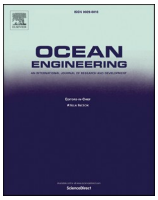
Ocean Engineering 291 (2024) 116403

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



An autonomous coverage path planning algorithm for maritime search and rescue of persons-in-water based on deep reinforcement learning

Jie Wu [a], Liang Cheng [a,b,c,d,*], Sensen Chu [a,b,c], Yanjie Song [e]

[a] School of Geography and Ocean Science, Nanjing University, Nanjing, China
[b] Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing, China
[c] Collaborative Innovation Center for the South Sea Studies, Nanjing University, Nanjing, China
[d] Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University, Nanjing, China
[e] College of Systems Engineering, National University of Defense Technology, Changsha, China

ARTICLE INFO

Handling Editor: Prof. A.I. Incecik

Keywords:
Persons-in-water
Drift trajectory prediction
Maritime search and rescue
Coverage path planning
Deep reinforcement learning

ABSTRACT

The prevalence of maritime transportation and operations is increasing, leading to a gradual increase in drowning accidents at sea. In the context of maritime search and rescue (SAR), it is essential to develop effective search plans to improve the survival probability of persons-in-water (PIWs). However, conventional SAR search plans typically use predetermined patterns to ensure complete coverage of the search area, disregarding the varying probabilities associated with the PIW distribution. To address this issue, this study has proposed a maritime SAR vessel coverage path planning framework (SARCPPF) suitable for multiple PIWs. This framework comprises three modules, namely, drift trajectory prediction, the establishment of a multilevel search area environment model, and coverage search. First, sea area-scale drift trajectory prediction models were employed using the random particle simulation method to forecast drift trajectories. A hierarchical probability environment map model was established to guide the SAR of multiple SAR units. Subsequently, we integrated deep reinforcement learning with a reward function that encompasses multiple variables to guide the navigation behavior of ship agents. We developed a coverage path planning algorithm aimed at maximizing the success rates within a limited timeframe. The experimental results have demonstrated that our model enables vessel agents to prioritize high-probability regions while avoiding repeated coverage.

1. Introduction

With continuous development of the global economy, maritime transportation has emerged as the primary mode for transporting international goods. With increasing human exploration and production activities at sea, offshore operations have become increasingly common. The marine environment is complex and dynamic, with natural disasters such as strong winds, high waves, storms, lightning strikes, and tsunamis potentially occurring. These events can increase the number of maritime accidents (Yang et al., 2020; Zhang et al., 2017; Zhou et al., 2020a, 2020b; Zhou, 2022). Maritime accidents often result in drowning and casualties. Therefore, the timely development of effective search plans and improvements in the efficiency of maritime search and rescue have become a key research focus (Koopman, 1956a, 1956b, 1957; Peng et al., 2022; Rani et al., 2022; Sendner, 2022). This is critical for enhancing the likelihood of survival among PIWs.

Maritime SAR comprises two essential components, that is, search and rescue, with search serving as a prerequisite for rescue (Carneiro, 1988; Haga and Svanberg, 2022; IAMSAR, 2016; International Maritime Organization, 1979). In maritime SAR, the scarcity of search resources and adverse meteorological conditions are the two primary factors that impede the SAR process (Zhou, 2022). This forces search planners to minimize the search area and maximize the chances of locating the search object (Tapkin and Temur, 2022). Search optimization is necessary to achieve maximum success rates while considering the correlation between time and resource constraints. Given the vulnerability of PIWs in maritime environments, rescuers must locate them immediately. Therefore, searching for PIWs includes three main tasks: (1) accurately and quickly predicting the drift trajectory of PIWs (Brushett et al., 2017; Chen et al., 2017, 2022; Wu et al., 2023); (2) determining the optimal search area to ensure full coverage of the possible distribution range; and (3) planning the search path for the SAR units and maximizing the

[*] Corresponding author. School of Geography and Ocean Science, Nanjing University, Nanjing, 210023, China.
E-mail addresses: 1936063643@qq.com (J. Wu), lcheng@nju.edu.cn (L. Cheng), chusensen@163.com (S. Chu), songyj_2017@163.com (Y. Song).

https://doi.org/10.1016/j.oceaneng.2023.116403
Received 26 April 2023; Received in revised form 20 October 2023; Accepted 17 November 2023
Available online 25 November 2023

cumulative probability of success (POS) of the entire search process (Brown, 1980; Kratzke et al., 2010; Lin and Goodrich, 2014; Mou et al., 2021; Washburn, 1983).

When the rescue units arrive at their initial position in maritime SAR, the PIWs continue to drift owing to the combined influence of surface currents, sea waves, and wind. The complexity of the maritime environment, along with numerous uncertain factors influencing the drifting process, amplifies the challenge of locating PIWs. It also increases the complexity of search path planning, thereby rendering the search more intricate. The prediction of drift trajectories involves the consideration and quantification of factors that affect the drift process, including the submersion scene, maritime conditions, and prediction modeling. In marine accidents involving PIWs, the drift characteristics vary depending on the posture, including whether the PIWs are upright, seated, or face down, or the load conditions (Wu et al., 2023). First proposed by Allen and Plourde (1999) to quantify the drift of objects, the Leeway model has been widely used to help plan national searches such as the French MOTHY (Daniel et al., 2003), Canadian CANSARP (Canadian Coast Guard College CANSARP Development Group Web site, 2009), and U.S. Coast SAROPS (Kratzke et al., 2010). Sea-based drift tests are widely recognized as the most commonly used and highly dependable approach for determining leeway coefficients (Breivik et al., 2012; Kasyk et al., 2021; Meng et al., 2021; Sutherland et al., 2020; Tu et al., 2021; Wu et al., 2023; Zhu et al., 2019).

Path-planning methods can be classified into two categories, that is, traditional and intelligent algorithmic. Traditional path planning algorithms include the dijkstra algorithm (Dijkstra, 1959; Wang et al., 2011), the A* algorithm and its improved versions (Chabini and Lan, 2002; Hart et al., 1972; Nash and Koenig, 2013), the D* algorithm and its improved versions (Koenig and Likhachev, 2005; Marija and Ivan, 2011; Stentz, 1994), the artificial potential field method (Zhang et al., 2012), the probabilistic path graph method (Kavraki et al., 1996), and the rapid exploration of random trees method (RRT) (Lavalle, 1998). Intelligent path planning algorithms include genetic algorithms (Prins, 2004), ant colony algorithms (Luo et al., 2020) and particle swarm algorithms (Masehian and Sedighizadeh, 2010). The reinforcement learning (RL) method (Wiering and Van, 2012) is an important approach in machine learning. In contrast with other intelligent algorithms for machine learning, RL focuses on the acquisition of system mapping from the environment to the behavior. It does not rely on labeled interactions as seen in supervised learning; instead, it learns from its own experiences. The objective of RL is not to discover hidden structures but rather to maximize rewards. The most used reinforcement learning methods include Q learning, SARSA learning, TD learning, and adaptive dynamic programming algorithms. Recently, significant advancements have been made in combining path planning with reinforcement learning (Busoniu et al., 2008; Xi et al., 2022; Xie et al., 2021).

Full-coverage path planning (CPP) is a specialized technique in robotics for generating a continuous path that passes through all accessible points within a given area with a minimum repetition rate and maximum coverage rate. This can be achieved using either random or environment-based models (Galceran and Carreras, 2013). To ensure comprehensive coverage, most existing CPP methods divide the target area and the surrounding space into cells using exact or approximate cell division techniques. CPPs have a wide range of applications in autonomous underwater vehicles (AUVs), including seabed mapping, mine detection, and oil spill cleanup (Englot and Hover, 2013; Shen et al., 2019; Song et al., 2013). CPP have also been extensively used in other fields, including photogrammetry for unmanned aerial vehicles (UAVs), agriculture, fire, disaster management, and vacuum-cleaning robots (Fevgas et al., 2022; Galceran and Carreras, 2013; Seraj et al., 2022). Recently, researchers have begun to consider using reinforcement learning in CPP. Theile et al. (2020) used deep reinforcement learning algorithms for UAV CPP under different power constraints. Kyaw et al. (2020) used a new approach for solving CPP problems in large complex environments based on the traveling salesman problem (TSP) and deep reinforcement learning. Xi et al. (2022) integrated ocean information for a regional ocean simulation system combined with RL to generate AUV path-planning solutions. Jonnarth et al. (2023) used an end-to-end RL approach based on a continuous state and action space to address online CPP problems in unknown environments.

In the field of maritime SAR path planning, the primary objective is to optimize the shortest route from the starting point to the destination while avoiding potential obstacles along the path (Cao et al., 2019; Li et al., 2021; Liu et al., 2017; Xi et al., 2022; Yang et al., 2020; Zhang et al., 2019, 2020). However, accurately determining the location of individuals in distress during maritime accidents is challenging because of the varying postures of PIW and complex and constantly changing marine environments. Therefore, it is crucial to establish a search area and plan a path that ensures full coverage of the entire region. This is known as maritime full coverage search path planning (Ai et al., 2021). Compared with traditional CPP problems, the maritime search and rescue coverage path planning (MCPP) problem presents unique challenges. In addition to achieving complete coverage of the search area and avoiding path overlaps and obstacles, priority must be given to searching for high-probability areas.

To achieve this objective, traditional SAR operations used methods such as parallel track, crawl line, extended square, and sector searches (IAMSAR, 2016; Koopman, 1957). Recently, there has been a surge in research aimed at enhancing traditional search methods. Ramirez et al. (2011) used a collaborative model of UAVs and unmanned boats for maritime rescue coordination, which proved to be highly effective in completing rescue missions. Karakaya (2014) used an ant colony system optimization algorithm for route planning, aiming to efficiently cover the maximum search area with a limited number of UAVs. Xiong et al. (2021) introduced a helicopter maritime SAR path-planning method based on the minimum outer rectangle and k-means clustering algorithm. Cho et al. (2021) presented a mixed-integer linear programming (MILP) model that used a hexagonal grid decomposition approach to efficiently generate search paths for multiple heterogeneous UAVs within the shortest possible timeframe. Ouelmokhtar et al. (2022) used a multi-objective evolutionary algorithm, namely, the non-dominated sorting genetic algorithm II (NSGA-II) and Pareto evolutionary strategy (PAES), to solve the dual-objective CPP problem, that is, minimizing energy consumption and maximizing coverage, for UAV maritime monitoring. However, these methods have not considered the variability in the probability distribution of personnel in distress (PIWs) within the search area. Given that rescue time is critical for ensuring personnel safety, incorporating the PIW probability distribution can substantially enhance survival rates. Therefore, it is imperative to devise a path that maximizes SAR cumulative success rates (Ai et al., 2021; Bourgault et al., 2003; Cho et al., 2021; Frost, 2001; Yao et al., 2019).

Most current studies have primarily focused on single drowning person scenarios in search and rescue (SAR) operations. However, it is crucial to consider SAR scenarios involving multiple individuals in varying postures, such as the upright and face down positions, particularly during maritime accidents. For large-scale drowning accidents, the range of maritime SAR is large and requires the establishment of a multidimensional search and rescue area and multi-agent coverage path planning. One strategy for promoting collaboration among agents is to partition regions into distinct blocks and assign each agent a responsibility to a specific block (Xiong et al., 2021). The second approach is cooperative path planning for multiple agents (Binney et al., 2010; Cho et al., 2021; Mou et al., 2021). Throughout the search process, UAVs encounter several limitations, including a restricted battery life, vulnerability to adverse environmental conditions, limited search ranges, and challenges in detecting diminutive targets within water bodies (Hou et al., 2020). The Automatic Identification System (AIS) can provide information on vessels in proximity to the distress area, facilitating the allocation of ship resources for search and rescue operations. Therefore, it is imperative to investigate the planning of maritime coverage paths for vessel agents.

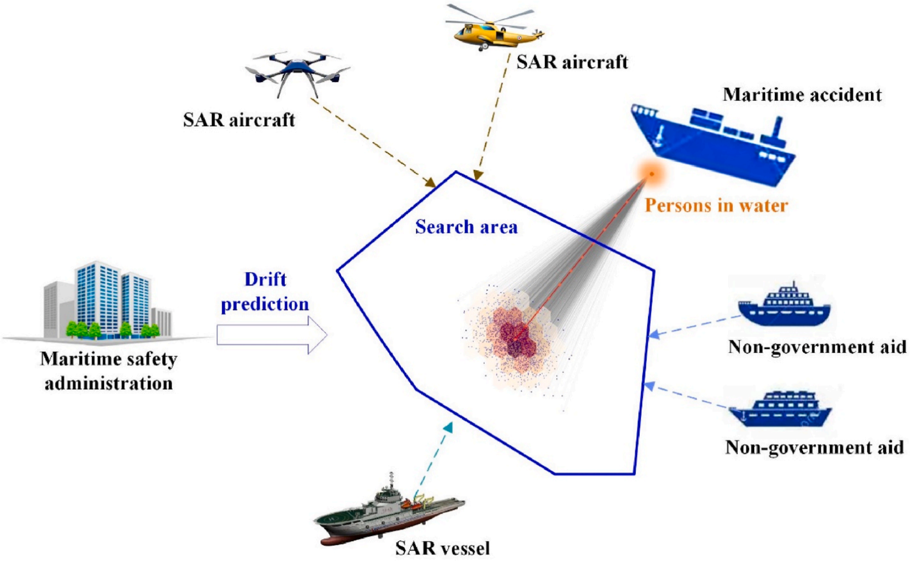


Fig. 1. Maritime SAR concept for PIWs.

This study integrated deep reinforcement learning techniques into the planning process of SAR coverage paths. First, sea-area-scale drift trajectory prediction models were used to predict the trajectories of persons in various sea areas. A random particle simulation algorithm was used to simulate the drift paths for different postures. Subsequently, a hierarchical probability map was established. By integrating deep reinforcement learning into the design of the covering path-planning algorithm, an improved success rate within a limited timeframe was achieved.

The main innovations of this study can be summarized as follows:

(1) A maritime search and rescue vessel path planning framework (SARCPPF) was proposed, which includes the prediction of the drift trajectory at the sea area scale, establishment of a hierarchical environment map of the search area for persons in water with multiple attitudes, and planning of the covering path.
(2) Developed a coverage path planning system with a multi-objective reward function based on deep reinforcement learning for maritime SAR. State and dynamically adjusted action-selection strategies applicable to specific maritime SAR scenarios were designed.
(3) For specific scenarios of maritime search and rescue, deep learning was introduced into search path planning, which achieves the goal of maximizing the cumulative success rate of search and rescue and provides a demonstration case for search path planning in high-dimensional state and action spaces.

The remainder of this study is organized as follows: In Section 2, the maritime optimal search theory, variables in search planning, and SARCPPF for this study are presented. Section 3 describes the sea-area-scale drift-trajectory prediction models and the drift prediction method. Section 4 introduces the modeling of the SAR environment. Section 5 introduces the SAR path planning algorithm combined with reinforcement learning. Section 6 presents a drift experiment of actual PIWs as a case study to perform a comparative analysis of the experimental results. Finally, Section 7 presents conclusions and prospects.

2. Maritime optimal search theory and maritime SARCPPF

Maritime optimal search theory serves as the foundation for determining search areas, dispatching SAR units, and assigning search tasks. Soza Company Ltd. (1996) and Frost (1997, 2001) distilled this theory into three critical components, that is, probability of containment (POC), probability of detection (POD), and probability of successful search (POS).

Maritime SAR aims to develop search plans and improve POS within the shortest possible time with limited search resources. POS relies mainly on POC and POD (Xiong et al., 2020):

$$\mathrm{POS} = \mathrm{POC} \times \mathrm{POD} \tag{1}$$

Therefore, maritime SAR aided decision-making involves two key issues: (1) optimal maritime SAR area determination, namely, full consideration and quantification of all influencing factors, such as distress waters, distress targets, and marine environmental conditions, in the drift process to predict the target trajectories and their final location probability distribution, and to determine the optimal search and rescue region, and (2) optimal planning of the maritime SAR, that is, based on SAR area determination, an optimal allocation scheme of SAR resources in time and space should be sought to improve the POS. The concept of maritime SAR for PIWs is illustrated in Fig. 1.

2.1. POC

Referring to the likelihood of an object being present within a search area, POC is a critical factor in search planning. Search planners need to allocate resources effectively to maximize their discovery potential. POC is expressed as a percentage and increases with larger search areas. When all particles can be contained within the region, POC reaches 100%. In actual maritime SAR missions, SAR units are often restricted in number, which requires SAR units to prioritize areas with high POC. Therefore, the search area is often subdivided into equally sized $A \times B$ square grids, with the size of the grid cells depending on the capability of the SAR detection equipment. The possibility of a SAR target being present in each subgrid was quantified by calculating the POC of each grid cell.

