N/$mm^2$ |

As it can be seen in Figure 5, the whole process is divided into three stages. The first is the observation period, which is mainly for obtaining the required data; the second is the training period, which trains the neural network weights according to the data; and the last is the online phase, which is to find the optimal action values at each moment by the well-trained neural network.

In general, the double DQN is similar to an agent with self-learning ability. It learns some rules through data, so that it can take optimal actions in different states. Therefore, the proposed method is of great significance for the towing system with a changeable and unknown environment.

## 4 | SIMULATION EXAMPLES

In this section, simulation experiments on the towing system with a specific cylindrical drilling platform are carried out, where the towed platform has no self-propelled capability. And the relevant parameters of the towing system are shown in Tables 1–3.

The course of the towing system is driven by the tugboat's rudder angle; therefore, the purpose of the control is to obtain the suitable rudder angle $\delta_1$, so that the course of the towing system could follow the prescribed value $\varphi_d$, which is softened according to Equation (27). In addition, for safety reasons, the actual rudder angle is subject to certain restrictions $|\delta_1| \leq 35°$.

$$\frac{\varphi_r}{\varphi_d} = \frac{0.046^2}{s^2 + 2 \times 0.95 \times 0.046 s + 0.046^2}, \tag{30}$$

where $\varphi_r$ is the planned heading angle after softening.

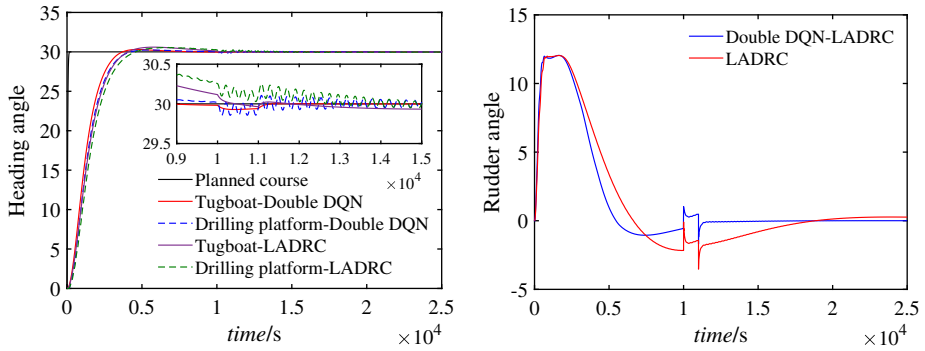
### 4.1 | Training of double DQN

The double DQN used in this article for adaptive adjustment is based on the manually obtained parameters. For the towing system, the action space is given in the format of Equation (31), where the parameters are selected as:

$$\begin{cases} \omega_{o\min} = 0.4, \omega_{o\max} = 0.9, h_1 = 0.001 \\ \omega_{c\min} = 0.001, \omega_{c\max} = 0.003, h_2 = 0.0001 \end{cases}. \tag{31}$$

**TABLE 4** Custom parameter values in double DQN

| Symbol meaning | Symbolic representation |
|---|---|
| Simulation time interval | $h = 1$ |
| Total simulation steps | $T = 25{,}000$ |
| Number of samples | $m = 84$ |
| Learning rate | $\alpha = 0.001$ |
| Discount factor | $\gamma = 0.99$ |
| Total step of observation period | $T_{\text{obs}} = 200$ |
| Total step of training period | $T_{\text{train}} = 100$ |
| Number of hidden layer neurons | $hl_1 = 20, hl_2 = 15, hl_3 = 8$ |




(A) Heading angle of the towing system (B) Rudder angle of the towing system

**FIGURE 6** Course control of towing system with wind disturbances

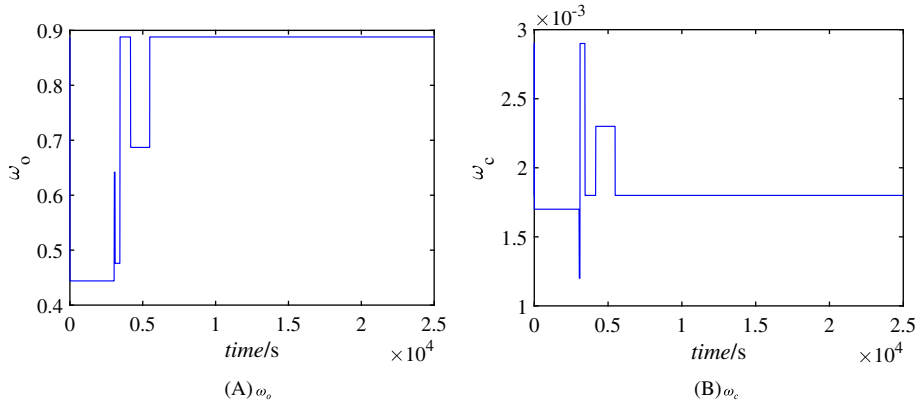
And $b_0$ is fixed to 0.00009. In addition, the parameters that need to be customized in the double DQN algorithm are given in Table 4.

