



OPEN

## A method for compressing AIS trajectory based on the adaptive core threshold difference Douglas–Peucker algorithm

Ting Zhang<sup>1,2,4</sup>, Zhiming Wang<sup>1,4</sup> & Peiliang Wang<sup>1,3</sup>✉

Traditional trajectory compression algorithms, such as the sliding window (SW) algorithm and the Douglas–Peucker (DP) algorithm, typically use static thresholds based on fixed parameters like ship dimensions or predetermined distances, which limits their adaptive capabilities. In this paper, the adaptive core threshold difference-DP (ACTD-DP) algorithm is proposed based on traditional DP algorithm. Firstly, according to the course value of automatic identification system (AIS) data, the original trajectory data is preprocessed and some redundant points are discarded. Then the number of compressed trajectory points corresponding to different thresholds is quantified. The function relationship between them is established by curve fitting method. The characteristics of the function curve are analyzed, and the core threshold and core threshold difference are solved. Finally, the compression factor is introduced to determine the optimal core threshold difference, which is the key parameter to control the accuracy and efficiency of the algorithm. Five different algorithms are used to compress the all ship trajectories in the experimental water area. The average compression ratio (ACR) of the ACTD-DP algorithm is 87.53%, the average length loss ratio (ALLR) is 23.20%, the AMSED (mean synchronous Euclidean distance of all trajectories) is 68.9747 mx, and the TIME is 25.6869 s. Compared with the other four algorithms, the ACTD-DP algorithm shows that the algorithm can not only achieve high compression ratio, but also maintain the integrity of trajectory shape. At the same time, the compression results of four different trajectories show that ACTD-DP algorithm has good robustness and applicability. Therefore, ACTD-DP algorithm has the best compression effect.

**Keywords** AIS, Ship trajectory, Trajectory compression, ACTD-DP algorithm

With the deepening of global economic integration, the maritime industry's role within the global economic system has become increasingly pivotal. The density of ship traffic in coastal ports, estuarial jetties, and other water areas continues to rise. The complex navigational environment presents novel challenges to maritime traffic supervision authorities and staff<sup>1,2</sup>. As an important carrier of ship motion information, AIS facilitates ship supervision and management. In accordance with the requirements of the SOLAS (International Convention for the Safety of Life at Sea), an increasing number of ships are mandated to be equipped with AIS devices to mitigate the risk of maritime collisions<sup>3</sup>.

The massive AIS data information provides significant support for research across various domains, including studies on ship behavior patterns<sup>4,5</sup>, maritime route planning<sup>6,7</sup>, and navigation safety<sup>8–10</sup>. However, the raw AIS data contains a plethora of redundant information, which can adversely affect the processing time of ship trajectories<sup>11</sup>. Therefore, studies on the characteristic direction of ship trajectory all need to compress AIS data, such as ship navigation prediction<sup>12</sup> and ship abnormal behavior detection<sup>13,14</sup>.

Compression algorithms commonly used for AIS data can be categorized into offline and online approaches. Offline compression algorithms include the Douglas–Peucker (DP) algorithm<sup>15</sup> and the TD-TR (top-down time ratio) algorithm<sup>16</sup> etc., while online compression algorithms encompass the Sliding Window (SW) algorithm<sup>17</sup> and the opening window time ratio (OPW-TR) algorithm<sup>18</sup> etc. Among these, the selection of the threshold value is most critical when utilizing algorithms such as DP and SW for the compression of AIS data. The literature<sup>19</sup>

<sup>1</sup>Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China. <sup>2</sup>Shandong Transport Vocational College, Weifang 261206, Shandong, China. <sup>3</sup>Shandong Vocation College of Science and Technology, Weifang 261053, Shandong, China. <sup>4</sup>These authors contributed equally: Ting Zhang, Zhiming Wang. ✉email: gfy5216@126.com

synthesized the ship's course deviation, position deviation and the spatiotemporal characteristics of AIS data to set the threshold value of SW algorithm. This approach facilitates the extraction of key feature points from the ship trajectory, thereby ship trajectory data is compressed. But the choice of distance and angle thresholds needs to be analyzed according to the experimental waters and the type of vessel, the distance threshold in paper was set as [0.731, 1.274] times the ship width, and the Angle threshold was set as [3.8°, 5.0°]. The literature<sup>20</sup> enhanced the traditional SW algorithm by combining speed values varied significantly. The threshold was dynamically adjusted to reduce the scale deviation caused by the speed difference. Consequently, it more effectively preserved the shape characteristics of the ship trajectory. However, this method is more suitable for trajectories with obvious velocity changes, and the improvement effect may not be as good as expected for trajectories with little velocity changes or simple motion modes. The literature<sup>21</sup> initialized and dynamically adjusted the threshold of DP algorithm to proportion value of the ship's length, based on the characteristics of ship trajectory. The literature statistically analyzed the relative azimuth differences between trajectory points to identify and order the features points, thereby yielded the final compressed trajectory. The literature uses the parameter self-selection method for the Silhouette Coefficient scores, but the determination of the optimal parameter combination will require extensive experiments and adjustments. Based on the traditional DP algorithm, the literature<sup>22</sup> used the minimum ship domain evaluation method to optimize the threshold setting method, and combined the ship parameters and maneuverability to set the threshold to 0.8 times the ship length. However, the selection of threshold value is based on the empirical method, and the generalization ability of the method is insufficient. Based on the traditional DP algorithm and SW algorithm, literature<sup>23</sup> combined with the space and motion characteristics of the trajectory, applied statistical theory to determine the algorithm threshold for compression. The paper recommended that the distance threshold value for DP algorithm was 0.8 times the ship length and the threshold coefficient of SW algorithm was 1.6. The threshold value in the literature is a fixed value and cannot provide an adaptive adjustment mechanism. The author<sup>24</sup> compressed ship trajectory using DP algorithm, and a novel metric known as ACS (Average Compression Score) was introduced as an evaluative criterion. But the choice of the most threshold is still based on experience, especially the determination of the optimal ACS value. The distance threshold of DP algorithm was set 0.87 nautical miles, the ACS reaches the optimal value of 0.1472, and the compression rate can reach 92.59%. The literature<sup>25</sup>, aiming to facilitate the rapid display of AIS trajectories on ECDIS (Electronic Chart Display and Information System), has utilized DP algorithm for compressing ship trajectory on nautical charts of varying scales. However, the threshold value is obtained based on the test of the research water area, and the generalization ability is weak. The results show that the recommended threshold range was [10, 20] meters and the compression ratio was [94%, 97%] for charts (scale 1:100,000 to 1:2990,000), while the recommended threshold was 20 meters and the compression ratio was 97% for charts (scale 1:3,000,000). The literature<sup>26</sup> introduced the ADPS (Adaptive Douglas-Peucker with Speed) algorithm, which combined trajectory characteristics, the rate of change of ship speed, and the distances between feature points to automatically calculate and set thresholds suitable for each trajectory. This approach ensured the retention of key features and pertinent information while the trajectory was compressed. However, the performance of the algorithm still depends on the initial setting of some parameters, such as the method of determining the baseline. The literature<sup>27</sup> used DP algorithm with distinct distance thresholds to compress the upbound and down-bound trajectories respectively. The DTW (Dynamic Time Warping) algorithm was used to evaluate the compression effect and solved the optimal compression threshold. However, the threshold determined by this method depends on the set of threshold values, so it is easy to fall into the local optimal value. The literature<sup>28</sup> proposed the MPDP (Multi-Objective Peak Douglas-Peucker) algorithm, which adopted a peak sampling strategy. The three optimization objectives of the trajectory, spatial characteristics, course, and speed, were considered. Additionally, the obstacle detection mechanism was added, aiming to achieve a compression algorithm more suited for curved trajectories. However, the threshold of this method is still a fixed value and lacks adaptability. The literature<sup>29</sup> proposed the PDP (Partition Douglas-Peucker) algorithm, which partitioned ship trajectory shapes and employs dynamic threshold settings to enhance the efficiency and accuracy of trajectory compression. This approach has successfully reduced the time required for compression and minimized data loss. However, the dynamic threshold determination within the PDP algorithm is still anchored to the ship's length, which limits its adaptiveness. The literature<sup>30</sup> proposed the ADP (Adaptive-threshold Douglas-Peucker) algorithm, which makes the threshold setting more flexible and accurate, and improves the computational efficiency of the algorithm by using matrix operations. However, when the cyclic ship trajectory was compressed by the ADP algorithm, the critical point information was lost more. At the same time, the complexity of ADP algorithm was higher to increase the calculation time and storage cost.

Summarizing the aforementioned research, the selection methods for threshold values in compressing ship trajectories based on DP and SW algorithms are primarily based on expert experience or trial-and-error approaches. The subjective of expert experience method is high, necessitating multiple iterations and experiments to ascertain the threshold values, resulting in low work efficiency. The thresholds set by the trial-and-error method are mostly based on the ship field or the ship static information, such as length and width, which are limited by the sailing waters and ship types. Furthermore, erroneous information within massive AIS datasets can also impact the setting of thresholds, thereby the compression effect can be diminished.

To overcome the deficiencies and limitations of DP algorithm in threshold setting, which is often based on expert experience or trial-and-error approaches, this paper proposes the ACTD-DP (Adaptive Threshold Difference-DP) algorithm. The ACTD-DP algorithm effectively reduces the reliance on experience or static ship information for threshold determination, demonstrating good applicability. Moreover, the ACTD-DP algorithm is suitable for compressing trajectory data of different types of water area and ship types, exhibiting strong robustness.

