Journal of Zhejiang University-SCIENCE A 2026 27(3):246-261
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.A2500146

Three-degree-of-freedom motion posture stabilization control of platform based on DTW-LSTM-MATD3 under high and low frequency disturbances of ships

Qin ZHANG, Jingyi ZHOU, Bangping GU, Xiong HU✉

School of Logistics Engineering, Shanghai Maritime University, Shanghai 201306, China

Abstract: In the complex and variable deep-sea environment, the compensation control of ship motion ensures the safety and efficiency of equipment installation and transportation in offshore wind farms. However, the ship motion posture compensation control system is severely affected by uncertainties, which significantly impact the accuracy of compensation control. In this paper, we propose a ship three-degree-of-freedom (3-DoF) motion posture stabilization control method based on the DTW-LSTM-MATD3 algorithm. We use the multi-agent twin delayed deep deterministic policy gradient (MATD3) to control a platform with six electric cylinders to achieve stable control. However, owing to random noise affecting the ship's motion posture, we use a dynamic time warping (DTW) algorithm to distinguish between high-frequency noise and low-frequency tracking signals. Further, we embed a long short-term memory (LSTM) network into the MATD3 network to better align the Critic network's training with the true Q-value. We use a combined reward function to enhance the agent's exploration capability in complex dynamic environments. Finally, verification was conducted under sixth-level, abrupt sea conditions with high-frequency noise, as well as under real abrupt sea conditions, and a generalization test was also carried out. Simulation results show that the proposed DTW-LSTM-MATD3 method has great compensation control ability.

Key words: Compensation control; Multi-agent twin delayed deep deterministic policy gradient (MATD3) algorithm; Dynamic time warping (DTW) algorithm; Long short-term memory (LSTM) network

1 Introduction

In recent years, with the increasing severity of global environmental issues, the demand for renewable energy in human society has grown extremely rapidly, and new energy sources are having a great impact on people's lives (Nathwani and Kammen, 2019). Wind power generation (Cauz et al., 2025), as a type of renewable energy, has attracted the attention of scholars and experts around the world. To stimulate innovation in the development and utilization of new energy, the Chinese government has advocated for the construction of offshore wind power bases to facilitate expansion into deep-water and offshore areas (Möllerström et al., 2025). Offshore wind power has the advantages of stable wind energy resources, no land occupation, and intense regenerative capacity (Li et al., 2022). Offshore wind power equipment operates in a more severe environment than on-shore equipment and faces challenges such as sudden changes in sea conditions and difficulties in maintenance (Kang et al., 2019). Specifically, owing to the interference of the marine environment, ships will experience random motion in six directions: roll, pitch, heave, sway, surge, and yaw, which affects offshore operations (Shao et al., 2022). The three-degree-of-freedom (3-DoF) motion of roll, pitch, and heave cause changes in the center of buoyancy of offshore operation platforms (Wang et al., 2025). In extreme cases, it may cause the platform to overturn, seriously affecting the safety and work efficiency of personnel on board. Therefore, this study focused on the compensation control of the 3-DoF motion of ships (Tang et al., 2023).

Traditional control methods include proportional-integral-derivative (PID) control, model predictive

✉ Xiong HU, huxiong@shmtu.edu.cn
Qin ZHANG, https://orcid.org/0000-0003-4733-6508

Received Apr. 26, 2025; Revision accepted Nov. 10, 2025;
Crosschecked Jan. 27, 2026; Online first Mar. 17, 2026

control (MPC), and linear quadratic regulator (LQR) systems. These control methods have been applied and improved in various fields (Woodacre et al., 2018; Yan et al., 2020; Winursito and Pratama, 2021). Zhang et al. (2020) proposed an adaptive self-tuning PID heading control scheme to solve the nonlinear control problem of unavailable time-varying environmental disturbances and parameter uncertainties in the ship heading control system. Their scheme reduces the impact of model parameters and position inputs on the control method. and effectively resists time-varying disturbances. Jimoh et al. (2021) proposed a disturbance observer-enhanced model predictive controller for the stability problem of ship roll motion. They used a speed model of ship motion to deal with external disturbances and formulated control inputs containing disturbance rates to attenuate the rate of change of disturbances caused by waves. Lv and Li (2023) used a strong fixed-time prescribed performance function (SFPPF) to ensure the prescribed performance with predefined convergence and proposed a dynamic inverse adaptive linear quadratic regulator (DI-ALQR) control strategy. Their approach realized the integrated control of ship position, roll, and pitch under the premise of balancing energy consumption and control accuracy, ensuring the prescribed performance. However, the effectiveness of these model-based traditional control methods is often limited when dealing with complex systems, presenting significant challenges in achieving satisfactory performance. When facing complex environments or systems, the limitations of traditional control methods will increase, affecting the accuracy of control.

Traditional reinforcement learning algorithms are usually studied based on single-agent scenarios, neglecting the impact of other complex motions. The ship motion control studied in this study involves multiple degrees of freedom, which can provide a more comprehensive analysis of a ship's dynamic characteristics, including the coupling between different degrees of freedom. Mou et al. (2021) studied the problem of covering 3D irregular terrain surfaces with hierarchical drone swarms and proposed a reinforcement learning algorithm based on swarm deep *Q*-learning network (SDQN). They designed an observation history model integrating convolutional neural networks (CNN) and mean embedding methods into SDQN to address the communication limitations of drones. Liu et al. (2022) proposed using a proximal policy optimization (PPO) algorithm to control the autonomous decision-making and coordinated operations of multiple drones, adopting the idea of centralized training and distributed implementation to enhance the decision-making capabilities of drones during operations. Wang and Zhao (2025) proposed a multi-agent deep deterministic policy gradient (MADDPG) algorithm with communication to handle the cooperation of multiple ships under partial observability. Agents established a unified communication protocol, sharing observations to compensate for missing information and achieve better coordinated control. Zhao et al. (2024) combined MADDPG with a long short-term memory (LSTM) network model to address the limitations in multi-agent navigation and obstacle avoidance. This algorithm can use more temporal observations as inputs for the policy network, improving the training efficiency of the algorithm. Qin et al. (2023) proposed and evaluated two approaches for multi-robot mobile airborne perception: a centralized soft actor-critic (SAC) method integrated with data augmentation, and a distributed multi-agent soft actor-critic (MASAC) algorithm, with the trajectory and power of the unmanned aircraft as the control objects to maximize the minimum weighted spectral efficiency. Experiments showed that both schemes are effective; SAC has better training speed and spectral efficiency, and MASAC has the best early training speed. Wu et al. (2025) proposed a computing offloading and energy optimization framework for reconfigurable intelligent surface (RIS), unmanned aerial vehicles (UAVs), and mobile edge computing (MEC), which uses the multi-agent twin delayed deep deterministic policy gradient (MATD3) algorithm to control task allocation, RIS phase shift, and UAV trajectories to maximize energy efficiency. This method is more adaptable to various scenarios and has higher energy efficiency than traditional algorithms. Zhou et al. (2024) proposed a new task decomposition MATD3 algorithm for the coordination and environmental uncertainty of multiple drones, decomposing the path planning task into two obstacle avoidance modules to guide drones to complete the overall planned path, and proposed a new reward function to enhance the algorithm's convergence. Hou et al. (2024) proposed a new online decision-making algorithm for MASAC to address the serious threat to maritime security posed by underwater vehicles. This method includes a

control-oriented framework for multi-agent reinforcement learning and combined curriculum learning to improve the success rate of multiple underwater vehicles during tracking.

Generally, reinforcement learning algorithms have a higher ability to perceive the environment than traditional control methods and do not rely on accurate models (Zhang et al., 2025), thereby achieving reasonable control effects. Among reinforcement learning algorithms, multi-agent reinforcement learning is an improvement based on single-agent reinforcement learning. Each agent enables cooperation and coordinates with others to complete complex tasks in the environment, improving the efficiency and accuracy of the overall system.

Most existing ship motion compensation control strategies struggle to balance rapid response and adaptability in complex and variable marine environments. Traditional control methods lack environmental perception capabilities, while conventional reinforcement learning methods often suffer from unstable training and operational oscillations. In this study, we developed a 3-DoF motion compensation control algorithm for ships under different sea conditions (Fig. S1 of the electronic supplementary materials (ESM)). The main contributions of this research are as follows:

1. A multi-agent reinforcement learning architecture based on dynamic time warping (DTW)-LSTM-MATD3 is proposed for ship 3-DoF motion compensation control. This architecture generates control actions through multi-agent collaboration, effectively reducing steady-state error and suppressing oscillations, thereby significantly improving control accuracy in dynamic environments.

2. A method integrating signal processing and network optimization is designed. This method introduces the DTW algorithm to distinguish high-frequency interference noise from low-frequency tracking signals; simultaneously, the LSTM network model is embedded into the Critic and Actor networks, enabling them to more accurately approximate the true Q-value, thereby enhancing the stability and convergence speed of policy learning.