The following will show the control effects of the proposed methods for the towing system under wind and wave disturbances, where the initial speed of the ships is taken as: $u_i = 6$ m/s, $v_i = 0$ m/s, $r_i = 0$ m/s. Moreover, the initial heading angles of the tugboat and the cylindrical drilling platform are both set to 0°.

## 4.2 | Course control of towing system with wind disturbances

In order to show the effectiveness of the proposed method, this article compares the control effect of LADRC with constant parameters and adaptive parameters optimized by double DQN. Assume that there is the wind with speed of 5.5 m/s and wind direction of 60° at $t$ =10,000–11,000 s. And the constant parameters are selected from the action space as $\omega_o = 0.4, \omega_c = 0.0015, b_0 = 0.00009$. The simulation results are shown in Figures 6 and 7.

It can be seen from Figure 6 that both the tugboat and the cylindrical drilling platform can finally reach the planned heading angle. On the one hand, compared with the LADRC with fixed parameters, the proposed double DQN-LADRC can reach the set value with smaller overshoot and undershoot and shorter settling time. This conclusion is also verified in Table 5. On the other hand, Figure 7 gives the adaptive parameters obtained by double DQN, which are selected according to the system states defined in Equation (27). As for the wind disturbance, the partial enlarged view in Figure 6 shows the response curves of the towing system when it is subjected to wind disturbance, in which the training parameters have not changed because the state values have not changed much. It can also be seen that the system has a longer response time, which is affected by the speed of the ship on the one hand, and is determined by the characteristics of the towing system itself on the other hand. Actually, without the drilling platform, the ship can quickly stabilize to the planned value. As for the entire towing system with unpowered cylindrical drilling platform, if the tugboat quickly stabilizes the planned value, the drilling platform will be difficult to stabilize.



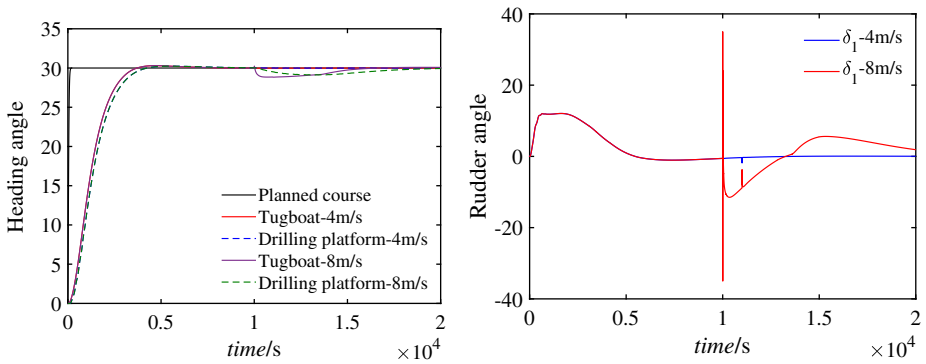

(A) $\omega_o$ (B) $\omega_c$

**FIGURE 7** Adaptive parameters optimized by double DQN

**TABLE 5** Comparison of performance indexes of towing system with wind disturbance

| Double DQN-LADRC | | Overshoot | Undershoot | $T_s$ | IAE·($10^4$) |
|---|---|---|---|---|---|
| | Tugboat | 0.01 | −0.0024 | 11,019 | 3.194 |
| | Drilling platform | 0.008 | −0.0048 | 14,417 | 4.358 |
| LADRC | | Overshoot | Undershoot | $T_s$ | IAE·($10^4$) |
| | Tugboat | 0.0256 | −0.0035 | 21,253 | 4.781 |
| | Drilling platform | 0.0241 | −0.003 | 24,538 | 5.237 |

*Note*: The adjustment time in the table refers to the shortest time required for the target value to reach and stay within ±0.05%. And IAE is the integral absolute error: IAE = $\int_0^\infty |e(t)|\, dt$