The remainder of this paper is organized as follows. In “Methods”, the traditional DP algorithm, relevant definitions and evaluation metrics are introduced, while the logic and flow of ACTD-DP algorithm are introduced.

In “[Experiment and analysis](#)”, the experimental results of different algorithms are compared and analyzed in the selected research area. In “[Discussion and conclusion](#)”, this paper is discussed and summarized, and the directions of future research are outlined.

## Methods

### Traditional DP algorithm

The DP algorithm, initially introduced by Douglas and Peucker in 1973<sup>31</sup>, is predicated on the concept of “straightening the curve” by approximating a curve with a series of points and reducing the number of points, which is commonly utilized for simplifying the motion trajectories of objects.

Assuming a curve is constituted by a set of points  $\{P_1, P_2, P_3, P_4, P_5, P_6\}$  as shown in Fig. 1a, with a given threshold. The compression process of the DP algorithm is proceeds as follows:

- (1) Connect the beginning and end of the curve  $P_1 - P_6$  as the baseline line (dashed line in Fig. 1a), calculate the vertical distance between the remaining points and the reference line, and obtain the maximum distance value  $d_{max1}$  and corresponding points  $P_4$ .
- (2) If  $d_{max1} < \epsilon$ , all the intermediate points are removed, and the curve point set is compressed as  $\{P_1, P_6\}$ ; otherwise, save  $P_4$  as the key point, and divide the curve into two sub-curve point sets:  $\{P_1, P_2, P_3, P_4\}$  and  $\{P_4, P_5, P_6\}$ , as shown in Fig. 1b. The reference lines are  $P_1 - P_4$  and  $P_4 - P_6$  respectively.
- (3) Repeat steps (1) and (2) for the above two sub-curves respectively, as shown in Fig. 1b, c. Finally, the compression curve is  $\{P_1, P_2, P_4, P_6\}$ , as shown in Fig. 1d.

### Relevant definition of trajectory compression

#### Projection

Ship trajectory is similar the curve shown in Fig. 1, thus by setting an appropriate threshold, the DP algorithm can detect and retain the critical points, discard non-critical points within the ship’s trajectory. During compressing the curve by DP algorithm, distances and thresholds are calculated based on the Cartesian coordinate system. However, ship trajectory point data are typically based on the geographic coordinate system, where the calculation of spherical distances is more complex, particularly when determining the distance between a trajectory point and a line.

Consequently, prior to the compression of ship trajectories by DP algorithm, a transformation from the geographic coordinate system to the Mercator projection coordinate system is required to facilitate the calculation of distance values, which is called projection.

$(\lambda, \varphi)$  denote the longitude and latitude value of the trajectory point in the geographical coordinate system, and  $(x, y)$  represents the coordinate value projected in the Mercator coordinate system. The conversion formula is as follows:

$$\begin{cases} R = \frac{\delta \cos \varphi_0}{\sqrt{1-e^2 \sin^2 \varphi_0}} \\ q = \ln\left(\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)\left(\frac{1-e \sin \varphi}{1+e \sin \varphi}\right)^{e/2}\right) \\ x = R\lambda \\ y = R\varphi \end{cases} \quad (1)$$

where,  $R$  represents the parallel circle radius of standard latitude;  $\delta$  represents the long radius of the earth’s ellipsoid;  $\varphi_0$  represents the standard latitude in the Mercator projection;  $e$  represents the first eccentricity of the earth ellipsoid;  $q$  represents isometric latitude.

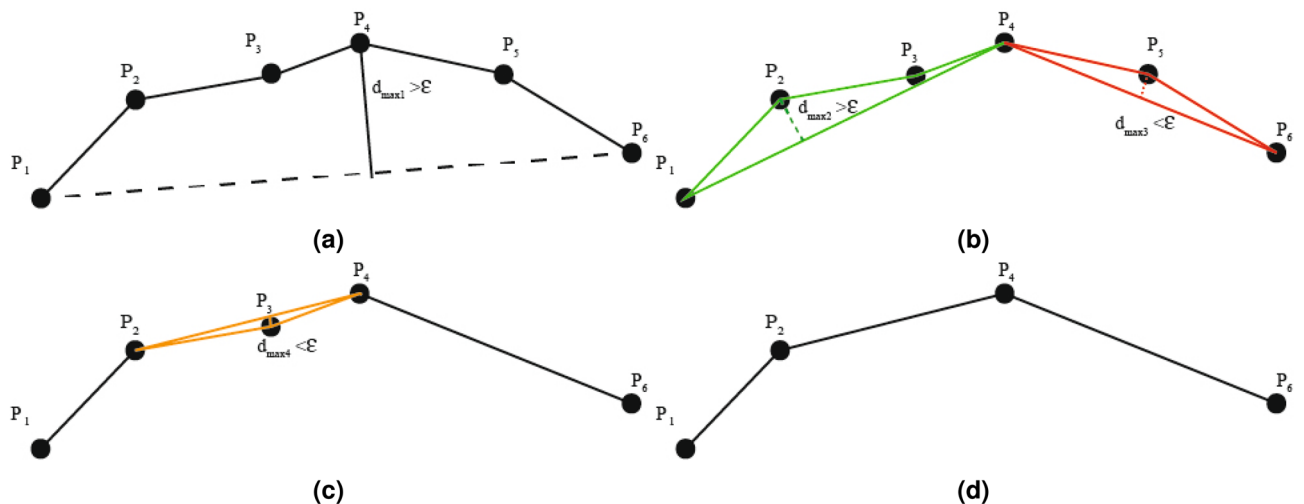


Figure 1. DP algorithm.

*Algorithm performance evaluation metrics*

As shown in Fig. 2, let  $Tra_{org} = \{P_1, P_2, \dots, P_n\}$  be the original ship trajectory,  $Tra_{cmp} = \{P_1, P_n\}$ , and  $n$  be the compressed trajectory and  $n$  be the number of trajectory points. Each trajectory point encompasses two fundamental attributes: coordinate values and a timestamp, that is,  $P_i = \{x_i, y_i, t_i\}$ .

**Definition 1** SED (synchronous euclidean distance). SED is generally used to evaluate the effectiveness of compression. SED denotes the Euclidian distance between the point  $P_1$  in the original trajectory  $Tra_{org}$  and the corresponding position  $P_{syn}$  in the compressed trajectory  $Tra_{cmp}$ , calculated in proportion to time.  $P_{ped}$  is the foot of  $P_i$  on the compression trajectory,  $P_i P_{ped}$  length is the vertical Euclidean distance corresponding to  $P_i$ , and  $P_i P_{syn}$  length is the corresponding to  $P_i$ .  $P_{syn}(x_{syn}, y_{syn})$  coordinates of trajectory points and corresponding SED formulas are as follows:

$$\begin{cases} x_{syn} = x_1 + \left(\frac{t_i - t_1}{t_n - t_1}\right)(x_n - x_1) & , t_m > t_i > t_1 \\ y_{syn} = y_1 + \left(\frac{t_i - t_1}{t_n - t_1}\right)(y_n - y_1) & , t_m > t_i > t_1 \\ SED_i = \sqrt{(x_i - x_{syn})^2 + (y_i - y_{syn})^2} \end{cases} \quad (2)$$

The mean SED of a single trajectory is denoted as MSED and the mean SED of all trajectories is expressed as AMSED. The formula is as follows:

$$\begin{cases} MSED = \frac{1}{n} \sum_{i=1}^n SED_i \\ AMSED = \frac{1}{N} \sum_{j=1}^N MSED_j \end{cases} \quad (3)$$

where,  $N$  represents the number of all trajectories.

MSED represents the location discrepancy between the original trajectory and the compressed trajectory, and the smaller the value, the smaller the discrepancy, the smaller the trajectory distortion, and the higher the integrity of the trajectory shape.

**Definition 2** CR (compression ratio). CR is the ratio of the number of points discarded during the compression process to the total number of points in the original trajectory. When only the CR is considered, the higher the CR, the better the compression effect of the compression algorithm. The formula is as follows:

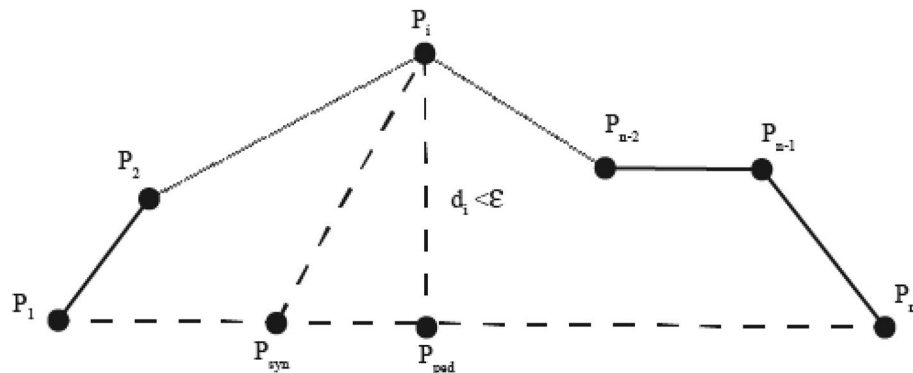
$$CR = 1 - a \quad (4)$$

ACR (average compression ratio) denotes the average compression ratio of all trajectories. The formula is as follows:

$$ACR = \frac{1}{N} \sum_{j=1}^N CR_j \quad (5)$$

The CR indicates the compression degree of the algorithm on the trajectory. The larger the value, the higher the compression degree of the trajectory, and the simpler the compression trajectory obtained.