The specific equation for the calculation can be described as follows:

$$\mathrm{POC} = m_i / M \tag{2}$$

where m_i is the number of particles falling in cell i, and M is the number of particles contained in the overall distribution area.

2.2. POD

POD represents the probability of detection, indicating the likelihood that a search unit can detect a SAR target, and it is a crucial metric for

Table 1
Weather correction coefficients for generic search targets.

	Objects	
Winds (km/h) or currents (m)	PIW, life raft, or ship <10 m (33 ft)	Other objects
0–28 km/h or 0–1 m	1.0	1.0
28–46 km/h or 1–1.5 m	0.5	0.9
>46 km/h or > 1.5 m	0.25	0.9

Table 2
Sweep width tables of vessels.

	Meteorological visibility (km)				
Objects	6	9	19	28	37
PIW	0.7	0.9	1.1	1.3	1.3

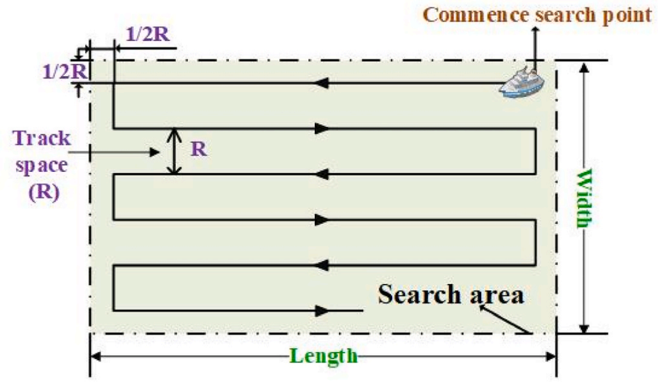


Fig. 2. Parallel line search for maritime search and rescue.

evaluating the effectiveness of an SAR detector in the search area (Abi-Zeid and Frost, 2005). Calculating the POD involves two important concepts, that is, sweep width and coverage rate. Sweep width refers to the effective distance within which the detector can locate the search object in a given search area. In this study, it serves as an indicator of the vessel's search capability.

Accurately determining the sweep width of the equipment requires statistical analysis of extensive experimental and practical samples, because different SAR targets in various search environments exhibit distinct horizontal range curves for each piece of equipment. Typically, a lateral curve can be plotted by analyzing large amounts of experimental data to assess the performance of a given device (Ai et al., 2019; Washburn and Kress, 2009; Wu and Zhou, 2015). Given that the actual sweep width can be affected by the performance of the SAR equipment, such as sensor performance, the characteristics of the search target, that is, physical characteristics such as size and color, and the maritime environment, such as wind, sea conditions, visibility, and sunlight reflection, it needs to be adjusted according to the actual situation.

IAMSAR (2016) provides the sweep width of the universal SAR equipment for generic search targets and the correction coefficient under different environmental conditions (Table 1). The sweep widths of the vessels are listed in Table 2. This table was compiled in the 1980s by the United States Coast Guard, which has conducted a large number of maritime SAR experiments according to the actual SAR environment. They measured the sweep widths of different search facilities under various conditions for different search targets (Anderson et al., 2006; Engel and Weisinger, 1988).

Coverage (C) is a measure of the degree to which a SAR unit's search area is covered during an operation (Burciu, 2010; Frost, 1997). This can be expressed as the effective coverage divided by the total search area. It is generally assumed that a vessel chooses to search using the parallel-line method, which requires fewer turns and is applicable to complex search scenarios (Fig. 2). The equation is as follows:

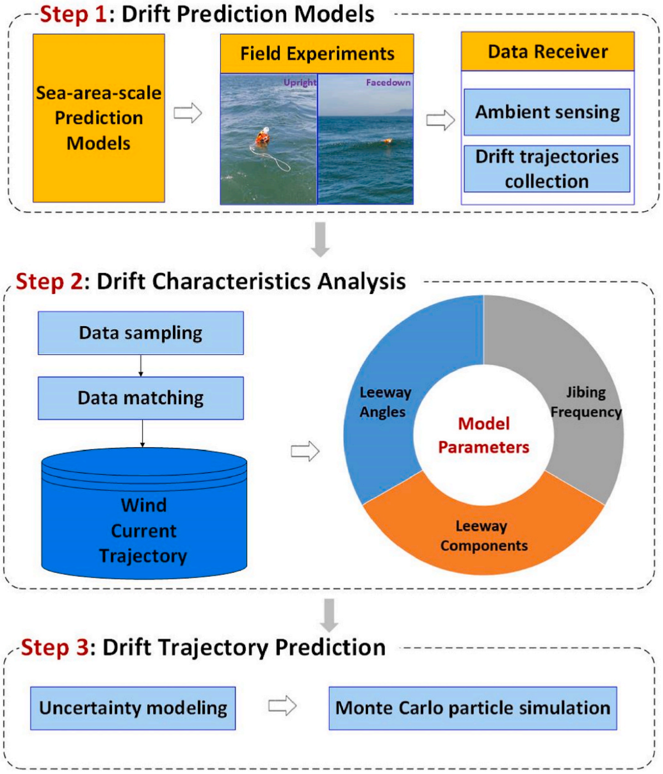

Step 1: Drift Prediction Models
Sea-area-scale Prediction Models
Field Experiments
Upright
Facedown
Data Receiver
Ambient sensing
Drift trajectories collection
Step 2: Drift Characteristics Analysis
Data sampling
Data matching
Wind
Current
Trajectory
Leeway Angles
Jibing Frequency
Model Parameters
Leeway Components
Step 3: Drift Trajectory Prediction
Uncertainty modeling
Monte Carlo particle simulation

Fig. 3. Framework for drift prediction.

$$C = \frac{W \times S}{A} = \frac{W}{R} \tag{3}$$

where W is the sweep width, S is the effective path length, A is the size of the search area, and R is the route spacing.

There is a close functional relationship between POD and coverage. Three models have been identified to describe this relationship, that is, fixed distance detection, inverse cube, and the random detection model (Abi-Zeid et al., 2011). Among them, the random detection model has been used to estimate the POD in a complex maritime SAR environment. Therefore, we used a random detection model in our study as follows:

$$\text{POD} = 1 - e^{-C} \tag{4}$$

2.3. Maritime SARCPPF

This paper presents a coverage path planning algorithm for search and rescue (SAR) vessels in maritime drowning accidents based on optimal SAR theory. The proposed SARCPPF consists of three modules, that is, drift trajectory prediction, SAR environment modeling, and coverage search. To predict the drift trajectories of PIWs with different postures in different sea areas, sea-area-scale drift trajectory prediction models (Wu et al., 2023) were used along with a random particle simulation method. All the predicted positions of the PIWs were then fused to generate a new prediction area. The search path planning region was determined based on the minimum bounding rectangle, and a hierarchical probability environment map was established to realize the SAR of multiple SAR units. A covering path planning algorithm combining deep reinforcement learning was proposed to enable rescue

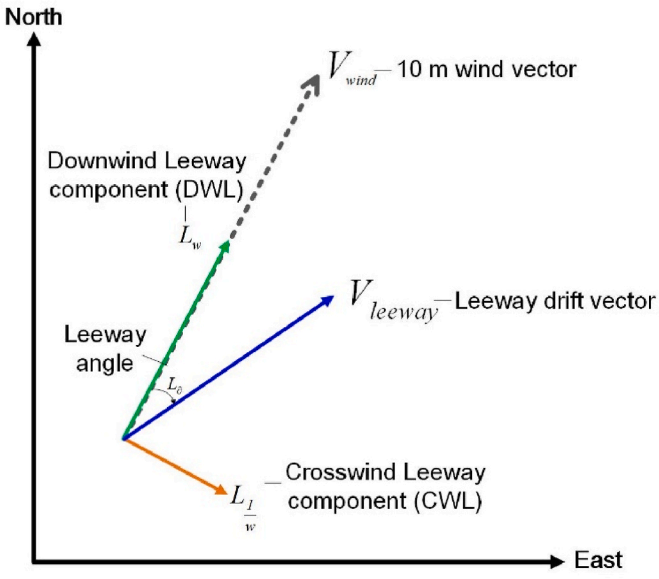

North
V_{wind} 10 m wind vector
Downwind Leeway component (DWL)
L_w
V_{leeway} Leeway drift vector
Leeway angle
L_α
Crosswind Leeway component (CWL)
$L_{\frac{1}{w}}$
East

Fig. 4. Relationships between the leeway, leeway angle, and the DWL and CWL components of the leeway (Breivik and Allen, 2008).

units to achieve fast arrival and high cumulative POS coverage, thereby increasing the SAR success rate in a limited time.

3. Predicting maritime drift trajectories

Drift theory investigates the impact of meteorological and oceanographic factors on object motion in marine environments and forms the basis for the mathematical methods used to determine search areas and routes in maritime SAR (Frost and Stone, 2001). In this study, sea-area-scale drift prediction models (Wu et al., 2023) were used to predict the trajectories of individuals with different postures in the Chinese sea areas, as illustrated in Fig. 3.

3.1. Drift prediction models for persons in the water

The Leeway model (Allen and Plourde, 1999; Allen, 2005; Allen et al., 2010; Breivik and Allen, 2008) was developed to analyze the impact of sea wind at a reference height of 10 m on the drift of various unpowered floating objects using Lagrangian particle simulation and probabilistic statistical analysis. Leeway is defined as “the motion of an object caused by wind and waves relative to currents (from depths ranging from 0.3 m to 1.0 m)” (Allen and Plourde, 1999; Breivik et al., 2013).

Marine environmental data can be used to compute the drift velocity vector of particles in water as follows:

$$\frac{dx}{dt} = v(x,t) \tag{5}$$

where dx is the change in the horizontal position of the floating object over time and $v(x,t)$ is the two-dimensional horizontal velocity.

$$v(x,t) = V_{F-current}(x,t) + V_{leeway}(x,t) + V_{F-wave}(x,t) \tag{6}$$

Among them, $V_{F-curren}(x,t)$ represents the velocity caused by the current, that is, the sea surface velocity. $V_{leeway}(x,t)$ represents the wind-induced drift velocity. $V_{F-wave}(x,t)$ is the wave-induced drift speed. Generally, the effect of wave forces is believed to be negligible for most targets in distress at less than 30 m in length (Breivik et al., 2011). Therefore, wave-induced drift velocity was excluded from this study.

The wind-induced drift speed can be decomposed into two components, that is, downwind speed (DWL) and crosswind speed (CWL), which are linearly correlated with wind speeds 10 m above sea level (Fig. 4), as demonstrated by Formula (7) in Allen (2005). The probability of +CWL and -CWL can be obtained from experimental statistics, where the CWL speed is deemed positive if it is to the right of the DWL. The direction of the crosswind speed changes from +CWL to -CWL or from -CWL to + CWL when the wind velocity falls within a specified range, which is referred to as jibing frequency (Allen and Plourde, 1999).

$$\begin{aligned} L_w &= c_w V_{wind} + b_w + \varepsilon_w \\ L_{\frac{1}{w}+} &= c_{\frac{1}{w}+} V_{wind} + b_{\frac{1}{w}+} + \varepsilon_{\frac{1}{w}+} \\ L_{\frac{1}{w}-} &= c_{\frac{1}{w}-} V_{wind} + b_{\frac{1}{w}-} + \varepsilon_{\frac{1}{w}-} \end{aligned} \tag{7}$$

Among them, L_w、 $L_{\frac{1}{w}+}$、 $L_{\frac{1}{w}-}$ represent the leeway components, c_w、 $c_{\frac{1}{w}+}$、 $c_{\frac{1}{w}-}$ are the linear regression slopes, b_w、 $b_{\frac{1}{w}+}$、 $b_{\frac{1}{w}-}$ are intercepts, ε_w、 $\varepsilon_{\frac{1}{w}+}$、 $\varepsilon_{\frac{1}{w}-}$ are the error terms. This equation is an unconstrained model, whereas the constrained method generates a linear fit with zero offset, as follows:

$$\begin{aligned} L_w &= c_w V_{wind} + b_w \\ L_{\frac{1}{w}+} &= c_{\frac{1}{w}+} V_{wind} + b_{\frac{1}{w}+} \\ L_{\frac{1}{w}-} &= c_{\frac{1}{w}-} V_{wind} + b_{\frac{1}{w}-} \end{aligned} \tag{8}$$

The leeway rate of each object at sea is highly specific and contingent on its exposure to wind, mass, and structures above and below the waterline. Simulating a person’s drift is an intricate process owing to the numerous uncertainties involved. Maritime meteorological conditions are often complex and are characterized by small-scale turbulence, vortices, stratification, and shear in near-surface currents. These issues are intricate, not easily discernible, and frequently pose challenges for resolution.

Therefore, in this study, we used the sea-area scale prediction models developed by Wu et al. (2023), which divided the Chinese coastline into distinct regions. We conducted drift tests that involved releasing manikins integrated with GPS devices and ship tracking to observe maritime environmental elements. Based on field experiment data, they modeled the drift trajectory at the sea area scale and generated predictive models for PIWs exhibiting different postures.

3.2. Drift trajectory prediction

In areas with complex marine environments, the accuracy of the Last-Known-Position (LKP) can be compromised, leading to significant deviations in the drift trajectory prediction. Therefore, alternative methods should be explored to improve the reliability of location information used as a starting point. To address this issue, we used the Monte Carlo simulation method to simulate the LKP error (Shchekinova and Kumkar, 2015). According to Breivik and Allen (2008), uncertainty modeling of the leeway parameters was performed. In the case of DWL:

$$L_w = (c_w + \varepsilon_w/20) \times V_{wind} + \left(b_w + \frac{\varepsilon_w}{2}\right) \tag{9}$$

where, $\varepsilon_W = S_{yx} \times Z$; S_{yx} is the standard deviation; Z is a random number which is normally distributed $N(0,1)$.

The maritime environmental data obtained may not accurately reflect real marine conditions owing to inherent limitations in measurement errors and other contributing factors. Therefore, a random walk model was used to effectively capture the uncertainty in marine environment data during the drift trajectory prediction process. Considering DWL as an example, the equation can be expressed as follows:

$$\begin{aligned} \mathbf{u}_m' &= (K)^{1/2} dw(t) \\ K &= \sigma_w^2 T \\ \mathbf{u}_m' &\equiv (u_m', v_m') \end{aligned} \tag{10}$$

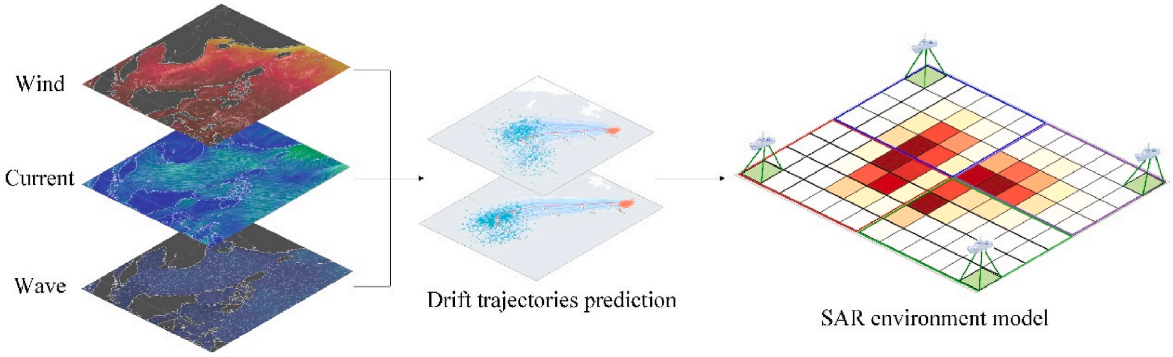


Fig. 5. Maritime search and rescue path planning environment modeling.

$$L_w = (c_w + \varepsilon_w/20) \times (\|V_{wind}| + \mathbf{u}_m^{'}|) + \left(b_w + \frac{\varepsilon_w}{2}\right) \tag{11}$$

where K is the diffusion coefficient, $dw(t)$ is a general random variable satisfying a normal distribution with mean 0 and a second moment of $2dt$, σ_w^2 is the variance of wind or current speed, and T is the integral time scale, usually $T = dt/2$.

The drift speed of the PIWs at any given time can be calculated based on the wind and flow data using Eq. (6). The drift trajectory can be determined by integrating the drift speed as follows:

$$loc_i(t) - loc_i(0) = \int_0^t \left[V_{drift}(t^{'})dt^{'}\right] = \int_0^t \left[V_{Leeway}(t^{'}) + V_{F-current}(t^{'})\right] dt^{'} \tag{12}$$

where $loc_i(t)$ is the location of the PIW at a given time t, $loc_i(0)$ are LKPs.