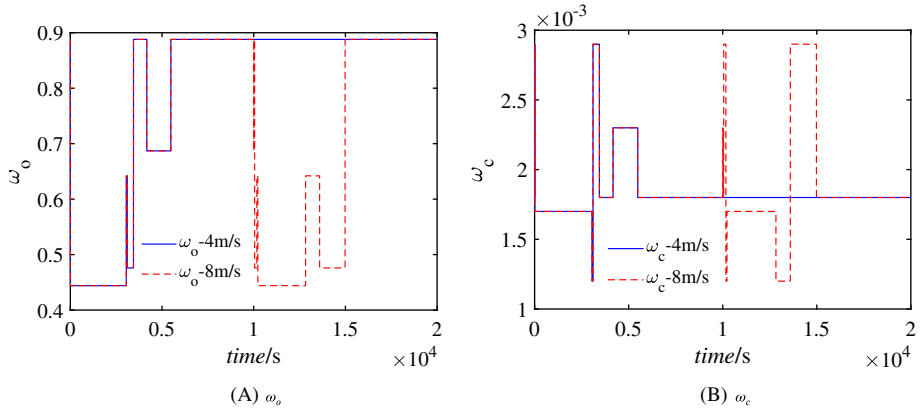

(A) Heading angle of the towing system (B) Rudder angle of the towing system

**FIGURE 8** Course control of the towing system under water flow disturbances with different water speed

## 4.3 | Course control of towing system with water flow disturbances

In this section, the towing system's simulation experiments with the influence of uniform water waves are carried out. Assume that the towing system is affected by water waves with a direction of 45° and a speed of 4 and 8 m/s, respectively, within a period of 10,000–11,000 s. Then the control results and parameters are illustrated in Figures 8 and 9.

It can be seen from Figure 8 that the greater the water flow velocity, the greater the disturbance to the towing system, and the more severe the navigational state changes. In addition, under the influence of larger disturbances, the agent adjusts the parameters according to the state values in the reinforcement learning, as shown in Figure 9. Comparing




(A) $\omega_o$ (B) $\omega_c$

**FIGURE 9** Adaptive parameters of the towing system under water flow disturbances with different water speed

**TABLE 6** Comparison of performance indexes of towing system with water wave disturbance

| 4 m/s | | Overshoot | Undershoot | $T_s$ | IAE·($10^4$) |
|---|---|---|---|---|---|
| | Tugboat | 0.01 | $-7.06 * 10^{-4}$ | 7604 | 3.915 |
| | Drilling platform | 0.008 | $-5.03 * 10^{-4}$ | 9094 | 4.331 |
| 8 m/s | | Overshoot | Undershoot | $T_s$ | IAE·($10^4$) |
| | Tugboat | 0.01 | −0.0389 | 19,585 | 4.382 |
| | Drilling platform | 0.008 | −0.0293 | 20,000 | 4.811 |

Figures 8 and 9, we can observe that in the case of wave disturbance of 8 m/s, the actual heading angle deviates greatly from the set value after 10,000 s. At this time, the parameters $\omega_o$ and $\omega_c$ have been adjusted in real time. Table 6 is the explanation of the performance index containing overshoot, undershoot, settling time $T_s$ and the IAE.

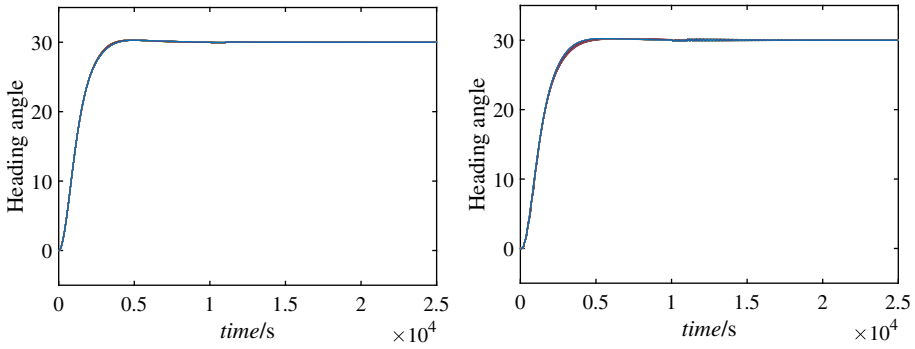
The above two experimental results show that, on the one hand, the proposed double DQN-LADRC overcomes the limitation that it is difficult to obtain the optimal results by manually adjusting the parameters, and it can find the optimal parameters by the reward function. The superiority of the proposed method is proved by comparison with the control effect of conventional LADRC. On the other hand, the proposed method can adjust the parameter value adaptively according to the defined system state value. Generally, the bionic algorithm can only get a set of optimal parameters under certain conditions, but when the system operating conditions change, the set of parameters cannot achieve the desired effect. The method proposed in this article only needs to initialize the system state randomly when training the agent, so that the agent can deal with more situations.

