



OPEN ACCESS

EDITED BY

Long Cheng,
North China Electric Power University, China

REVIEWED BY

Ali Abdelsalam,
National Authority for Remote Sensing and
Space Sciences, Egypt
Hanlin Zhou,
North China Electric Power University, China

*CORRESPONDENCE

Lei Wang
✉ wangl@hfuu.edu.cn

RECEIVED 29 August 2024

ACCEPTED 10 October 2024

PUBLISHED 22 October 2024

CITATION

Wang L, Liu D and Wang J (2024) A modified
A* algorithm combining remote sensing
technique to collect representative samples
from unmanned surface vehicles.
Front. Neurobot. 18:1488337.
doi: 10.3389/fnbot.2024.1488337

COPYRIGHT

© 2024 Wang, Liu and Wang. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

A modified A* algorithm combining remote sensing technique to collect representative samples from unmanned surface vehicles

Lei Wang*, Danping Liu and Jun Wang

School of Advanced Manufacturing Engineering, Hefei University, Hefei, China

Ensuring representativeness of collected samples is the most critical requirement of water sampling. Unmanned surface vehicles (USVs) have been widely adopted in water sampling, but current USV sampling path planning tend to overemphasize path optimization, neglecting the representative samples collection. This study proposed a modified A* algorithm that combined remote sensing technique while considering both path length and the representativeness of collected samples. Water quality parameters were initially retrieved using satellite remote sensing imagery and a deep belief network model, with the parameter value incorporated as coefficient Q in the heuristic function of A* algorithm. The adjustment coefficient k was then introduced into the coefficient Q to optimize the trade-off between sampling representativeness and path length. To evaluate the effectiveness of this algorithm, Chlorophyll-a concentration (Chl-a) was employed as the test parameter, with Chaohu Lake as the study area. Results showed that the algorithm was effective in collecting more representative samples in real-world conditions. As the coefficient k increased, the representativeness of collected samples enhanced, indicated by the Chl-a closely approximating the overall mean Chl-a and exhibiting a gradient distribution. This enhancement was also associated with increased path length. This study is significant in USV water sampling and water environment protection.

KEYWORDS

unmanned surface vehicles, A* algorithm, remote sensing technique, water sampling path planning, sampling representativeness, path length, chlorophyll-a

1 Introduction

Water pollution has long been a critical global concern (Schwarzenbach et al., 2010). Daily water safety has a direct and perpetual impact on human beings (Bhagwat, 2019). Water sampling and quality testing are essential components for safeguarding the water environment (Zhang, 2024; Behmel et al., 2016). Traditional water sampling methods primarily involve manual collection, which is time-consuming and labor-intensive (Pule et al., 2017). In recent years, the innovation of unmanned surface vehicles (USVs) has resulted in their widespread use in water sampling, effectively replacing traditional manual sampling methods (MahmoudZadeh et al., 2022). This advancement facilitates efficient and flexible water sample collection and environmental monitoring of aquatic ecosystems (Yu et al., 2022).

The initial USV-based sampling approach is typically collecting samples after reaching a designated water area through remote manipulation (Yu et al., 2019). While this technique is simple, effective control in vast water areas is challenging due to restrictions in observation and communication distance (Young and Peschel, 2020; Peng et al., 2021). Subsequently,

studies on path planning algorithms for USV sampling gradually emerged (Yu et al., 2022; Song et al., 2019; Liu and Bucknall, 2015), and USVs can navigate autonomously along the planned routes. Regardless of global, local or hybrid path planning algorithms, they all focus on the optimization of path length, time cost, smoothness, and collision avoidance (Xing et al., 2023; Yu et al., 2022; Vagale et al., 2021; Yu et al., 2019; Yu et al., 2024). However, these studies tend to overemphasize path optimization, neglecting the core objective of collecting representative samples that accurately reflect the condition of the water body.

In an USV water sampling mission, multi-goal path planning is another prevalent path planning algorithm (Ma et al., 2021; Luo et al., 2024). For example, USVs are employed to collect data and samples from multiple dispersed water quality monitoring stations (Liu and Bucknall, 2018). Unlike the aforementioned path planning algorithm that features a single goal point, the multi-goal path planning algorithm has multiple goal points, requiring the algorithm to design a route that efficiently visit each goal point with minimal cost (Liu and Bucknall, 2018). The task allocation in the multi-goal path planning can be summarized as a Travelling Salesman Problem. This type of algorithm is typically suitable for situations where the sampling points are predetermined, such as in the collection of samples from multiple monitoring stations or designated locations. Although samples collected by this method may be representative, its practicality is restricted and it is not suitable for situations where the sampling site is not predetermined. Therefore, there is an urgent requirement for an algorithm that meets the three crucial needs: minimizing cost, ensuring collecting representative samples, and adapting to situations where sampling points are not predetermined.

The A* algorithm is a well-known global path planning algorithm used to find the shortest route in a known global map. As a heuristic search algorithm, A* is more efficient than other deterministic methods due to its ability to consistently provide an optimal path with the minimum cost, thereby ensuring its completeness (Song et al., 2019). A* algorithm is widely used in the field of USV path planning due to its simplicity and high efficiency. It was initially introduced to this field by integrating with reactive navigation to propose an enhanced three-layered architecture for USV path planning within a harbor (Casalino et al., 2009). More studies were conducted on the basis of A* algorithm to adapt to the path planning in specific situations (Song et al., 2019; Sang et al., 2021; Singh et al., 2018; Casalino et al., 2009; Naeem et al., 2012). For instance, Sang et al. (2021) proposed a hybrid path planning algorithm that combines A* and artificial potential field to solve the path planning problem of USV formations. Singh et al. (2018) proposed a constrained A* algorithm to address the motion planning problem for USV navigating in a maritime environment containing dynamic obstacles.

Remote sensing refers to the science of identifying Earth's surface features and estimating their geo-biophysical properties through the use of electromagnetic radiation as a medium of interaction (Campbell and Wynne, 2011). With the advancement of remote sensing technology, it is now possible to obtain concentrations of water quality parameters in a non-contact manner, over large areas, and in quasi-real time (Ritchie et al., 2003; Wang and Yang, 2019; Gholizadeh et al., 2016). The concentrations of water quality parameters retrieved from remote sensing imagery can serve as prior knowledge for selecting sampling locations to address the representativeness of sampling points.

In this paper, a modified A* algorithm combining remote sensing technique was proposed, which simultaneously minimizes cost while ensuring that the sampling points are representative and not predetermined. Chlorophyll-a concentration (Chl-a), an important indicator of aquatic ecological status and can reflect the eutrophication level of the water body, was employed as the test water parameter to validate the proposed algorithm. Chaohu Lake was selected as the study area. The main contributions of this work are as follows:

- (1) To ensure the representativeness of collected samples, the heuristic function of the A* algorithm was modified by incorporating a sampling coefficient Q . The coefficient Q was derived from the Chl-a value retrieved by remote sensing technique through a negative exponential transformation of e . This adjustment ensured that Chl-a of the collected samples approximated the overall mean Chl-a of the entire water body and exhibited a gradient distribution.
- (2) To optimize the trade-off between sampling representativeness and path length, an adjustment coefficient k was introduced into the calculation formula of the sampling coefficient Q . As coefficient k increases, the modified A* algorithm plays less emphasis on path length but more emphasis on the Chl-a's spatial distribution. More representative samples will be collected at the expense of increasing path length, and vice versa.

2 Study area and materials

2.1 Overview of study area

Chaohu Lake, situated in central Anhui Province, is one of China's five prominent freshwater lakes (Figure 1). Covering approximately 770 km², the lake maintains an average depth of 2.7 m and an average water level of 8.0 m (Fang et al., 2019). Chaohu Lake receives inflow from approximately 33 rivers, with 8 primary tributaries including Nanfei River, Shiwuli River, Pai River, Hangbu River, Zhegao River, Shuangqiao River, Zhao River, and Baishitian River, and its principal outflow is the Yuxi River (Wang and Dou, 1998). The annual mean temperature in the Chaohu Lake basin is approximately 15 ~ 16°C, and the average annual rainfall varies between 500 and 1800 mm (Yang et al., 2020). The escalating population within the basin, coupled with the swift advancement of industrial and agricultural activities, has resulted in the discharge of substantial amounts of industrial effluents and municipal wastewater into the lake. This influx has notably heightened the levels of nutrients and organic substances and accelerated the process of lake eutrophication (Yang et al., 2013).