In this study, distress sea areas and different leeway coefficients of different search objects were introduced because drifting objects of different types and distress sea areas have different leeway coefficients. Search planners need to determine the types of distress targets when making search plans. However, in an actual operation, it is difficult to determine the type of distress object because the accident information obtained is not always sufficiently comprehensive. This requires the search planner to make strategic judgment based on existing accident information. This study has introduced multi-object leeway coefficients for various possible distress objects to further plan the search prediction areas.

The corresponding sea-scale drift trajectory prediction model was selected for the upright and facedown drowning personnel, and the drift trajectory prediction was performed accordingly. In maritime SAR, the target search range and the output probability distribution of the drift prediction model have proven to be of considerable importance in guiding search path planning. For each simulation, particle tracking was conducted using the Monte Carlo method to generate 1000 particles.

4. Maritime SAR environment modeling

The complexity and variability of the marine environment, coupled with the diverse postures of PIWs, have heightened the challenges in search-path planning. There is an urgent need for robust environmental modeling to facilitate future path planning based on the multi-posture PIW drift prediction results. Based on the simulation results, a new drift prediction area was generated by integrating all the PIW particles in different positions to simulate the drifting conditions of large-scale PIWs during maritime accidents. This area was then used to determine the search zone and generate a model for the maritime SAR environment. A path-planning algorithm was developed to enable searching among multiple vessels. The modeling process is illustrated in Fig. 5.

4.1. Establishment of the minimum bounding rectangle (MBR)

The Graham scanning algorithm (Graham, 1972; Kong et al., 1990) was used to generate the minimum convex hull. This algorithm consists of the following six steps.

(1) Find the bottom-left point from the point set that must be on the convex hull.
(2) Rank the remaining points according to the polar angle and compare the distance to the pole when the polar angles are the same, with the one closer to the pole taking precedence.
(3) Stack S was used to store the points on the convex hull, and the two smallest points sorted by pole angle and pole were pushed into the stack.
(4) Scan each point to check whether the line segment formed by the first two elements on the top of the stack and this point "turns" to the right (cross product ≤ 0).

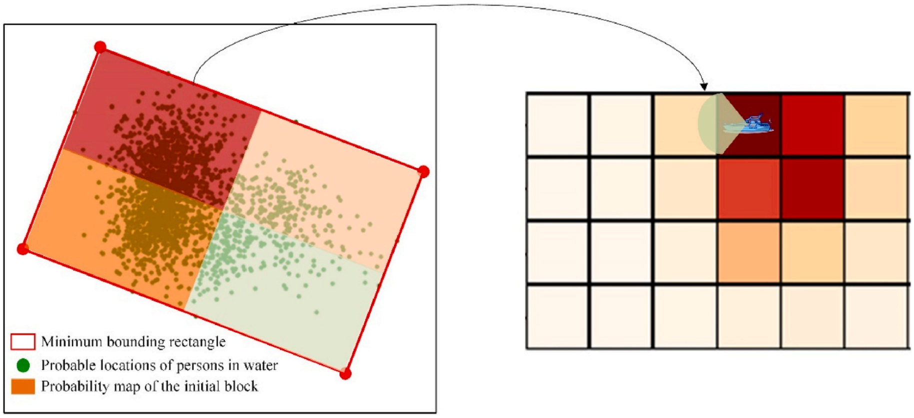

Minimum bounding rectangle
Probable locations of persons in water
Probability map of the initial block

Fig. 6. Hierarchical probability map modeling.

(5) If “yes,” pop up the top element of S and return to step (4) until “no,” then push this point onto S and proceed to step (5) for the other points.
(6) Vertex sequence of the convex hull is an element of the final stack.

The direct use of convex polygons as search areas is not conducive to search-path planning. Therefore, an MBR containing convex polygons should be generated to facilitate search-path planning (Cheng et al., 2008; Xiong et al., 2021). The procedure was as follows:

(1) Two points should be considered as the edge of the rectangle, using this edge as the base coordinate of the xy-axis.
(2) All the points are rotated around this base coordinate to find the minimum and maximum x-coordinates and the maximum y-coordinates of all points based on this edge; then, the area value of the range and the boundary data are obtained.
(3) Process (2) was repeated for each edge, and the MBR parameters with the minimum area were the output.

4.2. *Hierarchical probability map modeling*

In a scenario where multiple people fall into the water, the final search area may be relatively large, and multiple search and rescue forces are required to search simultaneously. Therefore, it is necessary to divide the search and rescue areas and conduct search and rescue path planning for each sub-area. In previous studies, the search area was primarily divided through a continuous expansion centered on the grid cell with the highest POC, potentially leading to locally optimal solutions. To address this issue, this study has proposed a new MBR and hierarchical path-planning algorithm based on the assumption of sufficient SAR units. The search area was determined based on the minimum bounding rectangle (MBR) by considering the integrated position distribution of the simulated particles at different PIW times. A hierarchical probability map was established, and each search and rescue unit proceeded directly to the highest probability area within its corresponding block for simultaneous search and rescue operations.

As shown in Fig. 6, the overall area was initially divided into large blocks of equal size, and the probability distribution of the particles in each block was calculated. A SAR unit was deployed for each block simultaneously. Subsequently, each block was divided into equal-sized $A \times B$ regular grids (Agbissoh Otote et al., 2019), with the grid cell size dependent on the sweeping width of the search ships (Galceran and Carreras, 2013). The POC was then calculated and colors were assigned to different grid cells according to their respective POC values, thereby generating a probability distribution map (Agbissoh Otote et al., 2019; Ai et al., 2019; Lee and Morrison, 2015; Xiong et al., 2020). In this study, we assumed that the environmental state was stable at a given time. Once the search area was defined, it remained unchanged with the development of the SAR process.

5. Maritime SAR coverage path planning based on deep reinforcement learning

In this study, we have proposed an autonomous coverage path-planning algorithm for multiship search and rescue (SAR) units based on deep reinforcement learning. Prior research (Wu et al., 2023) demonstrated the superior trajectory prediction accuracy and search area of the sea-area-scale drift trajectory prediction model. Therefore, in our study, each SAR unit navigates directly to the highest probability grid of its corresponding block using an environmental map established from the drift simulation results at a given time. Search path planning is the process of selecting navigation actions according to the current SAR environment information.

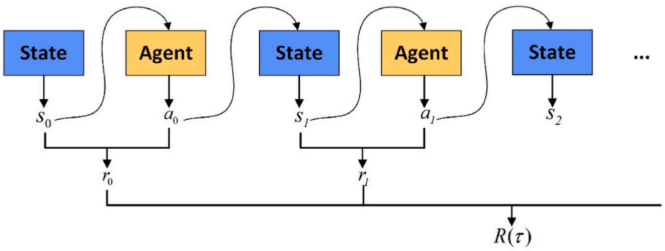


Fig. 7. Markov decision process diagram.

Table 3
The symbols and the corresponding explanations of the maritime SAR path planning model.

List of symbols	Explanations
S	The state space of the environment.
A	The action space of the SAR unit.
P	The state transition model.
R	The reward value function.
γ	The discount factor.
s_t	The state at time t.
S_{done}	The state when the agent passes the whole area.
a_t	The action taken at time t.
r_t	The reward value at time t.
R_t	The cumulative reward value.
R_{POC}	The reward value of POC.
R_{search}	The reward value if the next state $\mathrm{S}_{t+1} \notin H$.
R_{done}	The reward value that the agent passes the whole area.
POC_i	The POC value at step i.
$step_i$	The number of steps experienced at step i.
T	The maximum number of steps that the agent can take.
Π	The strategy of the agent.
$\Pi(\mathrm{a}\|\mathrm{s})$	The action selection strategy.
H	The set of grid units searched by the agents.
U	The set of grid units has not been searched by the agents.
F	The set of grid units whose POC value is 0.
ε	The probability that a random action selection is conducted.
A^*	The action with the highest Q value at the current state.
A_{choose}	The set of available actions under the current state.
$V_\Pi(s)$	The state value function.
$Q_\Pi(s,a)$	The action value function.

5.1. *Overall maritime SAR path planning process*

When planning maritime SAR routes, the next stage of the SAR unit depends only on the previous state and action, which can be expressed as a Markov Decision Process (MDP) (Sutton and Barto, 1998; Mnih et al., 2013). An MDP process is an interaction process between the environment and the agent, which includes three signals, namely, state (S), action (A), and reward (R). It provides direct feedback on the results generated by interactions with the environment (E). The agent receives the states at each discrete time step and selects the corresponding actions to transform them into new states. This transformation process then generates an evaluation value reward. The new state is acquired by the agent, and the cycle is repeated, as shown in Fig. 7.

5.2. *Algorithm structure*

5.2.1. *The Markov decision process of maritime SAR path planning*

The expression for maritime SAR path planning includes a vessel agent and two sets (state set S and action set A). By selecting and executing an action from the action set A, the agent completes a state transition. During reinforcement learning (RL), the vessel’s goal is to maximize the cumulative reward. This process mainly contains quintuples (s、a、p、r、γ). The symbols and their corresponding explanations are listed in Table 3.

S is the state space of the environment, that is, the limited state that the SAR unit can achieve (Minsky, 1967). A is the action space of the SAR unit, which consists of all possible actions that the vessel agent can choose through strategy selection in each environmental state. In this study, the action space of the vessel agent is discretized, meaning that, from one grid to another, there are only four actions to reduce the negative impact of irregular searches on the security of searches (IAMSAR, 2016). P is the state transition model; that is, the probability that in the current state s of the environment, where an agent causes this s to transfer to another state $s^{'}$. R is the reward value function fed back to the agent by the environment in the form of encouragement or punishment. The strategy of agent $\mathit{\Pi}$ is to map S to A. If state $s_t \in S$, the agent takes action $a_t \in A$, and moves to the next state s_{t+1} according to P, meanwhile receives a reward value $\mathrm{r}_t \in R$. The discount factor γ is used to calculate the cumulative value of returns over time. We have provided a detailed description of reward and action selection policies.

- Reward function

A suitable reward function is required to specify the training objective. The advantages and disadvantages of a vessel agent in the learning process can be determined using the reward function. This enables it to achieve its goal in the shortest time. In maritime SAR coverage path planning, a ship agent is required to search the overall SAR area under the conditions of prioritizing search areas with high probability and ensuring that no duplicate paths are taken.

Sets H and U are introduced to mark the position information of the grid units that the agent has searched for and has not searched for. At model initialization, set $H = \{s_0\}$ and $U = \{s_1, s_2, s_3, \cdots, s_n\}$. Grid units in the hierarchical environment map with a POC value of 0 were not searched and were denoted as set F to reduce the search time. Set F is then added to set H and it is removed simultaneously from set U.

After the ship's agent selects an action according to the action selection strategy, it arrives at state s_{t+1} and determines whether s_{t+1} is already in set H. If $\mathrm{s}_{t+1} \notin H$, positive rewards R_{search} is feedback, and the state s_{t+1} is then added to set H while being removed from set U. This can guide the agents to cover all the SAR areas. The reward function design should consider the priority search of high-probability grids, namely, the POC reward, which is calculated using the following equation:

$$R_{POC} = \frac{1}{step_i} POC_i \times 1000 i \in (0, T) \tag{13}$$

where T is the maximum number of steps that the vessel agent can take in each iteration; POC_i is the POC value of the SAR grid in the next state that the agent reaches; and $step_i$ is the number of steps experienced by the ship agent in the current state. As the number of steps increases, R_{POC} decreases, thereby guiding the ship agent to search the high-probability region first.

R_{done} is given once the ship agent passes through the entire search area, and they enter the termination state S_{done}. The reward function was set as follows:

$$\mathrm{R} = \begin{cases} R_{search} + R_{POC} & s_{t+1}! = S_{done} \&\& s_{t+1} \notin H \\ R_{done} & s_{t+1} = S_{done} \\ 0 & else \end{cases} \tag{14}$$

- Action selection policy

In reinforcement learning, the two crucial concepts of exploitation and exploration need to be balanced. Exploitation involves selecting the optimal action for the vessel agent by maximizing the value of all known state-action pairs. However, if the vessel agent chooses randomly from its set of actions, it is referred to as exploration. While exploitation helps maximize the expected rewards in real time, it may lead to local optima. By contrast, exploration helps maximize total returns in the long run.

In this study, we have proposed an action selection strategy that balances exploitation and exploration to achieve a global optimal solution. In the early stages of reinforcement learning, the vessel agent prioritizes exploration with a high probability. As the number of learning episodes increases, the probability of exploration gradually decreases, whereas the probability of exploitation increases. In this study, a ε-greedy strategy was used (Tokic, 2010). A random action selection is conducted with the probability of ε to explore the new environment. Meanwhile, action a with the highest Q value is selected with the probability of $1 - \varepsilon$. The equation used is as follows:

$$A^* \longleftarrow arg\ max_a Q(s, a) \tag{15}$$

$$\Pi(\mathrm{a}|\mathrm{s}) \longleftarrow \begin{cases} 1 - \varepsilon + \dfrac{\varepsilon}{|A(s)|} & if\ a = A^* \\ \dfrac{\varepsilon}{|A(s)|} & ifa \neq A^* \end{cases} \tag{16}$$

where $|A(s)|$ denotes the number of actions performed in the current state.

During learning, a random number $rand$, $rand \in (0, 1)$ is generated. If $rand < \varepsilon$, the action is selected at random, and if $rand > \varepsilon$, the action with the highest Q value in the current state is selected. To ensure model stability and obtain the global optimal solution, the value of ε is dynamically adjusted in the iterative calculation as follows:

$$\varepsilon = 1 - episode / L \tag{17}$$

where $episode$ is the current episode number and L is the maximum learning episode.

Boundary assessment, repeated search assessment, and termination conditions were added to the action selection process to prevent the model from looping endlessly. Each time a new state s_t is reached, the set of available actions A_{choose} under the current state is initialized as $[True, True, True, True]$, and the action a_t is selected using an ε-greedy approach. Upon reaching the next state s_{t+1}, if the current state s_t is located at the boundary of the search area and the state s_{t+1} is beyond the SAR area, or if $s_{t+1} \in H$, the action is reselected and the corresponding action in A_{choose} is marked as False. If $A_{choose} = [False, False, False, False]$, the termination condition is reached and the current episode ends.