**Define 3** LLR (length loss ratio). LLR is the ratio of the reduced length of the compressed trajectory to the length of the original track. The formula is as follows:



**Figure 2.** Trajectory compression diagram.

$$\begin{cases} Len = \sum_{i=1}^n \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \\ LLR = \frac{Len_{org} - Len_{cmp}}{Len_{org}} \end{cases} \quad (6)$$

where,  $Len_{org}$  represents the length of the original trajectory;  $Len_{cmp}$  represents the length of the compressed trajectory.

ALLR (average length loss ratio) denotes to the mean LLR of all trajectories. The formula is as follows:

$$ALLR = \frac{1}{N} \sum_{j=1}^N LLR_j \quad (7)$$

The LLR denotes the degree of loss on the track length. The larger the value, the more length information is lost during the compression process, and the greater the probability of deviation of some track features.

**Define 4 TIME** TIME is the total time taken by the algorithm to compress all trajectories in the current research area.

### ACTD-DP algorithm

The purpose of trajectory compression is to reduce the number of trajectory points while maintaining the integrity of the trajectory's shape, thereby enhancing the velocity of trajectory display and processing. The goal is to achieve an optimal balance between the quantity and the quality of the trajectory points. The compression quality of DP algorithm is predominantly determined by the threshold value. The larger the threshold value, the higher the compression rate and the larger the trajectory distortion; the smaller the threshold value, the lower the compression rate and the smaller the trajectory distortion.

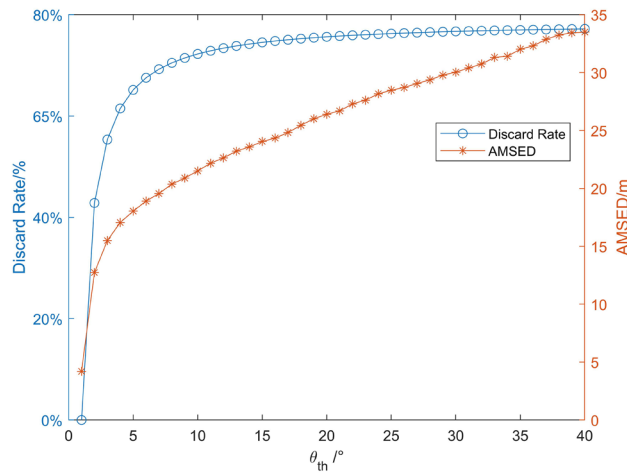
Traditional DP algorithms and the aforementioned studies in references<sup>22–25</sup> set threshold values based on static information such as ship length, ship width, and fixed distance values, which diminishes the algorithm's adaptive capacity. The ACTD-DP algorithm is proposed, as depicted in Fig. 6, drawing on the conceptual frameworks of the PDP algorithm<sup>29</sup> and the APD algorithm<sup>30</sup>. The ACTD-DP algorithm employs an adaptive threshold difference approach, aiming to diminish reliance on static information. Initially, the original trajectory data is preprocessed using the course attribute from AIS data, the certain trajectory points are discarded. Subsequently, a curve fitting method is used to establish a functional relationship between the threshold and the number of trajectory points. The characteristics of the function curve are then analyzed to determine the core threshold and the core threshold difference. Finally, the compression factor is introduced to ascertain the optimal threshold difference, which serves as the key parameter to control the accuracy and efficiency of the algorithm. In comparison with the PDP and ADP algorithms, the ACTD-DP algorithm is capable of achieving a higher compression rate while maintaining the integrity of the trajectory shape. Additionally, the ACTD-DP algorithm demonstrates adaptability across various maritime environments and ship types.

#### Preprocessing trajectory

Compared with the traditional DP algorithm, ACTD-DP algorithm needs to solve the optimal core threshold difference, which may increase the algorithm's execution time. To reduce the time consuming of the ACTD-DP algorithm, this paper processes the data and eliminate the noise data, drift data and other outliers in the data. After that, following the approach in reference<sup>17</sup>, this paper preprocesses the data based on the course differences between original trajectory points to reduce the number of trajectory points. The trajectory preprocessing method is as follows:

- Let  $Tra_{org}$  be the original trajectory (Fig. 2) and  $\theta_{th}$  be the course change threshold. Then  $\{P_1, P_n\}$  is saved in the preprocessing trajectory.
- Iteratively calculate the course difference between adjacent trajectory points as  $\Delta\theta$ . If  $\Delta\theta > \theta_{th}$ , the point is saved to middle  $Tra_{pre}$ ; Otherwise, the point is discarded.
- To ensure data integrity, if the number of points discarded between adjacent two points in  $Tra_{pre}$  is higher than that of  $\Delta n$  (the general value is [10, 30]), trajectory points are selected from  $Tra_{org}$  in equal proportion with equal proportional interpolation and saved to  $Tra_{pre}$ .

The DR (discard ratio) is the ratio of the number of points discarded during the preprocessing of the trajectory to the total number of points in the original trajectory. Based on the experimental AIS data (as described in "Description of experimental data"), the relationship between DR and AMSED values with respect to threshold  $\theta_{th}$  is depicted in Fig. 3. When threshold  $\theta_{th} < 10^\circ$  is set, both DR and AMSED exhibit more pronounced changes. At threshold  $\theta_{th} = 1^\circ$ , DR is 42.89% and AMSED is 8.5372 m. When threshold  $\theta_{th} = 10^\circ$  is used, DR increases to 72.30% and AMSED to 17.30 m. At threshold  $\theta_{th} > 10^\circ$ , the variation in DR is relatively minor. Compared to the data in Table 5, the AMSED values presented in Fig. 3 are consistently lower, indicating minimal distortion of the trajectory and a minor impact on the trajectory's shape features due to the discarded points. Consequently, the threshold  $\theta_{th} = 10^\circ$  and the number of points  $\Delta n = 10$  are selected for further analysis.



**Figure 3.** DR and AMSED values with respect to threshold  $\theta_{th}$ .

*Fitting thresholds-points function*

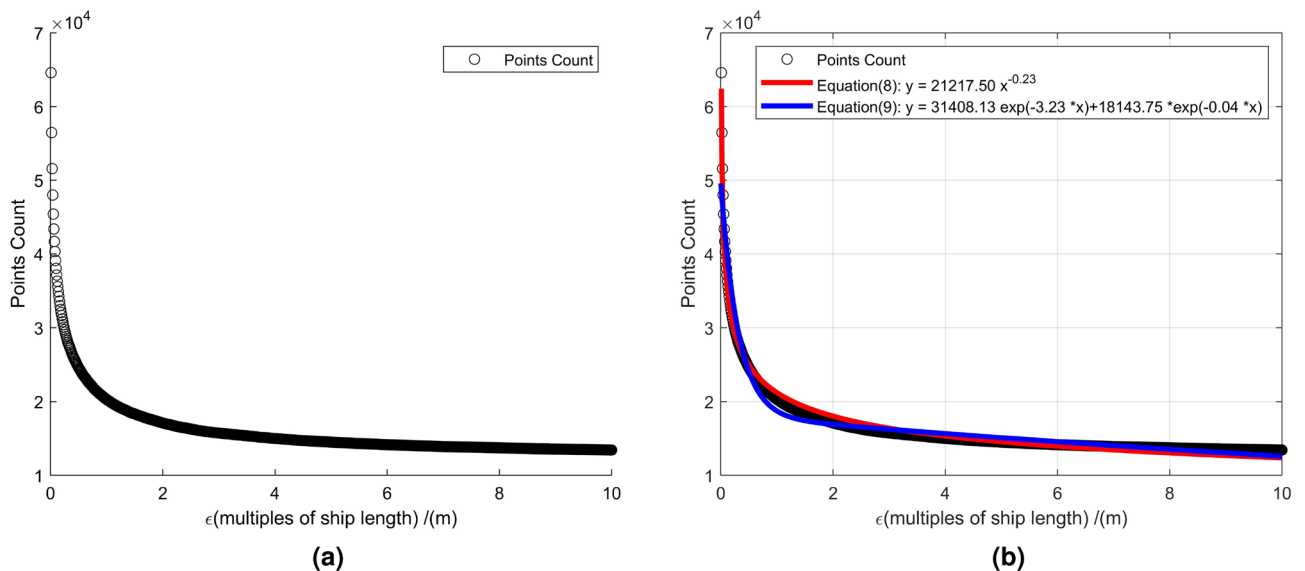
The threshold of trajectory compression algorithms based on DP algorithm is directly correlated with the number of points in the compressed trajectory, which subsequently affects the quality of the compressed trajectory. To quantify this relationship between the threshold values and the number of trajectory points for the ACTD-DP algorithm, this paper, based on the experimental AIS data (as detailed in “Description of experimental data”), presents a statistical analysis of the total number of compression trajectory points across a threshold range of [0.01, 10] times the ship’s length, as depicted in Fig. 4a. Statistical analysis demonstrates a nonlinear negative correlation between threshold values and the count of trajectory points. Initially, when the threshold is smaller, there is a significant and rapid decrease in the number of trajectory points. Subsequently, with an increase in the threshold, the rate of reduction in trajectory points attenuates, and in certain regions, there may be observed slight oscillations or a tendency towards rebound.