2.2 Materials

Samples in Chaohu Lake were collected from a manned vessel under clear weather conditions coinciding with the transit of Sentinel-2 satellites on June 15, 2022, August 14, 2022, and June 10, 2023. A total of 154, 80, and 49 samples were acquired on these respective dates, with the sampling sites delineated in Figure 1. At each sampling point, 1,000 mL of surface water samples were collected, and Chl-a was subsequently determined using spectrophotometric

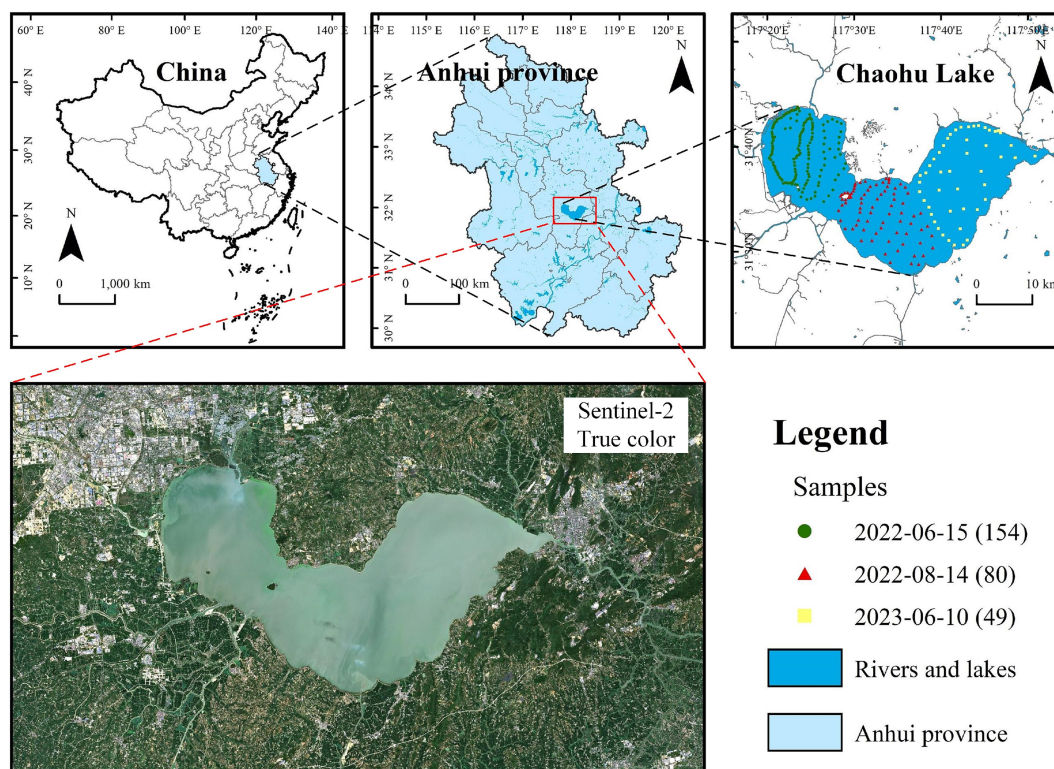


FIGURE 1
Location and true-color image of the study area, Chaohu Lake. The sample points in different colors indicate samples collected on different dates, while the numbers in brackets represent the quantity of sample points.

method once the samples were transported back to the laboratory. For all sampling points, the highest Chl-a was recorded on June 10, 2023, at the northeastern shoreline of Chaohu Lake, reaching 192.78 mg/m^3 . The lowest value was observed on June 15, 2022, at the western part of the lake, at 23.25 mg/m^3 . The average Chl-a across the sampling sites was 65.43 mg/m^3 , indicating a significant level of eutrophication in Chaohu Lake.

The harmonized Sentinel-2 Multispectral Instrument (MSI) Level-2A surface reflectance datasets from the aforementioned 3 days were gathered for retrieving Chl-a. The MSI dataset was downloaded from the Google Earth Engine platform,¹ with the cloud coverage remaining below 10%. Eleven bands from these images were selected, encompassing four 10-m resolution bands: 496 nm (B2), 560 nm (B3), 665 nm (B4) and 835 nm (B8), and six 20-m resolution bands: 704 nm (B5), 740 nm (B6), 782 nm (B7), 864 nm (B8A), 1,613 nm (B11) and 2,202 nm (B12), and one 60-m resolution band: 443 nm (B1).

3 Methods

3.1 Chlorophyll-a retrieval

In this study, a Deep Belief Network (DBN) model was utilized to retrieve Chl-a. DBN was proposed by Geoffrey Hinton in 2009

(Hinton, 2009). It is a generative model consisting of multiple layers of latent variables, constructed by several sequentially stacking Restricted Boltzmann Machines (RBMs; Mohamed et al., 2011). Each RBM undergoes layer-by-layer unsupervised pre-training, followed by supervised fine-tuning of the entire network using the backpropagation algorithm (Hua et al., 2015). DBN has found widespread applications in remote sensing retrieval (Yuan et al., 2020), such as water quality parameters (Sagan et al., 2020), leaf area index (Sun et al., 2021), and soil moisture (Yang et al., 2021). The training of the DBN in this study was conducted using an 8-layer network structure with 64 neurons in each layer and a total of 1,000 training iterations.

To fit the relationship between the *in situ* Chl-a and reflectance spectra (termed as $R_{rs}(\lambda)$, λ refers to wavelength) of MSI, 80% of the samples ($N=227$) were designated as the training dataset and input into the DBN model, while the remaining 20% ($N=56$) were reserved as the testing dataset for assessing the model's accuracy. The DBN model utilized in this study consists of 2 hidden layers, each with 128 neurons, and was trained for 500 epochs. The model used the *sigmoid* activation function, cross-entropy as the loss function, and stochastic gradient descent as the optimizer. The initial seven bands of MSI, with bandwidths ranging from 443 to 782 nm, were employed as input features (Li et al., 2021; Pahlevan et al., 2020). To further strengthen the robustness of the developed model, we included six band combinations derived from previously established high-performance Chl-a retrieval algorithms (Table 1). The bands specified in the reference model were substituted with the corresponding bands from the MSI. The output was the Chl-a (Figure 2).

¹ <https://earthengine.google.com/>

TABLE 1 The band combinations used in this study and their reference indexes.

Band combination used	Reference band combination	References
$B3(560)/B1(443)$	$R_{rs}(562)/R_{rs}(483)$	Ha et al. (2017)
$B5(705)/B4(665)$	$R_{rs}(709)/R_{rs}(665)$	Duan et al. (2012); Gurlin et al. (2011)
$B6(740)/B4(665)$	$R_{rs}(743 \sim 753)/R_{rs}(662 \sim 672)$	Gitelson et al. (2008)
$\frac{B5(705) - B4(665)}{B5(705) + B4(665)}$	$\frac{R_{rs}(709) - R_{rs}(665)}{R_{rs}(709) + R_{rs}(665)}$	Mishra and Mishra, 2012, Gitelson et al. (2008)
$(B4(665)^{-1} - B5(704)^{-1}) * B6(740)$	$(R_{rs}(\sim 670)^{-1} - R_{rs}(\sim 710)^{-1}) * R_{rs}(\sim 750)$	Gitelson et al. (2011)
$\frac{B4(665)^{-1} - B5(704)^{-1}}{B6(740)^{-1} - B5(704)^{-1}}$	$\frac{R_{rs}(665)^{-1} - R_{rs}(704)^{-1}}{R_{rs}(740)^{-1} - R_{rs}(704)^{-1}}$	Yang et al. (2010)

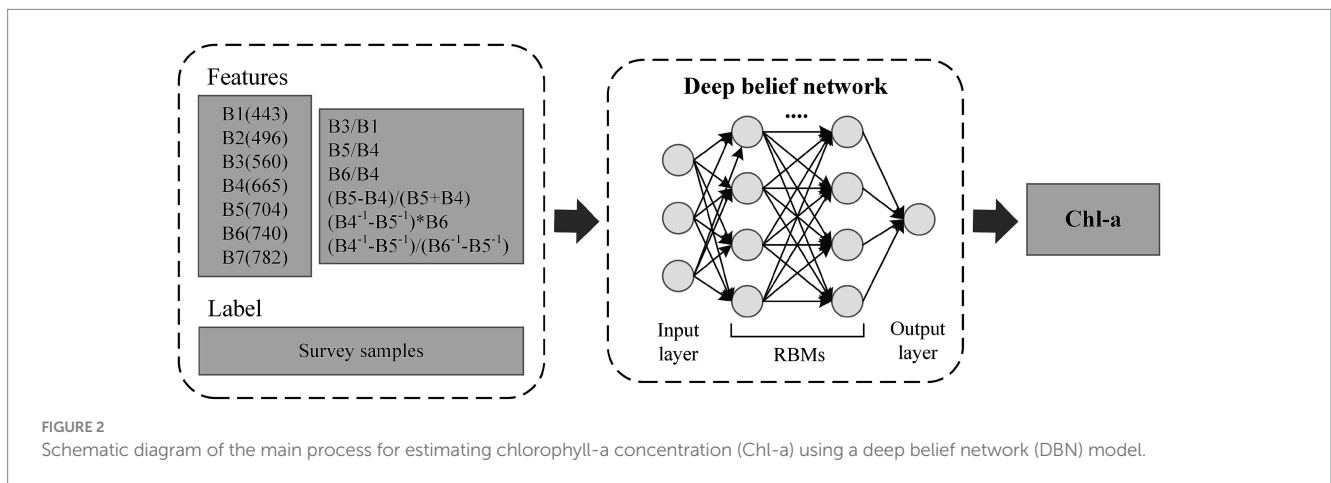


FIGURE 2 Schematic diagram of the main process for estimating chlorophyll-a concentration (Chl-a) using a deep belief network (DBN) model.

