

FULL PAPER

Open Access



ICESat-2 single photon laser point cloud denoising algorithm based on improved DBSCAN clustering

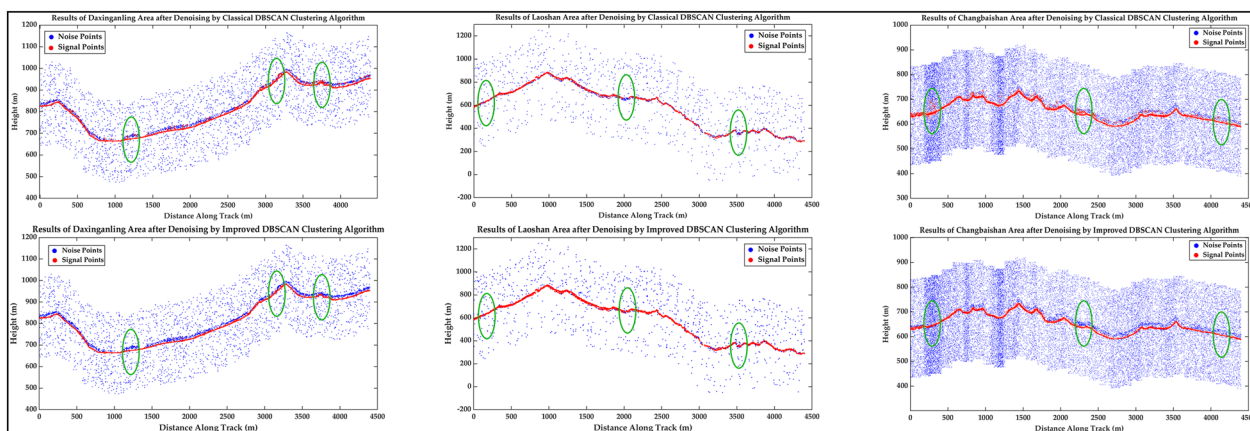
Dong Wang^{1,2*}, Jiachen Yu¹, Fengying Liu^{1,2} and Qinghua Li¹

Abstract

The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) has great potential for development due to its advantages of the use of multiple beams, low energy consumption, high repetition frequency, and high measurement sensitivity. However, the weak photon signal emitted by the photon counting lidar is susceptible to the background noise caused by the sun and the atmosphere, which can seriously affect the processing and application of laser data. This paper proposes an improved DBSCAN clustering algorithm for denoising single photon laser point clouds in mountainous areas. Firstly, a grouping method based on elevation and distance statistics is proposed to reduce the influence of terrain undulations on denoising accuracy. Finally, an automatic radius search method is put forward to determine clustering radius of each group, automatically find the optimal radius, and improve the existing DBSCAN clustering method. The method proposed in this paper is compared with the classical DBSCAN algorithm. The results show that the proposed algorithm significantly improves denoising accuracy in mountainous areas and effectively filters out most background noise.

Keywords ICESat-2, Laser altimetry, Point cloud denoising, DBSCAN, Distance statistics

Graphical Abstract



*Correspondence:

Dong Wang
sdustwd@163.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

1 Introduction

Spaceborne lidar is an emerging active remote sensing detection technology. It transmits a laser pulse with a certain frequency to the Earth's surface and receives the scattered echoes, finally obtaining the accurate three-dimensional space coordinates of the laser footprint point. Spaceborne lidar has the ability to actively obtain global surface elevation and features a wide observation range (Neumann et al. 2019; Xie et al. 2021). Satellite altimetry data have broad prospects and applications in forestry remote sensing, polar glacier monitoring, and global change research. ICESat-2 (Ice, Cloud and Land Elevation Satellite-2) was launched by NASA (National Aeronautics and Space Administration) in 2018. ICESat-2 is equipped with ATLAS (Advanced Topographic Laser Altimeter System), which uses a 6-beam micro-pulse photon counting lidar for measurement (Zhu et al. 2020). It also uses a laser with a high repetition frequency, low energy, and a highly sensitive laser detector. The single photon energy in the echo signal is used to obtain the distance information of a long-distance space target with a lower laser pulse energy compared to other lidar technologies. ICESat-2 can solve the problems of large volume, large mass, low reliability, energy intensity and repetition frequency that are seen in traditional lidars.

Currently, previous studies have used ICESat-2 data to monitor topographic changes, ice and vegetation. Vernimmen et al. (2020) used ICESat-2 data to create the Global Coastal Lowland DTM (Digital Terrain Model), which is much more accurate than other existing global digital elevation models. Michaelides et al. (2021) used ICESat-2 data to estimate ground surface-height changes due to the seasonal freezing and thawing of the active layer, and discussed several influencing factors and the future potential of ICESat-2 data in permafrost applications. Chen et al. (2022) verified the coverage performance of ICESat-2 on global reservoirs and further explored its potential for monitoring long-term changes in reservoir water level and water storage. A variety of marine data products provided by ICESat-2 can be used to study the changes in snow depth, thickness and volume in polar sea ice, sea ice surface classification, and surface height on complex ice surface (Kacimi and Kwok 2022; Petty et al. 2021; Herzfeld et al. 2022). Mulverhill et al. (2022) assessed the consistency of Canadian canopy height estimates obtained by using forest canopy height products from ICESat-2 and the NTEMS (National Terrestrial Ecosystem Monitoring System) at various ecological gradients. The coherence between ICESat-2 and NTEMS datasets suggests potential integration. Many scholars have shown that ATL08 (Advanced Topographic Laser Altimeter System Level 08) canopy height is more suitable for relatively dense canopy environments such

as coniferous forests and broad-leaved forests (Malambo and Popescu 2021). The quality of data acquired in winter is superior to that acquired in summer due to differences in vegetation structure and snow coverage (Zhu et al. 2022). The comprehensive evaluation of the quality of the ATL08 product can be useful for improving future versions of the product and for guiding the selection and use of ICESat-2 data (Tian and Shan 2021).

Because of the high sensitivity of the photon counting lidar, instrument and background noises impact on ATLAS data (McGarry et al. 2021). As a result, a large amount of noise is mixed into the observation results, which seriously restricts the use of ICESat-2. Therefore, denoising point cloud data quickly and efficiently have become a key issue for the extensive application of ICESat-2. The denoising algorithms for single photon point cloud data can be roughly classified into three types.

The first denoising algorithm is based on local statistical parameters, which is based on the local density information of photon point cloud data, and determines noise points according to the local density value or local density histogram of the signal points (Wang et al. 2016). Zhu et al. (2018) proposed a local statistical analysis method for denoising point cloud data. By calculating the maximum density of each photon in each direction as its true density, the noise points are denoised out by using the empirical threshold. But the algorithm is sensitive to the grid size when calculating the photon density, thus affecting the denoising accuracy.

The second denoising algorithm is based on raster image processing. This algorithm, which is based on morphological information from point cloud signals, rasterize two-dimensional images of point cloud sections. Also, image processing technology is applied to eliminate noise points. Aiming at the difficulty in measuring feature length in point cloud data, Li et al. (2021) proposed to convert the three-dimensional point cloud data into a two-dimensional elevation grid map in order to estimate the geometric features, which simplifies the problem. However, signal photons may cause partial data loss after rasterization, losing effectiveness.