According to the curve (Fig. 4a) analysis, the functional relationship between the thresholds and the number of trajectory points may conform to the characteristics of power function and exponential function. Therefore, the fitting function equations can be assumed to be:

$$y = \omega_1 x^{\omega_2} \tag{8}$$

$$y = \omega_3 * \exp(\omega_4 * x) + \omega_5 * \exp(\omega_6 * x) \tag{9}$$

where,  $\omega_i (i = 1, 2 \dots 6)$  represents the equation parameters.



**Figure 4.** Thresholds-points function fitting. (a) Thresholds-points statistics and (b) the fitting function curves.

Then, The equations are solved according to the statistical data, and the parameter values are obtained as shown in Table 1. The fitting function curves were shown in Fig. 4b.

Finally, the determination coefficient is selected as the Goodness of Fit for the two functions. The determination coefficient, as  $R^2$ , is the description of the degree of variation of the function independent variable and dependent variable, and it is an essential metric for assessing the fit of a regression equation. The formula for its calculation is as follows:

$$\begin{cases} S_r = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ S_t = \sum_{i=1}^n (y_i - \bar{y})^2 \\ R^2 = 1 - \frac{S_r}{S_t} \end{cases} \quad (10)$$

where,  $S_r$  represents residual sum of squares;  $S_t$  represents total sum of squares;  $y_i$  represents the actual value;  $\hat{y}_i$  represents the fitting curve value;  $\bar{y}$  represents the average value.

The determination coefficients for Eq. (8),  $R^2 = 0.9872$ , and for Eq. (9),  $R^2 = 0.9589$ , indicate that the values derived from Eq. (8) are closely aligned with the actual data, signifying a higher degree of fit for the curve equation and greater applicability. Consequently, the power function, the thresholds-points function, is selected to represent the functional relationship between the threshold and the number of trajectory points.

#### Optimal core threshold difference calculating

After the thresholds-points function has been established, the further analysis of the function curve characteristics is necessitated. This is essential for evaluating the compression effectiveness during the trajectory compression process. It is imperative to establish an intrinsic link between the curve and the ACTD-DP algorithm. Additionally, the role of thresholds or differences therein within the algorithm must be explicitly defined.

When the ACTD-DP algorithm compresses the trajectory, the maximum distance of all trajectory points to the baseline (similar to that depicted in Fig. 1) within each compressed trajectory segment is calculated. This maximum distance value is hypothesized as the core threshold for that trajectory segment, denoted as  $\epsilon_{core}$ . The difference in core thresholds obtained from two consecutive compressions is denoted as the core threshold difference. The formula is as follows:

$$\Delta\epsilon_k = \begin{cases} |\epsilon_{core}^k - \epsilon_{core}^{k+1}|, & \text{other} \\ \infty, & k = 1 \end{cases} \quad (11)$$

where,  $\Delta\epsilon_k$  represents the core threshold difference at the second compression;  $\epsilon_{core}^k$  represents the core threshold obtained during the second compression;  $k$  represents the compression order.

During the trajectory compression process, the overall trend of the core threshold difference is a gradual decrease (as shown in Fig. 5a). When  $\Delta\epsilon_k < \Delta\epsilon_o \leq \Delta\epsilon_{k-i}$  ( $i = 1, 2, \dots, k-1$ ) is true, it indicates that the change trend of core threshold difference tends to be stable, the trajectory shape changes little, and the trajectory compression ratio also tends to be stable. At this point,  $\Delta\epsilon_o$  is identified as the optimal core threshold difference, achieving the best balance between the quantity and the quality of the trajectory points, signifying the completion of trajectory compression. Therefore, the process of balancing between the quantity and the quality of the trajectory points is the process of searching the optimal core threshold difference.

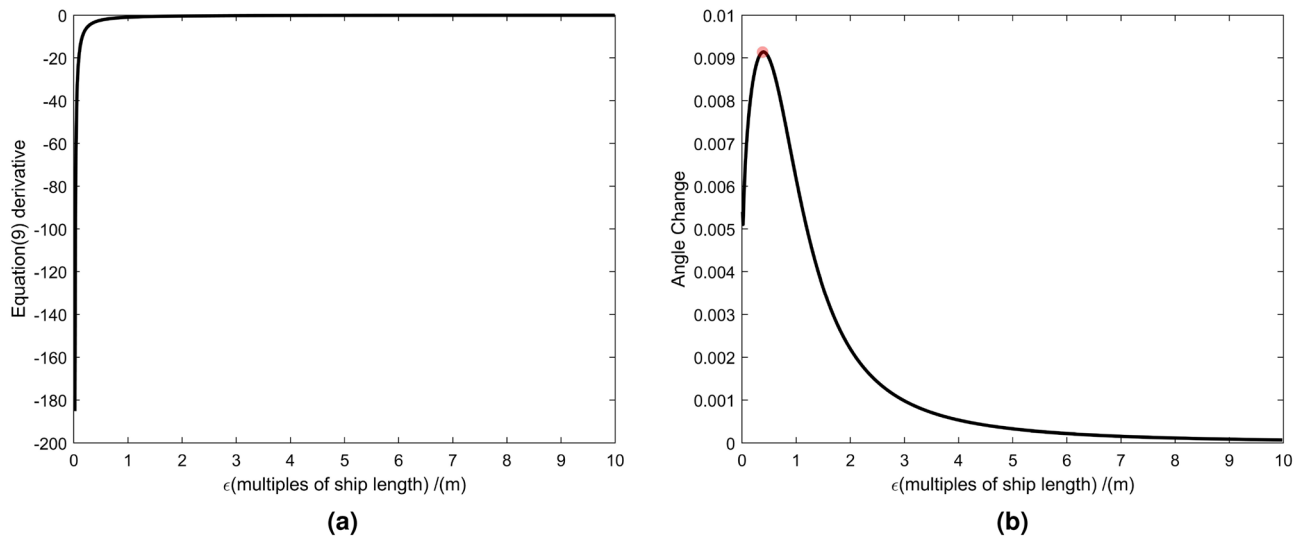
Combined with the curve (Fig. 4a) analysis of Equation (8), the core threshold difference corresponds to the derivative of the fitted curve (Fig. 5a). Consequently, the optimal core threshold difference can be calculated based on the angle difference between two points on the fitting curve (as illustrated in Fig. 5b), where the core threshold difference at the position of the maximum angle difference of the fitting curve is identified as the optimal core threshold difference. The formulas for calculating the derivative of the fitting curve and the angle difference are as follows:

$$\begin{cases} y' = \omega_1 * \omega_2 * x^{\omega_2-1} \\ \Delta r_k = \left| \tan^{-1} y'_k - \tan^{-1} y'_{k+1} \right| \end{cases} \quad (12)$$

The analysis (Fig. 5b) shows that when  $x = 0.4$ , the angle difference reaches the maximum, and the corresponding optimal core threshold difference is  $\Delta\epsilon_o = 3.097$ .

Parameter names	Parameter values	95% confidence
$\omega_1$	2.122e+04	(2.116e+04, 2.128e+04)
$\omega_2$	-0.2344	(-0.236, -0.2327)
$\omega_3$	3.141e+04	(3.07e+04, 3.211e+04)
$\omega_4$	-3.225	(-3.343, -3.107)
$\omega_5$	1.814e+04	(1.797e+04, 1.832e+04)
$\omega_6$	-0.03668	(-0.03839, -0.03496)

**Table 1.** Function equation parameter values.



**Figure 5.** The fitted curve analysis diagram. (a) The derivative of the fitted curve and (b) the angle difference of the fitted curve.

*Compression factor*

Different purposes of trajectory research require different compression effects. When studying the traffic flow state of water area, the higher compression ratio is preferred; whereas when examining the ship navigation state, the higher quality of compression is more desirable. To meet the diverse research objectives, this paper introduces a compression factor, denoted as  $\rho$ , based on the optimal core threshold difference, with a default value of 0.5.

$$\Delta \varepsilon_o = 3.097 * \left( \frac{\rho}{1 - \rho} \right), (0 < \rho < 1) \tag{13}$$

In summary, after the optimal core threshold difference is obtained, the trajectory compression can be carried out. The ACTD-DP Algorithm flow is shown in Fig. 6, and the algorithm code is shown in Algorithm 1.