3.2 Water sampling path planning

3.2.1 A* algorithm

The A* algorithm, first published in 1968 by Peter Hart, Nils Nilsson, and Bertram Raphael from the Stanford Research Institute, is regarded as an extension of Dijkstra’s algorithm (Hart et al., 1968). It is a prominent heuristic search algorithm in the domain of artificial intelligence. The cost function of A* algorithm is formulated as Equation 1

$$f(n) = g(n) + h(n) \tag{1}$$

where $g(n)$ denotes the cost from the initial node to node n , $h(n)$ represents the estimated cost from node n to the target node, i.e., heuristic function, and $f(n)$ equals to the sum of $g(n)$ and $h(n)$ (Hart et al., 1968). The selection of the cost function is heavily influenced by the specific scenario.

The heuristic function $h(n)$ has a profound impact on the performance of the A* algorithm. If $h(n)$ is always less than or equal to the cost from node n to the target node, the A* algorithm guarantees finding the shortest path. However, a smaller value of $h(n)$ means the algorithm will explore more nodes, leading to a slower performance. If $h(n)$ is greater than the actual cost from node n to the target node,

the A* algorithm may not guarantee finding the shortest path, but it will be efficient. Hence, through modifying the heuristic function, we can regulate the algorithm’s speed and accuracy. In certain instances, instead of seeking the shortest path, we may prioritize finding a path as quickly as possible.

For grid-based graphs, the heuristic function can utilize Manhattan distance, diagonal distance, and Euclidean distance. In this study, the eight-neighborhood node expansion method was employed in a raster map, with the Euclidean distance from the current node to the target node as the heuristic function. The $h(n)$ is formulated as Equation 2

$$h(n) = \sqrt{(x_t - x_n)^2 + (y_t - y_n)^2} \tag{2}$$

where x_t and y_t denote the horizontal and vertical coordinates of the target node, respectively, whereas x_n and y_n represent the coordinates of the current node n .

3.2.2 The modified a* algorithm for water sampling path planning

Environmental modeling is the foundation of USV path planning. In this study, Chl-a retrieved in Section 3.1 is resampled to a 200 m-resolution raster map. The Chl-a value is treated as a property

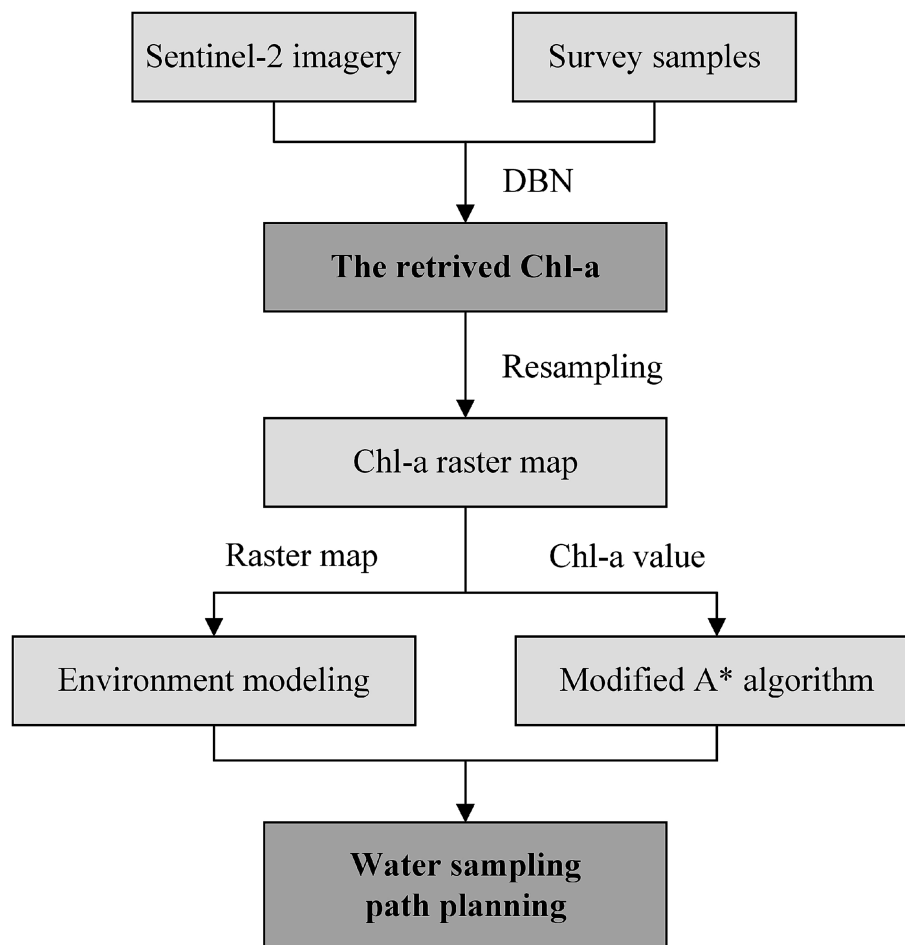


FIGURE 3
Technology roadmap of water sampling path planning for unmanned surface vehicles (USVs).

of the grid. The grid with a Chl-a value was assigned 1, while the grid with null value was assigned 0, serving as an obstacle. All the grids along a USV route are considered as sampling points. By reducing the movements of USVs to eight possible directions, the trajectory is decomposed into individual movements, each of which will be recorded. The actual grid coordinates are used as unique identifiers for each grid position, allowing the planned path to be used directly for real-world sampling. The technology roadmap of water sampling path planning for USVs is demonstrated in Figure 3.

This modified A* algorithm necessitates a route that is as brief as possible, while ensuring the collected samples are representative. To ensure the sampling representativeness, it is crucial to ensure that the average Chl-a of the collected samples closely approximates the average concentration of the entire water body. Simultaneously, maintaining a gradient in Chl-a among the collected samples is also necessary. This means that Chl-a of the collected samples should not be limited to a specific concentration range, but should instead exhibit maximum variance.

In order to achieve the above purpose, the retrieved Chl-a and path length were combined as the cost function of A* algorithm, instead of only considering the path length. The key steps of the A* algorithm involve computing the movement cost from the current

node to its eight neighboring nodes, and the node with the lowest cost is then removed from the OPEN_list (a list that contains all the nodes waiting for expansion) and added to the CLOSE_list (a list that contains all the nodes that have already been expanded). The improved cost function is $F(n)$ Equation 3

$$F(n) = f(n) * Q \quad (3)$$

$$Q = \exp(-k * |Chl - a - \overline{Chl - a}|)$$

where $f(n)$ is the cost function only considering the path length, Q is called the sampling coefficient, $Chl - a$ and $\overline{Chl - a}$ is Chl-a of an eight neighboring node and the average Chl-a of the entire water body, respectively, and $k \in [0, +\infty)$ is called the adjustment coefficient.

The adjustment coefficient k holds a pivotal position in the modified A* algorithm for water sampling path planning. By adjusting the value of k , we can modify the influence of Chl-a's spatial distribution on the routing of USVs. As k increases, the algorithm plays more emphasis on the Chl-a's spatial distribution. Thus, USV is more likely to navigate towards nodes where the Chl-a deviates from

the average concentration of the entire water body ($\overline{Chl-a}$). Conversely, a smaller k implies a greater influence of path length on route selection, with USVs more likely to choose a shorter route. In the extreme case, when the coefficient k equals to zero, the modified A* algorithm degenerates into a classic A* algorithm.

To validate the effectiveness of this algorithm, we recorded the path length, as well as the mean and variance of Chl-a in the collected samples, and analyzed their change trend as coefficient k increases. We also performed a one-way analysis of variance (ANOVA) on Chl-a of collected samples under different k values to assess whether the coefficient k has a significant effect on the variability of the collected samples.