The third denoising algorithm is based on density spatial clustering. This method makes clustering analysis using the discrete distribution of noise photons in space, thereby eliminating noise. The main density clustering methods include Bayesian, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and ordering points to identify the clustering structure (OPTICS). Among the three methods, DBSCAN is one of the most widely used photon denoising methods and can be extremely effective in clustering signal photons. Zhang et al. (2021) proposed an effective algorithm for the extraction of signal photons from the

weak beam data of ICESat-2 in mountainous areas. By using the slope–noise relationship in strong beam data to invert the track slope of the weak beam, an improved DBSCAN algorithm can be used to extract effective signal photons from weak beam data with a low signal-to-noise ratio. However, the classical DBSCAN algorithm is sensitive to the input parameters, so it has poor adaptability to different terrains. Inappropriate parameter settings in areas with large variations in terrain tend to cause excessive denoising, resulting in a lack of ground points that cannot reflect the surface morphology or filter out the noise points floating above the surface (Kui et al. 2023). Therefore, improving the accuracy of denoising is a key focus of research in altimetry observation.

Some scholars have proposed other types of denoising algorithms. An adaptive denoise that depends on the local slope is considered to be robust in the identification of signal points from points with high background noise, and appropriate for low-density data caused by slope (Xie et al. 2020). Zhang et al. (2022) proposed a noise removal algorithm with no input parameters based on the isolation of quadtrees, which can improve denoising accuracy and efficiency.

Due to the profile observation of the single photon laser radar system, the single photon point cloud data

are distributed in a narrowband, which is different from the conventional point cloud data that can generate a large area of 3D topographic map. The denoising effect is typically not good enough in the area with large terrain fluctuations. Therefore, it is necessary to perform fine denoising on point cloud data to filter out noise points suspended above or below the ground. However, owing to uneven point cloud distribution in the signal, the algorithm using a single threshold cannot effectively remove a large amount of noise. Targeting the presence of many noise points in single photon point cloud data and the similarity of the height of many noise points to that of signal points, denoising ground surface point cloud data can be formulated as a classification task. Ground points are considered signals, whereas instrument noise, background noise, and non-ground points such as vegetation are considered noise. However, the traditional denoising method is insufficient to filter out a large amount of background noise at significant variations in topography.

This paper proposes a refined DBSCAN algorithm to avoid the manual selection of parameters during clustering, and improve the efficiency and accuracy of data processing. Our main contributions are as follows:

- (1) Proposal of a DBSCAN denoising method coupled with height and distance factor. The DBSCAN algo-

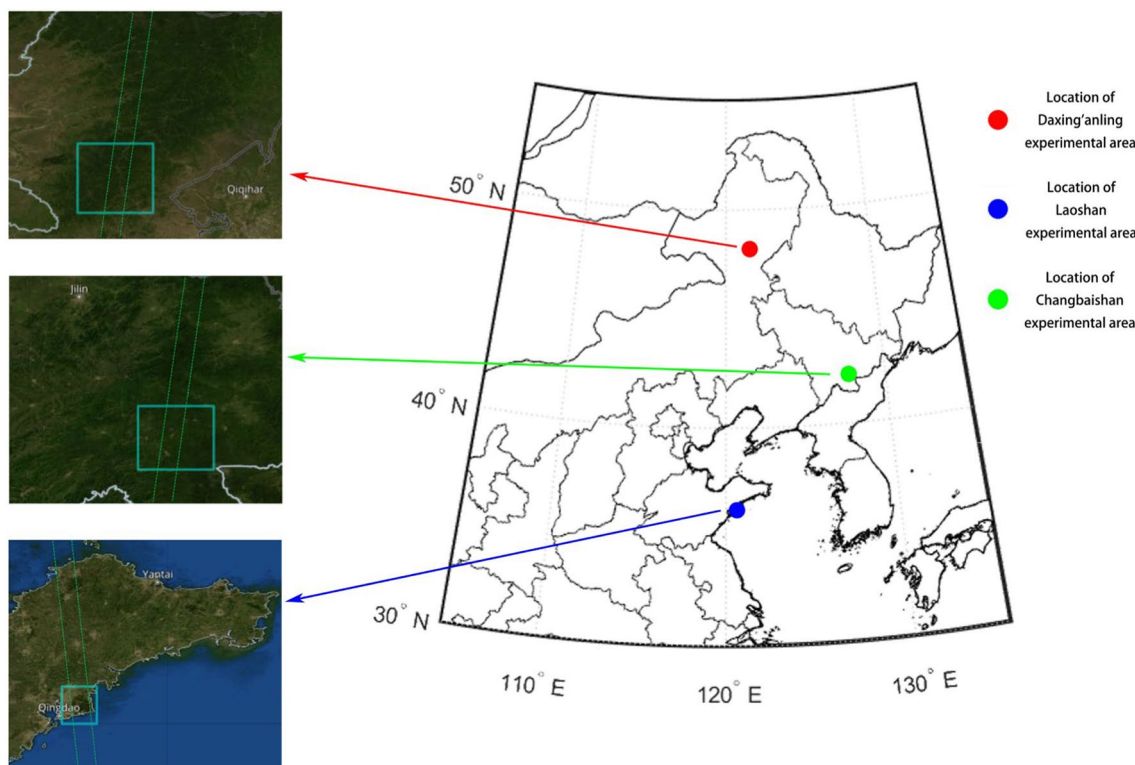


Fig. 1 Location of the experimental areas and ATLAS data trajectory

Table 1 Comparison of instrument parameters between ICESat-1 and ICESat-2

Key parameter	ICESat-1	ICESat-2
Rail type	Non-sun synchronization	Non-sun synchronization
Height (km)	600	500
Scope of coverage	86°N—86°S	88°N—88°S
Observation period (d)	91	91
Laser sensor	GLAS	ATLAS
Number of lasers	1 active and 2 standby	1
Number of laser beams	1	6
Measuring principle	Impulse type	Micro pulse photon-counting
Spot diameter (m)	70	17
Track spacing (m)	170	0.7
Wavelength (nm)	1064/532	532
Frequency (Hz)	40	10,000
Transmitted pulse width (ns)	4—6	About 1
Recording mode	Full-waveform	Photon point cloud

rithm is improved, grouping areas with large variation in topography based on elevation and distance statistics and automatically calculating the clustering radius Epsilon (Eps) according to the distance between point clouds.

- (2) The influence of different elevation and distance parameters on the results is analyzed and the best parameters are selected.
- (3) The denoising results of the classical DBSCAN algorithm and the improved DBSCAN algorithm are compared, which reflects the correct detection effect of the improved DBSCAN algorithm on signal points.

The rest of this paper is organized as follows: Sect. 2 describes the experimental areas, data sources and methods. Section 3 describes the parameters selection and the experiments. Section 4 discusses the results and potential future work. Section 5 presents general conclusions, highlights the main contributions.