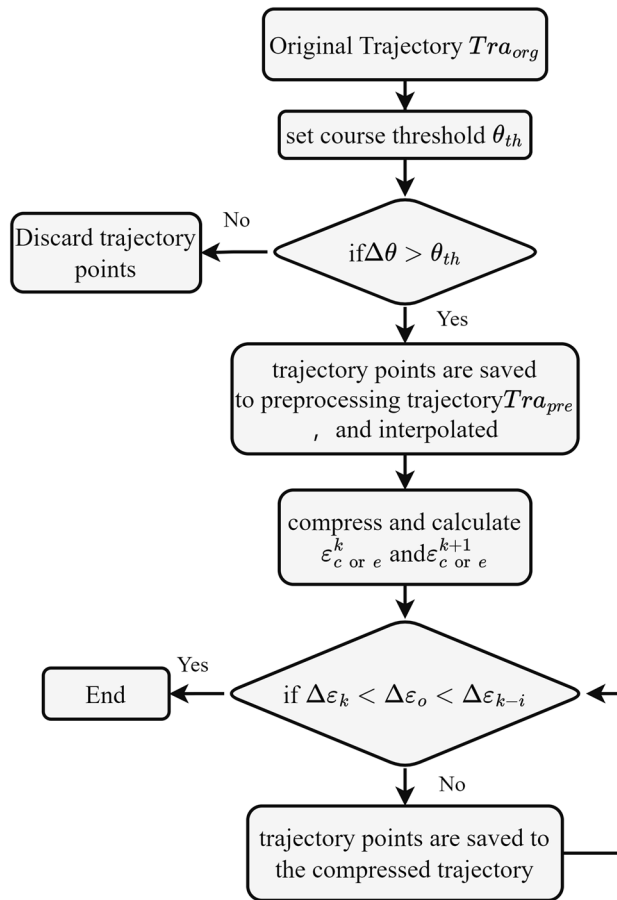


Figure 6. Algorithm flow.

---

```

1: Initialize:
2:  $\Delta n \leftarrow 20$ 
3:  $\Delta \theta \leftarrow 10$ 
4:  $Tra_{org} \leftarrow data_{org}.AIS$ 
5:  $Tra_{pre} \leftarrow Calculate\_Tra_{pre}(Tra_{org}, \Delta n, \Delta \theta)$ 
6:  $\Delta \epsilon \leftarrow 3.0970 \times \left( \frac{\rho}{1-\rho} \right)$ 
7:  $compressed\_track \leftarrow Adaptive\_Threshold\_Difference\_DP(Tra_{pre}, \Delta \epsilon, \infty)$ 
8: function  $Calculate\_Tra_{pre}(Tra_{org}, \Delta n, \Delta \theta)$ 
9:   if  $Tra_{org}$  has  $\leq 500$  rows then
10:     return  $Tra_{org}$ 
11:   end if
12:    $current\_Tra \leftarrow Tra_{org}(1, :)$ 
13:    $s \leftarrow 1$ 
14:    $e \leftarrow 2$ 
15:   while  $e \leq$  the number of rows in  $Tra_{org}$  do
16:      $course\_diff \leftarrow abs((Tra_{org}(s, 1) - Tra_{org}(e, 1) + 180) \bmod 360 - 180)$ 
17:     if  $course\_diff > \Delta \theta$  then
18:       if  $e - s > \Delta n$  then
19:          $insert\_counts \leftarrow \lfloor \frac{e-s}{\Delta n} \rfloor$ 
20:         for  $i \leftarrow 1$  to  $insert\_counts$  do
21:            $current\_Tra \leftarrow [current\_Tra; Tra_{org}(s + i \times \Delta n, :)]$ 
22:         end for
23:       end if
24:        $current\_Tra \leftarrow [current\_Tra; Tra_{org}(e, :)]$ 
25:        $s \leftarrow e$ 
26:     end if
27:      $e \leftarrow e + 1$ 
28:   end while
29:    $columnData \leftarrow$  the last element of each row in  $current\_Tra$ 
30:    $uniqueColumnData \leftarrow$  the set of unique values in  $columnData$ 
31:    $Tra \leftarrow Tra_{org}$  with second column values in  $uniqueColumnData$ 
32:   return  $Tra$ 
33: end function

```

▷ Parameters initialization  
 ▷ Turn point interval threshold  
 ▷ Turn angle threshold  
 ▷ Original trajectory data  
 ▷ Calculate optimal threshold difference for compression  
 ▷ Compress trajectory using adaptivethreshold-difference DP algorithm  
 ▷ Preprocessing trajectory  
 ▷ If the number of trajectory points is too small, preprocessing is not performed  
 ▷ Index of the current trajectory point  
 ▷ Index of the next trajectory point  
 ▷ The specific methods described in Section 2.3.1  
 ▷ Discard duplicate values

---

**Algorithm 1.** ATD-DP (Part 1).

---

```

34: function Adaptive_Threshold_Difference_DP(Tra,  $\epsilon$ , prev_maxdis_value_Tra)
35:   Input:
36:     Tra ▷ The current trajectory
37:      $\epsilon$  ▷ The current optimal core threshold difference
38:     prev_maxdis_value_Tra ▷ The core threshold difference of the last recursion
39:   if Tra has  $\leq 2$  rows then ▷ If the number of trajectory points is too small, preprocessing is not performed
40:     return Tra
41:   end if
42:   maxdis_id_pre  $\leftarrow$  Calculate_distance_matrix(Tra, Tra(1, :), Tra(end, :))
43:   maxdis_id_pos  $\leftarrow$  the index of maxdis_id_pre in Tra(:, 2)
44:   if |maxdis_value_TS - prev_maxdis_value_Tra|  $\geq \epsilon$  then ▷ Continue to segment the trajectory
45:     segment1_TS  $\leftarrow$  Tra(1 : maxdis_id_pos, :)
46:     segment2_TS  $\leftarrow$  Tra(maxdis_id_pos : end, :)
47:     compress1  $\leftarrow$  dp_auto_threshold(segment1_TS,  $\epsilon$ , prev_maxdis_value_Tra)
48:     compress2  $\leftarrow$  dp_auto_threshold(segment2_TS,  $\epsilon$ , maxdis_value_TS)
49:     return [compress1(1 : end - 1, :); compress2]
50:   else
51:     return [Tra(1, :); Tra(end, :)] ▷ Add begin and end points
52:   end if
53: end function

```

---

**Algorithm 2.** ATD-DP (Part 2).

## Experiment and analysis

### Description of experimental data

In order to verify the scientificity of the proposed algorithm, the AIS data of Zhoushan waters in China is selected as the experimental water area shown in Table 2, and the time range is from 00:00 to 24:00 on May 1st, 2021. According to the ship MMSI statistical analysis, the experimental waters shared 515 ships and 404,646 AIS data. The statistics of ship types are shown in Table 3, and the statistics of ship dimensions are shown in Table 4.

The experimental water area and the original ship trajectories are depicted in Fig. 7a. The raw data contains a significant amount of noise and redundant information. Preprocessing is conducted according to the method described in Section 2.3.1, and the results are presented in Fig. 7b. The experiment was implemented on the computer (Window64 Intel(R) Xeon(R) Gold 5218R CPU @ 2.10 GHz 2.10 GHz and 32.0 GB of RAM) using MATLAB(version R2024a<sup>32</sup>).

### Compression results of all trajectories

To compare the compression effects of different algorithms, the SW algorithm from the reference<sup>19</sup> (with the distance threshold of 0.8 times the ship width), the DP algorithm from the reference<sup>22</sup> (with a distance threshold of 0.8 times the ship length), and the PDP (Peak–Douglas–Peucker) algorithm from the reference<sup>31</sup> (with the distance threshold of 0.5 times the ship length and an angular threshold of 10°) and the ADP algorithm from

Boundary points	Longitude	Latitude
Upper left	121° 49' 18" E	30° 14' 50" N
Lower right	122° 16' 57" E	29° 55' 45" N

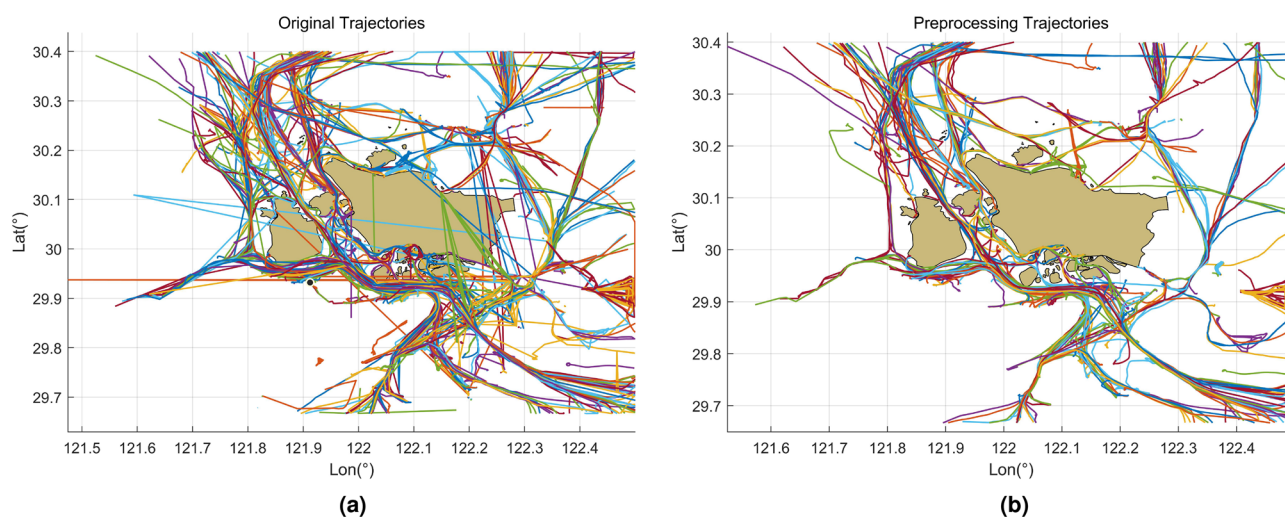
**Table 2.** The experimental water scope.