3. A combined reward function integrating linear and normal rewards is constructed. This function enhances the agent's exploration capability in complex dynamic environments, thereby strengthening the generalization and robustness of the control strategy. The feasibility and effectiveness of the proposed method are verified through simulation experiments under various sea conditions.

2 Ship 3-DoF motion compensation system model

2.1 Kinematic analysis

To effectively analyze the performance of the proposed control method, we establish the dynamic model in this section, which includes the servo motor model and the electric cylinder model, and analyze the principle and solution process of the inverse kinematics of the parallel 3-DoF platform.

Under the influence of ocean waves, an offshore operation platform is affected to varying degrees, producing complex and uncontrollable motions that affect the safety and stability of offshore wind turbine loading and unloading operations. Therefore, the wave compensation system must have high accuracy. Before studying the control method, it is necessary to establish an accurate compensation system model (Fig. 1). The wave compensation system is composed mainly of a load platform, a base platform, six electric cylinders, Hooke hinges and ball hinges. The load platform is a hexagonal base supported by six electric cylinders. Each electric cylinder is composed of an upper leg and a lower leg. The electric cylinders are connected to the base platform and enable telescopic motion, thereby achieving 6 DoFs of motion, allowing the load platform to be precisely positioned and adjusted in space.

Before realizing the entire motion of the electric cylinder, it is necessary to use inverse kinematics to transmit the 3-DoF ship motion information to the computer and convert it into the telescopic displacement of each electric cylinder. To determine the motion information of the load platform, according to the multi-rigid-body kinematics principle of the spatial parallel mechanism, a coordinate system is established to describe the position and posture of the electric cylinder. Fig. 1a shows the 3-DoF platform mechanism. We establish the coordinate system in the load platform o_p-xyz and the base platform o_b-$x'y'z'$ of the parallel 6-DoF platform, and establish a moving coordinate system in the load platform and a fixed coordinate system in the base platform. In the initial state, the axis of the fixed coordinate system and the axis of

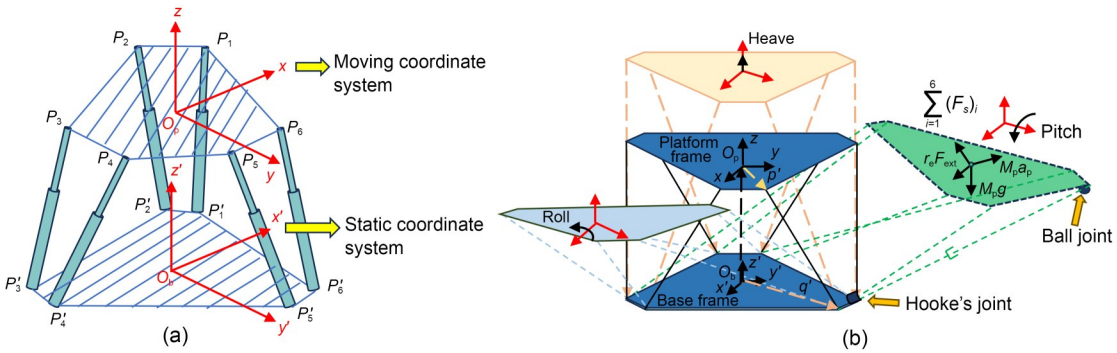

Fig. 1 Coordinate system of the parallel 3-DoF platform mechanism of the ship motion compensation system: (a) coordinate system of the parallel six-degree-of-freedom mechanism; (b) simulation of the motion of each electric cylinder of the Stewart platform under different force conditions

the moving coordinate system coincide. An analysis of the overall motion characteristics of the load platform is given in the electronic Section S2.1 of the ESM.

In this approach, we perform compensation for 3-DoF, and the load platform coordinates are simplified to $D_{\mathrm{L}}=(z,\varphi,\theta)$. The roll φ and pitch angles θ are substituted into the rotation matrix $\boldsymbol{r}$, that is, directly multiplied with $\boldsymbol{r}$, and then the z coordinate is added to or subtracted from the heave displacement to obtain the coordinates of the load platform in the fixed coordinate system. Finally, the length vector and displacement of the electric cylinder are calculated through inverse kinematics. The position of the load platform can be inversely solved from the pose D_{L} to find the extension lengths ΔL_i of the six electric cylinders, that is, by subtracting the initial length l_{i0} of the ith electric cylinder from the distance between the upper and lower Hooke hinge points:

$$\Delta L_i=\|\boldsymbol{l}_i\|_2-l_{i0}=\sqrt{(P_i'-B_i)^2}-l_{i0},\ i=1,2,\cdots,6, \quad (1)$$

where $\boldsymbol{l}_i$ represents the length of the electric cylinder in the inertial coordinate system, P_i' represents the coordinates of the connection point P_i on the load platform in the fixed coordinate system, B_i represents the coordinates of the connection point of the ith electric cylinder in the fixed coordinate system, and $\|\cdot\|_2$ represents the Euclidean norm of a vector.

2.2 Dynamic model

To achieve the telescopic motion of the six electric cylinders, it is necessary to analyze the dynamic principles of the compensation system. Fig. 1b illustrates the extension and retraction process of the six linear cylinders when the Stewart platform performs 3-DoF motion. The ship motion compensation system consists of mechanical and electrical control parts. The mechanical part is essentially the load platform, base platform, Hooke's joint, and ball joint. The electrical control part includes the industrial computer, motion control card, servo driver, servo motor, and other electrical components.

The balance equation is as follows:

$$-M_{\mathrm{p}}\boldsymbol{S}_{\mathrm{p}}\times a_{\mathrm{p}}+M_{\mathrm{p}}\boldsymbol{S}_{\mathrm{p}}\times g-\boldsymbol{r}_{\mathrm{e}}a_{\mathrm{p}}-\omega\times\boldsymbol{r}_{\mathrm{e}}\omega+\boldsymbol{r}_{\mathrm{e}}F_{\mathrm{ext}}-\sum_{i=1}^{6}\left[q_i\times(F_{\mathrm{s}})_i\right]+\sum_{i=1}^{6}f_i=0, \quad (2)$$

where $\boldsymbol{S}_{\mathrm{p}}$ is the position vector of the load platform's center of gravity, $\boldsymbol{r}_{\mathrm{e}}$ is the inertia matrix of the load platform, ω is the angular velocity of the load platform, F_{ext} is the external torque acting on the load platform, f_i is the frictional torque of the ball hinge of the ith electric cylinder, M_{p} is the mass of the load platform, $(F_{\mathrm{s}})_i$ is the constraint force at the connection point between the ith electric cylinder and the load platform, g represents the acceleration due to gravity, a_{p} is the load platform's center of mass acceleration, and q_i is the connection point between O_{b} and the base platform with the first lower leg. Eq. (2) represents the connection between the dynamics of the six electric cylinders and the load platform, which reflects the influence of the coordinated action of the six electric cylinders on the load platform's motion. By solving and analyzing these equations, the motion state and force conditions of the load platform under different circumstances can be obtained. The derivation of the balance equation for the electric cylinder is shown in Section S2.2 of the ESM.

Torques are generated by changes in the load platform's posture caused by the different displacements of the six electric cylinders. Servo motors drive the displacement changes of the electric cylinders, so it is also necessary to model and analyze the two-phase alternating current (AC) motor and the electric cylinder.

2.3 Motor system modeling

The two-phase AC servo motor consists of two mutually perpendicular stator coils and a rotor. The transfer function of the entire servo cylinder system is first obtained. The mathematical models of the servo motor and the electric cylinder are derived, and then the transfer function G_3 of the entire servo electric cylinder system is obtained. The specific derivation is shown in Section S2.3 of the ESM.

$$G_3(s)=\frac{X_{\mathrm{L}}(s)}{U_{\mathrm{a}}(s)}=\frac{1729.53}{s^2+810.23s+1616.44}, \quad (3)$$

where X_{L} represents the displacement of the telescopic rod, and U_{a} represents the input voltage.

The electric cylinder converts the rotational motion of the servo motor into linear motion. The motion characteristics of the six electric cylinders collectively determine the position and posture of the load platform in 3-DoF in space. The inverse kinematics converts the 3-DoF ship motion into the extension and retraction of the electric cylinders. Therefore, by controlling the extension and retraction of the electric cylinders, the changes in position of the load platform in space can be controlled, achieving wave compensation control.

3 Ship 3-DoF compensation control method based on DTW-LSTM-MATD3

3.1 Problem description

In this study, we investigated the compensation control of ship motion in 3-DoF: roll, pitch, and heave. The objective was to transform the acquired ship motion in these 3-DoF into the extension and contraction movements of the six electric cylinders in a Stewart platform and then compensate for the six electric cylinders. Unlike the compensation control of 1-DoF ship motion, the control object of 3-DoF ship motion is six actuators, which is more complex and poses greater challenges for compensation control. To handle this issue, we use the MATD3 algorithm, treating each electric cylinder as an agent. The six agents are used to control the ship's motion, ensuring effective compensation effects under different sea conditions.