2 Materials and methods

2.1 Data sources

The ATLAS and its auxiliary system (GPS and spaceborne camera) carried by ICESat-2 measure the photon round-trip time and determine the spatial position of the surface reflecting the photon (Zhu et al. 2021). ATLAS uses a 6-beam photon counting lidar with higher spatial coverage, which is more than three times that of ICESat-1. Laser pulses generate three pairs of ground trajectories, each of which is usually about 17 m wide. The left/right points of each pair of ground trajectories are about 90 m apart in the direction across the track, about 2.5 km apart in the direction along the track, and about 3.3 km apart between columns, as shown in Fig. 1. Each pair of beams has different emission energies, i.e., weak light and strong light, and the energy ratio between them is about 1:4 (Neuenschwander et al. 2020).

ATLAS is a lidar altimetry system based on micro-pulse photon counting. It emits and receives laser signals and records them at a band of 532 nm only, with a pulse width of about 1 ns. Since photon counting detects weak signals, in order to reduce the influence of background noise, the field angle of ICESat-2 is about 66 μrad (~ 40 m), smaller than that of ICESat-1 at the

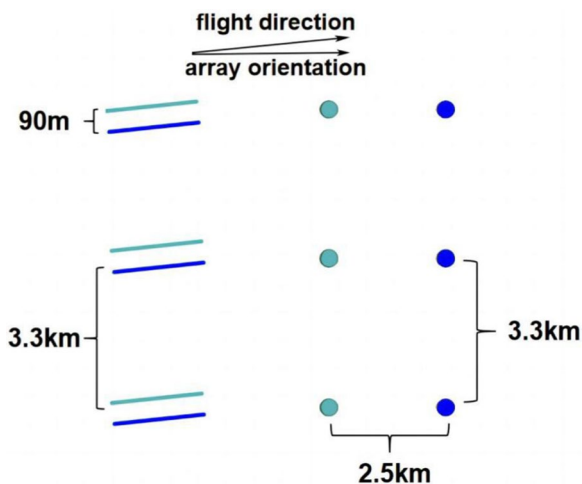


Fig. 2 The idealized beam and footprint pattern of ATLAS, with deep blue representing strong light and pale blue representing weak light

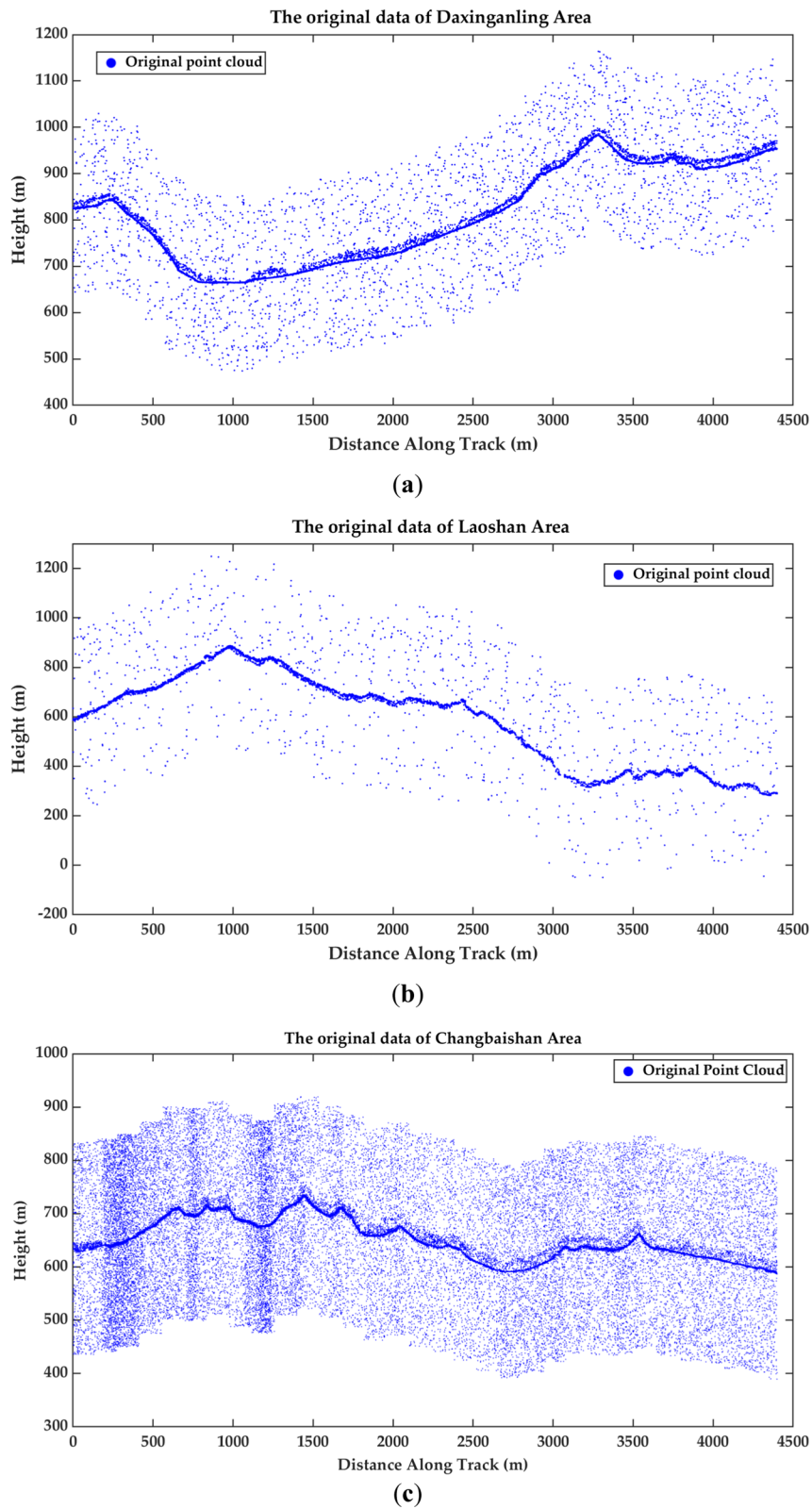


Fig. 3 Original point cloud in the distance and elevation coordinate system: **a** the original point cloud data of latitude 48.16° to 48.20° in Daxing'anling experimental area; **b** the original point cloud data of latitude 36.18° to 36.22° in Laoshan experimental area; **c** the original point cloud data of latitude 42.52° to 42.56° in Changbaishan experimental area

532 nm band. The footprint diameter of ICESat-2 is only about 17 m, which is the average of each pointing. ICESat-2 has a laser emission frequency of up to 10 kHz and a footprint spacing of approximately 0.7 m along the satellite's trajectory. Compared to ICESat-1's 170 m footprint spacing, the sampling density is greatly improved while the official photon sampling accuracy can reach the centimeter level (Xing et al. 2020). As a new generation of multi-beam lidar, compared with ICESat-1, its observation mode has varied greatly. Table 1 shows the comparison of the two instrument parameters.

The ATL03 (Advanced Topographic Laser Altimeter System Level 03) Gt11 photon laser point cloud data collected by the ATLAS pulsed beam during the operation of ICESat-2 around mid-October 2020 were used in this experiment. Due to the large data span of the whole experimental area, the point cloud density in the data segment changes significantly. Therefore, this paper selects the point cloud data within the latitude range of 48.16° to 48.20° in the Daxing'anling experimental area, 36.18° to 36.22° in the Laoshan area and 46.52° to 46.56° in the Changbaishan area for denoising. As shown in Fig. 3, the Laoshan experimental area has a more complex terrain, with an elevation gap of about 600 m and numerous hills. The elevation difference of Changbaishan experimental area is about 150 m. Changbaishan is covered with rich vegetation. The elevation difference of Daxing'anling experimental area is about 350 m, and this area contains a large number of forests. The point cloud profile generated is shown in Fig. 2.