The objective of reinforcement learning is to optimize the long-term cumulative reward for vessel navigation, rather than focusing on short-term rewards. With the introduction of $\gamma \in [0, 1)$, the feedback value can be described as follows:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ldots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{18}$$

The state value function $V_\Pi(s)$ is the evaluation of the quality of the current state. Each state's value depends not only on its current state, but also on its subsequent states. The value of the $V_\Pi(s)$ of the current s is obtained by calculating the expectation of the accumulated reward R_t of the state:

$$V_\Pi(s) = E_\Pi[R_t | s_t = s] \tag{19}$$

The action value function of the state-action couple (s, a), denoted as $Q_\Pi(s, a)$, evaluates the long-term payoff to the agent through the use of strategy $\mathit{\Pi}$:

$$Q_\Pi(s, a) = E_\Pi[R_t | s_t = s, a_t = a] \tag{20}$$

The optimal decision sequence of the MDP is solved using the Bellman equation, which is the transformation relation of the value function:

$$V_\Pi(s) = \sum_a \Pi(\mathrm{s}, \mathrm{a}) \sum_{s^{'}} P^a_{ss^{'}} \left[R^a_{ss^{'}} + \gamma V_\Pi(s^{'})\right] \tag{21}$$

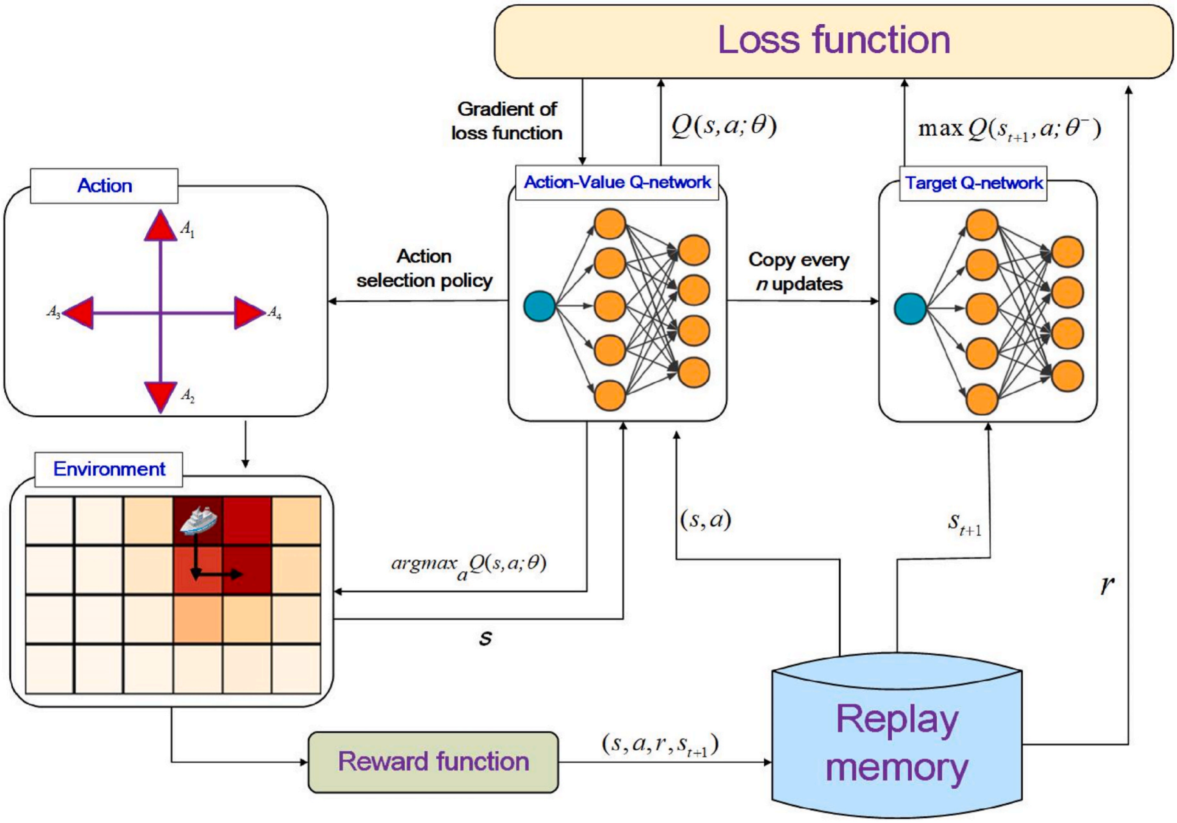


Fig. 8. The path planning model based on DQN.

$$Q_{\Pi}(s,a)=\sum_{s'}P_{ss'}^{a}\left[R_{ss'}^{a}+\gamma\sum_{a'}Q_{\Pi}(s',a')\right] \tag{22}$$

where, $P_{ss'}^{a} = P(s_{t+1} = s'|s_t = s, a_t = a)$, $R_{ss'}^{a} = E[r_{t+1}|s_t = s, a_t = a, s_{t+1} = s']$.

Monte Carlo learning and time difference learning (TD) are used to approximate the solution of the Bellman equation, while continuously optimizing of the value function to improve Π. Watkins first proposed the Q-learning algorithm (Watkins, 1989; Watkins and Dayan, 1992), combining the Bellman equation, MDP, and other theories with TD learning. TD Learning combines a Monte Carlo sampling method with a dynamic programming method, estimating the current value function from the value function of the subsequent state. The value function was computed as follows:

$$V_{\Pi}(s) \leftarrow V_{\Pi}(s) + \partial(R_{t+1} + \gamma V_{\Pi}(s') - V_{\Pi}(s)) \tag{23}$$

where $R_{t+1} + \gamma V_{\Pi}(s')$ is the TD target, $\delta_t = R_{t+1} + \gamma V_{\Pi}(s') - V_{\Pi}(s)$ is the TD bias, ∂ is the learning rate.

5.2.2. *Maritime coverage path planning model based on deep reinforcement learning*

RL has an edge in decision-making, and the deep learning approach combines low-level features to form more abstract high-level features or categories, approximating a nonlinear function, and excels in perception (Hinton et al., 2006). Combined with the characteristics of deep learning, the use of deep neural networks as function approximators can substantially improve RL performance of reinforcement learning. DRL integrates RL and deep learning to complement each other and provides a more effective solution to the perception and decision problems of the system. The Q-learning algorithm builds a Q-table to iterate over the values of all existing state-action pairs in the storage environment and then reads these values through queries. DQN uses a general function approximator (artificial neural network) to replace the stored Q value. The main idea is to replace the traditional Q table with a deep neural network trained from stored experience samples, build a “memory” of selected experiences, and train the Q network on a subset of states rather than on all states that the agent sees. Given that there is no network, Q-learning is too dependent on the state and may lead to insufficient learning (Cao et al., 2019; Fang et al., 2021; Zhu and Zhang, 2021; Meng et al., 2021).

Most current studies consider a single person falling into water as an example to determine the search area and plan the search path. This study has considered the integrated search and rescue area of people falling into water using different gestures. In this context, the state space is two-dimensional and the action space is discrete. In future studies, large-scale drowning accidents caused by ship collisions, as well as the collaborative search and rescue of drones and ships involving a three-dimensional state space and continuous action space, will be further considered. In the study of high-dimensional problems, the amount of computation required for the traditional RL algorithm increases sharply with an increase in the number of inputs, and it is difficult to determine an effective strategy. When the environment expands, it may cause a memory burden and lead to a failure in obtaining the optimal solution. A DQN uses neural networks to estimate values and overcomes the shortcomings of Q learning.

Mnih et al. (2015) published their work on DQN in Nature using a convolutional neural network (Lecun et al., 1998) to express the action value function and train it based on rewards. The Q-network approximation of the Q-value calculation is expressed as

$$\mathrm{Q}(\mathrm{s},\mathrm{a};\theta) \approx Q_{\Pi}(s,a) \tag{24}$$

The DQN used in this study consists of two fully connected layers. Feature extraction and nonlinear combinations are performed to obtain the Q-value evaluated by the network for each action. The main characteristics of a DQN are as follows (Zhu and Zhang, 2021):

Table 4
Pseudo-code of the DQN training.

Algorithm. Training algorithm of DQN

```
Input: initial state $s_0$, maximum learning episode $L$, maximum step size $T$, threshold $\varepsilon$, size
       of batch $B$, memory length $M$, number of steps to copy the target Q-network $n$
Output: trained Q-network
Initialize the experience buffer pool D
Initialize the current Q-network, generate random parameters $\theta$.
Initialize the target Q-network, parameters $\theta^{-} \leftarrow \theta$.
Set episode = 1
While (episode⩽L) do
    Initialize state list $s_0$
    Set time step $t = 0$
    While ($t$⩽$T$) do
        $rand \leftarrow$ random()
        Q ← predict ($\mathrm{s}_t$)
        Get the available task list AL under state $\mathrm{s}_t$
        $\mathrm{a}_t \leftarrow \begin{cases} \text{Randomly taken a action from AL} & rand \leq \varepsilon \\ \mathrm{argmaxQ}(A)|\mathrm{A} \in \mathrm{AL} & \text{others} \end{cases}$
        Calculate reward $\mathrm{r}_t$ by Eq. 14
        Transfer to state $\mathrm{s}_{t+1}$
        Add the experience set of ($\mathrm{s}_t$, $\mathrm{a}_t$, $\mathrm{r}_t$, $\mathrm{s}_{t+1}$) to the experience buffer pool D.
        $t \leftarrow t + 1$
        For $i$ = 1: min ($M$, $B$) do
            load a record ($\mathrm{s}_j$, $\mathrm{a}_j$, $\mathrm{r}_j$, $\mathrm{s}_{j+1}$) from experience buffer pool D
            add the state $\mathrm{s}_j$ to the set H
            Q (s, a) ← max (predict ($\mathrm{s}_{j+1}$))
            If $\mathrm{s}_{j+1}$ is the terminal state, then
                targets ←$\mathrm{r}_j$;
            Else
                targets ←$\mathrm{r}_j + \gamma * \mathrm{Q}(s, \mathrm{a})$;
            End If
            Calculate loss using mean square loss function
            Update the current Q-network $\theta$ using gradient descent algorithm.
        End For
        Copy the target Q-network every $n$ step, update parameters $\theta^{-} \leftarrow \theta$.
    End While
End While
Return: Q-network
```

Table 5
The experimental overview of maritime SAR simulation.

No.	Posture	Start position	Start time (UTC+8)	End time (UTC+8)	Wind speed (m/s)	Current speed (m/s)
1	upright	119.885°E 25.455°N	2021/04/17 09:45	2021/04/17 16:00	2.7–7.1	0.17–0.63
2	facedown	119.885°E 25.455°N	2021/04/17 09:45	2021/04/17 16:00	2.7–7.1	0.17–0.63

We used Taiwan Strait drift prediction models for PIW with upright and facedown postures, namely TS_ I and TS_ II. In this section, unconstrained models are used to predict drift trajectories.

(1) Target network

To mitigate the instability that arises during Q-function updates, a target network is introduced to obtain the Q value before updating the Q function. The new Q function is then used to update the target network; that is, every n steps, the parameter θ_i of the current Q network will be copied to the target Q network $Q(s^{'}, a^{'}; \theta_i^{-})$.

(2) Experience pool U (D)

The experience pool constructs a replay buffer, also known as a replay memory, to store and manage samples (s_t, a_t, r_t, s_{t+1}). The experience replay mechanism was used, and minibatch (B) samples were randomly selected for training the Q-network.

The current Q-network parameters are updated using the gradient descent method, and their loss function is expressed as:

$$L(\theta_i) = \frac{1}{2}\left(r + \gamma \max_{a^{'}} Q\left(s^{'}, a^{'}; \theta_i^{-}\right) - Q(s, a; \theta_i)\right)^2 \tag{25}$$

The derivative $\nabla_\theta L$ of the parameter θ in the loss function is calculated as follows:

$$\nabla_\theta L = \left[r + \gamma \max_{a^{'}} Q\left(s^{'}, a^{'}; \theta_i^{-}\right) - Q(s, a; \theta_i)\right] \nabla_\theta Q(s, a; \theta_i) \tag{26}$$

The proposed DQN-based search path planning model obtains the optimal search path through interactive learning between agents and a maritime SAR environment model. The vessel agent arrives at the highest POC grid unit of each block and begins searching. The path planning model is illustrated in Fig. 8. Table 4 lists these algorithms.

Maritime SAR coverage path planning based on deep RL algorithms is divided into two phases, that is, training and testing. In the training phase, the time complexity of the DQN is $O(n_s * n_d)$ when forward propagation is performed. Here, n_s is the number of states and n_d denotes the state feature dimension. When backward propagation is performed, the time complexity of the DQN is also $O(n_s * n_d)$. Therefore, the time complexity is $O(n_s * n_d)$. When training *epoch* rounds, the forward propagation time complexity is $O(epoch * n_s * n_d)$. In the testing phase, the time complexity of DQN is also $O(n_s * n_d)$.

6. Results and discussion

6.1. Maritime SAR experimental settings

To verify the validity of the proposed SARCPPF, a real offshore drift experiment was used as a case study for the calculations and analysis. On April 16–17, 2021, sea-drift experiments with manikins in different postures (upright and facedown) were conducted in the Pingtan waters of the Taiwan Strait. The experimental overview is presented in Table 5.

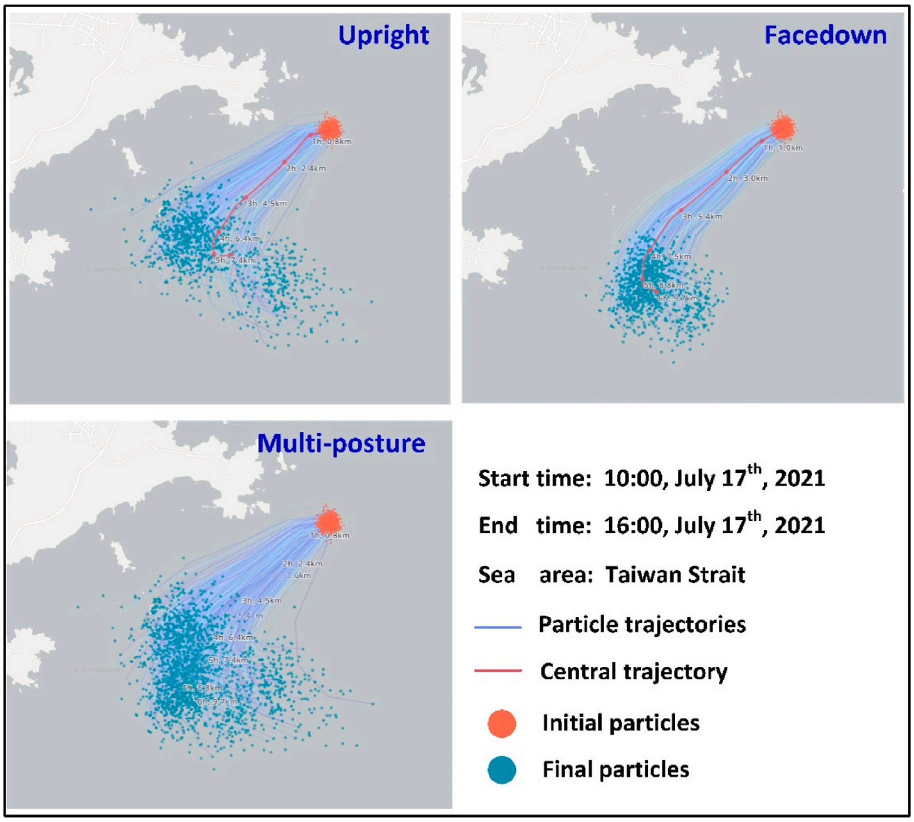
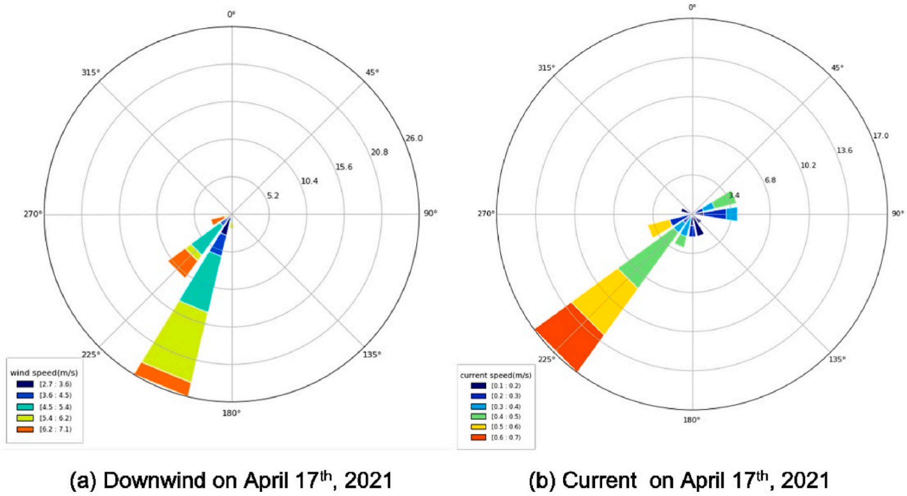
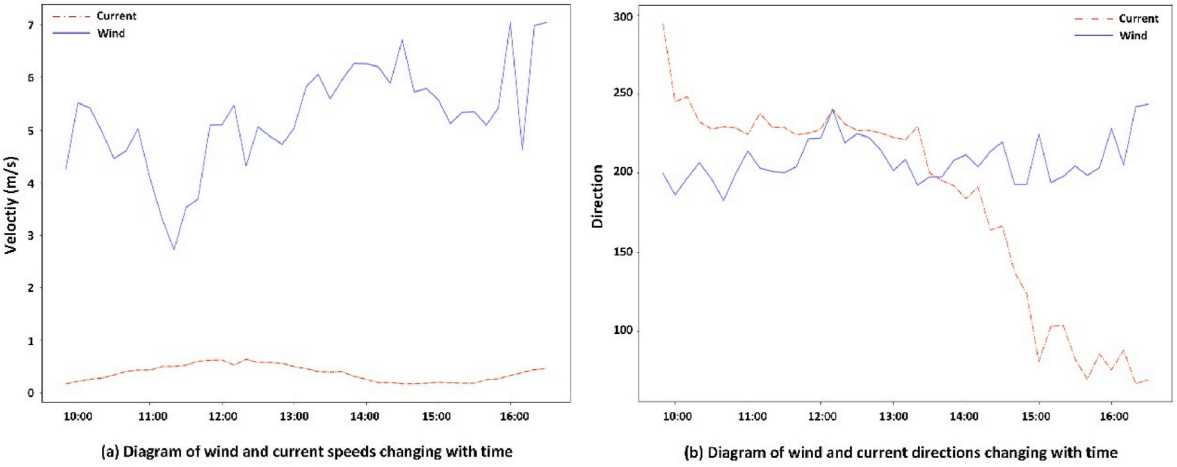


Fig. 9. Drift trajectory prediction results.