4 Results

4.1 Chlorophyll-a retrieval

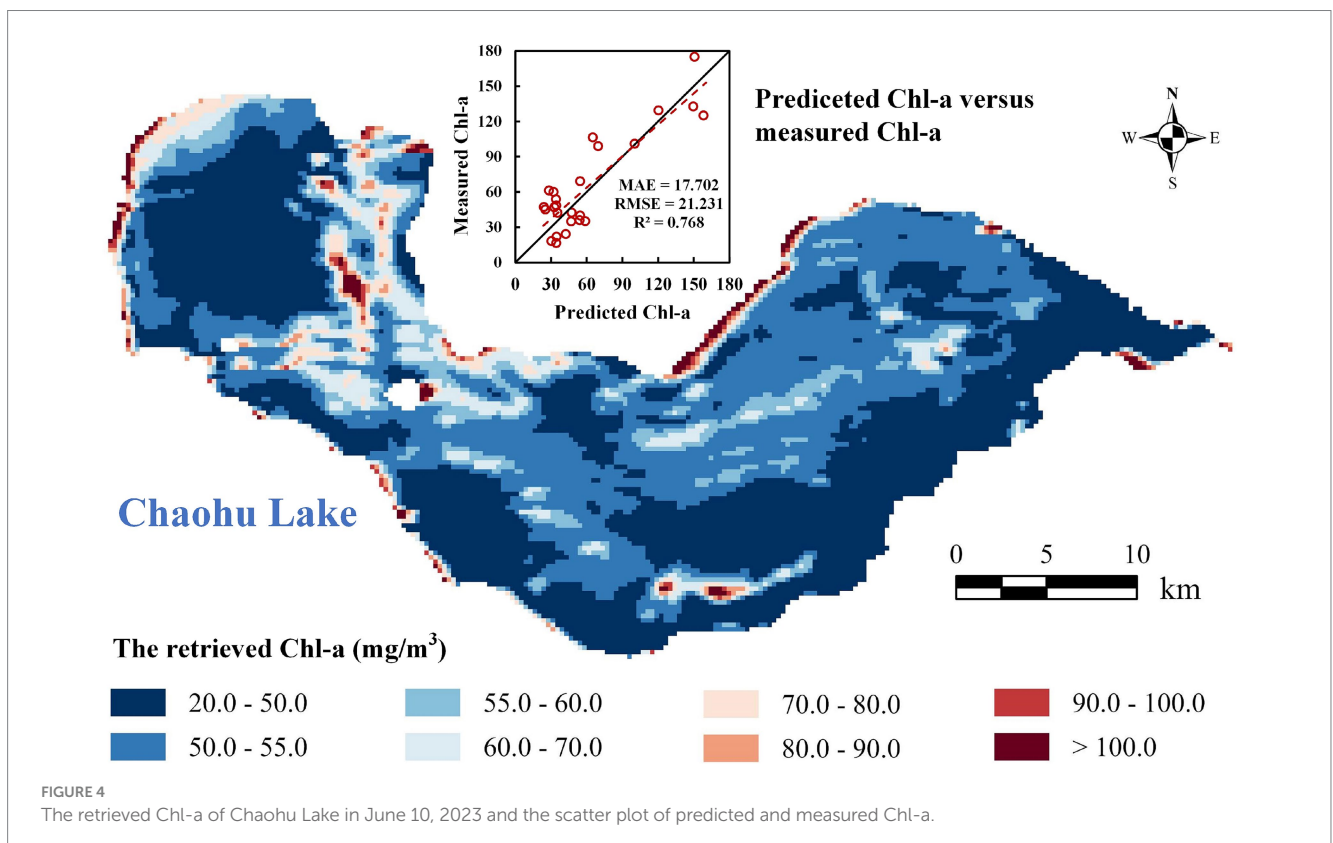
The statistical metrics obtained from the training and testing dataset indicate good performance for the DBN model, with an R^2 of 0.883 and 0.821, respectively. The retrieved Chl-a of Chaohu Lake in June 10, 2023 is demonstrated in Figure 4 after resampling to 200-m resolution. The retrieval results reveal that Chl-a is higher in the shoreline of Chaohu Lake and the eastern coast of its islands, usually exceeding 100 mg/m^3 , and even achieving 180 mg/m^3 . Except for these areas, the majority of regions exhibit low Chl-a, not exceeding 50 mg/m^3 . Moreover, Chl-a in the western part of Chaohu Lake is higher compared to the eastern part. This is primarily due to the proximity of the western part of Chaohu Lake to urban areas, where domestic and industrial wastewater is discharged.

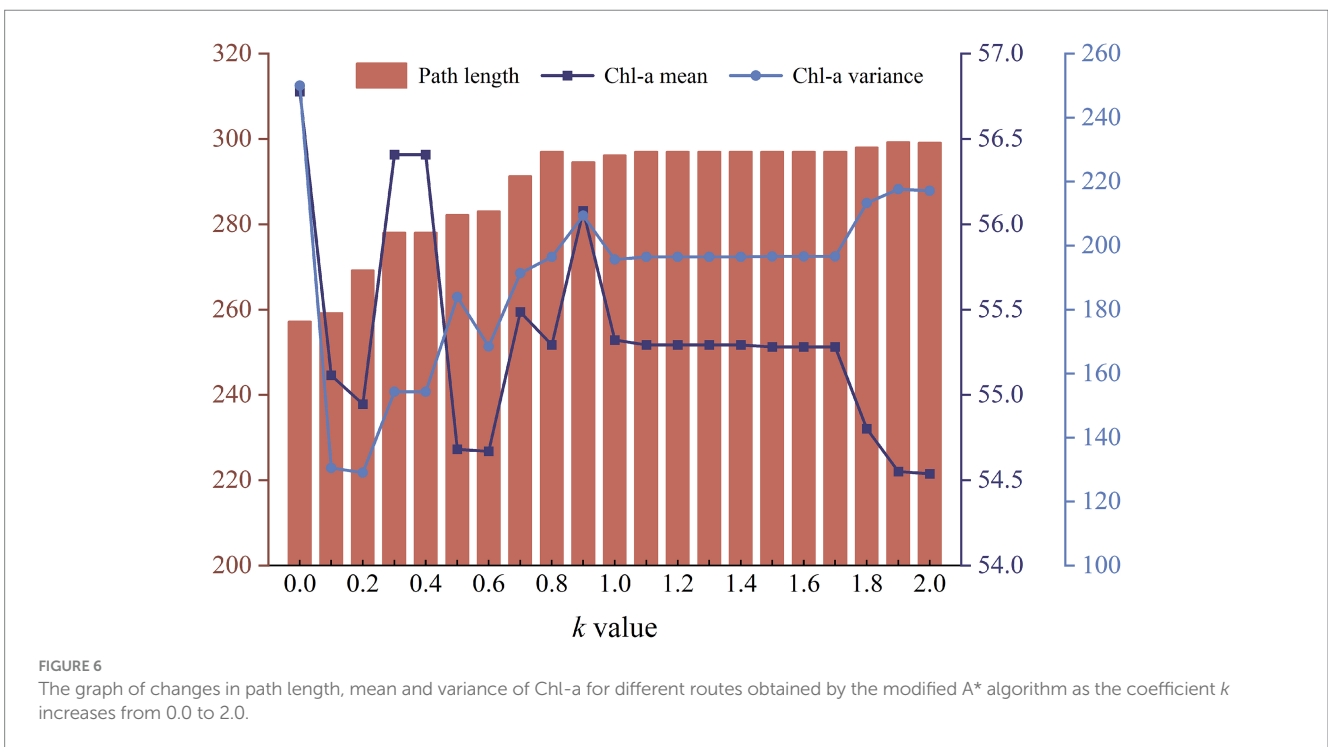
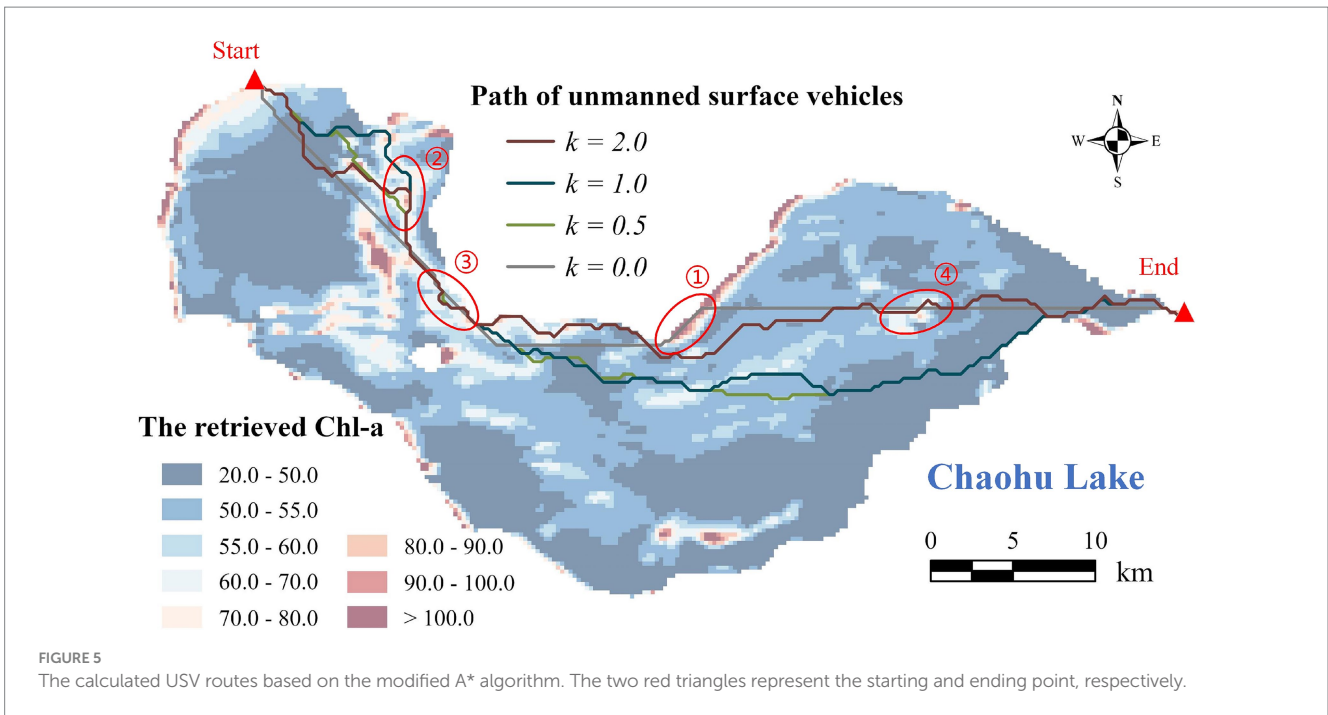
The measured Chl-a values at 49 points on June 10, 2023, are compared with the retrieval values, and the resulting scatter plot is shown in Figure 4. The mean relative error (MAE), root mean square error (RMSE) and R^2 are 17.702, 21.231, and 0.768, respectively. This indicates that the retrieved Chl-a demonstrates high accuracy, providing a robust foundation for subsequent path planning. Based on the resampled Chl-a raster map, we completed the environmental modeling for Chaohu Lake, as demonstrated in Supplementary Figure S1 online.