2.2 Methods

2.2.1 Classical DBSCAN clustering method

DBSCAN is a classical density-based clustering algorithm (Ester et al. 1996). Its principle is as follows: firstly, an initial radius and a minimum number of neighborhoods are determined. Starting from any point in the data set, if there is more than the minimum number of neighborhoods within the radius distance from this point (including the original point itself), it is considered that they are all part of the "cluster". Then, the cluster is

expanded by checking all points, and whether they also have points exceeding the minimum number of neighborhoods within the radius distance. Then, the cluster is recursively expanded. Finally, points that exceed the minimum number of neighborhoods are added to the cluster. A new arbitrary point is selected, and the above process is repeated.

For a given dataset $D = \{x_1, x_2, \dots, x_m\}$, several concepts involved in the DBSCAN clustering algorithm are defined as follows:

Epsilon-neighborhood (N_{Eps}): For $x_j \in D$, its N_{Eps} contains the points in D whose distance from x_j is not greater than Eps , namely $N_{Eps}(x_j) = \{x_i \in D \mid dist(x_i, x_j) \leq Eps\}$;

Core object: If the N_{Eps} of x_j includes at least $MinPts$ (minimum points) samples, i. e. $|N_{Eps}(x_j)| \geq MinPts$, then x_j is the core object;

Density direct: If x_j is located in the N_{Eps} of x_i and x_i is the core object, x_j is called density direct from x_i ;

Density-reachable: For x_i and x_j , if there is a sample sequence P_1, P_2, \dots, P_n , where $P_1 = x_i, P_2 = x_j$ and P_{i+1} are density direct from P_i , then x_j is called density-reachable from x_i ;

Density-based: For x_i and x_j , if there is x_k so that x_i and x_j are both density-reachable from x_k , then x_i is called density-based with x_j .

The basic concept of the DBSCAN clustering algorithm is briefly described through four points x_1, x_2, x_3, x_4 and their spatial distribution characteristics (Fig. 3). The dotted lines in Fig. 3 express Epsilon-neighborhoods. x_1 contains more samples than $MinPts$ and is the core object. x_2 is density-reachable from x_1 directly, x_3 is density-reachable from x_2 directly and is density-reachable from x_1 indirectly. x_3 is density-based with x_4 .

2.2.2 Improved DBSCAN clustering method

To improve denoising accuracy and efficiency for point cloud data in large topographic gradients, we propose a revised DBSCAN method using elevation and distance statistics with an automatic radius search. The specifics of the method are as follows (Fig. 4).

- (1) Point cloud data grouping method based on elevation and distance statistics

For mountainous forest areas with large topographic relief, a single Eps and $MinPts$ parameter setting cannot simultaneously meet different terrain conditions in the experimental area. Therefore, the data of the experimental area are grouped to realize the adaptability of the parameters and improve the overall denoising accuracy. The grouping process is described as follows:

- ① The original point cloud data are arranged along the track; the average elevation is calculated and

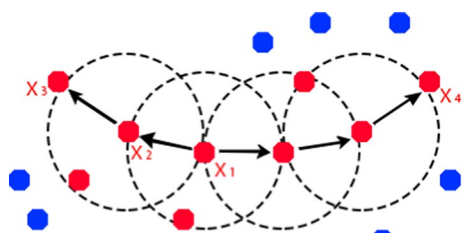


Fig. 4 DBSCAN basic concept diagram ($MinPts=3$)

second-order curve fitting is performed using the least squares method. According to Eqs. (1):

$$h = ax^2 + bx + c. \tag{1}$$

② Taking the fitting curve of the average elevation as a reference, the height difference h along-track direction is calculated and the height difference parameters h_a and distance parameters l_a are set as grouping conditions. Each group needs to meet the following conditions: $h_i \leq h_a$ and $l_i \leq l_a$ (where, h_i, l_i is the height difference and horizontal distance along the orbit of group i).

(2) Calculating Eps parameter by group

After grouping, Eps is calculated in each group. In each group of raw data, the distance between any two points is calculated. According to Eqs. (2) to (5), the Euclidean distance, the maximum distances, the minimum distances and the differences between the maximum and minimum distances for any two points are calculated, respectively:

$$dist(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \tag{2}$$

$$max = MAX\{dist(i, j) | 0 \leq i \leq n, 0 \leq j \leq D\}, \tag{3}$$

$$min = MIN\{dist(i, j) | 0 \leq i \leq n, 0 \leq j \leq D\}, \tag{4}$$

$$Dist = max - min, \tag{5}$$

where D is a group of sample data sets, and n is the number of point clouds in the group of data sets.

The search radius (Eps) is further refined. Based on the difference between the maximum and minimum values of the distance, a statistical interval value is set. Using the minimum distance as the starting point and the maximum distance as the ending point, the distances between points in each section are plotted based on a histogram. The interval with the highest frequency appears as the Eps of this dataset, which is the search radius for clustering.

The statistical Eps are substituted into the DBSCAN clustering algorithm. MinPts is set based on the point cloud density and the size of Eps, combined with experience, to obtain the clustering results of a data set. The results after grouping and clustering are integrated to obtain the final denoising results. The points in the cluster are saved as signal points, and the points free from the point cloud cluster are denoised as noise points.

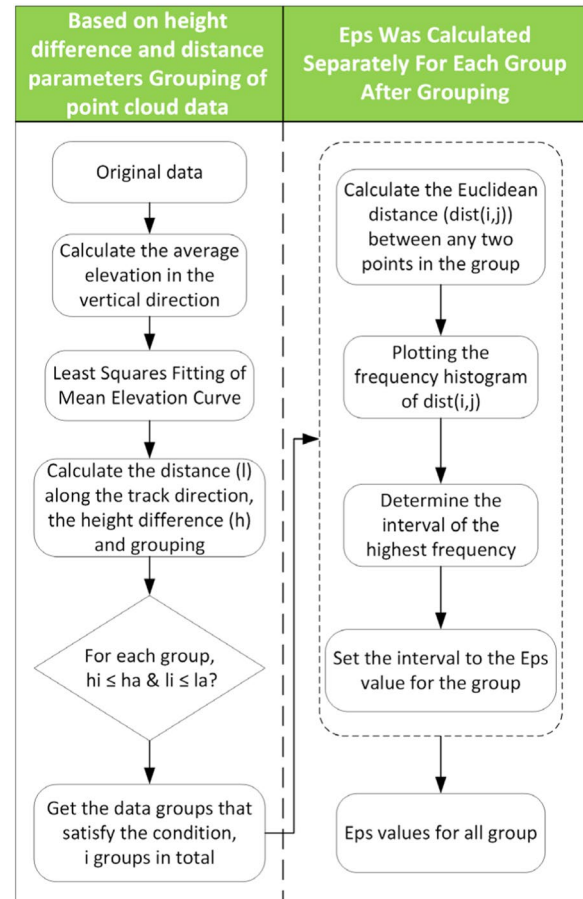


Fig. 5 Schematic diagram of grouping to determine Eps parameters

2.3 Accuracy evaluation

The confusion matrix of binary classification data is used to describe some common error indicators. These error indicators reflect the accuracy of point cloud denoising from different aspects.