The Markov decision process (MDP) in reinforcement learning describes an entirely observable environment, where the observed state contains all the features required for decision-making. At any given moment, the future state and reward depend only on the current state and action. The reinforcement learning agent can perceive the ship's motion attitude and continuously update its policy to achieve more precise compensation control of the ship's 3-DoF. MDP describes the decision-making process of a single agent in a certain environment, fundamentally composed of a five-tuple $\langle S, A, P, R, \gamma \rangle$. The state space S describes the environment in which all agents are located, the action A describes the possible actions that each agent can take, R is the set of reward functions for all agents, and P is the state transition probability, indicating the probability of the next state when the agent is in the current state S and takes action A. γ is the discount factor, used to evaluate the impact of future rewards on the current policy. Finally, reinforcement learning enables each agent to learn each policy to maximize its own cumulative reward; that is, the agent finds the optimal policy.

(1) State space S: the state space contains all the observations of the agents. For the ship, the environment state space observed by the agent includes the displacement h_m of the electric cylinder after inverse solution in the environment, the movement speed v_{1m} of the electric cylinder, the movement position x_m of the compensation platform, and the movement speed v_{2m}. T represents the total time step size. The state space of the agent is:

$$S=\{h_m(t), v_{1m}(t), x_m(t), v_{2m}(t)\}_{m=1,2,\cdots,T}. \quad (4)$$

(2) Action space A: the compensation control of ship motion needs to consider the continuity of the action space. In this study, the action space refers to the actions a_m of each electric cylinder in the compensation platform. Since the control of the ship's 3-DoF motion compensation system adopts position control, the input of the neural network is the action space.

$$A=\{a_m(t)\}_{m=1,\,2,\,\cdots,\,T}. \tag{5}$$

(3) Reward function R: owing to the complexity and diversity of the ship environment, it is difficult for agents to obtain positive rewards during exploration, leading to slow learning and the problem of sparse rewards. To handle these issues, a method combining linear reward functions and normal reward functions is adopted to increase the exploratory nature of agents and improve learning efficiency.

$$R=\begin{cases}-5|h_{1m}-h_{2m}|, & |h_{1m}-h_{2m}|>0.001,\\ \mathrm{e}^{\frac{|h_{1m}-h_{2m}|^2}{-9\times10^{-6}}}, & |h_{1m}-h_{2m}|\leqslant 0.001.\end{cases} \tag{6}$$

To enable the six agents to explore better, we adopt a composite reward function. For the linear reward function, the slope of the reward function is a constant, implying that the exploration speed of the agents remains unchanged under all error values, which is not conducive to improving the training speed. For the normal reward function, when the compensation error is large, the reward value is small; when the compensation error is minimal, the reward value becomes larger. When the error is extremely large, although the corresponding reward value obtained is 0, the distinction between different error magnitudes is not obvious, which is not conducive to training. Therefore, a composite reward function is adopted, which combines the advantages of both reward functions and helps improve the training effect. A Lyapunov analysis of the combined reward function is shown in Section S3.1 of the ESM.

3.2 Noise discrimination based on the DTW algorithm

The DTW algorithm is used to describe the similarity matching relationship of time series with nonlinear deformations on the time axis. It adjusts mainly the corresponding relationship on the time axis elastically to calculate the minimum alignment cost between two time series of different lengths, thus more accurately measuring the similarity between time series.

The attitude time series of a ship during continuous ship posture motion is divided into two segments: $a=\{a_1,a_2,\cdots,a_{n1}\}$ and $b=\{b_1,b_2,\cdots,b_{n2}\}$. Firstly, the cost matrix $\boldsymbol{D}$ of $n_1\times n_2$, representing the degree of difference between corresponding elements in the two attitude time series is constructed. Any element $D(i,j)$ in the matrix $\boldsymbol{D}$ represents the local cost between element a_i in sequence a and element b_j in sequence b, as follows:

$$\boldsymbol{D}=\begin{bmatrix}\sqrt{(a_1-b_1)^2} & \cdots & \sqrt{(a_1-b_{n2})^2}\\ \cdots & \sqrt{(a_i-b_j)^2} & \cdots\\ (a_{n1}-b_1)^2 & \cdots & \sqrt{(a_{n1}-b_{n2})^2}\end{bmatrix},\quad i=1,2,\cdots,n_1,\ j=1,2,\cdots,n_2. \tag{7}$$

Secondly, the cumulative cost matrix $\boldsymbol{M}$ is initialized to record the minimum cumulative cost from the starting point to each point in the matrix, and the optimal alignment path is calculated gradually, as follows:

$$\boldsymbol{M}=\begin{bmatrix}M(1,1) & \cdots & M(n_1,1)\\ \cdots & M(i,j) & \cdots\\ M(n_2,1) & \cdots & M(n_1,n_2)\end{bmatrix}, \tag{8}$$

where $M(i,j)=D(i,j)+\min\{M(i-1,j),M(i,j-1),M(i-1,j-1)\}$. Let $M(i,j)$ denote any element in the matrix $\boldsymbol{M}$. The last element $M(n_1,n_2)$ in the cumulative cost matrix $\boldsymbol{M}$ is the DTW value between the two attitude time series a and b.

During the ship's 3-DoF motion, owing to the effect of different types of noise signals, the ship's attitude time series exhibits characteristics different from those under normal conditions. Therefore, it is necessary to find the boundary point between high-frequency and low-frequency noise and take different compensation actions for different types of noise signals on the compensation platform. At this time, the DTW value of the ship posture motion sequence is quite different from that under normal conditions. The DTW value is transformed by fast Fourier transform (FFT), converting the time-domain signal into a frequency-domain signal, and a DTW method for discriminating noise signals is constructed. The criterion is as follows:

$$\mathrm{FFT}(D)>\mathrm{FFT}(D_{\mathrm{n}}=D_{\mathrm{softmax}}), \tag{9}$$

$$D_{\text{softmax}} = \text{Soft max}(D) = \frac{e^{D}}{\sum_{j=1}^{n} e^{D_j}}. \tag{10}$$

In Eq. (9), D represents the DTW value of two ship attitude time series. D_n represents the DTW value at the boundary point between high-frequency noise and low-frequency signal. D_n is set in Eq. (9). When the noise signal frequency f is less than 30 Hz, D_{softmax} approaches zero (Fig. 2). The specific expansion and derivation of the softmax formula are shown in Section S3.4 of the ESM. When f is greater than 30 Hz, D_{softmax} is greater than zero and shows a significant increase, indicating that the boundary point is at f_c= 30 Hz, where the DTW value is 25. Therefore, D_n is set to the DTW value when the noise signal frequency is 30 Hz.

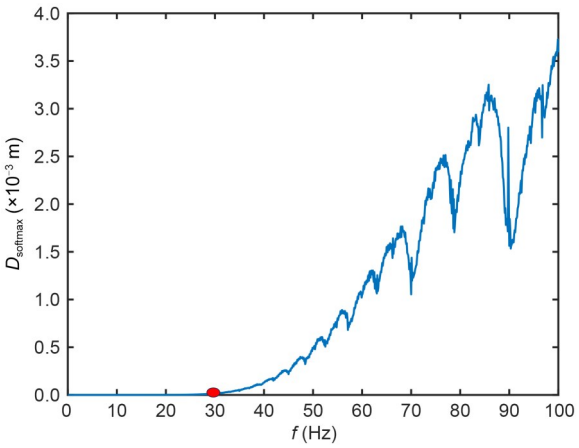

Fig. 2 Variation of noise signal frequency ranging from 0 to 100 Hz

Upon analysis, when $f > f_c$, the DTW values of the two ship attitude time series are greater than 25, and the similarity is low. This indicates that during the ship's motion, high-frequency vibrations caused by waves, mechanical vibrations from the main engine or propeller, and other high-frequency noise factors are present. The compensation platform needs to identify random signals in this region and explore a relatively stable operating state for the ship. When $f \leqslant f_c$, the DTW values of the two ship attitude time series are less than 25, and the similarity is high. This suggests that low-frequency noise factors such as the response inertia delay of steering and fluctuations in propeller thrust are present in this region. At this time, the compensation platform should implement tracking and compensation control. Therefore, in this study we used an LSTM network, which enables the agent to learn long-term dependencies in the data and reduce the problems of gradient vanishing and gradient explosion in sequence data. The structure and function of the algorithm are detailed in the next section.