4.2 Water sampling path planning

By adjusting the coefficient k from 0.0 to 2.0, various USV routes are derived (Figure 5; Supplementary Figure S2). The grey route ($k=0$) represents the shortest trajectory, with the path length of 257.17 grid units, where the spatial resolution of the raster Chl-a is 200 m. The routes depicted in green, cyan, and maroon correspond to k values of 0.5, 1.0, and 2.0, respectively, with path lengths of 282.18, 296.18, and 299.08 grid units accordingly.

As the coefficient k increases, the modified A* algorithm demonstrates a decreasing emphasis on path length but a growing emphasis on the Chl-a's spatial distribution (Figure 6). The path length gradually increases from 257.17 ($k=0$) to 297.00 grid units ($k=0.8$), until reaching a plateau between 297.00 and 299.08 grid units ($k=2.0$). The average Chl-a of the collected samples exhibits a fluctuating decline from 56.78 to 54.54 mg/m^3 , gradually approaching the mean concentration of the entire water body ($\overline{Chl-a} = 53.70 \text{ mg/m}^3$). The variance of Chl-a for the collected samples increases from 130.51 ($k=0.1$) to 217.13 ($k=2.0$). When $k=0$, the variance of Chl-a is





extremely high (Figure 6), because the route travels through an area with exceptionally high concentration (Figure 5 ①).

Chl-a of all collected samples along USV routes with different coefficient k is represented in Figure 7. The peak in the gray line ($k = 0$) marked with a red circle indicates the area with unusually high Chl-a (Figure 5 ①). As the coefficient k rises from 0.5 to 2.0, the line chart of Chl-a exhibits increasingly pronounced fluctuations, with corresponding variances of 183.97, 195.67, and

217.13, respectively. This phenomenon can also be observed in Figure 5 ②–④, where the three colored routes, as opposed to the grey route (the shortest route), tend to pass through areas where Chl-a deviates from the overall mean concentration. The results of one-way ANOVA revealed that coefficient k has a significant effect on the variability of Chl-a in the collected samples, with an F value of 1.781 exceeding the critical value of 1.573, and a p value of 0.017 (<0.05).

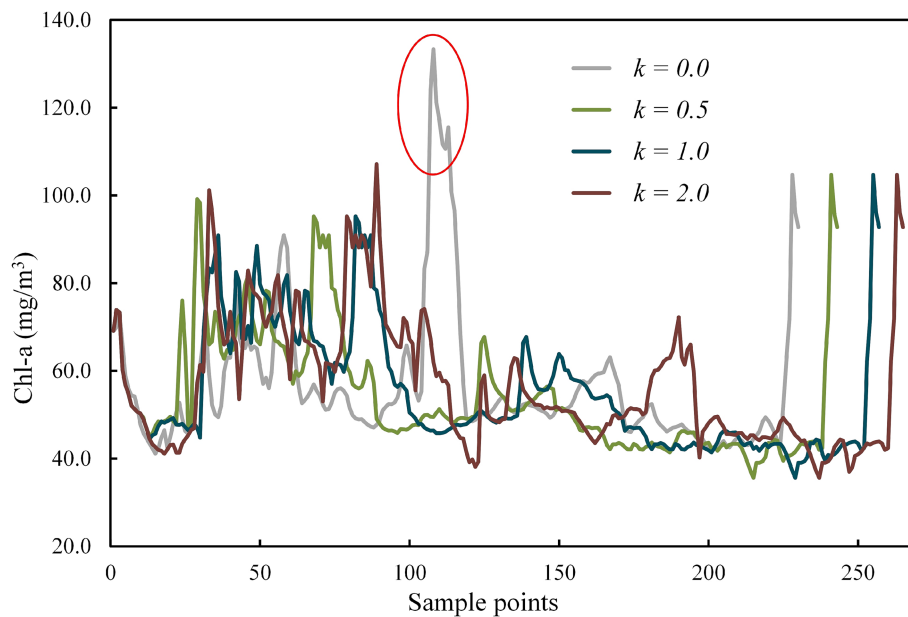


FIGURE 7

Chl-a of collected samples for routes with different coefficient k obtained by the modified A* algorithm.

5 Discussion

5.1 Comparisons with previous studies

This study proposed a modified A* algorithm that takes into account not only the path length, but also the sampling representativeness in the water quality sampling path planning of USVs. The verification results showed that the proposed algorithm was effective in real-world conditions and achieved the trade-off between the representativeness of sampling points and the path length. As the coefficient k increased, the representativeness of the collected samples enhanced, but this came at the cost of increased path length, and vice versa.

This study differs significantly from previous researches, which focused primarily on factors such as path length, time consumption, path smoothness, and collision avoidance. Until now, there have been no observed studies on USV-based water quality sampling path planning that consider the representativeness of collected samples. Table 2 lists several typical studies on USV path planning, including studies on water quality sampling path planning of USVs. Researches on path planning with a single target point (referred to here as single-goal path planning) can be broadly categorized into global, local, and hybrid algorithms. In all three types of studies, path design typically takes into account both path length and time consumption, while some studies also consider path smoothness. Global path planning also focuses on global obstacle avoidance, whereas local path planning usually emphasizes dynamic obstacle avoidance. Hybrid algorithms typically take into account all the aforementioned four factors. For instance, Yu et al. (2019) introduced a hybrid approach that combined A* and artificial potential field. This method took into account the four factors and demonstrated effectiveness in surface water environment with unknown districts. Multi-goal path planning algorithms typically focus on path length and time consumption, with few studies also considering path smoothness, such as Zhao et al. (2023) and Yu et al. (2024).

Another distinguishing aspect of our path planning research with current studies is its integration of remote sensing technology, which is rarely seen in current USV path planning field. Although, some recent studies have employed satellite imagery for environmental modelling in USV path planning, demonstrating the practicality of their algorithms in the real world (Yang et al., 2015; Su et al., 2023; Song et al., 2019), these studies fail to fully harness the information contained within the imagery. Our study, for instance, utilized Chl-a values retrieved from remote sensing imagery as prior knowledge to facilitate the collection of representative samples.

5.2 Possible factors influencing the adjustment amplitude of coefficient k

As mentioned above, the modified A* algorithm can alter the emphasis placed on path length and Chl-a's spatial distribution by adjusting the coefficient k . Therefore, after obtaining the retrieved Chl-a and determining the starting and ending points of USVs, our only approach is to incrementally increase the k value to achieve the optimal routing. However, we still do not know the appropriate adjustment amplitude for a specific water body. The distance between the starting and ending points and the data distribution of Chl-a are two potential factors influencing the adjustment amplitude of coefficient k .

5.2.1 The distance between the starting and ending points

The distance between the starting and end points of USVs has a significant impact the adjustment amplitude of coefficient k , with the adjustment amplitude of k generally increasing as the distance between these two points decreases. In practical experiments, it is common for the starting and end points of USVs to be very close or even coinciding. In this case, selecting routes with more representative samples, as opposed to the shortest route, will lead to a significant

TABLE 2 Factors in consideration for previous studies on USV path planning.