Accuracy was evaluated by defining *Accuracy* and *Sensitivity*. The equations are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{7}$$

where TP is the actual target signal, identified as the target signal. FN is the actual target signal, but identified as a non-target signal. FP is the actual non-target signal, but identified as target signal. TN is the actual

non-target signal, identified as a non-target signal. Accuracy can reflect the overall effect of processing the algorithm on experimental data. A higher accuracy indicates a better classification effect. Sensitivity can reflect the classification effect of the algorithm for positive class. In this paper, denoising ground surface point cloud data is formulated as a classification task, where ground points are considered as signals and the

remaining points are considered as noise. The reference signal photon obtained by visual interpretation is used as the actual target signal.

2.4 Study areas

The study areas are located in the Daxing'anling area in the northwest of Heilongjiang, the Laoshan area in

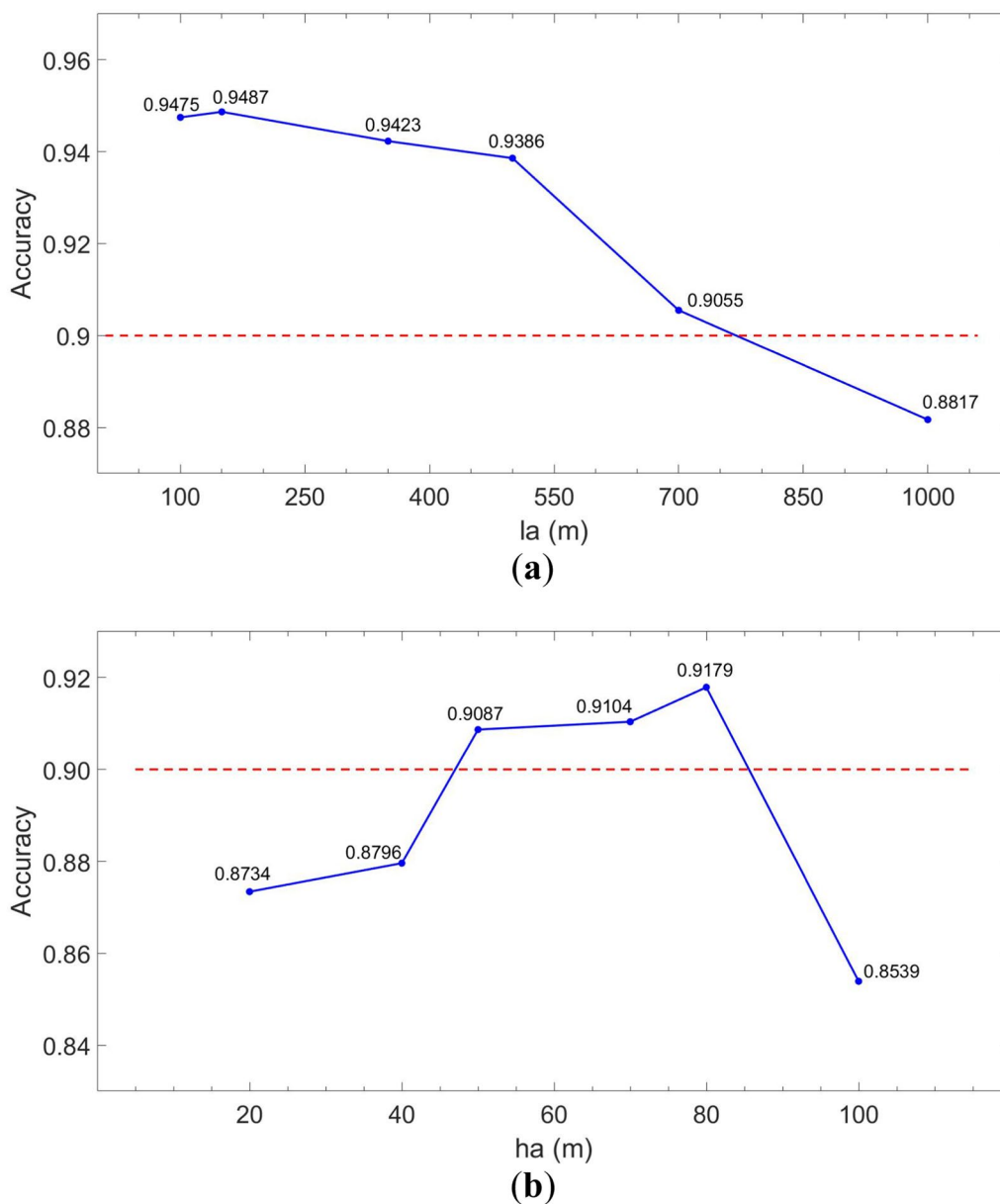


Fig. 6 Schematic diagram of parameter analysis: **a** accuracy changes for different l_a ; **b** accuracy changes for different h_a

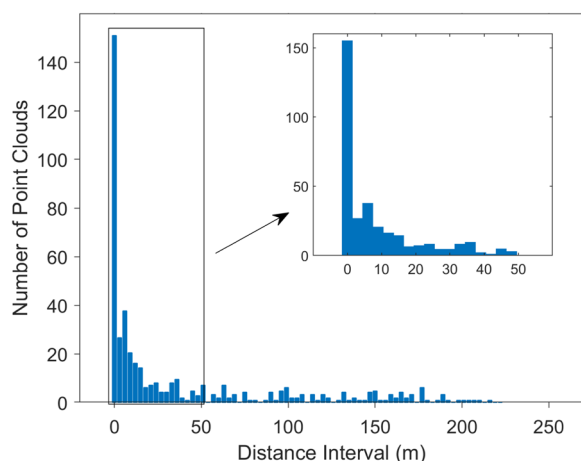


Fig. 7 Histogram of distance statistics. (Taking the third group data of the Daxing'anling experimental area as an example.)

Qingdao, Shandong Province and Changbaishan area in Antu County, Jilin Province, as shown in Fig. 5.

The Daxing'anling Mountains are home to plenty of mountainous forests, which are well-preserved and the largest primitive forests in China. The Daxing'anling area is one of the important forestry bases in China, with complex and diverse forest features. The Daxing'anling Mountains range from the north of the Heilongjiang River in the northern part of Mohe City, Heilongjiang Province, to the upper reaches of the Xilamulun River in the northern part of Chifeng City, Inner Mongolia Autonomous Region. With a northeast–southwest trend, the mountains are more than 1,400 km long, with an average width of about 200 km and an altitude of 1,100–1,400 m. The Daxing'anling area is covered by primitive forest, with the southern section being temperate hardwood forest and the northern section being cold temperate coniferous forest.