(a) Downwind on April 17th, 2021 (b) Current on April 17th, 2021

Fig. 10. The marine environment data used in the modeling process ((a) downwind speed and direction, (b) current speed and direction).



(a) Diagram of wind and current speeds changing with time

(b) Diagram of wind and current directions changing with time

Fig. 11. The marine environment data and their changes with time used in the modeling process ((a) velocity changes, (b) direction changes).

6.2. *Results of drift trajectory prediction*

We took the release point of the manikins as the initial point and used the hydrometeorological data measured on April 17, 2021, the possible 6 h drift trajectories of PIWs in vertical and facedown postures were calculated, as shown in Fig. 9. A 6 h drift trajectory and particle distribution map of PIWs was generated with the fusion of multiple postures. This was used to simulate the possible position distribution of survivors when SAR vessels arrive at the scene of multiple drownings in a maritime accident. In this study, it was assumed that the data of the possible position distribution were constant at the 6 h drift time.

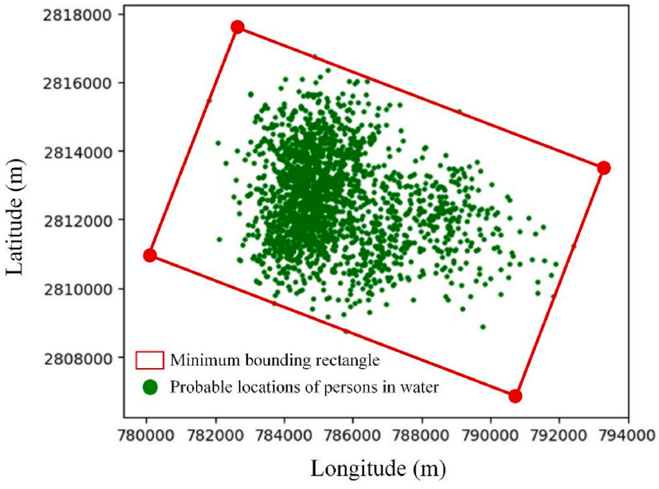


Fig. 12. The MBR of the area for SAR path planning.

The ranges of the wind and flow velocities measured (10 min average) during the experiment were counted, and a detailed statistical

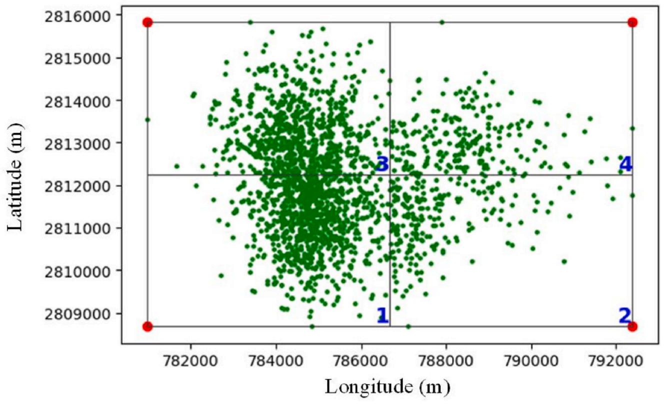


Fig. 13. The initial block division results.

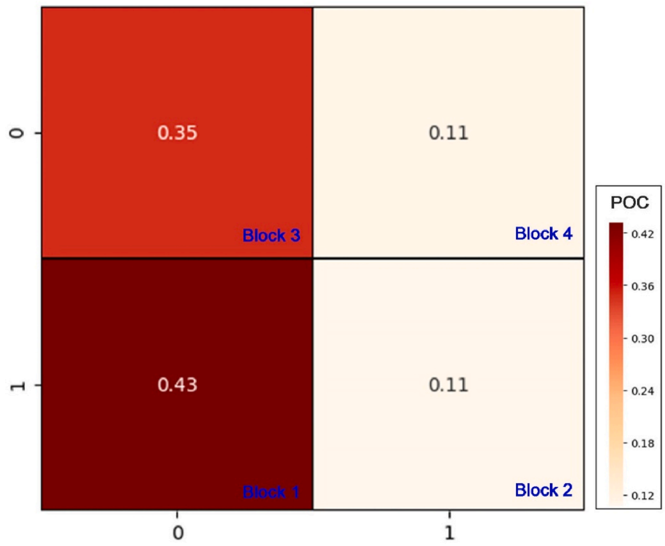


Fig. 14. The one-level hierarchical environment map.

analysis of the sea state was conducted. The wind speed range was 2.7–7.1 m/s, and the downwind direction varied in the southwest direction, and the variation range was less than 90° with relatively little fluctuation. The wind speed was divided according to wind conditions, and a rose diagram of the wind direction was drawn. The flow speed range was 0.17–0.63 m/s, and the flow direction varied from southwest to northeast with counterclockwise deflection over time. Based on the flow conditions, the flow speeds were divided, and a flow rose diagram was drawn. As shown in Figs. 10 and 11, the marine environment data and changes used in the modeling process are depicted.

Table 6
The parameters of the algorithm proposed in this study.

Parameter	Value	Description
L	5000	maximum learning episode
T	200	maximum step size
γ	0.5	discount factor
LR	0.1	learning rate
n	50	target network update frequency
ε	0.9	initial action selection strategy
B	32	batch size
M	1000	memory length
Layers	10	the number of neurons in each hidden layer
N_STATES	2	the input neurons
N_ACTIONS	4	the output neurons
Optimizer	Adam	optimizer

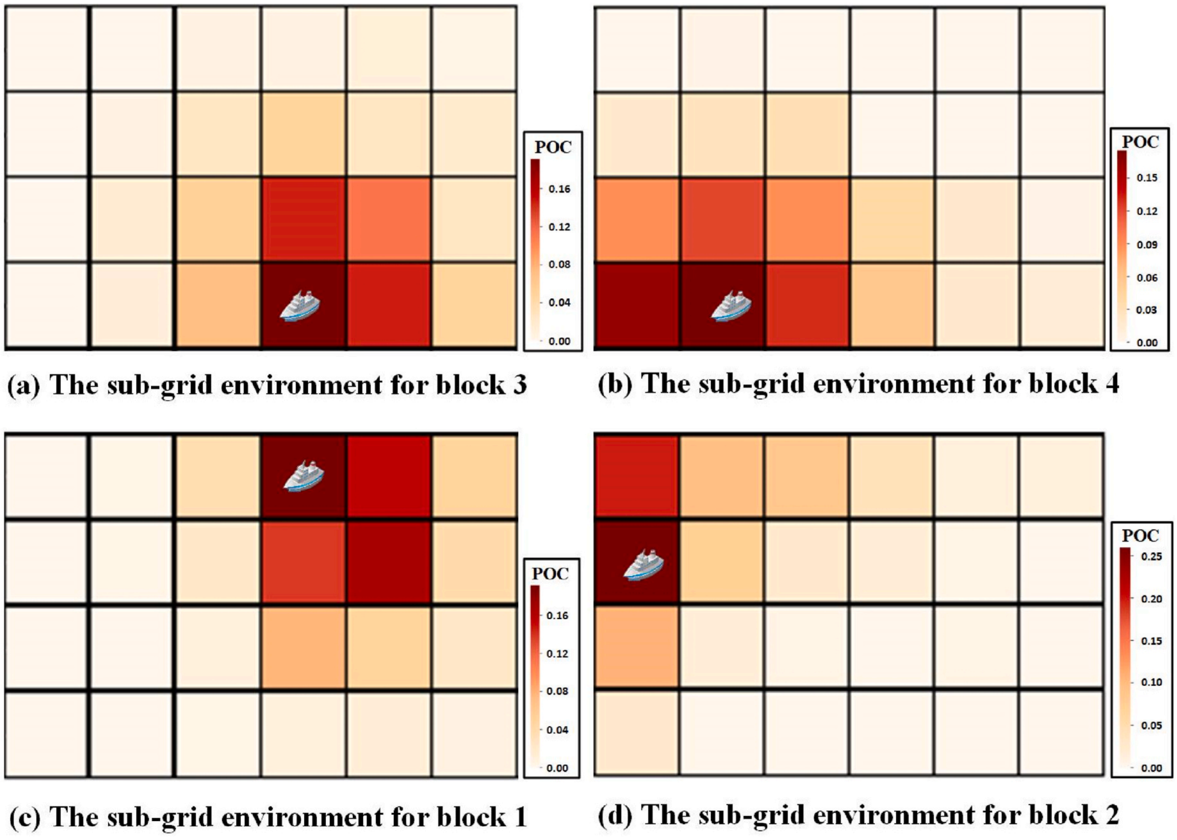
6.3. *Results of maritime SAR environment modeling*

6.3.1. *Results of the establishment of MBR*

The MBR of the area for SAR path planning according to the final particle distribution of the multi-posture PIWs is shown in Fig. 12. Assuming that four nearby vessels can be engaged in simultaneous search and rescue after rotation, the initial block-division results are as shown in Fig. 13.

6.3.2. *Results of the hierarchical probability map*

A one-level hierarchical environment map was generated. According to visibility reanalysis data from the European Center for Medium-Range Weather Forecasts (ECMWF), visibility in Pingtan was 18.5 km on April 17, 2021. Based on the sweep width table (Table 2), the sweeping width of the vessel agent was set to 1 km. From the measured marine environment data (Table 5 and Fig. 10), the value of the weather correction coefficient was 1. Therefore, the corrected sweep width of the vessel is the same as the unadjusted sweep width.



(a) The sub-grid environment for block 3

(b) The sub-grid environment for block 4

(c) The sub-grid environment for block 1

(d) The sub-grid environment for block 2

Fig. 15. The sub-grid hierarchical environment map.

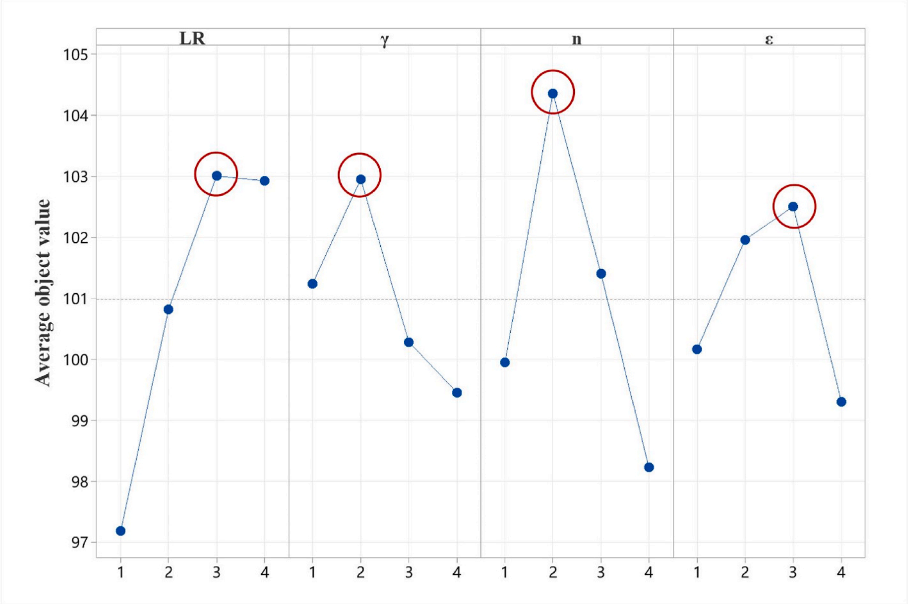
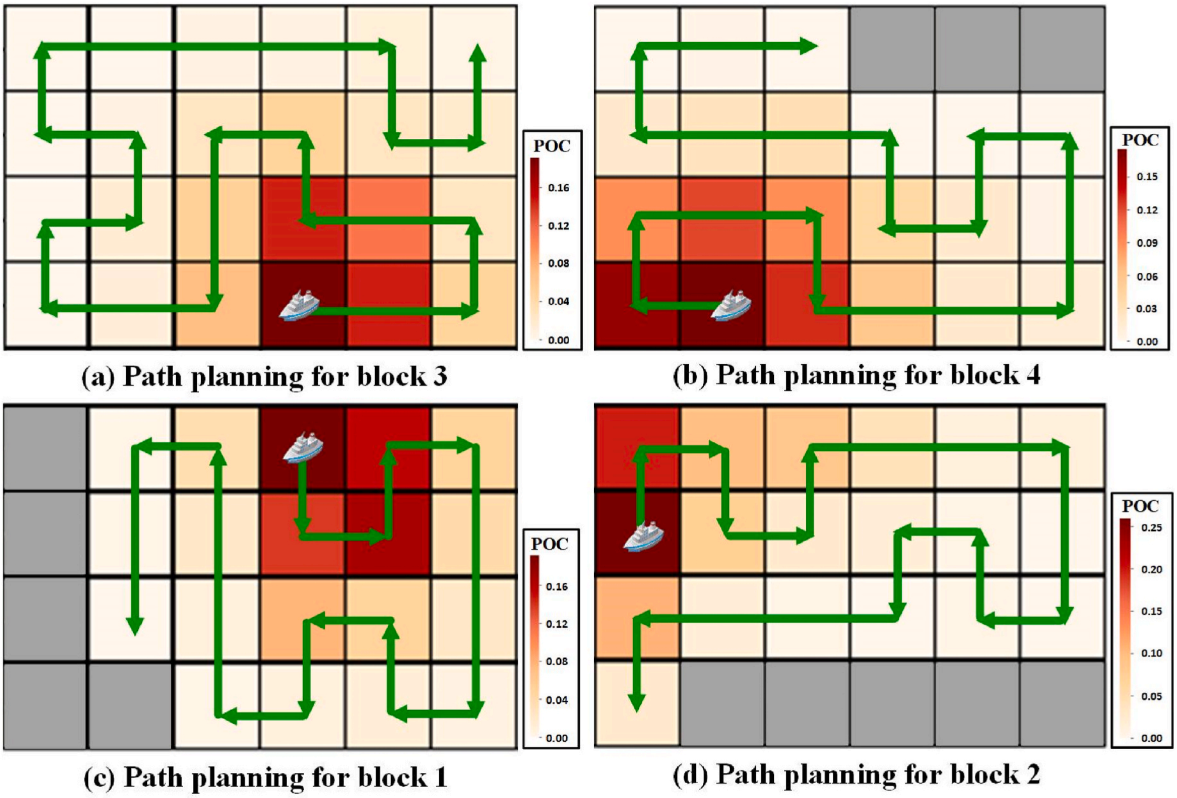


Fig. 16. The average results of our object for different parameter settings.



(a) Path planning for block 3

(b) Path planning for block 4

(c) Path planning for block 1

(d) Path planning for block 2

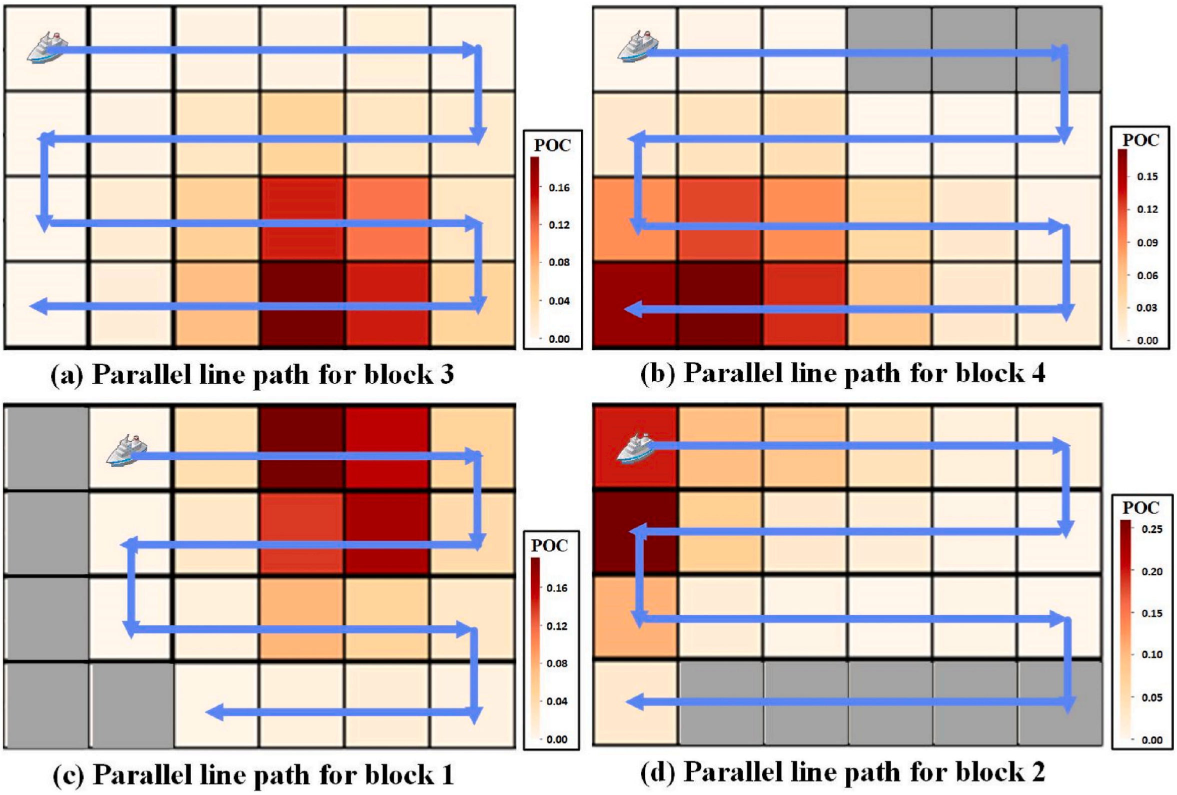
Fig. 17. The path planning results obtained by the model proposed in this study.