Type	Number	Proportion
Fishing vessel	122	23.69 %
Dry cargo ship	88	17.09%
Bulk carrier	46	8.93%
Oil tanker, chemical tanker	39	7.57%
Container ship	20	3.88%
Others	200	38.83%

**Table 3.** The statistics of ship type.

Length	Number	Proportion (%)	Width	Number	Proportion (%)
<50	103	20.00	<10	242	46.99
50–100	166	32.23	10–20	107	20.78
100–200	161	31.26	20–30	22	4.27
200–300	55	10.68	30–40	24	4.66
>300	24	4.66	>40	17	3.30
Error data	6	1.17	Error data	103	20.00

**Table 4.** The statistics of ship dimension.



**Figure 7.** The experimental water area and AIS data (The color curves are ship trajectories. The yellow polygon area is the land part of the experimental area. The white background indicates the experimental water area. The figure is drawn using MATLAB, which version is R2024a<sup>32</sup>). (a) The original trajectories and (b) the preprocessing trajectories.

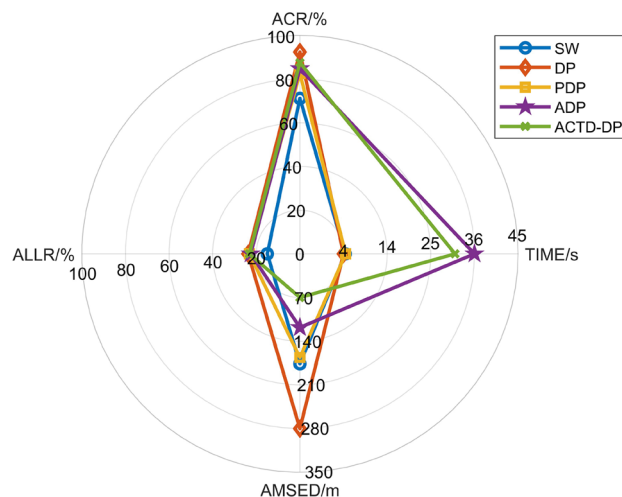
Algorithm	ACR (%)	ALLR (%)	AMSED (m)	TIME (s)
SW	71.17	14.95	173.5568	3.5389
DP	92.36	23.75	275.6075	3.2139
PDP	82.54	21.61	163.2476	3.6378
ADP	84.77	22.57	116.1772	29.5981
ACTD-DP	87.53	23.20	68.9747	35.7439

**Table 5.** The statistic results of all compressed trajectories. *ACR* average compression ratio, *ALLR* average length loss ratio, *AMSED* mean synchronous Euclidean distance of all trajectories.

the reference<sup>30</sup> (with the optimal threshold change rate of 1.36) are selected as the comparative algorithms. All trajectories are compressed by different algorithms, and the evaluation metrics values obtained are shown in Table 5 and Fig. 8.

By analyzing Table 5 and Fig. 8, it is obvious that the evaluation metric values corresponding to the five types of algorithms are different. The TIME value of the SW, DP, and PDP algorithms are obviously smaller than that of the ADP and ACTD-DP algorithms, indicating that the former trio exhibits lower complexity and superior computational efficiency, while the latter two algorithms, due to their trajectory division requirements, exhibit higher complexity and thus comparatively reduced computational efficiency. ADP algorithm has the lowest computational efficiency.

The SW algorithm exhibits the lowest values in terms of ACR and ALLR, yet it presents a notably high AMSED. This observation indicates that the SW algorithm, during the trajectory compression process, retains a relatively large number of non-critical points, leading to minimal discrepancies in length between the compressed and original trajectories. However, the location discrepancy is significantly pronounced, thereby yielding the least effective trajectory compression outcome among the evaluated methods.



**Figure 8.** Evaluation metric values comparison chart.

The DP algorithm is characterized by the highest values in ACR, ALLR, and AMSED. These metrics suggest that the DP algorithm achieves a significant reduction in trajectory data, indicative of its pronounced compression efficacy. However, the substantial elimination of critical points results in considerable discrepancies in both length and location discrepancy between the compressed and original trajectories. Consequently, the DP algorithm's trajectory compression performance is deemed to be of moderate quality.

The PDP algorithm exhibits a high ACR and the lowest ALLR among the evaluated methods, while its AMSED is relatively elevated yet notably lower than that of the DP algorithm. These observations indicate that the PDP algorithm exhibits a marked optimization over the traditional DP algorithm, yielding a more favorable trajectory compression outcome.

The ADP algorithm is distinguished by its higher ACR and ALLR, both of which surpass the corresponding values of the PDP algorithm. Additionally, the ADP algorithm exhibits a lower AMSED, signifying a reduced location discrepancy between the compressed and original trajectories. Collectively, these metric values underscore the ADP algorithm's notable effectiveness in trajectory compression.

The ACTD-DP algorithm is notable for its minimal AMSED, which is significantly lower than the values observed in other algorithms. This algorithm also presents a higher ACR and a lower ALLR. These metric values indicate that the ACTD-DP algorithm excels in retaining critical points during the trajectory compression process, leading to a relatively higher length discrepancy but a markedly reduced location discrepancy when compared to the original trajectories. Consequently, the ACTD-DP algorithm is recognized for its superior compression performance.

### Compression results of different types trajectories

To further demonstrate the robustness and applicability of the ACTD-DP algorithm, four ship trajectories were randomly selected for compression. The information for the chosen ships and their trajectories is presented in Table 6.

Different algorithms compress the above trajectories respectively, and statistical evaluation metric values are shown in Table 7. The TIME values corresponding to the compression of different ship trajectories by various algorithms align with the trends observed in Section 3.2, demonstrating that the TIME values for the ADP and ACTD-DP algorithms are significantly higher than those for the other algorithms.

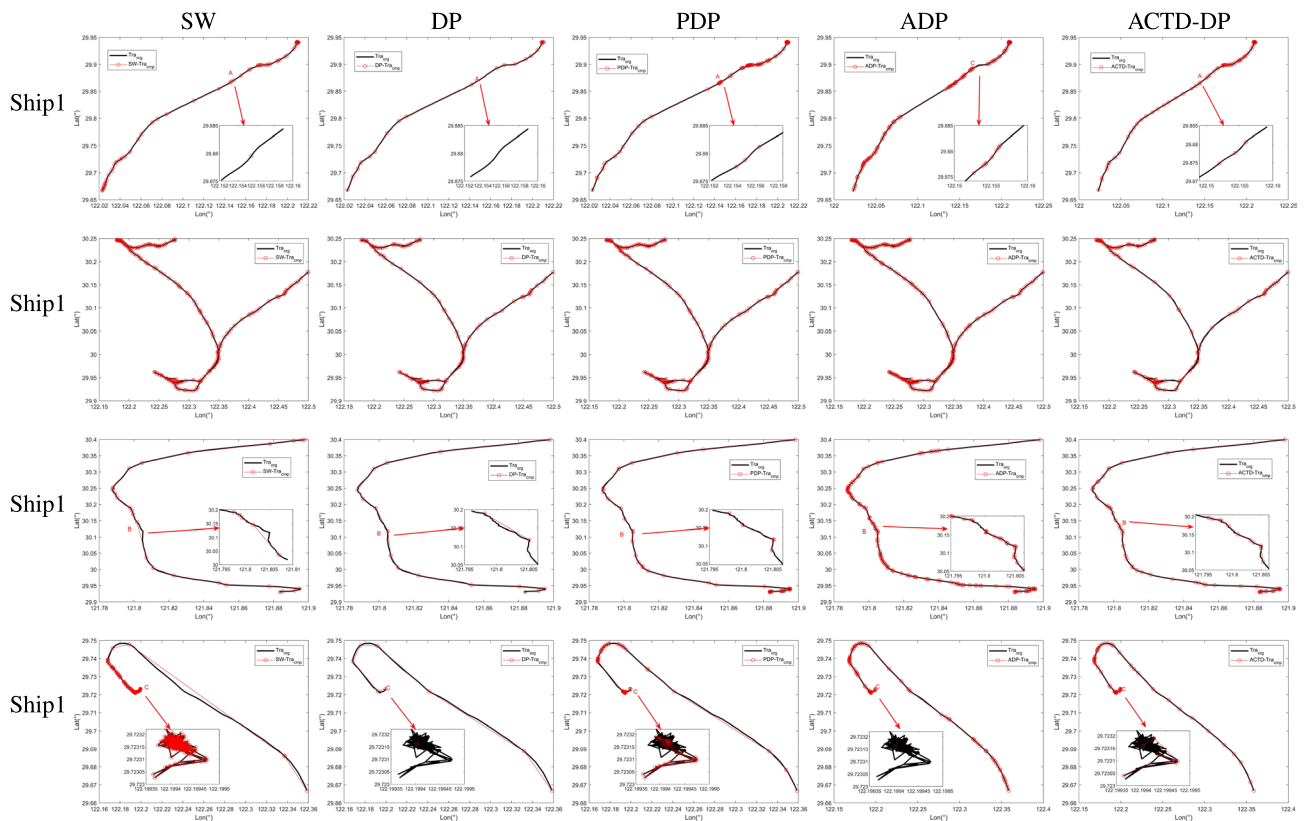
Ship1's course changes are small, the trajectory shape is simple, and the voyage is almost straight. MSED value of SW algorithm is the largest (62.7271 m, approximately 1.1 times the ship length), and CR is suboptimal. The SW algorithm demonstrates the weaker capability in detecting and retaining critical points, particularly in areas where the trajectory's curvature is low (region A in the Fig. 9, where only one point is retained), leading to significant distortion in the trajectory. The DP algorithm exhibits the highest values for both CR and LLR, with