Single-/Multi-goal path planning	Algorithm	Factors in consideration				References	
		Path length	Time consumption	Path smoothness	Dynamic obstacle avoidance		
Single-goal	Global	A* algorithm	✓	✓	✓	Song et al. (2019)	
		Constrained A* algorithm	✓	✓		✓	Singh et al. (2018)
		Complete coverage neural network	✓	✓	✓	✓	Xu et al. (2021)
		Finite Angle A* algorithm	✓	✓	✓		Yang et al. (2015)
	Local	Constrained fast marching	✓	✓		✓	Liu and Bucknall, 2015
		Locking sweeping method	✓	✓	✓	✓	Su et al. (2023)
		Improved artificial potential field	✓	✓		✓	Song et al. (2018)
		Model predictive artificial potential field	✓	✓	✓	✓	He et al. (2023)
	Hybrid	A* + Artificial potential field	✓	✓	✓	✓	Yu et al. (2019)
		B-Spline data frame + Particle swarm optimization	✓	✓	✓	✓	MahmoudZadeh et al. (2022)
		Improved A* + Artificial potential field	✓	✓	✓	✓	Sang et al. (2021)
		Non-uniform Theta* + Improved dynamic window approach	✓	✓	✓	✓	Han et al. (2022)
	Multi-goal	Greedy Partheno Genetic Algorithm	✓	✓	✓		Zhao et al. (2023)
Genetic algorithm based on improved crossover operators		✓	✓			Chen et al. (2021)	
Self-organizing map + Artificial repulsive force field		✓	✓			Liu and Bucknall, 2018	
D*Lite algorithm + Dubins search tree algorithm + Reeds-Shepp curves		✓	✓	✓		Yu et al. (2024)	

increase in path length. Therefore, after calculating the total cost, the modified A* algorithm tends to select a route with a shorter length rather than one with more representative samples. Thus, to address this issue, increasing the value of coefficient k is essential to place greater emphasis on the spatial distribution of Chl-a within the algorithm. This adjustment will compel the algorithm to overcome the cost associated with increased path length and to consider the spatial distribution of Chl-a.

Figure 8 suggests that when the starting and ending points are in close proximity, the adjustment amplitude of coefficient k is 20 times greater in order to significantly alter the routes compared to when these two points are at a greater distance (Figure 5). Additionally, the routes calculated with an adjustment amplitude of 1.0 are depicted in Supplementary Figure S3. It is evident that the routes generated by the modified A* algorithm ($k > 0$) primarily focus on path length rather than the Chl-a's spatial distribution, resulting in slight variations compared to the shortest path ($k = 0$).

5.2.2 The data distribution of Chl-a

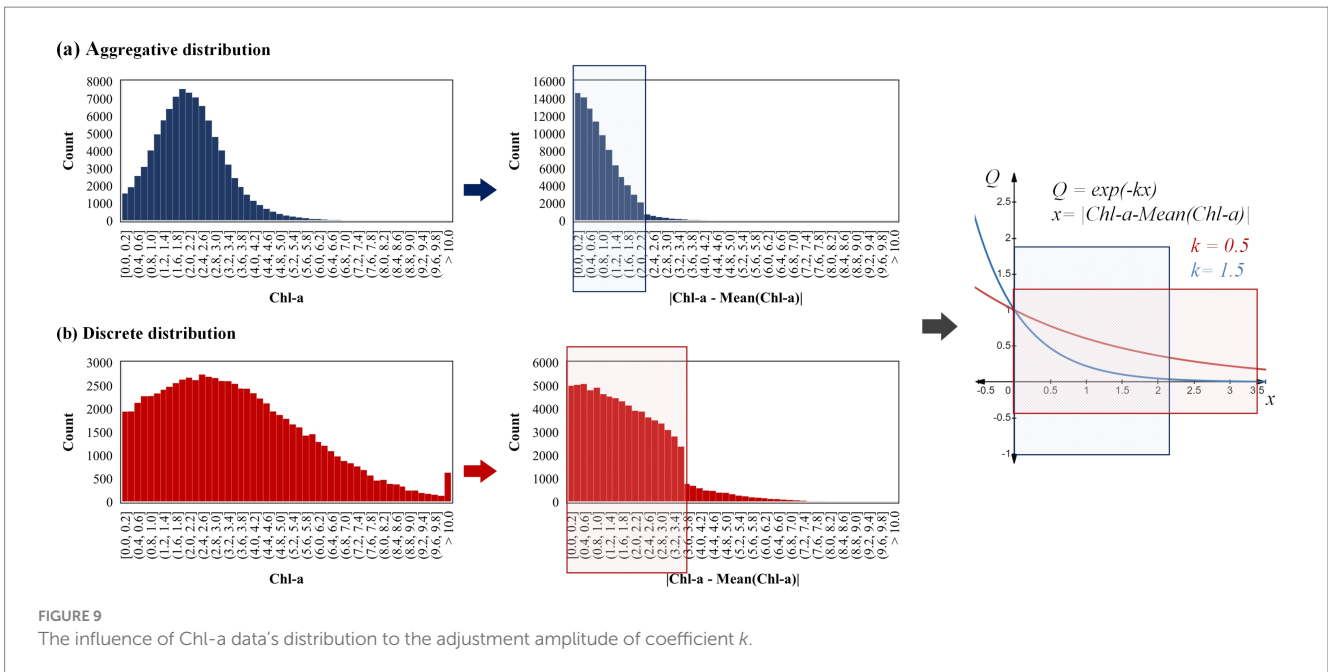
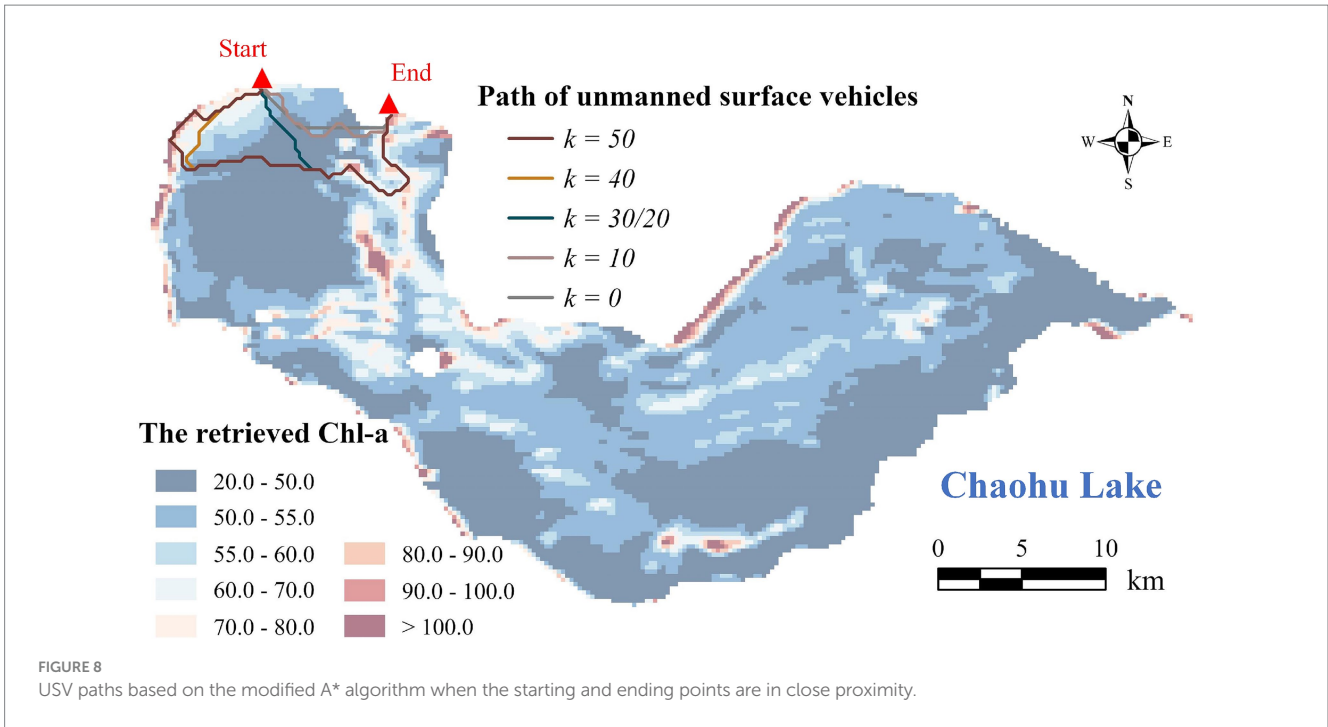
As the aggregation degree of Chl-a's data distribution increases, a larger adjustment amplitude of coefficient k is typically required. For the sampling coefficient $Q = \exp(-kx)$, the convergence rate to zero accelerates with an increase of the adjustment coefficient k ($k \geq 0$). As the distribution of Chl-a data becomes more aggregated, the value of $|Chl - a - Chl - a|$ will tend to concentrate to zero (Figure 9). This

necessitates that the function $Q = \exp(-kx)$ converges rapidly, resulting in a broadened distribution range of Q values (avoiding excessive concentration to 1). The expansion of Q will further affect the distribution of the cost function $F(n)$ (Equation 3), thereby amplifying the influence of Chl-a's spatial distribution on route regulation.

The two datasets presented in Figure 9 are used as a reference. Dataset A, which is more aggregated with a smaller range of $|Chl - a - Chl - a|$ is more suitable for the function $Q = \exp(-1.5x)$ with faster convergence speed (colored with blue). Conversely, Dataset B better suited for the function $Q = \exp(-0.5x)$ with slower convergence speed (colored with red). However, if the function $Q = \exp(-0.5x)$ is employed for the aggregated dataset A, the obtained Q value will tend to concentrate to 1. Conversely, if the function $Q = \exp(-1.5x)$ is employed for dataset B, the range of Q values obtained will be concentrated to 0. Therefore, when the aggregation degree of Chl-a's data distribution increases, a larger adjustment amplitude of coefficient k is typically more appropriate.