3.3 DTW-LSTM-MATD3 Algorithm

Owing to the presence of six highly coupled actuators in the Stewart platform, the policy network of the MATD3 algorithm needs to learn complex nonlinear relationships and deal with the challenges of non-stationary environments and other aspects. The introduction of the TD3 and MATD3 algorithms, along with noise analysis, is given in Sections S3.2 and S3.3 of the ESM. In this study, LSTM networks were introduced into the Critic and Actor networks of the MATD3 algorithm. LSTM is a deep learning model commonly used to process sequential data. Compared with traditional recurrent neural networks, LSTM introduces three gates, namely the input gate, forget gate, and output gate, as well as memory cells with the same shape as the hidden state. It helps to capture the long-term and short-term dependencies in time series during network training.

The current state space S is input into the Actor network, and the current state space S and current action A are concatenated and input into the Critic network (Fig. 3). The fully connected layer is used to learn the complex relationships between input features, and the features are transformed linearly or nonlinearly through weights and biases. The third layer of the Critic and Actor networks is set as the LSTM layer. The output value H_t of the LSTM layer is used as the input for the next layer. This significantly enhances the ability of the Critic and Actor network models to handle complex sequential data. The use of cell states and control gates in LSTM allows for reasonable retention of earlier information in the time series and also enables the capture of long-range dependencies within the sequence information. Additionally, the presence of cell states can prevent the vanishing gradient problem. This allows the agent to better utilize effective sequential information during learning, enhance the rewards obtained by the agent, enhance exploration, and ultimately enable the actuators to output optimal actions.

Fig. 4 is a structure diagram of the DTW-LSTM-MATD3 algorithm used in this study. We input the

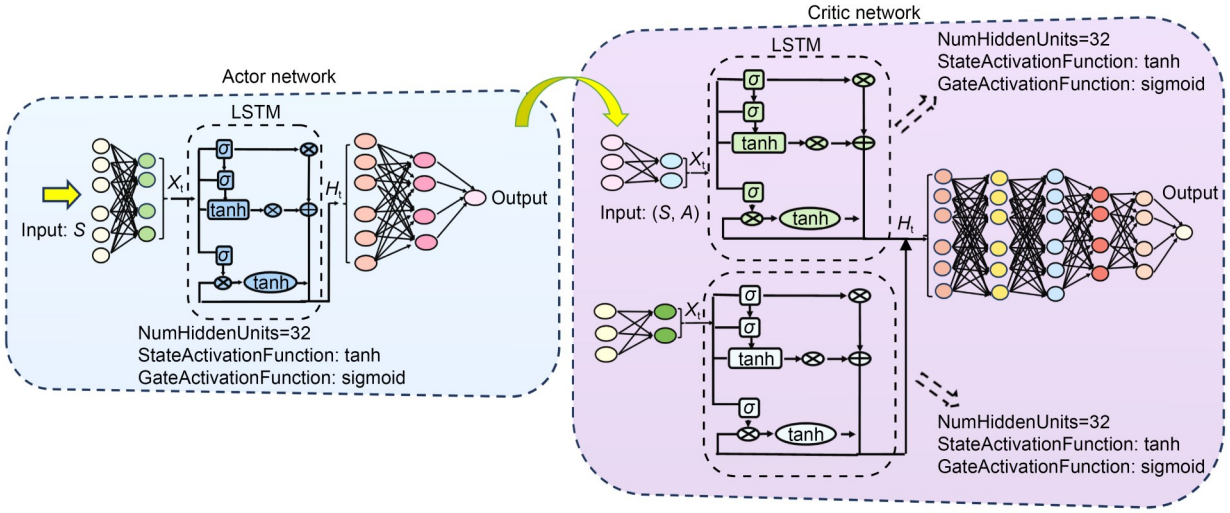

Fig. 3 Structure of the Critic and Actor networks in the DTW-LSTM-MATD3 algorithm

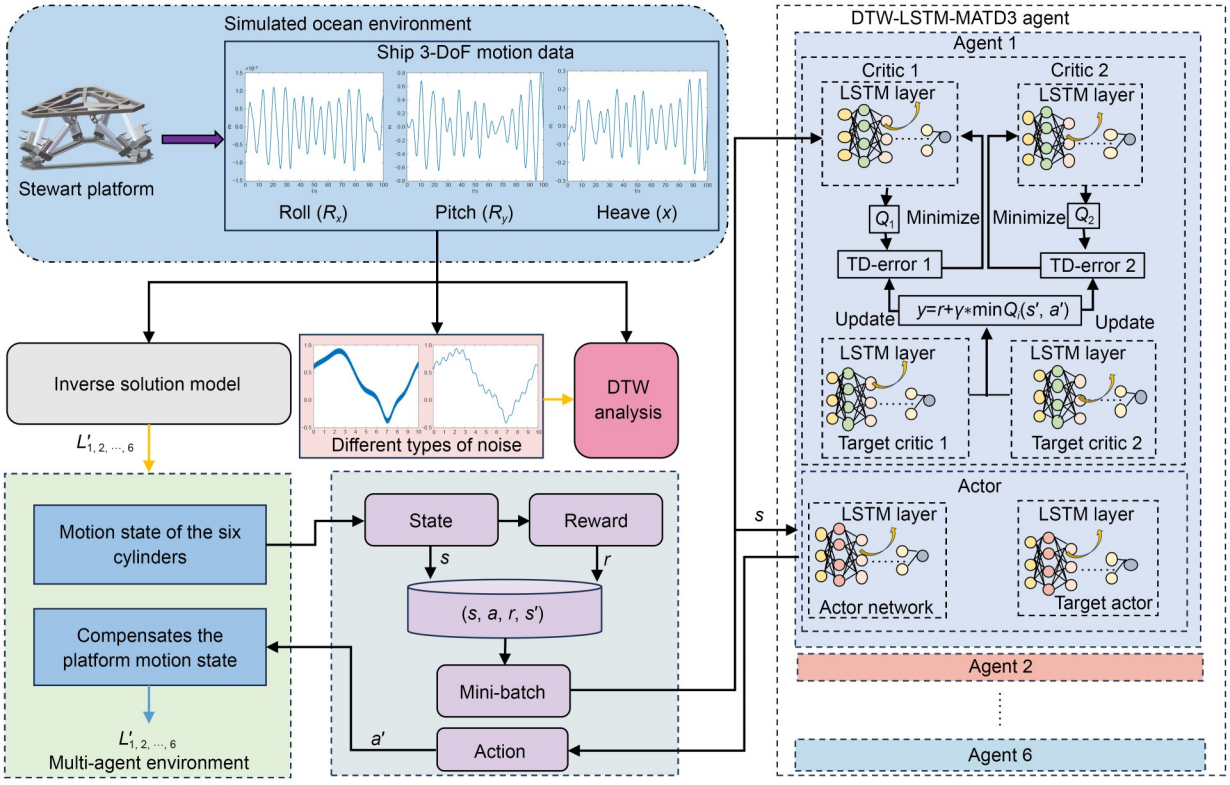

Fig. 4 Structure diagram of the DTW-LSTM-MATD3 algorithm

3-DoF ship motion data into the Simulink model. The extension and retraction amount of each actuator on the lower platform are obtained through kinematic inversion, and are used as the input signals for the wave-compensation system control. Then, the states, actions, transition probabilities, and reward function in the environment are determined, and the architecture of the agent is designed. The Critic and Actor networks with LSTM layers were established. These networks can estimate target Q-values and optimize strategies based on the next state s' and corresponding action a', to handle time-series correlation issues in ship attitude motion compensation control. Finally, the agent needs to continuously explore and improve the accuracy of

its policy through interaction with the environment. It learns through trial and error and ultimately finds the optimal control policy. Analyses of the impact of noise on DTW-LSTM-MATD3 are shown in Section S3.5 of the ESM.

4 Simulation results comparison and analysis

The simulation environment was based on the 3-DoF Stewart model in MATLAB/Simulink. The Stewart model can control the output of continuous actions and can quickly compensate for the target points generated by ship motion. However, since it is difficult to train the Stewart model with large-amplitude inputs, the input data were first scaled down by a ratio of 10:3 to facilitate the training of the entire model. We trained the model in Simulink in MATLAB, and the Adam optimizer was used to update the network model. Through continuous iteration, the network parameters of the neural network were continuously adjusted to ultimately achieve the optimal control policy output. In this simulation experiment, the neural network parameters of the MATD3 algorithm were set (Table 1).

Table 1 Experimental parameter settings

Parameter	Value
Data volume	10000
Sampling time (s)	0.01
Mini batch size	64
Critic network leaning rate	0.0001
Actor network learning rate	0.0001
Discount factor	0.995
Experience replay buffer capacity	2×10^6
Sequence length	20

To verify the effectiveness of the DTW-LSTM-MATD3 algorithm, we added comparisons with the TD3, MADDPG, and MATD3 control methods. Simulation experiments were conducted in scenarios such as sixth-level sea conditions under the P-M spectrum, sudden changes from the fourth-level to sixth-level sea conditions, generalization from the fourth-level to sixth-level sea conditions, and real-world changing sea. Compensation performance was evaluated using root mean square error (RMSE), mean absolute error (MAE), and compensation efficiency (η).