The Laoshan mountainous area, with a size of 446 km², is the first peak along China's coastline. The mountains are centered on Laoding and extend in all directions, with longer branches in the northwest and southwest directions. The Laoshan branch extends to the north of Jimo District along the east coast and to the Jiaozhou Bay in the west and the southwest branch extends to the city of Qingdao, forming more than ten hilltops and the ups and downs of the hilly terrain. Laoshan is more than 100 km long, with an altitude of about 600–1,100 m. Due to the complexity of the terrain and a wide variety of plants, various vegetation cover such as forests, shrubs, grasses, desert plants, halophytes, and agricultural cultivation have been formed.

Changbaishan is located in the southeastern part of Jilin Province. The northern and southwestern parts are bordered by Heilongjiang and Liaoning provinces, respectively. The southeastern and eastern parts border North Korea and Russia, respectively. The terrain of Changbaishan gradually slopes from southeast to northwest. The southeast is mainly a middle mountain range with an altitude of more than 1000 m, which gradually decreases to low mountains and hills in the northwest, until the platform with an altitude of about 300 m. The forest cover is 87.9%. It is a natural complex nature reserve with forest ecosystem as the main object of protection.

3 3 Results and analysis

3.1 Parameter selection analysis

In order to select the optimal parameters of height difference and distance, the accuracy of different parameters is calculated. As shown in Fig. 6, when the distance parameter l_a is 150–500 m, the accuracy is higher than $l_a > 500$ m. However, it is found in the verification that if the distance parameter $l_a < 300$ m, it will lead to more groups, much workload and reduced efficiency. Considering the need to balance the work efficiency and high accuracy, the distance parameter l_a is set to 400 m. The accuracy is above 0.9 when the height difference parameter h_a is 50–80 m, although overall the highest accuracy was achieved at $h_a = 80$ m. However, when grouping, the steepness of the terrain at some locations and the large drop in elevation causes the distance along the track for a group to become less than 50 m at $h_a = 80$ m. The number of points in this group is low, which leads to an accuracy of less than 0.5 for this group, which can result in missing ground points so $h_a = 50$ m is set to group the original data.

For the grouped data, the Eps parameter is separately calculated in each group. Using the third ground of data in the Daxing'anling experimental area, the distance along the track is 300 to 416 m, encompassing a total of 441 points. For the determination of Eps of DBSCAN, by counting the distance between any point cloud data, the highest number of occurrences in an interval is chosen as the Eps. According to the statistics, the maximum distance between any two points is 285.4 m, and the minimum distance between any two points is 0.1 m. The statistical interval value is set to 3 m, and the distance between the point clouds counted as shown in Fig. 7. It can be seen that the distance between point clouds is the greatest at the intervals of 0 to 3 m. Therefore, the Eps of this group of

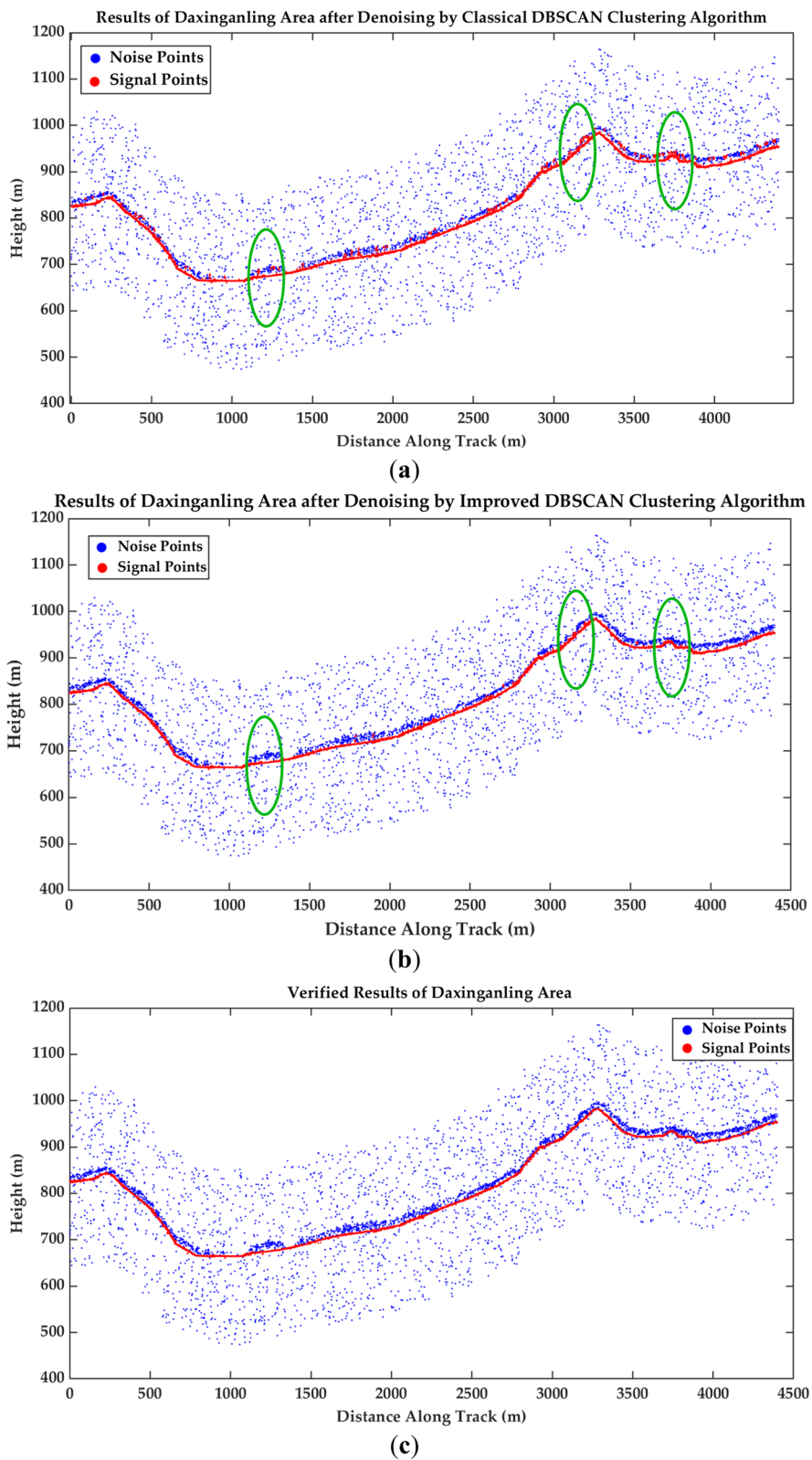


Fig. 8 Comparison of denoising results in the Daxinganling area. **a** Denoising results using classical DBSCAN, **b** denoising results using improved DBSCAN, and **c** verification results from visual interpretation

Table 2 Confusion matrix of point cloud denoising results

		Number of signal points (7785)	Number of noise points (4558)	Accuracy (%)	Sensitivity (%)
Classical DBSCAN clustering algorithm	Number of signal points (7669)	5304	2365	60.73	69.16
	Number of noise points (4674)	2481	2193		
Improved DBSCAN clustering algorithm	Number of signal points (7623)	7426	197	95.49	97.42
	Number of noise points (4720)	359	4361		

data is 3 m, and MinPts is typically determined as an empirical threshold.