## 4.4 | Robustness test

As the main marine transportation, the mass of ships will inevitably undergo certain changes. The robustness is an important indicator to evaluate the performance of the proposed method.

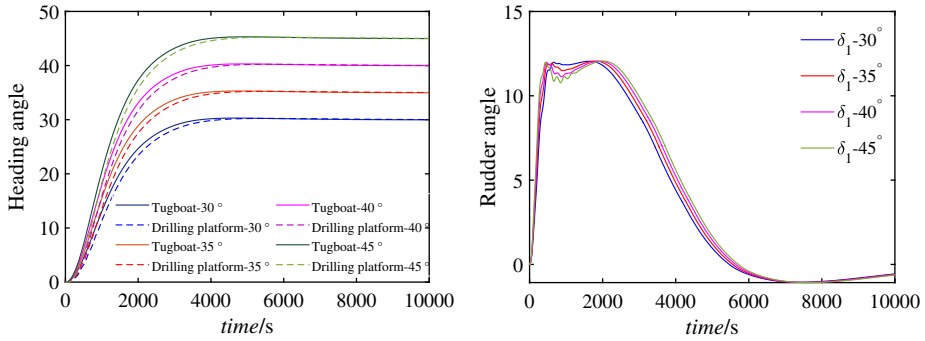
### 4.4.1 | System model parameter uncertainty

Take the towing system with wind disturbances as an example, keep the controller parameters shown in Figure 7 unchanged, and randomly change the mass of the tug and the drilling platform at the same time between ±30% of their masses for 50 times. The simulation results are shown in Figure 10. It can be seen that the heading angle of the tugboat and the cylindrical drilling platform can be stabilized to the planned value within this range, which proves the robustness of the controller.



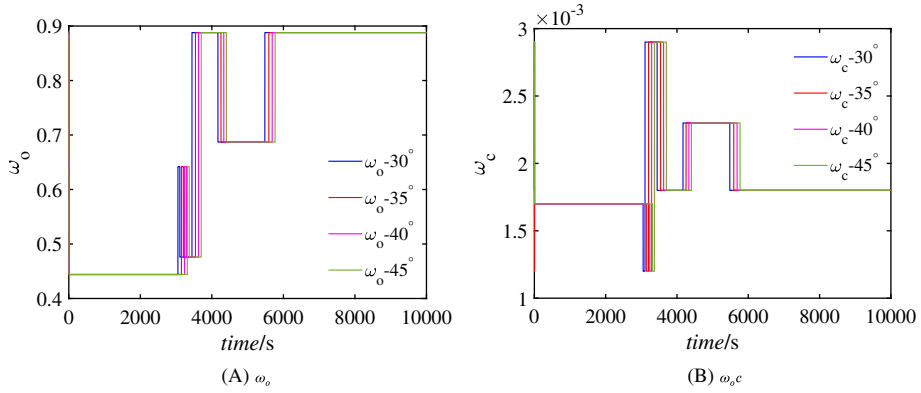

(A) Heading angle of tugboat (B) Heading angle of drilling platform

**FIGURE 10** Monte Carlo experiment of randomly changing the mass of tugboat and drilling platform




(A) Heading angle (B) Rudder angle

**FIGURE 11** Response curves of different target heading angles




(A) $\omega_o$ (B) $\omega_o c$

**FIGURE 12** Adaptive controller parameters corresponding to different target heading angles

### 4.4.2 | Target uncertainty

In the above experiment, the planned course is all 30°. In order to show that it cannot only reach this setting value, we develop the simulation under different target setting values, as shown in Figures 11 and 12. The following results are also a display of another perspective about adjusting parameters based on errors. Meanwhile, the robustness of the proposed method on different heading angles have been proved.

The results show that the controller parameters obtained by double DQN can still keep the course stable even though the tugboat and the platform's mass is changed, and the towing system with various planned heading

angles realizes adaptive control. That is to say, the designed double DQN optimized LADRC controller has good robustness.