4.1 DTW-LSTM-MATD3 compensation control based on P-M spectrum under sixth-level sea conditions with high-frequency noise

In this section, we describe comparative simulation experiments conducted on the compensation control of 3-DoF ship motion. We simulated and analyzed four algorithms: TD3, MADDPG, MATD3, and DTW-LSTM-MATD3. Firstly, we used the marine hydrodynamics simulation software to generate the 3-DoF data of a certain engineering ship under sixth-level sea conditions based on the PM spectrum.

We added high-frequency noise to Agent 5 within the 30–40 s interval and conducted simulation comparisons. Fig. 5 illustrates the compensation error of the six electric cylinder agents under sixth-level sea conditions. The sea conditions were more severe, the extension and retraction of the electric cylinders were greater, and the compensation difficulty was also increased. The compensation effect of the TD3 algorithm was not optimal. The compensation error amplitude of Agents 1 and 2 exceeded 1 m, and the magnitude of the compensation error for Agents 2 to 6 exceeded 0.4 m, making compensation control impossible. Among the other three algorithms, the compensation error amplitude of the MADDPG algorithm was clearly greater than that of the MATD3 and DTW-LSTM-MATD3 algorithms. Within the 30–40 s interval, when Agent 5 was subjected to high-frequency noise, the compensation error of DTW-LSTM-MATD3 was closer to the zero-error level line, which was the best result among the four algorithms.

Table 2 shows the performance metrics RMSE, MAE, and η for the TD3, MADDPG, MATD3, and DTW-LSTM-MATD3 algorithms under sixth-level sea conditions. Under the more severe sixth-level sea condition, the TD3 algorithm had the lowest compensation efficiency of only 49.96% (Agent 2), significantly lower than the other algorithms. The MADDPG algorithm showed some improvement compared to TD3, with a compensation efficiency of 98%, but was still lower than that of DTW-LSTM-MATD3, which achieved the highest compensation efficiency of up to 99.53%. The maximum efficiency of the MATD3 algorithm was only 96.76%, indicating a certain level of improvement but still lower than that of DTW-LSTM-MATD3.

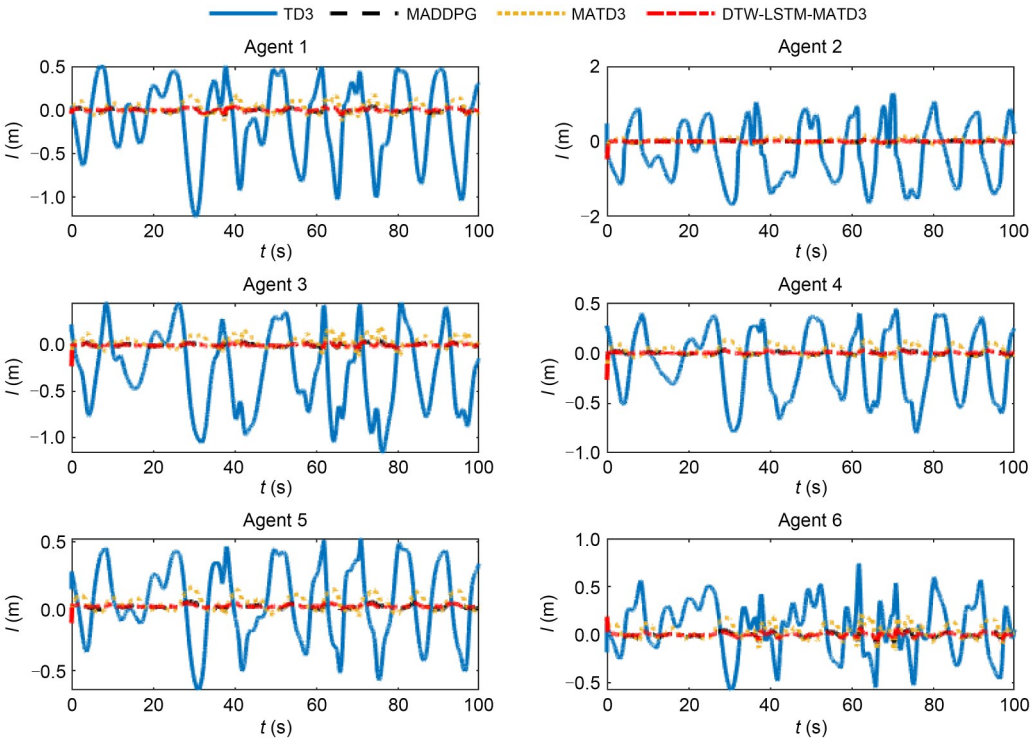

Fig. 5 Compensation control error of each algorithm under sixth-level sea conditions

Table 2 Comparison of compensation control performance parameters under sixth-level sea conditions

Agent	TD3			MADDPG			MATD3			DTW-LSTM-MATD3		
	RMSE	MAE	η	RMSE	MAE	η	RMSE	MAE	η	RMSE	MAE	η
1	0.4633	0.3776	78.27%	0.0241	0.0209	98.80%	0.0714	0.0580	96.66%	0.0159	0.0116	99.33%
2	0.8309	0.7223	49.96%	0.0271	0.0181	98.74%	0.0645	0.0515	96.43%	0.0215	0.0089	99.39%
3	0.5217	0.4303	72.72%	0.0213	0.0166	98.95%	0.0700	0.0542	96.56%	0.0142	0.0091	99.42%
4	0.3552	0.3038	78.06%	0.0193	0.0154	98.89%	0.0584	0.0484	96.51%	0.0120	0.0066	99.52%
5	0.3030	0.2623	83.02%	0.0191	0.0165	98.93%	0.0617	0.0501	96.76%	0.0101	0.0073	99.53%
6	0.2878	0.2429	86.64%	0.0267	0.0215	98.82%	0.0836	0.0664	96.35%	0.0195	0.0133	99.27%

4.2 Compensation control of DTW-LSTM-MATD3 under sudden change sea conditions

Robustness in control systems refers to the ability to maintain reasonable performance when the model is subjected to uncertain external changes or parameter variations. It is a critical indicator for measuring control accuracy. In this study, we conducted simulation experiments under sudden changes from the fourth- to sixth-level sea conditions, with a high-frequency noise signal of 100 Hz present in Agent 4 between 30 and 40 s. Under sudden change sea conditions, the amplitude of ship motion underwent a drastic change at the 50-s mark, with a sharp increase, making the compensation control of ship motion more challenging (Fig. 6). The compensation control error of MADDPG, MATD3, and DTW-LSTM-MATD3 was close to the zero-error line, without any spikes or fluctuations. The DTW-LSTM-MATD3 algorithm had the smallest error amplitude and higher compensation accuracy. Table 3 compares compensation control performance parameters under sudden change sea conditions. The maximum compensation efficiency of the TD3 algorithm was only 93.37%. For the MADDPG algorithm, the compensation efficiency of Agents 1 to 6 ranged between 97.20% and 99.24%. The maximum compensation efficiency of the MATD3 algorithm was only 98.74%, indicating poor compensation performance. In contrast, the DTW-LSTM-MATD3 algorithm had the highest compensation efficiency of up to 99.60%, with all compensation efficiencies above 99%. The

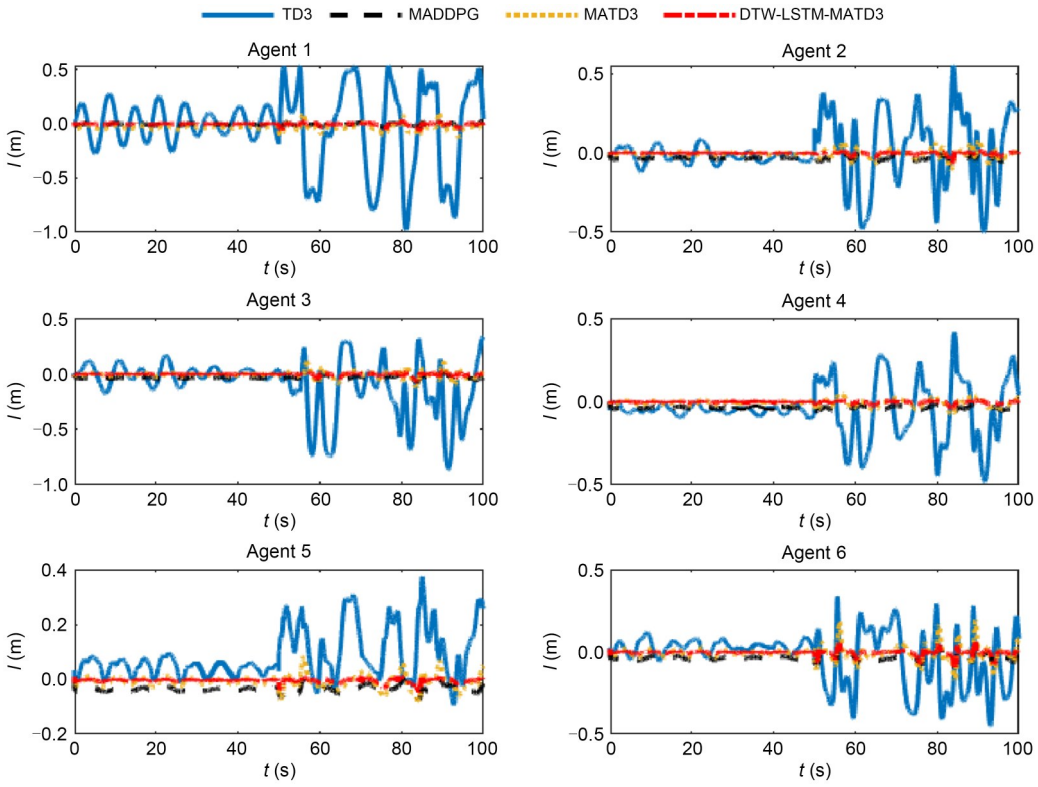

Fig. 6 Compensation control error of each algorithm under sudden change from the fourth- to sixth-level sea conditions