No.	MMSI	Type	Length	Width	Numbe
Ship1	413404840	Oil tanker	55	9	967
Ship2	412424446	Fishing vessel	36	6	562
Ship3	412053000	Bulk carrier	149	22	717
Ship4	414750000	Bulk carrier	225	33	1415

**Table 6.** Ship and trajectory information.

No.	Metrics	SW	DP	PDP	ADP	ACTD-DP
Ship1	CR (%)	92.76	97.62	84.18	84.07	91.42
	LLR (%)	31.05	35.78	29.42	29.85	32.46
	MSED (m)	62.7271	60.1836	50.2737	48.3278	42.907
	TIME (s)	0.0179	0.0105	0.0149	0.2844	0.0825
Ship2	CR (%)	61.75	77.58	72.06	54.63	74.38
	LLR (%)	5.87	6.22	5.05	4.44	7.59
	MSED (m)	66.0019	155.6476	127.1705	77.0435	71.0208
	TIME (s)	0.0086	0.0101	0.0159	0.3625	0.1168
Ship3	CR (%)	96.23	97.77	86.47	77.41	90.24
	LLR (%)	3.27	3.32	2.94	1.77	2.88
	MSED (m)	105.8984	199.1665	33.4355	12.3542	20.507
	TIME (s)	0.0081	0.0095	0.0165	0.2735	0.0625
Ship4	CR (%)	22.05	99.43	88.41	92.86	93.29
	LLR (%)	47.45	54.93	48.09	47.81	49.06
	MSED (m)	63.6979	373.5619	43.4620	45.3802	24.9138
	TIME (s)	0.0139	0.0135	0.0166	0.1712	0.0799

**Table 7.** The statistic results of each trajectory. *CR* compression ratio, *LLR* length loss ratio, *MSED* mean synchronous Euclidean distance.



**Figure 9.** Trajectory compression comparison.

the suboptimal MSED value. The DP algorithm performs the poorest in detecting and retaining critical points (region A in the Fig. 9, where no critical points are retained), leading to the least effective compression effect. The PDP algorithm demonstrates the lowest values for LLR, with the MSED and CR being both smaller. The PDP algorithm demonstrates the stronger capability in detecting and retaining critical points, particularly handling the junctions between straight lines and curves more effectively than the DP algorithm (region A in the Fig. 9, where three points are retained). The ADP algorithm exhibits the lowest value for CR, and LLR and MSED value are both smaller. ADP algorithm has the strongest ability to retain critical points, resulting in an excessive number of trajectory points at the junctions between straight lines and curves (region A in the Fig. 9, where ten points are retained). The ACTD-DP algorithm exhibits the lowest MSED value (42.907 m, approximately 0.8 times the ship

length), CR and LLR are better. The ACTD-DP algorithm exhibits the strongest capability in detecting and retaining critical points, handling the transitions between straight lines and curves effectively (region A in the Fig. 9, where three points are retained too), resulting in minimal trajectory distortion and the best compression effect.

Ship2 is the fishing vessel with the lowest number of track points. However, the course changes are large and frequent, there are large angle turning and U-turn phenomena, and the navigation trajectory is the most complicated. Compared with the other three trajectories, the five algorithms yield lower CR and LLR values for this trajectory (with the best CR being 77.58% from the DP algorithm and the best LLR being 4.44% from the ADP algorithm), and the performance of MSED values is poor. The SW algorithm exhibits the lower MSED value (66.0019 m, approximately 1.8 times the ship length), while the DP algorithm exhibits the highest MSED value (155.6476 m, about 4.3 times the ship length). CR and MSED of the ACTD-DP algorithm are suboptimal. Consequently, the ACTD-DP algorithm maintains a good compression effect with the high CR and the low MSED. However, from the evaluation metrics analysis of the four trajectories compressed by ACTD-DP algorithm, the compression effect of this trajectory is the worst.

Ship3's trajectory is relatively complex, with large course changes and sharp turns. However, due to the large proportion of straight line segments in the trajectory, the compression effect of each algorithm is similar to that of ship1. The CR Value of the SW algorithm is the highest, while its LLR and MSED values are the second highest among the evaluated methods. But the compression effect is poor for the location where the curvature of the trajectory curve changes frequently (region B in the Fig. 9). The DP algorithm demonstrates the highest values for CR, LLR and MSED (199.1665 m, approximately 1.3 times the ship length). This indicates a weaker capability in detecting critical points, resulting in the fewest number of critical points retained (region B in the Fig. 9). Consequently, the DP algorithm is associated with the poorest compression performance among the evaluated methods. The PDP algorithm exhibits the second lower CR, indicative of a relatively higher retention of trajectory points. Its LLR and MSED values are moderate, reflecting a balanced performance in terms of trajectory fidelity and compression efficiency. The PDP algorithm effectively captures critical points in the trajectory with significant changes in course, yet the fixed threshold discards many critical points (region B in the Fig. 9). Despite this, the overall compression performance of the PDP algorithm is commendably effective. The ADP algorithm exhibits the lowest values for CR, LLR and MSED (12.3542 meters, approximately 0.08 times the ship length). Compared to the PDP algorithm, the ADP algorithm demonstrates a heightened ability to detect critical points within the curved segments of the trajectory. However, it is noted that the ADP algorithm retains an excessive number of critical points (region B in the Fig. 9). This over-retention, while enhancing detection, may lead to a less efficient compression outcome. The ACTD-DP algorithm is distinguished by a high CR and is second only in terms of the lowest LLR and MSED, with the MSED (20.507 m, approximately 0.14 times the ship length). Comparatively, the ACTD-DP algorithm outperforms the ADP algorithm in the detection of critical points. During the compression of linear segments of trajectories, the algorithm retains a greater number of critical points. In the compression of curved trajectory segments, the number of retained critical points is moderate (region B in the Fig. 9). Therefore, the ACTD-DP algorithm is recognized for its superior compression efficacy.

Ship4's trajectory is relatively simple with distinct boundaries between straight and curved segments. However, there is a high number of anchoring points, constituting 72% of the total number of points. As a result, the compression of the anchoring paths is significant, leading to the higher LLR values. The SW algorithm exhibits the lowest values for both CR (only 22.05%) and LLR, with the relatively large MSED. The ability of SW algorithm to detect the critical points of the line segments and the curve segments is quite different. And the SW algorithm fails to effectively process critical points in anchoring trajectories (region C in the Fig. 9), with the CR of less than 1% for such trajectories, thus resulting in the poorest compression performance. The DP algorithm demonstrates the highest values for CR, LLR and MSED (373.5619 meters, approximately 1.7 times the ship length). This indicates a significant reduction in trajectory detail, which adversely affects the detection capability for critical points, particularly in anchoring trajectories (region C in the Fig. 9). The DP algorithm's approach compresses the entire anchoring trajectory into a single critical point, which fails to meet the research requirements for analyzing anchorage stay patterns and utilization rates. The PDP algorithm exhibits a lower CR and a suboptimal MSED, suggesting a more conservative compression strategy. While the PDP algorithm performs well in compressing linear and typical curved trajectories, its heightened sensitivity to transitional segments leads to a less effective compression of anchoring trajectories (region C in the Fig. 9). This overemphasis on detecting connections between trajectory segments may compromise the fidelity of the compressed trajectory in representing the original anchoring behavior. The ADP algorithm exhibits moderate values for CR, LLR and MSED. Compared to the PDP algorithm, the ADP algorithm demonstrates enhanced capabilities in detecting and retaining critical points, particularly at the junctions between linear and curved segments of trajectories. However, akin to the DP algorithm, the ADP algorithm struggles to effectively process anchoring trajectories (region C in the Fig. 9), thereby limiting its utility in accurately capturing the nuances of such trajectories. The ACTD-DP algorithm is distinguished by the second-highest values for CR and LLR and the lowest MSED (24.9138 m, approximately 0.11 times the ship's length). This algorithm excels in the detection and retention of critical points, particularly with a uniform distribution of critical points in linear trajectory segments. In comparison to the ADP algorithm, the ACTD-DP algorithm also demonstrates efficacy in handling anchoring trajectories, capturing the typical positions and trajectories during the ship's anchoring process (region C in the Fig. 9). Consequently, upon comprehensive analysis of the compression effects, the ACTD-DP algorithm is concluded to provide the optimal compression performance.

## Discussion and conclusion

Building upon the traditional DP algorithm<sup>31</sup> and drawing inspiration from the methodology of the PDP algorithm<sup>29</sup> and the ADP algorithm<sup>30</sup>, the ACTD-DP algorithm is proposed and experimental validation is conducted. The threshold values of other comparison algorithms are based on static information (ship length, ship width, fixed distance value, etc.), and the trajectory compression effects are greatly affected by the external environment, and the algorithms have poor adaptive ability. In contrast, the ACTD-DP algorithm employs the optimal threshold difference method, reducing reliance on fixed thresholds and enhancing the robustness and applicability of the algorithm. From the overall analysis of compression effects (Table 5 and Fig. 8), compared to the other four algorithms, the ACTD-DP algorithm demonstrates the strongest capability in detecting and retaining key points. It maintains the smallest AMSED value while preserving the higher ACR value, resulting in the best compression performance. Analyzing the evaluation metrics for the four trajectories, the ACTD-DP algorithm exhibits the best compression performance for all trajectories except Ship2's trajectory, demonstrating strong adaptability to different trajectories.