3.2 Evaluation of denoising results

3.2.1 A case study in Daxing'anling, China

As shown in Fig. 8, in the Daxing'anling area, the denoising effect of the improved DBSCAN algorithm is better than that of the classical DBSCAN algorithm and is basically consistent with the verified result. Most of the noise points far from the ground can be removed by the classical DBSCAN algorithm, but in the Daxing'anling area, which contains primitive forest areas, the vegetation near the ground is identified as signal points and not filtered out. However, the improved DBSCAN algorithm can not only remove the noise points far from the ground points, but also filter out the vegetation points near the ground, leaving only surface elevation. The surface morphology of the experimental area is clearly visible, but there are a few noise points that are not filtered out. Table 2 establishes the confusion matrix between the denoising results of the classic and improved algorithms and the verified results of the Daxing'anling area. The number of points identified as true signals after processing by the improved DBSCAN clustering algorithm is 7,623, of which the number of noise points identified as signals is 197. After processing with the classical DBSCAN clustering algorithm, the number of points identified as signals is 7,669, but the number of noise points identified as signals is 2,365, and the number of noises among signal points is reduced by 2,168. This indicates that the improved DBSCAN clustering algorithm effectively reduces the number of noise points. The improved DBSCAN obtained a denoising accuracy is 95.49% and the sensitivity is 97.42%.

3.2.2 A case study in Laoshan, China

As shown in Fig. 9, the improved DBSCAN algorithm produces an obvious clustering effect on point cloud data in the Laoshan experimental area, and can remove more than 90% of noise points. In the classical DBSCAN clustering algorithm, some of the signal points are misclassified as noise points, resulting in thin, discontinuous ground points. The vegetation in this area is mostly low shrubs, so some near-surface vegetation points are confused with ground points, thus affecting the noise classification accuracy. However, according to the confusion matrix of the denoising results and the verified results of the two algorithms established in Table 3, the improved DBSCAN clustering algorithm correctly identifies 3,703 signal points, with 92 signal points misclassified as noise points. Compared to the classical algorithm, 3,399 signal points were correctly identified after processing, while 396 signal points were misclassified as noise points. This indicates that the improved algorithm has shown a significant enhancement in the correct identification of signal points. The improved accuracy is 94.22% and the sensitivity is 97.58%.

3.2.3 A case study in Changbaishan, China

In the Changbaishan area shown in Fig. 10, due to the large amount of data at this area, the classical DBSCAN algorithm misclassifies the noise points into signal points when encountering dense noise. From the confusion matrix of the denoising results established in Table 4, the number of misclassified points of the classical DBSCAN algorithm is 10228, and the number of misclassified points of the improved algorithm is 3063, indicating that the proposed DBSCAN algorithm improved the clustering of signal points and noise points (Table 4). The classification effect is effectively improved. The improved

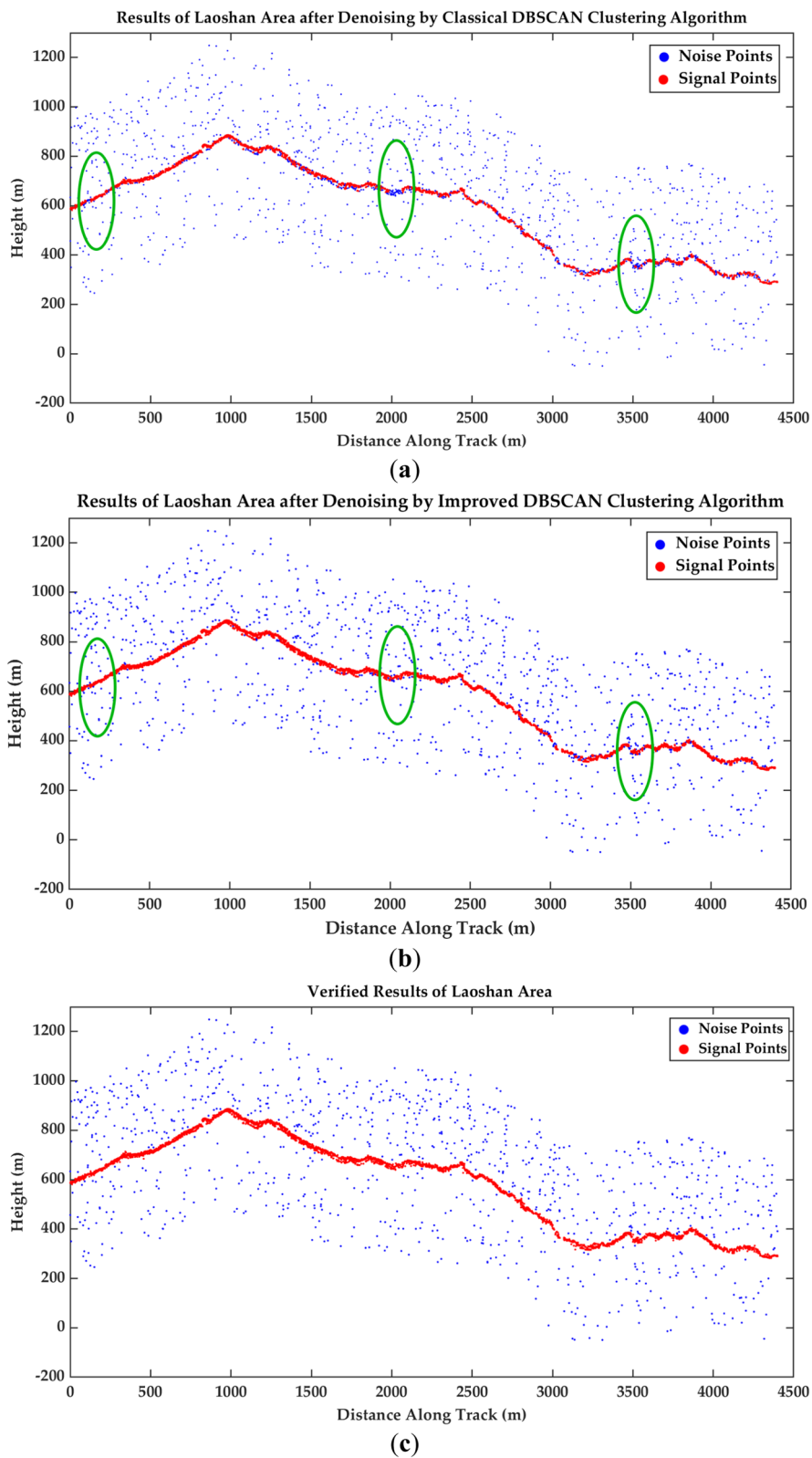


Fig. 9 Comparison of denoising results in the Laoshan area. **a** Denoising results using classical DBSCAN, **b** denoising results using improved DBSCAN, **c** verification results from visual interpretation

Table 3 Confusion matrix of point cloud denoising results in the Laoshan area

		Number of signal points (3624)	Number of noise points (1477)	Accuracy (%)	Sensitivity (%)
Classical DBSCAN clustering algorithm	Number of signal points (3795)	3399	396	87.83	89.57
	Number of noise points (1306)	225	1081		
Improved DBSCAN clustering algorithm	Number of signal points (3795)	3703	92	94.22	97.58
	Number of noise points (1306)	203	1103		

classification accuracy is 94.89% and the sensitivity is 96.50%.