Table 3 Comparison of compensation control performance parameters under sudden change from the fourth- to sixth-level sea conditions

Agent	TD3			MADDPG			MATD3			DTW-LSTM-MATD3		
	RMSE	MAE	η	RMSE	MAE	η	RMSE	MAE	η	RMSE	MAE	η
1	0.3398	0.2597	82.79%	0.0151	0.0115	99.24%	0.0463	0.0404	97.33%	0.0127	0.0074	99.51%
2	0.1841	0.1304	90.63%	0.0386	0.0368	97.36%	0.0265	0.0181	98.70%	0.0104	0.0062	99.56%
3	0.2566	0.1675	89.62%	0.0376	0.0355	97.80%	0.0311	0.0205	98.73%	0.0127	0.0073	99.55%
4	0.1602	0.1206	90.72%	0.0379	0.0365	97.20%	0.0221	0.0164	98.74%	0.0079	0.0052	99.60%
5	0.1210	0.0879	93.37%	0.0380	0.0363	97.26%	0.0251	0.0187	98.59%	0.0088	0.0057	99.57%
6	0.1566	0.1145	93.13%	0.0391	0.0362	97.83%	0.0429	0.0284	98.30%	0.0179	0.0097	99.42%

RMSE and MAE values decreased, and during the 30–40 s interval when Agent 4 was subjected to high-frequency noise, the compensation error was the lowest among the four algorithms, demonstrating the ability to resist high-frequency noise.

4.3 Generalization test of DTW-LSTM-MATD3 with a high-frequency noise

Considering the variability of sea conditions, the model must exhibit reasonable generalization. Therefore, we selected the 3-DoF motion data of the ship under fourth-level sea conditions for simulation verification. We trained the compensation platform model with the data of the fourth-level sea conditions, and the network parameters of the trained model were saved. Then, the data of the sixth-level sea conditions with a high-frequency noise of 100 Hz were imported for verification to conduct the generalization test. Fig. 7 shows the compensation error of each electric cylinder generalized from the fourth- to the sixth-level. With the TD3 algorithm, the compensation error of Agents 1 and 2 both reached 1 m, and the compensation effect of the compensation platform was inadequate. The compensation error of MADDPG reached 0.2 m, with low compensation control accuracy. The compensation error amplitude of the MATD3 algorithm was

lower than that of MADDPG, while the compensation error of the DTW-LSTM-MATD3 algorithm was closer to zero, with the best compensation effect. Within the 30–40 s interval, when Agent 4 was subjected to a high-frequency noise signal of 100 Hz, the compensation error of the DTW-LSTM-MATD3 algorithm approached the zero-scale line, while the other three algorithms had larger amplitude fluctuations in their compensation errors. Table 4 compares compensation control performance parameters generalized from the fourth to the sixth level sea conditions. The maximum compensation efficiency of the TD3 algorithm was only 89.15%, and the maximum compensation efficiency of the MADDPG algorithm was only 93.64%. The compensation efficiency of DTW-LSTM-MATD3 was above 99.18%, with a maximum of 99.45%, higher than the MATD3 algorithm. In addition, the RMSE and MAE decreased, indicating that the DTW-LSTM-MATD3 algorithm has better compensation performance. It not only performed well on the training set but also had good generalization ability.

4.4 Compensation control experiment under real sea conditions

In this experiment, we used the actual motion data of the 'Yuming' ship at sea. Fig. 8 presents the

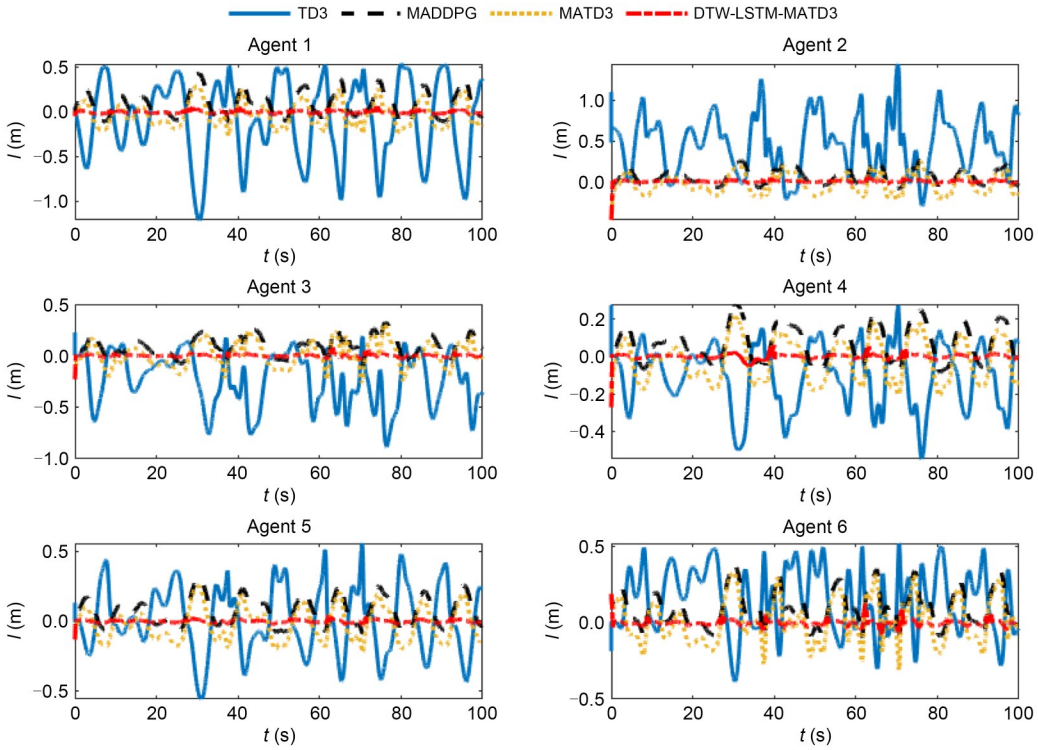

Fig. 7 Compensation control error of each algorithm generalized from the fourth- to sixth-level sea conditions

Table 4 Comparison of compensation control performance parameters generalized from the fourth- to sixth-level sea conditions

Agent	TD3			MADDPG			MATD3			DTW-LSTM-MATD3		
	RMSE	MAE	η	RMSE	MAE	η	RMSE	MAE	η	RMSE	MAE	η
1	0.4541	0.3755	78.39%	0.1643	0.1272	92.68%	0.1378	0.1190	93.15%	0.0173	0.0142	99.18%
2	0.5834	0.4913	65.96%	0.1182	0.0957	93.37%	0.1242	0.1090	92.45%	0.0218	0.0103	99.29%
3	0.3454	0.2642	83.25%	0.1385	0.1126	92.86%	0.1185	0.0983	93.77%	0.0143	0.0093	99.41%
4	0.1981	0.1503	89.15%	0.1183	0.0956	93.10%	0.1139	0.1002	92.76%	0.0152	0.0095	99.31%
5	0.2392	0.1959	87.32%	0.1235	0.0983	93.64%	0.1197	0.1032	93.32%	0.0107	0.0085	99.45%
6	0.2562	0.2107	88.41%	0.1555	0.1235	93.21%	0.1413	0.1156	93.64%	0.0214	0.0138	99.24%

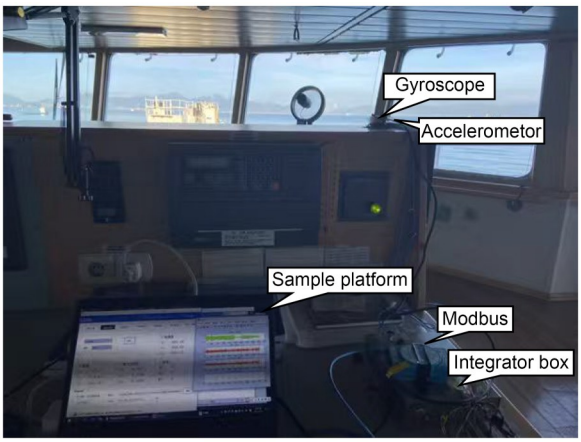

Fig. 8 On-site diagram of the 'Yuming' ship's motion attitude test

test site of ship motion attitude. Firstly, the angular velocity and linear acceleration raw data of the ship were measured by gyroscopes and accelerometers, respectively. Subsequently, these data were sent to the integration box, where the internal processor performed preliminary filtering and processing on the raw data. Then, the integration box packages the processed ship motion attitude data into standard data frames through the Modbus communication protocol. Finally, the data acquisition platform on the computer, acting as the Modbus master station, actively queried or received these data frames and recorded them. This provided a true and reliable data source for subsequent research on control algorithms.