5.3 Shortcomings and future work

This study proposed a modified A* algorithm that concurrently considers the spatial distribution of water quality parameters and path length. While this algorithm exhibits perfect results, it also has several shortcomings.



Firstly, the use of deep learning constrains the practicality of this approach. Actually, the algorithm proposed in this study does not necessitate high precision in the retrieval of water quality parameters, and it solely requires an approximate determination of their spatial distribution. Thus, empirical models (Woo Kim et al., 2022; Li et al., 2021; Gurlin et al., 2011; Gitelson et al., 2008) can also be effectively employed in the parameter retrieval. Moreover, the readily available Chl-a products from satellites like Sentinel-3 can also be used directly, which could provide comprehensive coverage and consistent data quality (Pahlevan et al., 2020).

Secondly, the algorithm is not applicable when the starting and ending points are completely coincident. Nevertheless, effectively

addressing the routing problem can be achieved by maintaining an adequate distance between these two points and increasing the adjustment amplitude of coefficient k .

Thirdly, considering that water quality parameters often display a right-skewed distribution (with the distribution of retrieved Chl-a in Chaohu Lake presented in Supplementary Figure S4), it is worth further exploring whether using the median instead of the mean ($\text{Chl}-\bar{a}$) would be more suitable in cases of extreme right skewness.

Fourthly, this study utilized Chl-a as the water quality parameter to test the proposed algorithm, and additional water parameters such as turbidity, colored dissolved organic matter (CDOM), and total

suspended matter (TSM) could also be employed for further testing purposes.

Fifthly, the proposed modified A* algorithm is an initial version aimed at highlighting the importance of representativeness of sampling points in water sampling path planning. Future optimizations could include collision avoidance, increasing path smoothness, and improving efficiency etc.

6 Conclusion

To address the easily overlooked problem of representative sampling in the water quality sampling path planning of USVs, a modified A* algorithm that integrates remote sensing technique has been proposed. Specifically, the heuristic function of the A* algorithm was modified by incorporating a sampling coefficient Q to ensure the representativeness of collected samples. Subsequently, an adjustment coefficient k was introduced into the coefficient Q to optimize the trade-off between the representativeness of sampling points and the path length. The verification results showed that the proposed algorithm was effective in collecting more representative samples in real-world conditions. As the coefficient k increased, the representativeness of the collected samples enhanced, indicated by the Chl-a closely approximating the overall mean Chl-a and exhibiting a gradient distribution. However, this enhancement was also associated with increased path length, and vice versa. The distance between the starting and ending points and the data distribution of water quality parameters are two potential factors influencing the adjustment amplitude of coefficient k . A reduced distance between the starting and ending points or an increased degree of aggregation in the distribution of parameter data will result in a larger adjustment amplitude of coefficient k .

Regarding future research, Chl-a values retrieved from empirical models or Chl-a products can serve as prior knowledge for water quality sampling path planning. Besides Chl-a, water parameters like turbidity, CDOM, and TSM can be employed for algorithm testing. The proposed algorithm should also be optimized to avoid collision, increase path smoothness, and improve efficiency etc.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration,

Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. DL: Data curation, Investigation, Software, Visualization, Writing – review & editing. JW: Data curation, Investigation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work is supported by the Research on Decision-making Mechanisms of Unmanned Water Quality Monitoring Boats Based on Multi-Sensor Fusion Technology (KJ2021A0982) and End-to-End Autonomous Driving Research Based on Visual Perception Multi-Task Learning in Small Sample Scenarios (KJ2021A0978).

Acknowledgments

We express our appreciation to Xue Yang from Hefei USTAR Technology Co., Ltd. for providing invaluable support in terms of unmanned surface vehicles and sampling data. Additionally, we are grateful to Dechao Zhai from Peking University for their dedicated efforts in imparting the field knowledge of remote sensing and enhancing the outcomes of remote sensing retrieval.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbot.2024.1488337/full#supplementary-material>

References

- Behmel, S., Damour, M., Ludwig, R., and Rodriguez, M. J. (2016). Water quality monitoring strategies - a review and future perspectives. *Sci. Total Environ.* 571, 1312–1329. doi: 10.1016/j.scitotenv.2016.06.235
- Bhagwat, V. R. (2019). Safety of water used in food production. *Food Safety Hum. Health* 2019, 219–247. doi: 10.1016/B978-0-12-816333-7.00009-6
- Campbell, J. B., and Wynne, R. H. (2011). Introduction to remote sensing. New York: Guilford Press.
- Casalino, G., Turetta, A., and Simetti, E. (2009). A three-layered architecture for real time path planning and obstacle avoidance for surveillance USVs operating in harbour fields. Oceans 2009-Europe, Bremen, Germany. pp. 1–8.