Based on the corrected sweep width, each initial block was partitioned into grid units to generate a hierarchical environmental map. The grid cell size and track spacing were determined using the corrected sweep-width. The simulated scene was situated in an open sea without natural or artificial obstacles. The hierarchical probability environment map and the initial position of the SAR vessels are shown in Figs. 14 and 15.

6.4. *SAR coverage path planning results*

- Experimental settings

Simulation Environment: All the simulation experiments in our study were conducted on a desktop computer with an Intel (R) Core (TM) i7-1260P 2.10 GHz CPU, 16 GB of RAM, and Windows 11 operating system, using the Python programming language.



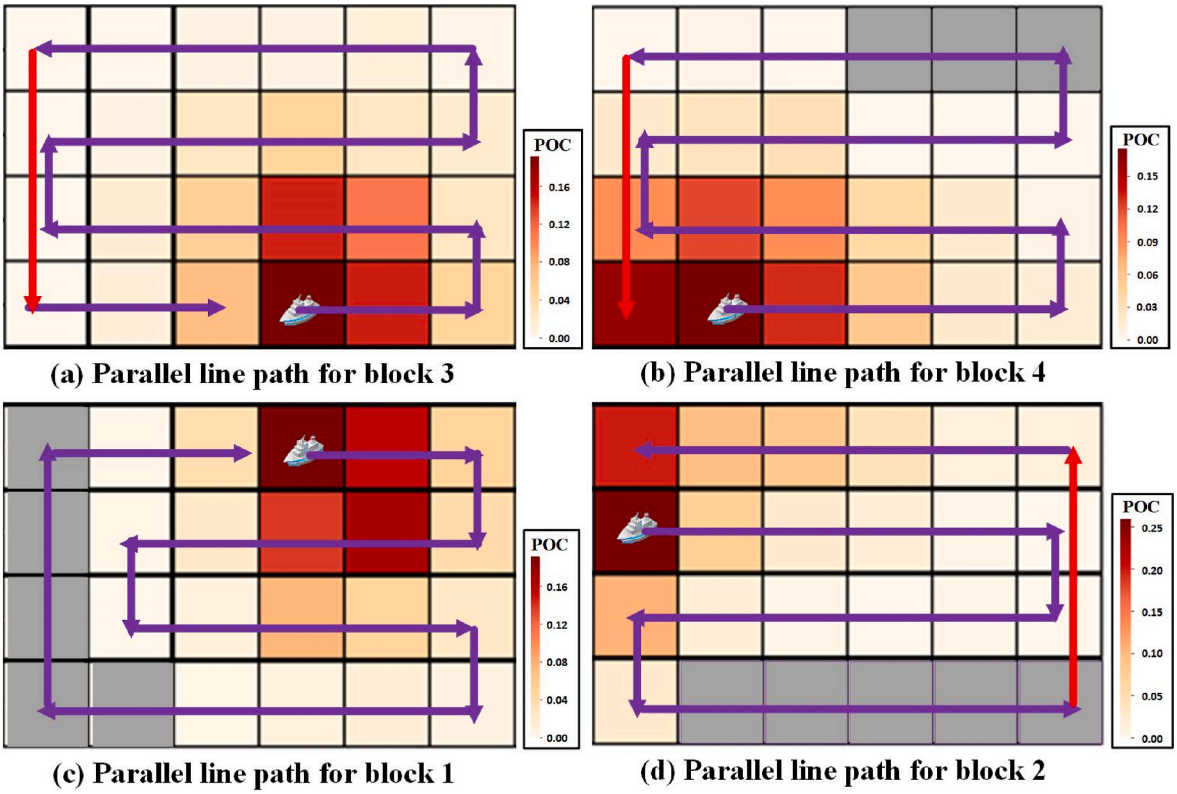
(a) Parallel line path for block 3

(b) Parallel line path for block 4

(c) Parallel line path for block 1

(d) Parallel line path for block 2

Fig. 18. Search starting points and routes of the traditional parallel line scanning algorithm (PA).



(a) Parallel line path for block 3

(b) Parallel line path for block 4

(c) Parallel line path for block 1

(d) Parallel line path for block 2

Fig. 19. Search starting points and routes of the traditional parallel line scanning algorithm starting from the highest heat grid (SPA) (the red line represents the overlapping path).

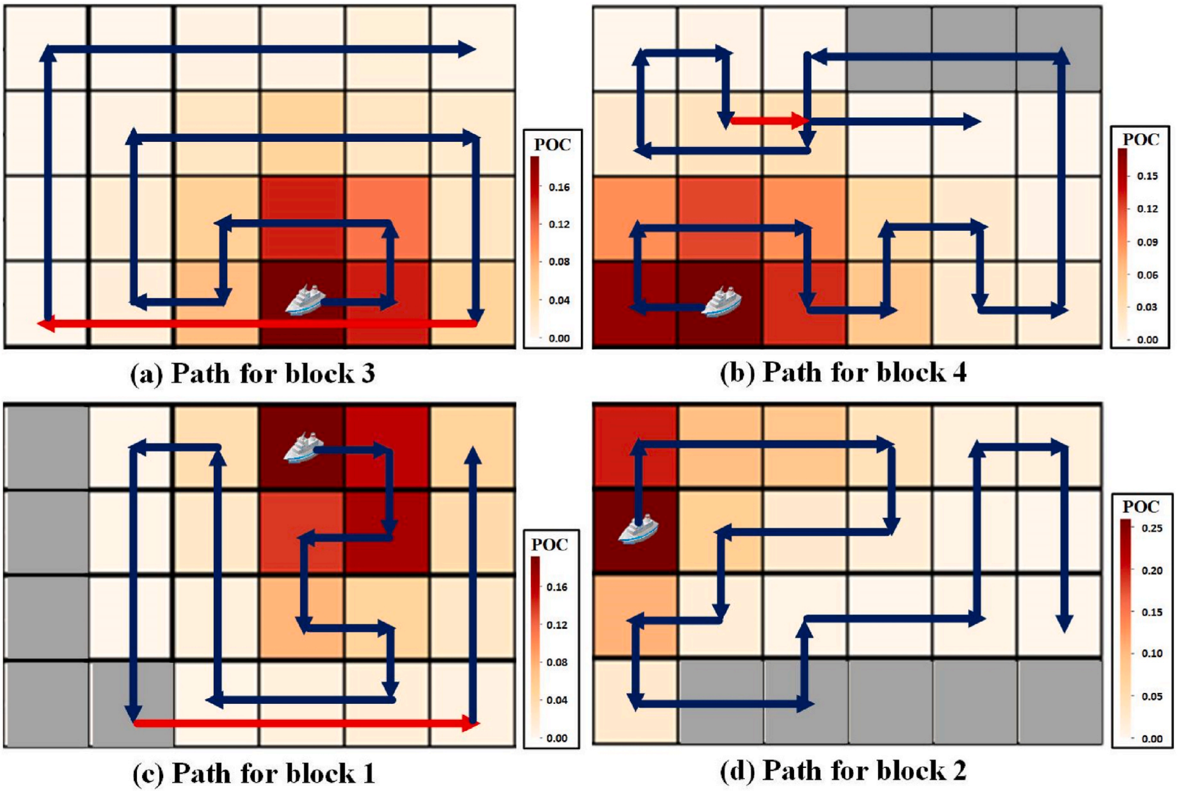


Fig. 20. Search starting points and routes of the BA* algorithm (the red line represents the overlapping path).

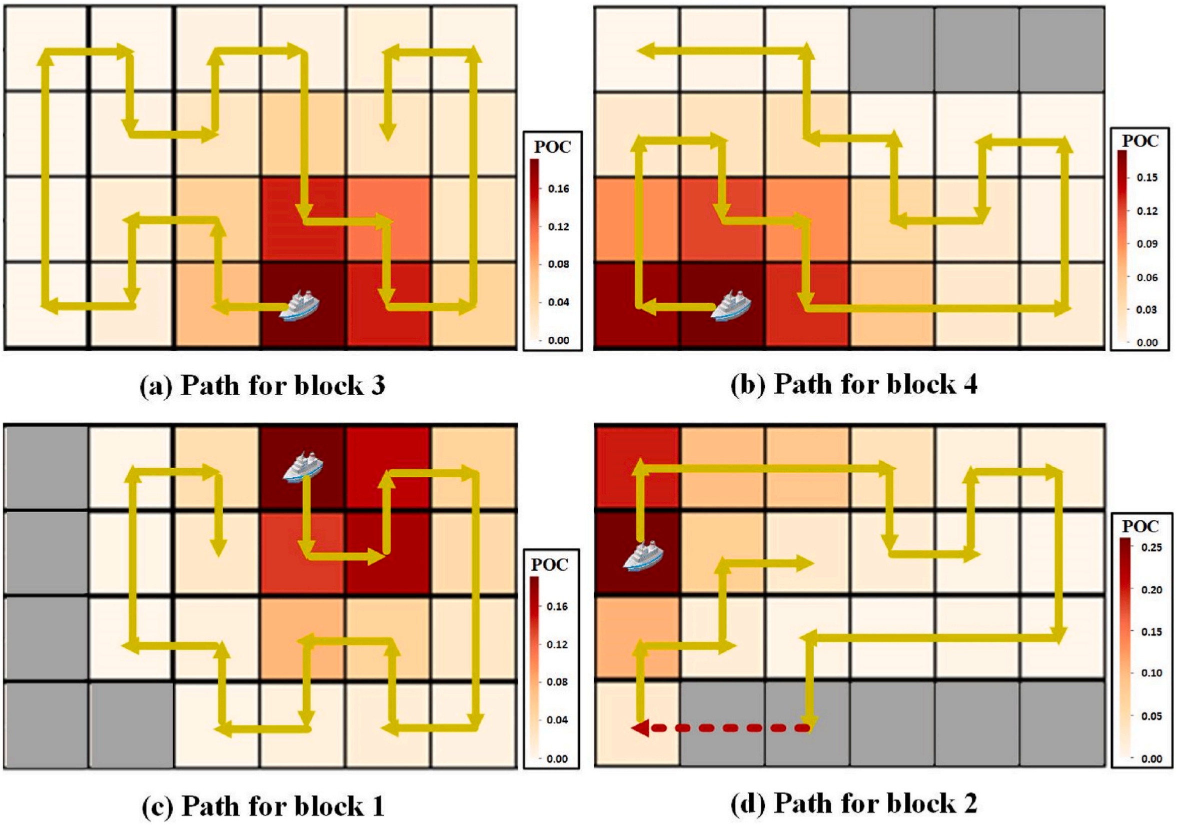


Fig. 21. Search start points and routes of the Q-learning algorithm.

Table 7
The repetition rate, coverage ratio, and the number of steps of SAR planning results.

Block	Algorithms	Repeated coverage (%)	Coverage (%)	Step
Block 1	Ours	0	100	18
	Q-learning	0	100	18
	PA	0	100	18
	SPA	0	100	23
	BA*	13.6	100	22
Block 2	Ours	0	100	18
	Q-learning	0	100	20
	PA	0	100	23
	SPA	8	100	25
	BA*	0	100	20
Block 3	Ours	0	100	23
	Q-learning	0	100	23
	PA	0	100	23
	SPA	8	100	25
	BA*	14.8	100	27
Block 4	Ours	0	100	20
	Q-learning	0	100	20
	PA	0	100	23
	SPA	8	100	25
	BA*	4	100	25

Algorithm Comparison: The proposed algorithm was compared with a search method commonly used in maritime SAR and a more advanced path-planning algorithm. Including the traditional parallel line scanning algorithm (PA) (IAMSAR, 2016), the parallel line scanning algorithm starts from the highest heat grid (SPA), the BA* algorithm (Viet et al., 2013), and the Q-learning method (Ai et al., 2021). The BA* algorithm was used to solve an online complete coverage task for an autonomous cleaning robot in an unknown workspace based on boustrophedon motion and an A* search algorithm. The robot performed a boustrophedon motion to cover the unvisited area until it reached the critical point. The robot then detected the backtracking point based on its accumulated knowledge, determined the best backtracking point as the starting point for the next boustrophedon motion, and continued to cover the next unvisited region, thereby achieving complete coverage. We added a search strategy for prioritizing high-probability areas to the BA* algorithm and Q-learning method to ensure fairness.

Algorithm parameter settings: The parameters of the proposed algorithm are listed in Table 6. Among them, the Taguchi experimental design methodology was used to determine the four control parameters, that is, learning rate LR, discount factor γ, target network update frequency n, and the initial action selection strategy ε, in our maritime coverage path planning model. The levels of the four parameters are as follows:

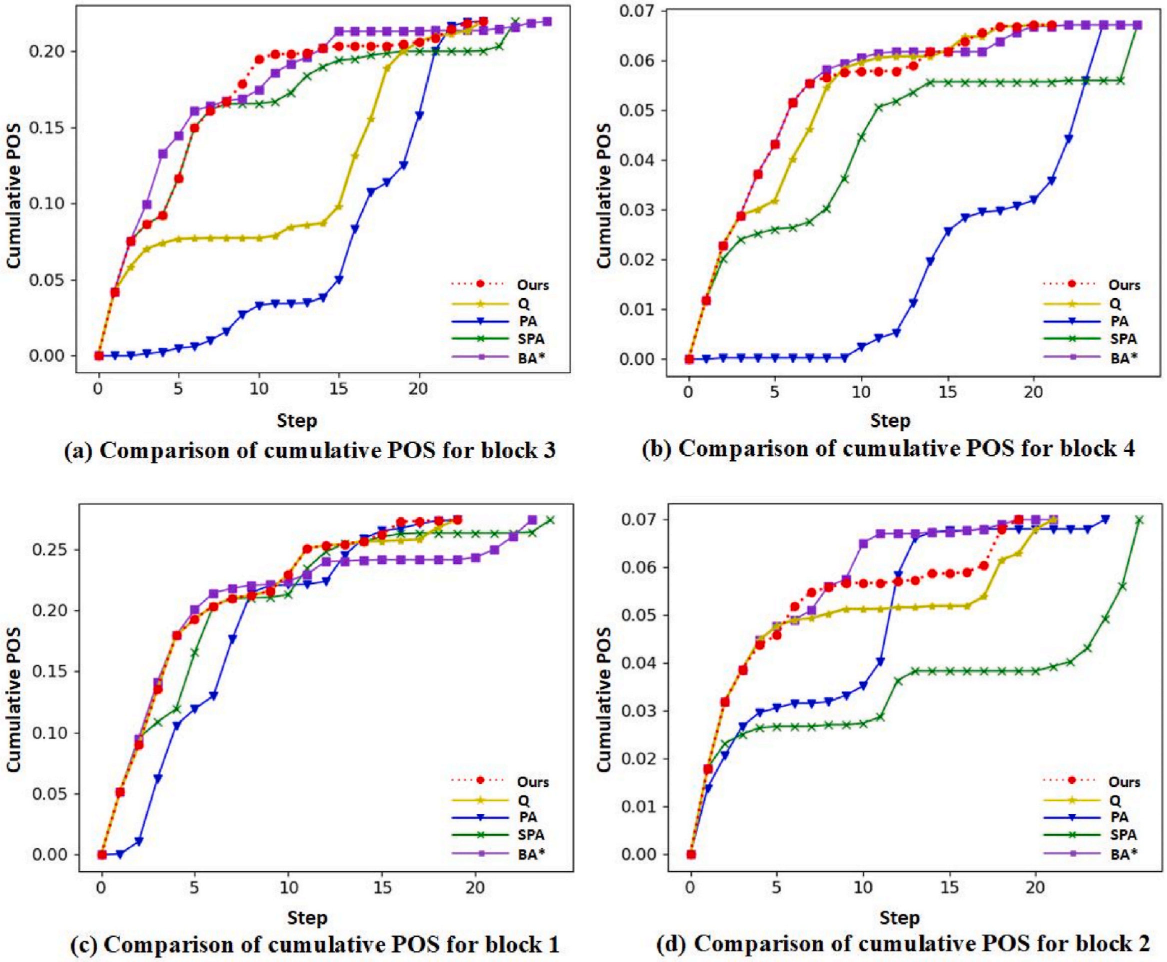
$$LR = \{0.005, 0.001, 0.1, 0.15\}$$

$$\gamma = \{0.45, 0.5, 0.6, 0.8\}$$

$$n = \{20, 50, 80, 100\}$$

$$\varepsilon = \{0.8, 0.85, 0.9, 0.95\}$$

At these parameter levels, orthogonal matrix $L_{16}(4^4)$ was used for the



(a) Comparison of cumulative POS for block 3

(b) Comparison of cumulative POS for block 4

(c) Comparison of cumulative POS for block 1

(d) Comparison of cumulative POS for block 2

Fig. 22. Comparison of the cumulative changes in the POS between different methods for four hierarchical probability environments.