Fig. 9 presents the 100-s roll motion of the 'Yuming' ship in a real marine environment. After the kinematic inverse solution, low-frequency noise was detected in Agent 1 from 10 to 20 s. Error comparisons were conducted using four algorithms: TD3, MADDPG, MATD3, and DTW-LSTM-MATD3. The results show that DTW-LSTM-MATD3 had a smaller compensation error and better performance. In addition, we collected 100-s mutation sea condition data with real noise and conducted simulation experiments again.

Fig. 10 presents the compensation error diagram of the four algorithms. The errors of TD3 and MADDPG were relatively large, while those of MATD3 and DTW-LSTM-MATD3 were closer to the zero-scale line. The error amplitude of DTW-LSTM-MATD3 was lower, closer to the zero-scale line, and thus had a better compensation effect.

Fig. 11 illustrates the roll motion diagram of the 'Yuming' ship under sudden sea conditions with real

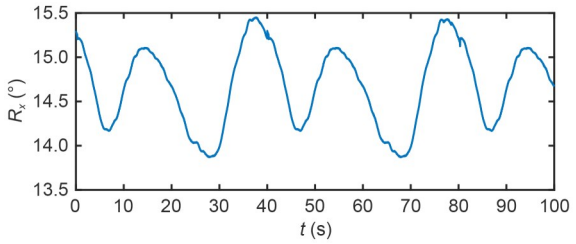

Fig. 9 Roll motion diagram of the 'Yuming' ship in a real marine environment

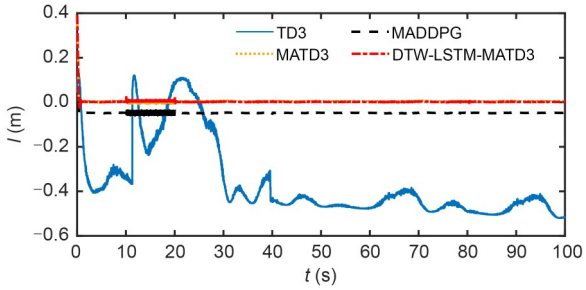

Fig. 10 Compensation control error diagram of various algorithms

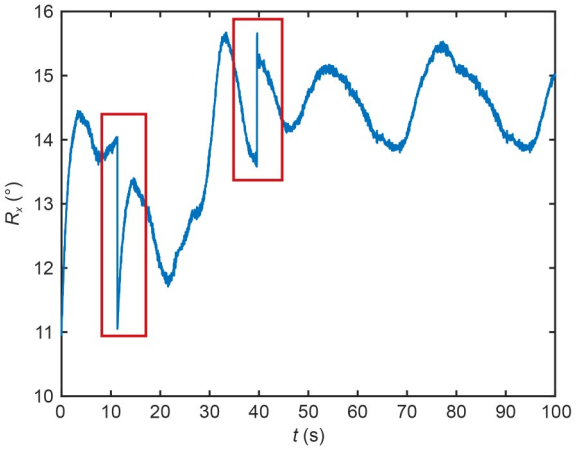

Fig. 11 Roll motion diagram of the 'Yuming' ship under sudden changes in sea conditions with real noise

noise. From 0 to 10 s, the ship's roll motion remained within the range of 11.0° to 14.5°, maintaining a stable noisy state. At the 10-s mark, the ship's motion was affected by external disturbances, causing a sudden change, and the motion attitude dropped sharply. Between 20 and 40 s, it re-entered a new state and reached a higher roll motion near 35 s. At the 40-s mark, the ship's motion changed again, with the roll angle rising. From 40 to 100 s, the ship maintained a stable noisy motion at a higher roll angle. The ship's motion in real sea conditions was highly sensitive to

external disturbances, and its dynamic changes often exhibited nonlinearity. Therefore, the controller needs to have strong robustness and generalization, capable of maintaining stable and effective compensation control even under conditions of model parameter uncertainty and unknown environmental disturbances.

With the TD3 algorithm, the error amplitude of Agents 1 and 2 reached 0.5 m, with significant fluctuations, resulting in unstable compensation control effects (Fig. 12). The compensation effect of MADDPG was worse than TD3 but better than MATD3 and DTW-LSTM-MATD3. The error of Agent 6 produced sudden changes at 10 and 40 s, with errors exceeding 0.2 m. With Agent 3, the DTW-LSTM-MATD3 algorithm had the lowest compensation error amplitude. However, at the mutation moments, all four algorithms showed unstable compensation phenomena. Therefore, owing to the sudden and high-frequency disturbances in real sea conditions, the algorithm proposed in this study struggled to achieve high-precision compensation control at the peak and trough points of the ship's sudden changes.

Table 5 compares compensation control performance parameters under real mutation scenarios with noisy sea conditions, where the compensation efficiency of four algorithms was evaluated using RMSE, MAE, and η indicators. The η values of MATD3 and DTW-LSTM-MATD3 were all higher than those of TD3 and MADDPG. The compensation efficiencies of MATD3 and DTW-LSTM-MATD3 were greater than

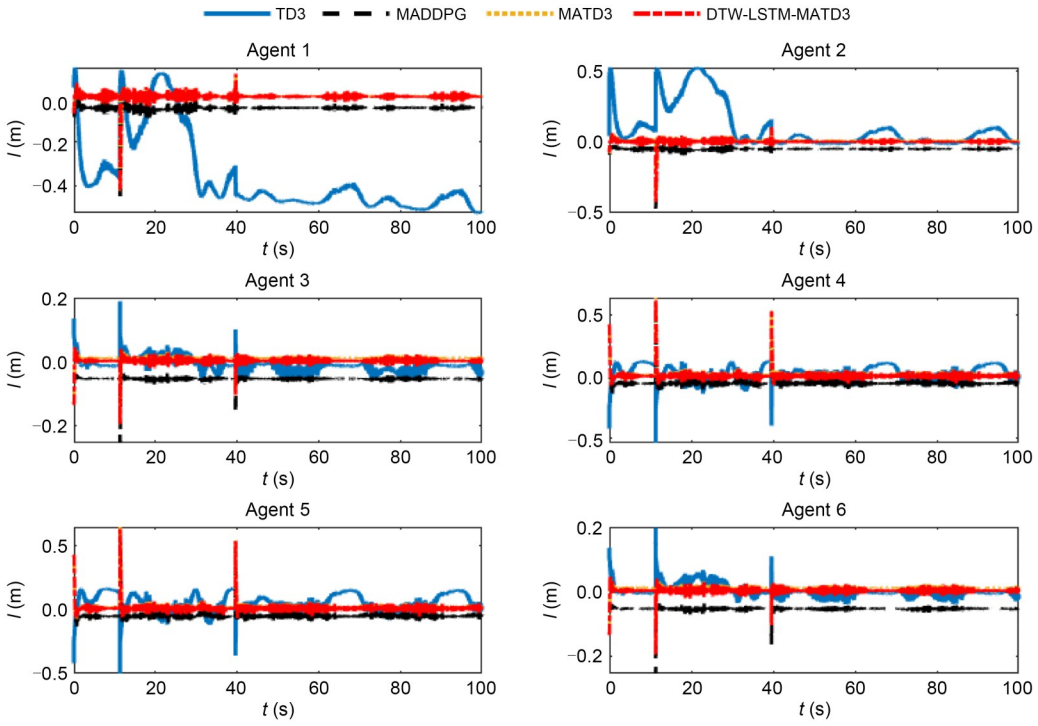

Fig. 12 Roll motion compensation error diagram of 'Yuming' ship under real-time noisy abnormal sea conditions

Table 5 Comparison of compensation control performance parameters under real-time noisy abnormal sea conditions