- Chen, W., Hao, X., Yan, K., Lu, J., Liu, J., He, C., et al. (2021). The Mobile water quality monitoring system based on low-power wide area network and unmanned surface vehicle. *Wirel. Commun. Mob. Comput.* 2021:1609612. doi: 10.1155/2021/1609612
- Duan, H., Ma, R., and Hu, C. (2012). Evaluation of remote sensing algorithms for cyanobacterial pigment retrievals during spring bloom formation in several lakes of East China. *Remote Sens. Environ.* 126, 126–135. doi: 10.1016/j.rse.2012.08.011
- Fang, T., Yang, K., Lu, W., Cui, K., Li, J., Liang, Y., et al. (2019). An overview of heavy metal pollution in Chaohu Lake, China: enrichment, distribution, speciation, and associated risk under natural and anthropogenic changes. *Environ. Sci. Pollut. Res.* 26, 29585–29596. doi: 10.1007/s11356-019-06210-x
- Gholizadeh, M. H., Melesse, A. M., and Reddi, L. (2016). A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors* 16:1298. doi: 10.3390/s16081298
- Gitelson, A. A., Dall'olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., et al. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: validation. *Remote Sens. Environ.* 112, 3582–3593. doi: 10.1016/j.rse.2008.04.015
- Gitelson, A. A., Gao, B. C., Li, R. R., Berdnikov, S., and Saprygin, V. (2011). Estimation of chlorophyll-a concentration in productive turbid waters using a hyperspectral imager for the Coastal Ocean—the Azov Sea case study. *Environ. Res. Lett.* 6:024023. doi: 10.1088/1748-9326/6/2/024023
- Gurlin, D., Gitelson, A. A., and Moses, W. J. (2011). Remote estimation of chl-a concentration in turbid productive waters—return to a simple two-band NIR-red model? *Remote Sens. Environ.* 115, 3479–3490. doi: 10.1016/j.rse.2011.08.011
- Ha, N. T. T., Koike, K., Nhuan, M. T., Canh, B. D., Thao, N. T. P., and Parsons, M. (2017). Landsat 8/OLI two bands ratio algorithm for chlorophyll-a concentration mapping in hypertrophic waters: an application to West Lake in Hanoi (Vietnam). *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 10, 4919–4929. doi: 10.1109/jstars.2017.2739184
- Han, S., Wang, L., Wang, Y., and He, H. (2022). A dynamically hybrid path planning for unmanned surface vehicles based on non-uniform Theta* and improved dynamic windows approach. *Ocean Eng.* 257:111655. doi: 10.1016/j.oceaneng.2022.111655
- Hart, P. E., Nilsson, N. J., and Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans. Syst. Sci. Cybern.* 4, 100–107. doi: 10.1109/TSSC.1968.300136
- He, Z., Chu, X., Liu, C., and Wu, W. (2023). A novel model predictive artificial potential field based ship motion planning method considering COLREGs for complex encounter scenarios. *ISA Trans.* 134, 58–73. doi: 10.1016/j.isatra.2022.09.007
- Hinton, G. E. (2009). Deep belief networks. *Scholarpedia* 4:5947. doi: 10.4249/scholarpedia.5947
- Hua, Y., Guo, J., and Zhao, H. (2015). “Deep belief networks and deep learning.” in *Proceedings of 2015 international conference on intelligent computing and internet of things, Harbin, China*. pp. 1–4.
- Li, S., Song, K., Wang, S., Liu, G., Wen, Z., Shang, Y., et al. (2021). Quantification of chlorophyll-a in typical lakes across China using Sentinel-2 MSI imagery with machine learning algorithm. *Sci. Total Environ.* 778:146271. doi: 10.1016/j.scitotenv.2021.146271
- Liu, Y., and Bucknall, R. (2015). Path planning algorithm for unmanned surface vehicle formations in a practical maritime environment. *Ocean Eng.* 97, 126–144. doi: 10.1016/j.oceaneng.2015.01.008
- Liu, Y., and Bucknall, R. (2018). Efficient multi-task allocation and path planning for unmanned surface vehicle in support of ocean operations. *Neurocomputing* 275, 1550–1566. doi: 10.1016/j.neucom.2017.09.088
- Luo, J., Zhang, Y., Zhuang, J., and Su, Y. (2024). Intelligent task allocation and planning for unmanned surface vehicle (USV) using self-attention mechanism and locking sweeping method. *J. Mar. Sci. Eng.* 12:179. doi: 10.3390/jmse12010179
- Ma, S., Guo, W., Song, R., and Liu, Y. (2021). Unsupervised learning based coordinated multi-task allocation for unmanned surface vehicles. *Neurocomputing* 420, 227–245. doi: 10.1016/j.neucom.2020.09.031
- Mahmoudzadeh, S., Abbasi, A., Yazdani, A., Wang, H., and Liu, Y. (2022). Uninterrupted path planning system for multi-USV sampling mission in a cluttered ocean environment. *Ocean Eng.* 254:111328. doi: 10.1016/j.oceaneng.2022.111328
- Mishra, S., and Mishra, D. R. (2012). Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sens. Environ.* 117, 394–406. doi: 10.1016/j.rse.2011.10.016
- Mohamed, A. R., Dahl, G. E., and Hinton, G. (2011). Acoustic modeling using deep belief networks. *IEEE Trans. Audio Speech Lang. Process.* 20, 14–22. doi: 10.1109/TASL.2011.2109382
- Naeem, W., Irwin, G. W., and Yang, A. (2012). COLREGs-based collision avoidance strategies for unmanned surface vehicles. *Mechatronics* 22, 669–678. doi: 10.1016/j.mechatronics.2011.09.012
- Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., et al. (2020). Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: a machine-learning approach. *Remote Sens. Environ.* 240:111604. doi: 10.1016/j.rse.2019.111604
- Peng, Z., Wang, J., Wang, D., and Han, Q. (2021). An overview of recent advances in coordinated control of multiple autonomous surface vehicles. *IEEE Trans. Ind. Inform.* 17, 732–745. doi: 10.1109/TII.2020.3004343
- Pule, M., Yahya, A., and Chuma, J. (2017). Wireless sensor networks: a survey on monitoring water quality. *J. Appl. Res. Technol.* 15, 562–570. doi: 10.1016/j.jart.2017.07.004
- Ritchie, J. C., Zimba, P. V., and Everitt, J. H. (2003). Remote sensing techniques to assess water quality. *Photogramm. Eng. Remote. Sens.* 69, 695–704. doi: 10.14358/PERS.69.6.695
- Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., et al. (2020). Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Sci. Rev.* 205:103187. doi: 10.1016/j.earscirev.2020.103187
- Sang, H., You, Y., Sun, X., Zhou, Y., and Liu, F. (2021). The hybrid path planning algorithm based on improved A* and artificial potential field for unmanned surface vehicle formations. *Ocean Eng.* 223:108709. doi: 10.1016/j.oceaneng.2021.108709
- Schwarzenbach, R. P., Egli, T., Hofstetter, T. B., Von Gunten, U., and Wehrli, B. (2010). Global water pollution and human health. *Annu. Rev. Environ. Resour.* 35, 109–136. doi: 10.1146/annurev-environ-100809-125342
- Singh, Y., Sharma, S., Sutton, R., Hatton, D., and Khan, A. (2018). A constrained A* approach towards optimal path planning for an unmanned surface vehicle in a maritime environment containing dynamic obstacles and ocean currents. *Ocean Eng.* 169, 187–201. doi: 10.1016/j.oceaneng.2018.09.016
- Song, R., Liu, Y., and Bucknall, R. (2019). Smoothed A* algorithm for practical unmanned surface vehicle path planning. *Appl. Ocean Res.* 83, 9–20. doi: 10.1016/j.apor.2018.12.001
- Song, A. L., Su, B. Y., Dong, C. Z., Shen, D. W., Xiang, E. Z., and Mao, F. P. (2018). A two-level dynamic obstacle avoidance algorithm for unmanned surface vehicles. *Ocean Eng.* 170, 351–360. doi: 10.1016/j.oceaneng.2018.10.008
- Su, Y., Luo, J., Zhuang, J., Song, S., Huang, B., and Zhang, L. (2023). A constrained locking sweeping method and velocity obstacle based path planning algorithm for unmanned surface vehicles in complex maritime traffic scenarios. *Ocean Eng.* 279:113538. doi: 10.1016/j.oceaneng.2022.113538
- Sun, L., Wang, W., Jia, C., and Liu, X. (2021). Leaf area index remote sensing based on deep belief network supported by simulation data. *Int. J. Remote Sens.* 42, 7637–7661. doi: 10.1080/01431161.2021.1942584
- Vagale, A., Ouicheikh, R., Bye, R. T., Osen, O. L., and Fossen, T. I. (2021). Path planning and collision avoidance for autonomous surface vehicles I: a review. *J. Mar. Sci. Technol.* 26, 1292–1306. doi: 10.1007/s00773-020-00787-6
- Wang, S., and Dou, H. (1998). *China lakes*. Beijing: Science Press.
- Wang, X., and Yang, W. (2019). Water quality monitoring and evaluation using remote sensing techniques in China: a systematic review. *Ecosyst. Health Sustain.* 5, 47–56. doi: 10.1080/20964129.2019.1571443
- Woo Kim, Y., Kim, T., Shin, J., Lee, D. S., Park, Y. S., Kim, Y., et al. (2022). Validity evaluation of a machine-learning model for chlorophyll-a retrieval using Sentinel-2 from inland and coastal waters. *Ecol. Indic.* 137:108737. doi: 10.1016/j.ecolind.2022.108737
- Xing, B., Yu, M., Liu, Z., Tan, Y., Sun, Y., and Li, B. (2023). A review of path planning for unmanned surface vehicles. *J. Mar. Sci. Eng.* 11:1556. doi: 10.3390/jmse11081556
- Xu, P., Ding, Y., and Luo, J. (2021). Complete coverage path planning of an unmanned surface vehicle based on a complete coverage neural network algorithm. *J. Mar. Sci. Eng.* 9:1163. doi: 10.3390/jmse9111163
- Yang, X., Cui, H., Liu, X., Wu, Q., and Zhang, H. (2020). Water pollution characteristics and analysis of Chaohu Lake basin by using different assessment methods. *Environ. Sci. Pollut. Res.* 27, 18168–18181. doi: 10.1007/s11356-020-08189-2
- Yang, L., Lei, K., Meng, W., Fu, G., and Yan, W. (2013). Temporal and spatial changes in nutrients and chlorophyll-a in a shallow lake, Lake Chaohu, China: an 11-year investigation. *J. Environ. Sci.* 25, 1117–1123. doi: 10.1016/S1001-0742(12)60171-5
- Yang, W., Matsushita, B., Chen, J., Fukushima, T., and Ma, R. (2010). An enhanced three-band index for estimating chlorophyll-a in turbid case-II waters: case studies of Lake Kasumigaura, Japan, and Lake Dianchi, China. *IEEE Geosci. Remote Sens. Lett.* 7, 655–659. doi: 10.1109/LGRS.2010.2044364
- Yang, J. M., Tseng, C. M., and Tseng, P. S. (2015). Path planning on satellite images for unmanned surface vehicles. *Int. J. Nav. Archit. Ocean Eng.* 7, 87–99. doi: 10.1515/ijnaoe-2015-0007
- Yang, Z., Zhao, J., Liu, J., Wen, Y., and Wang, Y. (2021). Soil moisture retrieval using microwave remote sensing data and a deep belief network in the Naqu region of the Tibetan plateau. *Sustain. For.* 13:12635. doi: 10.3390/su132212635
- Young, S. N., and Peschel, J. M. (2020). Review of human-machine interfaces for small unmanned systems with robotic manipulators. *IEEE Trans. Hum. Mach. Syst.* 50, 131–143. doi: 10.1109/THMS.2020.2969380
- Yu, J., Chen, Y., Yang, M., Chen, Z., Xu, J., Lu, Y., et al. (2024). A path planning algorithm for unmanned surface vessel with pose constraints in an unknown environment. *Int. J. Nav. Archit. Ocean Eng.* 16:100602. doi: 10.1016/j.ijnaoe.2024.100602
- Yu, J., Deng, W., Zhao, Z., Wang, X., Xu, J., Wang, L., et al. (2019). A hybrid path planning method for an unmanned cruise ship in water quality sampling. *IEEE Access* 7, 87127–87140. doi: 10.1109/ACCESS.2019.2925894

Yu, J., Yang, M., Zhao, Z., Wang, X., Bai, Y., Wu, J., et al. (2022). Path planning of unmanned surface vessel in an unknown environment based on improved D*lite algorithm. *Ocean Eng.* 266:112873. doi: 10.1016/j.oceaneng.2022.112873

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., et al. (2020). Deep learning in environmental remote sensing: achievements and challenges. *Remote Sens. Environ.* 241:111716. doi: 10.1016/j.rse.2020.111716

Zhang, C. (2024). *Fundamentals of environmental sampling and analysis*. Hoboken: John Wiley & Sons.

Zhao, L., Bai, Y., and Paik, J. K. (2023). Global path planning and waypoint following for heterogeneous unmanned surface vehicles assisting inland water monitoring. *J. Ocean Eng. Sci.* 2023, 1–21. doi: 10.1016/j.joes.2023.07.002