However, the ACTD-DP algorithm also has a notable drawback. The ACTD-DP algorithm requires curve fitting and the calculation of core thresholds and optimal threshold difference. Consequently, the computational complexity is relatively high, leading to increased algorithmic execution time. Concurrently, the ACTD-DP algorithm yields the lower CR for trajectories with a limited number of points and abrupt changes in course, such as Ship2. The compression performance for these types of trajectories could be further enhanced. These observations also provide directions for future research endeavors.

## Data availability

The data that support the findings of this study are available from Shanghai Maritime Safety Administration but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. The datasets generated and analysed during the current study are not publicly available due data sensitivity but are available from the corresponding author on reasonable request.

Received: 1 May 2024; Accepted: 30 August 2024

Published online: 13 September 2024

## References

- Huang, C. et al. A simulation model for marine traffic environment risk assessment in the traffic separation scheme. In *2021 6th International Conference on Transportation Information and Safety (ICTIS)*. 213–221 <https://doi.org/10.1109/ICTIS54573.2021.9798623> (IEEE, 2021).
- Kim, E. et al. Sensitive resource and traffic density risk analysis of marine spill accidents using automated identification system big data. *J. Mar. Sci. Appl.* **19**, 173–181. <https://doi.org/10.1007/s11804-020-00138-2> (2020).
- An, K. E-navigation services for non-SOLAS ships. *Int. J. e-Navigat. Maritime Econ.* **4**, 13–22. <https://doi.org/10.1016/j.enavi.2016.06.002> (2016).
- Barco, S. G., Lockhart, G. G. & Swingle, W. M. Using RADAR & AIS to investigate ship behavior in the Chesapeake Bay ocean approach off of Virginia, USA. In *2012 Oceans*. 1–8 <https://doi.org/10.1109/OCEANS.2012.6404872> (IEEE, 2012).
- Lei, P.-R., Yu, P.-R. & Peng, W.-C. A Framework for maritime anti-collision pattern discovery from AIS network. In *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. 1–4 <https://doi.org/10.23919/APNOMS.2019.8892899> (IEEE, 2019).
- Han, P. & Yang, X. Big data-driven automatic generation of ship route planning in complex maritime environments. *Acta Oceanol. Sin.* **39**, 113–120. <https://doi.org/10.1007/s13131-020-1638-5> (2020).
- Hao, Y., Zheng, P. & Han, Z. Automatic generation of water route based on AIS big data and ECDIS. *Arab. J. Geosci.* **14**, 1–8 <https://doi.org/10.1007/s12517-021-06930-w> (Springer, 2021).
- Blindheim, S., Johansen, T. A. & Utne, I. B. Risk-based supervisory control for autonomous ship navigation. *J. Mar. Sci. Technol.* **28**, 624–648. <https://doi.org/10.1007/s00773-023-00945-6> (2023).
- Guo, Y. & Ding, Z. Application of big data in analyzing the impact of explosive cyclone on ship navigation safety. In *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*. 910–913 <https://doi.org/10.1109/ICOSEC51865.2021.9591958> (IEEE, 2021).
- Liu, Z., Li, Y., Zhang, Z., Yu, W. & Du, Y. Spatial modeling and analysis based on spatial information of the ship encounters for intelligent navigation safety. *Reliabil. Eng. Syst. Saf.* **238**, 109489. <https://doi.org/10.1016/j.res.2023.109489> (2023).
- Makris, A. et al. Evaluating the effect of compressing algorithms for trajectory similarity and classification problems. *GeoInformatica* **25**, 679–711. <https://doi.org/10.1007/s10707-021-00434-1> (2021).
- Chen, J., Zhang, J., Chen, H., Zhao, Y. & Wang, H. A TDV attention-based BiGRU network for AIS-based vessel trajectory prediction. *iScience* **26**, 106383 <https://doi.org/10.1016/j.isci.2023.106383> (2023).
- Liu, H., Liu, Y. & Zong, Z. Research on ship abnormal behavior detection method based on graph neural network. In *2022 IEEE International Conference on Mechatronics and Automation (ICMA)*. 834–838 <https://doi.org/10.1109/ICMA54519.2022.9856198> (IEEE, 2022).
- Murray, B. & Perera, L. P. Ship behavior prediction via trajectory extraction-based clustering for maritime situation awareness. *J. Ocean Eng. Sci.* **7**, 1–13. <https://doi.org/10.1016/j.joes.2021.03.001> (2022).
- Jeong, S. & Kim, T.-W. Generating a path-search graph based on ship-trajectory data: Route search via dynamic programming for autonomous ships. *Ocean Eng.* **283**, 114503. <https://doi.org/10.1016/j.oceaneng.2023.114503> (2023).
- Meratnia, N. & De By, R. A. Spatiotemporal compression techniques for moving point objects. In (Goos, G. et al. eds.) *Advances in Database Technology-EDBT 2004. Lecture Notes in Computer Science*. Vol. 2992. 765–782 [https://doi.org/10.1007/978-3-540-24741-8\\_44](https://doi.org/10.1007/978-3-540-24741-8_44) (Springer, 2004).
- Jie, G., Xin, S., Xiaowei, S. & Daofang, C. Online compression algorithm of fishing ship trajectories based on improved sliding window. *J. Shanghai Maritime Univ.* **44**, 17–24 <https://doi.org/10.13340/j.jsmu.202208070214> (2023).
- Zhu, F. & Ma, Z. Ship trajectory online compression algorithm considering handling patterns. *IEEE Access* **9**, 70182–70191. <https://doi.org/10.1109/ACCESS.2021.3078642> (2021).
- Gao, M. & Shi, G.-Y. Ship spatiotemporal key feature point online extraction based on AIS multi-sensor data using an improved sliding window algorithm. *Sensors* **19**, 2706. <https://doi.org/10.3390/s19122706> (2019).
- Qi, Z., Yi, C., Li, X. & Wen, G. Improved sliding window trajectory compression algorithm considering motion characteristics. *J. Geomat. Sci. Technol.* **37**, 622–627 (2020) (5 citations (CNKI)[2024-3-13]).

21. Bai, X., Xie, Z., Xu, X. & Xiao, Y. An adaptive threshold fast DBSCAN algorithm with preserved trajectory feature points for vessel trajectory clustering. *Ocean Eng.* **280**, 114930. <https://doi.org/10.1016/j.oceaneng.2023.114930> (2023).
22. Zhang, S.-K., Shi, G.-Y., Liu, Z.-J., Zhao, Z.-W. & Wu, Z.-L. Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity. *Ocean Eng.* **155**, 240–250. <https://doi.org/10.1016/j.oceaneng.2018.02.060> (2018).
23. Wei, Z., Xie, X. & Zhang, X. AIS trajectory simplification algorithm considering ship behaviours. *Ocean Eng.* **216**, 108086. <https://doi.org/10.1016/j.oceaneng.2020.108086> (2020).
24. Huang, C., Qi, X., Zheng, J., Zhu, R. & Shen, J. A maritime traffic route extraction method based on density-based spatial clustering of applications with noise for multi-dimensional data. *Ocean Eng.* **268**, 113036. <https://doi.org/10.1016/j.oceaneng.2022.113036> (2023).
25. Shukai, Z., Zhengjiang, L., Xianku, Z., Guoyou, S. & Yao, C. A method for AIS track data compression based on Douglas–Peucker algorithm. *J. Harbin Eng. Univ.* **36**, 595–599 (2015).
26. Li, H. *et al.* Unsupervised hierarchical methodology of maritime traffic pattern extraction for knowledge discovery. *Transport. Res. Part C Emerg. Technol.* **143**, 103856. <https://doi.org/10.1016/j.trc.2022.103856> (2022).
27. Cui, C. & Dong, Z. Ship space-time AIS trajectory data compression method. In *2022 7th International Conference on Big Data Analytics (ICBDA)*. 40–44 <https://doi.org/10.1109/ICBDA55095.2022.9760355> (IEEE, 2022).
28. Zhou, Z., Zhang, Y., Yuan, X. & Wang, H. Compressing AIS trajectory data based on the multi-objective peak Douglas–Peucker algorithm. *IEEE Access* **11**, 6802–6821. <https://doi.org/10.1109/ACCESS.2023.3234121> (2023).
29. Zhao, L. & Shi, G. A method for simplifying ship trajectory based on improved Douglas–Peucker algorithm. *Ocean Eng.* **166**, 37–46. <https://doi.org/10.1016/j.oceaneng.2018.08.005> (2018).
30. Tang, C. *et al.* A method for compressing AIS trajectory data based on the adaptive-threshold Douglas–Peucker algorithm. *Ocean Eng.* **232**, 109041. <https://doi.org/10.1016/j.oceaneng.2021.109041> (2021).
31. Douglas, D. H. & Peucker, T. K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. In *Classics in Cartography* (Dodge, M. ed.). 1 Ed. 15–28 <https://doi.org/10.1002/9780470669488.ch2> (Wiley, 2011).
32. MathWorks, I. *Matlab*. <https://www.mathworks.com/products/matlab/> (2024).

## Acknowledgements

This study is supported by the research project “Science and Technology Commission of Shanghai Municipality” (Grant Nos. 22010502000 and 23010501900).

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to P.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024