Agent	TD3			MADDPG			MATD3			DTW-LSTM-MATD3		
	RMSE	MAE	η	RMSE	MAE	η	RMSE	MAE	η	RMSE	MAE	η
1	0.4000	0.3734	32.05%	0.0533	0.0497	90.96%	0.0215	0.0084	98.47%	0.0207	0.0080	98.54%
2	0.0185	0.1089	80.18%	0.0533	0.0497	90.96%	0.0234	0.0126	97.71%	0.0207	0.0080	98.54%
3	0.0227	0.0184	91.03%	0.0543	0.0538	73.85%	0.0146	0.0124	93.95%	0.0099	0.0056	97.26%
4	0.0641	0.0503	92.59%	0.0643	0.0575	91.52%	0.0387	0.0177	97.38%	0.0374	0.0154	97.73%
5	0.0825	0.0644	90.50%	0.0643	0.0575	91.52%	0.0387	0.0177	97.38%	0.0374	0.0154	97.73%
6	0.0189	0.0129	93.75%	0.0543	0.0538	73.85%	0.0146	0.0124	93.95%	0.0099	0.0056	97.26%

93%, and the compensation efficiency of DTW-LSTM-MATD3 was higher than that of MATD3, with lower RMSE and MAE values. However, the improvement in compensation efficiency of Agents 1, 2, 4, and 5 was less than 1%, and the compensation efficiency of Agents 1 to 6 was below 99%. Therefore, our proposed improved algorithm had difficulty achieving high-precision compensation control throughout the process when facing real, dynamic, instantaneous mutation in complex environments with noise.

In the future, we will address these deficiencies, aiming to enhance the algorithm's robustness in dealing with dynamic unknown environments and enhance the overall compensation control effect of the system.

5 Conclusions

In this study, we developed a compensation control strategy suitable for the 3-DoF motion of ships using multi-agent reinforcement learning methods, targeting the 3-DoF motion of ships in complex sea conditions. First, we constructed a reinforcement learning environment for the ship's 3-DoF motion compensation system. Next, we applied the DTW algorithm to distinguish between high-frequency noise and low-frequency tracking signals, and trained the model using the MATD3 algorithm, incorporating the LSTM neural network and a combined reward function to improve the compensation efficiency of traditional MATD3. Finally, simulation results showed that, compared to the TD3 algorithm, MADDPG algorithm, and traditional MATD3 algorithm, the DTW-LSTM-MATD3 trained wave compensation platform model achieved higher compensation efficiency under sixth-level sea conditions, sudden change sea conditions, and noisy signals. The model's generalization ability was also verified, demonstrating the superior compensation performance of the proposed method for the 3-DoF motion control of the ship's posture.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 52105466).

Author contributions

Qin ZHANG: writing–review & editing, writing–original draft, supervision, methodology, and conceptualization. Jingyi ZHOU: writing–review & editing, writing–original draft, validation, data curation. Bangping GU: resources and funding acquisition. Xiong HU: resources and project administration.

Conflict of interest

Qin ZHANG, Jingyi ZHOU, Bangping GU, and Xiong HU declare that they have no conflict of interest.

References

Cauz M, Wyrsch N, Perret L, et al., 2025. Embracing wind power in the solar PV-dominated Swiss landscape. *Energy Reports*, 13:3341-3350. https://doi.org/10.1016/j.egyr.2025.02.059

Hou Y, Han GJ, Zhang F, et al., 2024. Distributional soft actor-critic-based multi-AUV cooperative pursuit for maritime security protection. *IEEE Transactions on Intelligent Transportation Systems*, 25(6):6049-6060. https://doi.org/10.1109/TITS.2023.3341034

Jimoh IA, Küçükdemiral IB, Bevan G, 2021. Fin control for ship roll motion stabilisation based on observer enhanced MPC with disturbance rate compensation. *Ocean Engineering*, 224:108706. https://doi.org/10.1016/j.oceaneng.2021.108706

Kang JC, Sun LP, Guedes Soares C, 2019. Fault tree analysis of floating offshore wind turbines. *Renewable Energy*, 133:1455-1467. https://doi.org/10.1016/j.renene.2018.08.097

Li C, Mogollón JM, Tukker A, et al., 2022. Environmental impacts of global offshore wind energy development until 2040. *Environmental Science & Technology*, 56(16): 11567-11577. https://doi.org/10.1021/acs.est.2c02183

Liu XX, Yin Y, Su YZ, et al., 2022. A multi-UCAV cooperative decision-making method based on an MAPPO algorithm for beyond-visual-range air combat. *Aerospace*, 9(10):563. https://doi.org/10.3390/aerospace9100563

Lv YY, Li H, 2023. Strong fixed-time dynamic inverse adaptive LQR integrated control strategy for dynamic positioning of ship. *Ocean Engineering*, 288:115969. https://doi.org/10.1016/j.oceaneng.2023.115969

Möllerström E, Gipe P, Ottermo F, 2025. Wind power development: a historical review. *Wind Engineering*, 49(2): 499-512. https://doi.org/10.1177/0309524X241260061

Mou ZY, Zhang Y, Gao FF, et al., 2021. Deep reinforcement learning based three-dimensional area coverage with UAV swarm. *IEEE Journal on Selected Areas in Communications*, 39(10):3160-3176. https://doi.org/10.1109/JSAC.2021.3088718

Nathwani J, Kammen DM, 2019. Affordable energy for humanity: a global movement to support universal clean energy access. *Proceedings of the IEEE*, 107(9):1780-1789. https://doi.org/10.1109/JPROC.2019.2918758

Qin YH, Zhang ZS, Li XL, et al., 2023. Deep reinforcement learning based resource allocation and trajectory planning in integrated sensing and communications UAV network.

IEEE Transactions on Wireless Communications, 22(11): 8158-8169.
https://doi.org/10.1109/TWC.2023.3260304

Shao S, Liu HW, Zhang L, et al., 2022. Integration of super-resolution ISAR imaging and fine motion compensation for complex maneuvering ship targets under high sea state. *IEEE Transactions on Geoscience and Remote Sensing*, 60:5222820.
https://doi.org/10.1109/TGRS.2022.3147266

Tang G, Lei JM, Li FR, et al., 2023. A modified 6-DOF hybrid serial–parallel platform for ship wave compensation. *Ocean Engineering*, 280:114336.
https://doi.org/10.1016/j.oceaneng.2023.114336

Wang WX, Ning YH, Zhang Y, et al., 2025. Linear active disturbance rejection control with linear quadratic regulator for Stewart platform in active wave compensation system. *Applied Ocean Research*, 156:104469.
https://doi.org/10.1016/j.apor.2025.104469

Wang YF, Zhao Y, 2025. Multiple ships cooperative navigation and collision avoidance using multi-agent reinforcement learning with communication. *Ocean Engineering*, 320:120244.
https://doi.org/10.1016/j.oceaneng.2024.120244

Winursito A, Pratama GNP, 2021. LQR state feedback controller with precompensator for magnetic levitation system. *Journal of Physics: Conference Series*, 2111(1):012004.
https://doi.org/10.1088/1742-6596/2111/1/012004

Woodacre JK, Bauer RJ, Irani R, 2018. Hydraulic valve-based active-heave compensation using a model-predictive controller with non-linear valve compensations. *Ocean Engineering*, 152:47-56.
https://doi.org/10.1016/j.oceaneng.2018.01.030

Wu LS, Zhang C, Zhang B, et al., 2025. Toward energy-efficiency: integrating MATD3 reinforcement learning method for computational offloading in RIS-aided UAV-MEC environments. *IEEE Internet of Things Journal*, 12(14):26582-26595.
https://doi.org/10.1109/JIOT.2025.3560835

Yan F, Fan K, Yan XC, et al., 2020. Constant tension control of hybrid active-passive heave compensator based on adaptive integral sliding mode method. *IEEE Access*, 8: 103782-103791.
https://doi.org/10.1109/ACCESS.2020.2995651

Zhang Q, Ding ZY, Zhang MJ, 2020. Adaptive self-regulation PID control of course-keeping for ships. *Polish Maritime Research*, 27(1):39-45.
https://doi.org/10.2478/pomr-2020-0004

Zhang YH, Li GX, Tian Y, et al., 2025. Model-free reinforcement learning-based transient power control of vehicle fuel cell systems. *Applied Energy*, 388:125614.
https://doi.org/10.1016/j.apenergy.2025.125614

Zhao EY, Zhou N, Liu CJ, et al., 2024. Time-aware MADDPG with LSTM for multi-agent obstacle avoidance: a comparative study. *Complex & Intelligent Systems*, 10(3): 4141-4155.
https://doi.org/10.1007/s40747-024-01389-0

Zhou YT, Kong XR, Lin KP, et al., 2024. Novel task decomposed multi-agent twin delayed deep deterministic policy gradient algorithm for multi-UAV autonomous path planning. *Knowledge-Based Systems*, 287:111462.
https://doi.org/10.1016/j.knosys.2024.111462

Electronic supplementary materials

Fig. S1, Sections S1–S3