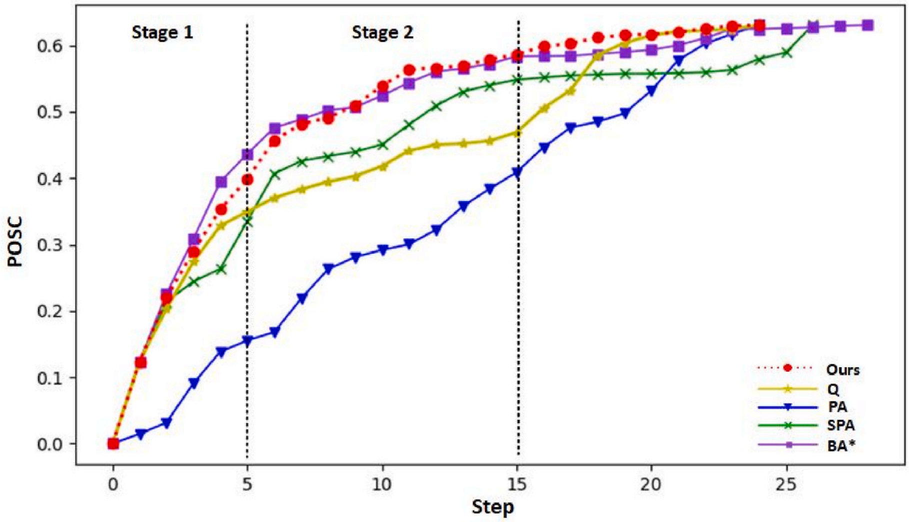


Fig. 23. Comparison of the cumulative changes in the POSC between different methods for the whole search area.

calibration experiment. To ensure equity, the algorithm was run 20 times independently for each parameter setting. The average results for our object for different parameter settings are shown in Fig. 16. The optimal control parameter settings were determined to be: $LR = 0.1$, $\gamma = 0.5$, $n = 50$, $\varepsilon = 0.9$.

Therefore, the control parameters used in this study were set as follows. The learning rate is 0.1, the γ coefficient is 0.5, the target network update frequency is 50 steps, the initial action selection strategy is 0.9, and the maximum learning episode is 5000.

- Results and analysis

The proposed method was tested in four hierarchical environments (Fig. 15), and the path-planning results are shown in Figs. 17–21. These algorithms can achieve full coverage of the search and rescue areas by setting search rules. However, the search path planning results of the SPA and BA* algorithms contained overlapping paths, whereas both the Q-learning algorithm and our model have the ability to achieve no duplicate path coverage. However, in some environments (grey grids in Fig. 21d), the Q-learning algorithm passes through the region where the grid unit POC value is zero in the hierarchical probabilistic environment map, which may lead to an increase in search time.

To quantitatively assess the above methods, the repetition rate, coverage ratio, and number of path planning steps were computed, as shown in Table 7. The results have shown that the navigation route generated by our model performs more effectively than the other algorithms in terms of repetition and the number of steps. According to the calculation methods of POC and POD described in Section 2, they were evaluated from the perspective of the cumulative POS. As shown in Fig. 22, in the four sub environments, our model reached the maximum cumulative POS with the fewest steps, and the cumulative POS growth rate of our model was faster, indicating that the vessel agent in our model can prioritize covering high-probability regions.

The cumulative POS of the parallel line scanning algorithm and BA* algorithm increased rapidly at some point. However, the proposed algorithm still achieved the maximum cumulative POS with a relatively small number of steps, indicating that the proposed algorithm can still find higher-quality solutions after reaching a certain degree of search, demonstrating its superior exploitation ability. Although the BA* algorithm can achieve full coverage and rapid growth of cumulative POS in the short term, it performed poorly in balancing the two goals of non-duplicate paths and prioritizing the search for high probability zones. In some cases, it produced more duplicated searches in pursuit of high probability zone coverage (block 3), making it difficult to meet SAR requirements in complex scenarios with large search areas. Although the Q-learning algorithm also showed strong performance in reaching the maximum cumulative POS with fewer steps, in some environments the cumulative POS growth rate was slower than that of our algorithm, and our algorithm showed better performance in preferentially covering high-probability regions.

Based on the hierarchical probability environment map, the cumulative POS for the overall SAR area (POSC) of each path-planning method was calculated based on the assumption that ships in the four subgrid hierarchical environments perform simultaneous searches at the same time and search speed. POSC can be calculated as follows:

$$POC_{mn} = POCT_{mn} * POC_m \tag{27}$$

$$POSC = \sum_{k=0}^{N} POS_k \tag{28}$$

$$POS_k = \sum_{m=1}^{M} POST_m \tag{29}$$

where POC_{mn} is the POC of each grid cell in the overall environment map, m is the block number, n is the grid cell number, $POCT_{mn}$ is the POC of grid cell n in the subgrid hierarchical environment, POC_m is the POC of block m in the one-level hierarchical environment map, k is the number of search steps of the current ship, N is the maximum number of search steps of each subgrid hierarchical environment map, POS_k is the POS of the entire SAR region in the current step, $POST_m$ is the POS of each subgrid hierarchical environment in the current step.

The POSC of the different path-planning methods are shown in Fig. 23. For the entire search and rescue area, both the Q-learning algorithm and our algorithm achieved the maximum POSC with the shortest number of steps and showed better search performance and convergence ability. The Q-learning algorithm also showed a high level of performance in the initial stage of the search (Stage 1). In the middle stage of the search (Stage 2), the cumulative POS growth was slow, indicating that our algorithm is superior to the Q-learning algorithm in exploration and can be used more effectively in conducting priority search in high-probability regions. Although the traditional PA algorithm can also achieve the fastest speed to complete the coverage search of the entire SAR area, it performs poorly in terms of the growth rate of POSC. The SPA algorithm achieves the improvement of search capability based on PA, and the BA* algorithm performed better in terms of prioritizing the search of high probability areas but performed poorly in terms of the time to complete the full-coverage search.

7. Conclusions

This study has integrated reinforcement learning into maritime SAR coverage path planning and establishes a maritime SARCPPF suitable for PIWs scenarios. This system comprises three modules, namely, drift trajectory prediction, hierarchical environment map modeling, and coverage search. Sea-area-scale drift prediction models of PIWs in the Chinese sea area were used based on the variations in PIWs across different sea areas and postures. A minimum bounding rectangle was used to establish a hierarchical probability environment map, facilitating the search for multiple SAR units. A coverage path planning algorithm that leverages deep reinforcement learning was devised. Comparative experiments have demonstrated that the proposed algorithm significantly enhances POS within a constrained timeframe.

However, this study has certain limitations, including that it was assumed that the search environment remained constant during path planning. In future studies, the algorithm should dynamically update the search environment based on the SAR task performance and drifting conditions to improve the search accuracy. This study was conducted based on the assumption that the number of SAR units was sufficient. In the future, a detailed analysis will be conducted on the number, location, search and rescue capabilities, as well as other characteristics of SAR forces, to optimize the division of search and rescue areas. In addition, coordinated searches using multiple SAR forces will also be studied.

Funding

This work was supported by the National Key Research and Development Program of China (2022YFB3903603) and National Natural Science Foundation of China (42001401).

CRediT authorship contribution statement

Jie Wu: Data curation, Methodology, Writing – original draft, Software. **Liang Cheng:** Conceptualization, Resources, Funding acquisition, Supervision. **Sensen Chu:** Visualization, Investigation. **Yanjie Song:** Validation, Writing – review & editing.

Declaration of competing 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

Data will be made available on request.

References

Abi-Zeid, I., Frost, J.R., 2005. SARPlan: a decision support system for Canadian search and rescue operations. Eur. J. Oper. Res. 162, 630–653. https://doi.org/10.1016/j.ejor.2003.10.029.

Abi-Zeid, I., Nilo, O., Lamontagne, L., 2011. A constraint optimization approach for the allocation of multiple search units in search and rescue operations. Info. R. 49, 15–30. https://doi.org/10.3138/infor.49.1.015.

Agbissoh Otote, D., Li, B., Ai, B., et al., 2019. A decision-making algorithm for maritime search and rescue plan. Sustainability 11, 2084. https://doi.org/10.3390/su11072084.

Ai, B., Jia, M., Xu, H., et al., 2021. Coverage path planning for maritime search and rescue using reinforcement learning. Ocean Eng. 241, 110098 https://doi.org/10.1016/j.oceaneng.2021.110098.

Ai, B., Li, B., Gao, S., Xu, J., Shang, H., 2019. An intelligent decision algorithm for the generation of maritime search and rescue emergency response plans. IEEE Access 7, 155835–155850. https://doi.org/10.1109/ACCESS.2019.2949366.

Allen, A.A., 2005. Leeway Divergence. U.S. Coast Guard Rep. CG-D-05-05, p. 128.

Allen, A.A., Plourde, J.V., 1999. Review of Leeway: Field Experiments and Implementation. U.S. Coast Guard Rep. CG-D-08-99, p. 351.

Allen, A.A., Roth, J.C., Maisondieu, C., et al., 2010. Field Determination of the Leeway of Drifting Objects. Norwegian Meteorological Institute, Oslo.

Anderson, E., Greenbaum, H., McClay, T., et al., 2006. NSMRL. S.E.E./Rescue Project Target Detectability Testing, Modeling, and Analysis, vol. 2006. nsmrl see/rescue project target detectability testing modeling & analysis.

Binney, J., Krause, A., Sukhatme, G.S., 2010. Informative Path Planning for an Autonomous Underwater Vehicle. IEEE Publications, pp. 4791–4796. https://doi.org/10.1109/ROBOT.2010.5509714.

Bourgault, F., Furukawa, T., Durrant-Whyte, H.F., 2003. Coordinated decentralized search for a lost target in a Bayesian world. Int. Conf. Intell. Robots Syst., Las Vegas 48–53. https://doi.org/10.1109/IROS.2003.1250604.

Breivik, Ø., Allen, A.A., 2008. An operational search and rescue model for the Norwegian Sea and the North Sea. J. Mar. Syst. 69, 99–113. https://doi.org/10.1016/j.jmarsys.2007.02.010.

Breivik, Ø., Allen, A.A., Maisondieu, C., Olagnon, M., 2013. Advances in search and rescue at sea. Ocean Dynam. 63, 83–88. https://doi.org/10.1007/s10236-012-0581-1.

Breivik, Ø., Allen, A.A., Maisondieu, C., Roth, J., Forest, B., 2012. The leeway of shipping containers at different immersion levels. Ocean Dynam. 62, 741–752. https://doi.org/10.1007/s10236-012-0522-z.

Breivik, Ø., Allen, A.A., Maisondieu, C., Roth, J.C., 2011. Wind-induced drift of objects at sea: the leeway field method. Appl. Ocean Res. 33, 100–109. https://doi.org/10.1016/j.apor.2011.01.005.

Brown, S.S., 1980. Optimal search for a moving target in discrete time and space. Oper. Res. 28, 1275–1289. https://doi.org/10.1287/opre.28.6.1275.

Brushett, B.A., Allen, A.A., King, B.A., Lemckert, C.J., 2017. Application of leeway drift data to predict the drift of panga skiffs: case study of maritime search and rescue in the tropical pacific. Appl. Ocean Res. 67, 109–124. https://doi.org/10.1016/j.apor.2017.07.004.

Burciu, Z., 2010. Bayesian methods in reliability of search and rescue action. Pol. Marit. Res. 17, 72–78. https://doi.org/10.2478/v10012-010-0039-7.

Busoniu, L., Babuska, R., De Schutter, B.A., 2008. A Comprehensive survey of multiagent reinforcement learning. IEEE Trans. Syst. Man Cybern. C. 38, 156–172. https://doi.org/10.1109/TSMCC.2007.913919.

Cao, X., Sun, C., Yan, M., 2019. Target search control of AUV in underwater environment with deep reinforcement learning. IEEE Access 7, 96549–96559. https://doi.org/10.1109/ACCESS.2019.2929120.

Canadian Coast Guard, Canadian Coast Guard College CANSARP Development Group Web site, 2009. CANSARP User Manual. http://loki.cgc.gc.ca/cansarp/cansarpmanualsept1609.pdf.

Carneiro, J.P., 1988. Maritime search and rescue. IETE Tech. Rev. 5, 111–114. https://doi.org/10.1080/02564602.1988.11438248.

Chabini, I., Lan, S., 2002. Adaptations of the A* algorithm for the computation of fastest paths in deterministic discrete-time dynamic networks. Intel. Transport. Syst. IEEE Trans. 3 (1), 60–74. https://doi.org/10.1109/6979.994796.

Chen, H.T., Xu, W.M., Sun, L.Y., et al., 2017. An example of AP98 Leeway drift model application: drift experiment of DongFangHong 2. Trans. Oceanol. Limnol. 6, 46–51.

Chen, Y., Zhu, S., Zhang, W., Zhu, Z., Bao, M., 2022. The model of tracing drift targets and its application in the South China Sea. Acta Oceanol. Sin. 41, 109–118. https://doi.org/10.1007/s13131-021-1943-7.

Cheng, P.F., Yan, H.W., Han, Z.H., 2008. An algorithm for computing the minimum area bounding rectangle of an arbitrary polygo. J. Eng. Graph. 29 (1), 122–126. https://doi.org/10.1007/s11123-007-0067-1, 2008.

Cho, S.W., Park, H.J., Lee, H., Shim, D.H., Kim, S., 2021. Coverage path planning for multiple unmanned aerial vehicles in maritime search and rescue operations. Comput. Ind. Eng. 161, 107612 https://doi.org/10.1016/j.cie.2021.107612.

Daniel, P., Marty, F., Josse, P., Skandrani, C., Benshila, R., 2003. Improvement of drift calculation in MOTHY operational oil spill prediction system. In: International Oil Spill Conference Proceedings International Oil Spill Conference, Vancouver, vol. 2003, pp. 1067–1072. https://doi.org/10.7901/2169-3358-2003-1-1067, 2003.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numer. Math. 1, 269–271. https://doi.org/10.1007/BF01386390.

Engel, D.D., Weisinger, J.R., 1988. OR practice—estimating visual detection performance at sea. Oper. Res. 36, 651–659. https://doi.org/10.1287/opre.36.5.651.

Englot, B., Hover, F.S., 2013. Three-dimensional coverage planning for an underwater inspection robot. Int. J. Robot Res. 32, 1048–1073. https://doi.org/10.1177/0278364913490046.

Fang, Y., Pu, J., Zhou, H., et al., 2021. Attitude Control Based Autonomous Underwater Vehicle Multi-Mission Motion Control with Deep Reinforcement Learning. In: 5th International Conference on Automation, Control and Robots (ICACR), Nanning, China, pp. 120–129. https://doi.org/10.1109/ICACR53472.2021.9605171, 2021.

Fevgas, G., Lagkas, T., Argyriou, V., Sarigiannidis, P., 2022. Coverage path planning methods focusing on energy efficient and cooperative strategies for unmanned aerial vehicles. Sensors (Basel) 22, 1235. https://doi.org/10.3390/s22031235.

Frost, J.R., 1997. The Theory of Search: a Simplified Explanation. Soza Limited.

Frost, J.R., 2001. Review of Search Theory. U. S. Coast Guard Office of Operations (G-OPR), 2001.

Frost, J.R., Stone, L.D., 2001. Review of Search Theory: Advances and Applications to Search and Rescue Decision Support. U.S. Coast Guard Research and Development Center. Report No. CG-D-15-01.