## 5 | CONCLUSIONS

In this article, the course control for the towing system of unpowered cylindrical drilling platform based on double deep Q-network (DQN) optimized linear active disturbance rejection control (LADRC) was studied. Firstly, a three DOF dynamic model of the towing system under various disturbance conditions based on the MMG model and the catenary model was proposed. Then, the LADRC was designed for the tugboat, which controls the rudder angle of the tugboat to drive the towing system following the planned course. To determinate the parameters of LADRC, we applied the double DQN algorithm for the real-time tuning of its controller parameters. Unlike the bionic algorithms that get parameters offline or the fuzzy control algorithm that changes parameters online, the proposed method adopts the Markov idea of reinforcement learning so that the optimal parameters of the controller can be trained without determining the model and controller information. At last, we performed simulation experiments to validate the proposed method. The simulation results show that the double DQD-LADRC can make the towing system reach a better course control performance under the condition of wind disturbances compared with the LADRC with fixed parameters. Moreover, the simulation results under different velocities of water flow disturbances reflect the proposed method's ability to deal with disturbances. In addition, we verify the robustness of the proposed method through model parameter perturbation and different target values. At present, ship path tracking control is also a hot research issue, but the path tracking control of the towing system is still in urgent need of research, which will be considered in our future work.

### ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No.61973172, 61973175, 62003175, and 62003177), the National Key Research and Development Project (Grant No. 2019YFC1510900), the key Technologies Research and Development Program of Tianjin (Grant No.19JCZDJC32800), this project also funded by China Postdoctoral Science Foundation (Grant No.2020M670633), and Academy of Finland (Grant No.315660).

### CONFLICT OF INTEREST

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

### DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

### ORCID

*Yuemin Zheng* https://orcid.org/0000-0002-2815-092X
*Jin Tao* https://orcid.org/0000-0003-1066-1809
*Hao Sun* https://orcid.org/0000-0002-9730-0063
*Mingwei Sun* https://orcid.org/0000-0002-0974-6525
*Zengqiang Chen* https://orcid.org/0000-0002-1415-4073

### REFERENCES

1. Fitriadhy A, Yasukawa H. Course stability of a ship towing system. *Ship Technol Res*. 2011;58(1):4-23.
2. Fang M, Ju J. The dynamic simulations of the ship towing system in random waves. *Marine Technol SNAME News*. 2009;46(2):107-115.
3. Li O, Zhou Y. Precise trajectory tracking control of ship towing systems via a dynamiccal tracking target. *Mathematics*. 2021;9:974.
4. Yoshimura Y. Mathematical model for manoeuvring ship motion (MMG Model). Proceedings of the Mathematical Models for Operations involving Ship-Ship Interaction; August, 2005; Tokyo.
5. Zhang Q, Zhang X, Im N. Ship nonlinear-feedback course keeping algorithm based on MMG model driven by bipolar sigmoid function for berthing. *Int J Nav Archit Ocean Eng*. 2017;9(5):525-536.
6. Guo H, Zou Z. System-based investigation on 4-DOF ship maneuvering with hydrodynamic derivatives determined by RANS simulation of captive model tests. *Appl Ocean Res*. 2017;68:11-25.
7. Woo NB. Numerical investigation on nonlinear dynamic responses of a towed vessel in calm water. *J Marine Sci Eng*. 2020;8(3):219.