Galceran, E., Carreras, M., 2013. A survey on coverage path planning for robotics. Robot. Autonom. Syst. 61, 1258–1276. https://doi.org/10.1016/j.robot.2013.09.004.

Graham, R.L., 1972. An efficient algorithm for determining the convex hull of a finite planar set. Inf. Process. Lett. 1 (1972), 132–133.

Haga, J.M., Svanberg, K.L., 2022. Search and Rescue. Textbook on Maritime Health. Norwegian Center FOR Maritime Medicine.

Hart, P.E., Nilsson, N.J., Raphael, B.A., 1972. A Formal basis for the heuristic determination of minimum cost paths. ACM SIGART Bulletin. IEEE Trans. Syst. Sci. Cyber. 4, 100–107. https://doi.org/10.1109/TSSC.1968.300136.

Hinton, G.E., Osindero, S., Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18, 1527–1554. https://doi.org/10.1162/NECO.2006.18.7.1527. MIT Press.

Hou, X., Ren, Z., Wang, J., et al., 2020. Distributed fog computing for latency and reliability guaranteed Swarm of drones. IEEE Access 8, 7117–7130. https://doi.org/10.1109/ACCESS.2020.2964073.

IAMSAR, 2016. International Aeronautical and Maritime Search and Rescue Manual. II. Mission Coordination. IMO/International Civil Aviation Organization publications, London/Montreal.

International Maritime Organization, 1979. https://www.imo.org/en/OurWork/Safety/Pages/SearchandRescue-Default.aspx.

Jonnarth, A., Zhao, J., Felsberg, M., 2023. End-to-End Reinforcement Learning for Online Coverage Path Planning in Unknown Environments, p. 2023 arXiv Preprint arXiv:2306.16978.

Karakaya, M., 2014. UAV route planning for maximum target coverage. Comput. Sci. Eng. 4 (4) https://doi.org/10.5121/cseij.2014.4103.

Kasyk, L., Pleskacz, K., Kapuściński, T., 2021. Analysis of wind and drifter movement parameters in terms of navigation safety: the example of Szczecin lagoon. Eur. Res. Stud. J. XXIV, 541–559. https://doi.org/10.35808/ersj/2370.

Kavraki, L.E., Svestka, P., Latombe, J.-C., Overmars, M.H., 1996. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. IEEE Trans. Robot. Automat. IEEE Int. Conf. Robot. Aotomation. 12, 566–580. https://doi.org/10.1109/70.508439.

Koenig, S., Likhachev, M., 2005. Fast replanning for navigation in unknown terrain. IEEE Trans. Robot. 21, 354–363. https://doi.org/10.1109/TRO.2004.838026.

Kong, X., Everett, H., Toussaint, G., 1990. The Graham scan triangulates simple polygons. Pattern Recogn. Lett. 11, 713–716. https://doi.org/10.1016/0167-8655(90)90089-K.

Koopman, B.O., 1956a. The theory of search. I. Kinematic bases. Oper. Res. 4, 324–346. https://doi.org/10.1287/opre.4.3.324.

Koopman, B.O., 1956b. The theory of search. II. Target detection. Oper. Res. 4, 503–531. https://doi.org/10.1287/opre.4.5.503.

Koopman, B.O., 1957. The theory of search: III. The optimum distribution of searching effort. Oper. Res. 5, 613–626. https://doi.org/10.1287/opre.5.5.613.

Kratzke, T.M., Stone, L.D., Frost, J.R., 2010. Search and Rescue Optimal Planning System 13th International Conference on Information Fusion, vol. 2010. IEEE Publications, pp. 1–8. https://doi.org/10.1109/ICIF.2010.5712114.

Kyaw, P.T., Paing, A., Thu, T.T., et al., 2020. Coverage path planning for decomposition reconfigurable grid-maps using deep reinforcement learning based travelling salesman problem. IEEE Access 8, 225945–225956. https://doi.org/10.1109/ACCESS.2020.3045027.

Lavalle, S., 1998. Rapidly-Exploring Random Trees: a New Tool for Path Planning. Research Report, pp. 293–308.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791.

Lee, S., Morrison, J.R., 2015. Decision support scheduling for maritime search and rescue planning with a system of UAVs and fuel service stations. In: Proceeding of 2015 International Conference on Unmanned Aircraft Systems (ICUAS). UAS, Denver, Colorado, Jun, pp. 1168–1177. https://doi.org/10.1109/ICUAS.2015.7152409, 2015.

Li, L., Wu, D., Huang, Y., Yuan, Z., 2021. A path planning strategy unified with a COLREGS collision avoidance function based on deep reinforcement learning and artificial potential field. Appl. Ocean Res. 113, 102759 https://doi.org/10.1016/j.apor.2021.102759.

Lin, L., Goodrich, M.A., 2014. Hierarchical heuristic search using a Gaussian mixture model for UAV coverage planning. IEEE Trans. Cybern. 44, 2532–2544. https://doi.org/10.1109/TCYB.2014.2309898.

Liu, Y., Bucknall, R., Zhang, X., 2017. The fast marching method based intelligent navigation of an unmanned surface vehicle. Ocean Eng. 142, 363–376. https://doi.org/10.1016/j.oceaneng.2017.07.021.

Luo, Q., Wang, H., Zheng, Y., He, J., 2020. Research on path planning of mobile robot based on improved ant colony algorithm. Neural Comput. Appl. 32, 1555–1566. https://doi.org/10.1007/s00521-019-04172-2.

Marija, D., Ivan, P., 2011. Two-way D* algorithm for path planning and replanning. Robot. Autonom. Syst. 59 (5), 329–342. https://doi.org/10.1016/j.robot.2011.02.007.

Masehian, E., Sedighizadeh, D., 2010. A multi-objective PSO-based algorithm for robot path planning. In: IEEE International Conference on Industrial Technology, Via del Mar, Chile, vol. 2010, pp. 465–470. https://doi.org/10.1109/ICIT.2010.5472755, 2010.

Meng, S.J., Lu, W., Li, Y., Wang, H., Jiang, L., 2021. A study on the leeway drift characteristic of a typical fishing vessel common in the northern South China Sea. Appl. Ocean Res. 109 https://doi.org/10.1016/j.apor.2020.102498.

Minsky, M.L., 1967. Computation: Finite and Infinite Machines. Prentice-Hall, Inc.

Mnih, V., Kavukcuoglu, K., Silver, D., et al., 2013. Playing atari with deep reinforcement learning. Comp. Sci. https://doi.org/10.48550/arXiv.1312.5602, 2013.

Mnih, V., Kavukcuoglu, K., Silver, D., et al., 2015. Human-level control through deep reinforcement learning. Nature 518, 529–533. https://doi.org/10.1038/nature14236.

Mou, J., Hu, T., Chen, P., Chen, L., 2021. Cooperative MASS path planning for marine man overboard search. Ocean Eng. 235, 109376 https://doi.org/10.1016/j.oceaneng.2021.109376.

Nash, A., Koenig, S., 2013. Any-angle path planning. AI Mag. 34, 85–107.

Ouelmokhtar, H., Benmoussa, Y., Benazzouz, D., Ait-Chikh, M.A., Lemarchand, L., 2022. Energy-based USV maritime monitoring using multi-objective evolutionary algorithms. Ocean Eng. 253, 111182 https://doi.org/10.1016/j.oceaneng.2022.111182.

Peng, Z., Wang, C., Xu, W., Zhang, J., 2022. Research on location-routing problem of maritime emergency materials distribution based on bi-level programming. Mathematics 10. https://doi.org/10.3390/math10081243.

Prins, C., 2004. A simple and effective evolutionary algorithm for the vehicle routing problem. Comput. Oper. Res. 31, 1985–2002. https://doi.org/10.1016/S0305-0548(03)00158-8.

Ramirez, F.F., Benitez, D.S., Portas, E.B., Orozco, J.A.L., 2011. Coordinated sea rescue system based on unmanned air vehicles and surface vessels. Oceans 2011, 1–10. https://doi.org/10.1109/Oceans-Spain.2011.6003509. IEEE Publications, Spain. IEEE Publications.

Rani, S., Babbar, H., Kaur, P., Alshehri, M.D., Shah, S.H.A., 2022. An optimized approach of dynamic target nodes in wireless sensor network using bio inspired algorithms for maritime rescue. IEEE Trans. Intell. Transport. Syst. 99, 1–8. https://doi.org/10.1109/TITS.2021.3129914.

Sendner, F.M., 2022. An energy-autonomous UAV swarm concept to support sea-rescue and maritime patrol missions in the Mediterranean Sea. Aircraft Eng. Aero. Technol. 94, 112–123. https://doi.org/10.1108/AEAT-12-2020-0316.

Seraj, E., Silva, A., Gombolay, M., 2022. Multi-UAV planning for cooperative wildfire coverage and tracking with quality-of-service guarantees. Aut. Agents Multi-Agent Syst. 36, 39. https://doi.org/10.1007/s10458-022-09566-6.

Shchekinova, E.Y., Kumkar, Y., 2015. Stochastic modeling for trajectories drift in the ocean: application of density clustering algorithm. Physics 2015. https://doi.org/10.48550/arXiv.1505.04736.

Shen, Z., Wilson, J.P., Gupta, S., 2019. An online coverage path planning algorithm for curvature constrained AUVs. Oceans. MTS/IEEE SEATTLE. IEEE. 2019, 1–5. https://doi.org/10.23919/OCEANS40490.2019.8962629.

Song, J., Gupta, S., Hare, J., et al., 2013. Adaptive cleaning of oil spills by autonomous vehicles under partial information. Oceans 2013. https://doi.org/10.23919/OCEANS.2013.6741246. IEEE.

Soza Company, Ltd, 1996. The theory of search a simplified explanation. Office of Search and Rescue U.S. Coast Guard, Washington. www.aiai.ed.ac.uk.

Stentz, A., 1994. Optimal and efficient path planning for partially-known environments. IEEE Int. Conf. Robot. Automat. 4 (2002), 3310–3317. https://doi.org/10.1109/ROBOT.1994.351061.

Sutherland, G., Soontiens, N., Davidson, F., et al., 2020. Evaluating the Leeway Coefficient for Different Ocean Drifters Using Operational Models. Arxiv e-Prints. https://doi.org/10.48550/arXiv.2005.09527, 2020.

Sutton, R.S., Barto, A.G., 1998. Reinforcement learning: an introduction. IEEE Trans. Neural Network. 9 (5), 1054. https://doi.org/10.1109/TNN.1998.712192, 1998.

Tapkin, S., Temur, S., 2022. Determining the Wind Effect on Person-In-Water: A New Datum Calculation Method. Available at: SSRN 4147573.

Theile, M., Bayerlein, H., Nai, R., Gesbert, D., Caccamo, M., 2020. UAV Coverage Path Planning under Varying Power Constraints Using Deep Reinforcement Learning, vol. 2020. IEEE Publications, pp. 1444–1449. https://doi.org/10.1109/IROS45743.2020.9340934.

Tokic, M., 2010. Adaptive ε-greedy exploration in reinforcement learning based on value differences. In: Annual Conference on Artificial Intelligence. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-16111-7_23, 2010.

Tu, H., Wang, X., Mu, L., Xia, K., 2021. Predicting drift characteristics of persons-in-the-water in the South China Sea, 242-Dec.15 Ocean Eng. 242. https://doi.org/10.1016/j.oceaneng.2021.110134.

Viet, H.H., Dang, V.-H., Laskar, M.N.U., Chung, T., 2013. BA*: an online complete coverage algorithm for cleaning robots. Appl. Intell. 39, 217–235. https://doi.org/10.1007/s10489-012-0406-4.

Wang, H., Yu, Y., Yuan, Q., 2011. Application of Dijkstra algorithm in robot path-planning, 2011. 5987118. In: International Conference on Mechanic Automation & Control Engineering. IEEE. https://doi.org/10.1109/MACE.

Washburn, A.R., 1983. Search for a moving target: the fab algorithm. Oper. Res. 31, 739–751. https://doi.org/10.1287/opre.31.4.739.

Washburn, A.R., Kress, M., 2009. Combat Modeling. Springer, New York. https://doi.org/10.1002/9780470498620.ch9.

Watkins, C.J.C.H., 1989. Learning from Delayed Rewards, PhD Dissertation. University of Cambridge, Cambridge, England.

Watkins, C.J.C.H., Dayan, P., 1992. Q-learning. Mach. Learn. 8, 279–292. https://doi.org/10.1007/BF00992698.

Wiering, M.A., Van, O.M., 2012. Reinforcement learning. Adapt. Learn. Optim. 12, 729.

Wu, J., Cheng, L., Chu, S.S., 2023. Modeling the leeway drift characteristics of persons-in-water at a sea-area scale in the seas of China. Ocean Eng. 270, 113444 https://doi.org/10.1016/j.oceaneng.2022.113444.

Wu, X., Zhou, J.H., 2015. Study on probability of detection in marine search and rescue. J. Saf. Sci. Technol. 11, 28–33.

Xi, M., Yang, J., Wen, J., et al., 2022. Comprehensive ocean information-enabled AUV path planning via reinforcement learning. IEEE Internet Things J. 9, 17440–17451. https://doi.org/10.1109/JIOT.2022.3155697.

Xie, R., Meng, Z., Wang, L., et al., 2021. Unmanned aerial vehicle path planning algorithm based on deep reinforcement learning in large-scale and dynamic environments. IEEE Access 9, 24884–24900. https://doi.org/10.1109/ACCESS.2021.3057485.

Xiong, P., Liu, H., Tian, Y., et al., 2021. Helicopter maritime search area planning based on a minimum bounding rectangle and K-means clustering. Chin. J. Aeronaut. 34, 554–562. https://doi.org/10.1016/j.cja.2020.08.047.

Xiong, W., van Gelder, P.H.A.J.M., Yang, K., 2020. A decision support method for design and operationalization of search and rescue in maritime emergency. Ocean Eng. 207, 107399 https://doi.org/10.1016/j.oceaneng.2020.107399.

Yang, T., Jiang, Z., Sun, R., Cheng, N., Feng, H., 2020. Maritime search and rescue based on group mobile computing for unmanned aerial vehicles and unmanned surface vehicles. IEEE Trans. Ind. Inf. 16, 7700–7708. https://doi.org/10.1109/TII.2020.2974047.

Yao, P., Xie, Z., Ren, P., 2019. Optimal UAV route planning for coverage search of stationary target in river. IEEE Trans. Control Syst. Technol. 27, 822–829. https://doi.org/10.1109/TCST.2017.2781655.

Zhang, H., Sun, J., Yang, B., Shi, Y., Li, Z., 2020. Optimal search and rescue route design using an improved ant colony optimization. Inf. Technol. Control 49, 438–447. https://doi.org/10.5755/j01.itc.49.3.25295.

Zhang, J.F., Teixeira, Â.P., Guedes Soares, C.G., Yan, X., 2017. Probabilistic modelling of the drifting trajectory of an object under the effect of wind and current for maritime search and rescue. Ocean Eng. 129, 253–264. https://doi.org/10.1016/j.oceaneng.2016.11.002.

Zhang, Q., Chen, D., Chen, T., 2012. An obstacle avoidance method of soccer robot based on evolutionary artificial potential field. Energy Proc. 16 (5), 1792–1798. https://doi.org/10.1016/j.egypro.2012.01.276.

Zhang, X., Wang, C., Liu, Y., Chen, X., 2019. Decision-making for the autonomous navigation of maritime autonomous surface ships based on scene division and deep reinforcement learning. Sensors (Basel). 19, 4055. https://doi.org/10.3390/s19184055.

Zhou, X., 2022. A comprehensive framework for assessing navigation risk and deploying maritime emergency resources in the South China Sea. Ocean Eng. 248, 110797 https://doi.org/10.1016/j.oceaneng.2022.110797.

Zhou, X., Cheng, L., Li, W.D., et al., 2020a. A comprehensive path planning framework for patrolling marine environment. Appl. Ocean Res. 100, 102155 https://doi.org/10.1016/j.apor.2020.102155.

Zhou, X., Cheng, L., Min, K., et al., 2020b. A framework for assessing the capability of maritime search and rescue in the South China Sea. Int. J. Disaster Risk Reduc. 47, 10568 https://doi.org/10.1016/j.ijdrr.2020.101568.

Zhu, K., Mu, L., Tu, H.W., 2019. Exploration of the wind-induced drift characteristics of typical Chinese offshore fishing vessels. Appl. Ocean Res. 92, 101916 https://doi.org/10.1016/j.apor.2019.101916.

Zhu, K., Zhang, T., 2021. Deep reinforcement learning based mobile robot navigation: a review. Tsinghua Sci. Technol. 26, 674–691. https://doi.org/10.26599/TST.2021.9010012.