8. Sinibaldi M, Bulian G. Towing simulation in wind through a nonlinear 4-DOF model: bifurcation analysis and occurrence of fishtailing. *Ocean Eng*. 2014;88:366-392.
9. Sun H, Chen G, Lin W. A hydrodynamic model of bridle towed systems. *J Mar Sci Technol*. 2018;24:200-207.
10. Liang K, Deng D, Huang G. A study of towing system maneuvering motion simulation. *Navigat China*. 2007;71(2):10-13,29.
11. Pang S, Liu J, Yu S, Mao L, Yi H. Track-keeping and positioning control in vessel towing operation. Proceedings of the 27th International Ocean and Polar Engineering Conference; June 25-30, 2017; San Francisco, CA.
12. Han J. Auto-disturbance-rejection controller and its applications. *Control Decis*. 1998;13(1):19-23.
13. Han J. From PID to active disturbance rejection control. *IEEE Trans Ind Electron*. 2003;56(3):900-906.
14. Gao Z. On the foundation of active disturbance rejection control. *Control Theory Appl*. 2013;30(12):1498-1510.
15. Gao Z. On the centrality of disturbance rejection in automatic control. *ISA Trans*. 2014;53(4):850-857.
16. Tao J, Du L, Dehmer M, Wen Y, Xie G. Path following control for towing system of cylindrical drilling platform in presence of disturbance and uncertainties. *ISA Trans*. 2019;95:185-193.
17. Zhuang H, Sun Q, Jiang Y, Chen Z. Back-stepping sliding mode control for pressure regulation of oxygen mask based on an extended state observer. *Automatica*. 2020;119:109106.
18. Sun H, Sun Q, Wu W, Chen Z, Tao J. Altitude control for flexible wing unmanned aerial vehicle based on active disturbance rejection control and feedforward compensation. *Int J Robust Nonlinear Control*. 2020;30:222-245.
19. Huang Z, Chen Z, Zheng Y, Sun M, Sun Q. Optimal design of load frequency active disturbance rejection control via double-chains quantum genetic algorithm. *Neural Comput Appl*. 2021;33:3325-3345.
20. Zhou X, Gao H, Zhao B, Zhao L. A GA-based parameters tuning method for an ADRC controller of ISP for aerial remote sensing applications. *ISA Trans*. 2018;81:318-328.
21. Xu B, Cheng Z, Zhang R, Gong C, Huang L. Pso optimization of ladrc for the stabilization of a quad-rotor. Proceedings of the 2020 12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA); February 28–29, 2020; Phuket, Thailand.
22. Yu Y, Wang H, Li N, Su Z, Wu J. Automatic carrier landing system based on active disturbance rejection control with a novel parameters optimizer. *Aerosp Sci Technol*. 2017;69:149-160.
23. Sun C, Liu M, Liu C, Feng X, Wu H. An industrial quadrotor UAV control method based on fuzzy adaptive linear active disturbance rejection control. *Electronics*. 2021;10(4):376.
24. He C, Wan Y, Gu Y, Lewis FL. Integral reinforcement learning-based approximate minimum time-energy path planning in an unknown environment. *Int J Robust Nonlinear Control*. 2020. https://doi.org/10.1002/rnc.5122
25. Xu Z, Ni H, Karimi HR, Zhang D. A Markovian jump system approach to consensus of heterogeneous multiagent systems with partially unknown and uncertain attack strategies. *Int J Robust Nonlinear Control*. 2020;30(7):3039-3053.
26. Li X, Lv Z, Wang S, Wei Z, Wu L. A reinforcement learning model based on temporal difference algorithm. *IEEE Access*. 2019;7:121922-121930.
27. Chen Z, Qin B, Sun M, Sun Q. Q-Learning-based parameters adaptive algorithm for active disturbance rejection control and its application to ship course control. *Neurocomputing*. 2020;408:51-63.
28. Zheng Y, Chen Z, Huang Z, Sun M, Sun Q. Active disturbance rejection controller for multi-area interconnected power system based on reinforcement learning. *Neurocomputing*. 2021;425:149-159.
29. Mnih V, Kavukvuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG. Human-level control through deep reinforcement learning. *Nature*. 2015;518:518,529-518,533.
30. Pan J, Wang X, Cheng Y, Yu Q. Multisource transfer double DQN based on actor learning. *IEEE Trans Neural Netw Learn Syst*. 2018;29(6):2227-2238.
31. Meng W, Zheng Q, Yang L, Li P, Pan G. Qualitative measurements of policy discrepancy for return-based deep Q-Network. *IEEE Trans Neural Netw Learn Syst*. 2019;31(10):4374-4380.
32. Hessel M, Modayil J, Hasselt H, et al. Rainbow: combining improvements in deep reinforcement learning. Proceedings of the 32nd AAAI Conference on Artificial Intelligence; Vol. 32: February 2–7, 2018; New Orleans.
33. Wu X. *Ship Maneuverability and Newaveity*. China Communications Press; 1988.
34. Jia X, Yang Y. *Ship Motion Mathematical Model-Mechanism Modeling and Identification Modeling*. Dalian Maritime University Press; 1999.
35. Gao Z. Active disturbance rejection control: a paradigm shift in feedback control system design. Proceedings of the 2006 American Control Conference; June 14–16, 2006; Minneapolis, MN.
36. Xue W, Huang Y. Performance analysis of active disturbance rejection tracking control for a class of uncertain LTI systems. *ISA Trans*. 2015;58:133-154.


**How to cite this article:** Zheng Y, Tao J, Sun Q, Sun H, Sun M, Chen Z. An intelligent course keeping active disturbance rejection controller based on double deep Q-network for towing system of unpowered cylindrical drilling platform. *Int J Robust Nonlinear Control*. 2021;1–18. https://doi.org/10.1002/rnc.5740