4 Discussion

In the classical DBSCAN clustering algorithm, the threshold setting is limited by the terrain which is prone to excessive denoising (Zhang et al. 2021). Due to the single threshold setting, the noise points near the ground will be incorrectly recognized as signal points. The improved DBSCAN clustering algorithm solves the single threshold problem. By setting the grouping parameters appropriately and applying the automatic radius search, the parameters' adaptability is improved, thus enhancing the ability to extract signals from the ground. In the experiment of the Daxing'anling area, the difference between the two algorithms is more evident because the experimental area has a large virgin forest coverage. The classical DBSCAN clustering algorithm can only remove more obvious noise points away from the ground, and the vegetation points above the ground are misclassified as ground points. In the experiment of the Laoshan area, because the vegetation in the experimental area is mostly composed of low bushes and shrubs, the classical DBSCAN clustering algorithm has the problem of classifying low vegetation as ground points.. Due to the large amount of data in the Changbaishan area, the improved DBSCAN clustering algorithm can enhance the denoising effect in the area with dense noise points through the adaptive parameters. It has strong adaptability to single photon laser point cloud data similar to ICESat-2 data type, and can quickly and efficiently filter out noise points in point cloud data.

In our study, we found that the height difference parameter h_a in the improved algorithm contributes more to the improvement of the denoising accuracy. Because after setting the height difference parameter grouping, the data can be grouped according to the

terrain characteristics. Also different terrain areas can be avoided to have the same single threshold and result in low denoising accuracy. The distance parameter l_a is set mainly to avoid long distances along the trajectory of the grouping and a large amount of data for a single group in areas with gentle terrain. And a small l_a will result in more groups and lower work efficiency.

As a satellite with a global observation range, ICESat-2 can collect global ground elevation information (Xing et al. 2020). However, the global terrain is complex and changeable. For areas with large undulating terrain, point cloud denoising has low accuracy. In the follow-up study, the slope parameters in the ICESat-2 data should be used to improve the denoising method so that complex terrain can be processed (Hao et al. 2022). The improved DBSCAN clustering algorithm can filter out most of the noise points, but there is still confusion between the near-surface vegetation noise and the ground point signal since these are difficult to denoise via geometry alone. If the point cloud data are forcefully denoised under these conditions, the resulting point cloud data will be sparse, thus affecting the further application.

Our next steps involve conducting a more thorough exploration of signal classification within point cloud data. The goal is to improve the identification of near-ground noise points and ground points, thereby enhancing denoising accuracy. To achieve this, we plan to employ advanced machine learning and deep learning techniques, creating more intelligent and adaptive point cloud classification models (W. Li., et al., 2020). By conducting finer feature extraction and analysis of point cloud data, we aim to enhance the accuracy of identifying various ground features, such as buildings, vegetation, and terrain. Utilizing deep learning algorithms, we will seek to train models on large and diverse datasets to enhance their adaptability and generalization capabilities across different environmental conditions (Meng

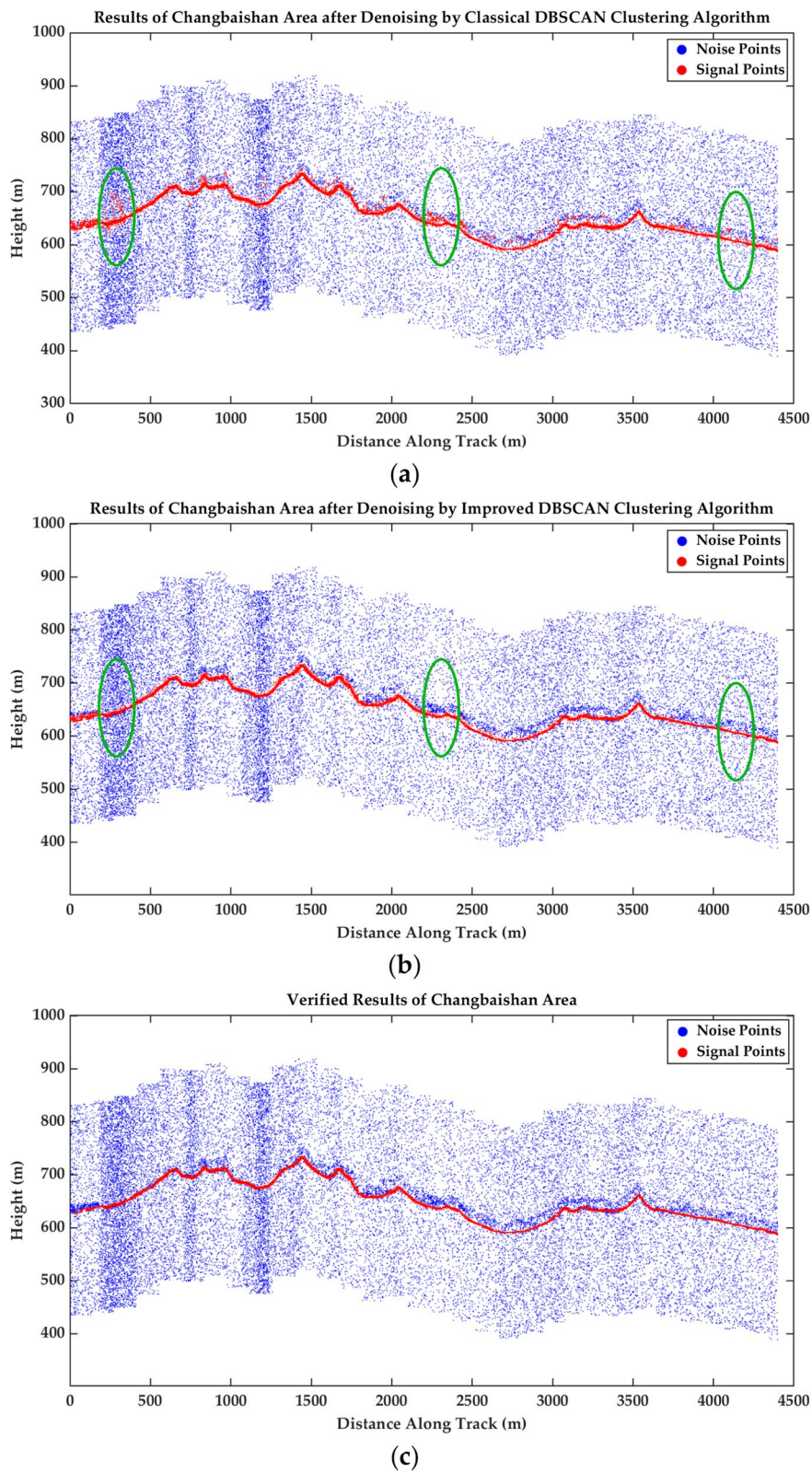


Fig. 10 Comparison of denoising results the Changbaishan area. **a** Denoising results using classical DBSCAN, **b** denoising results using improved DBSCAN, **c** verification results from visual interpretation

Table 4 Confusion matrix of point cloud denoising results in the Changbaishan area

		Number of signal points (27037)	Number of noise points (32891)	Accuracy (%)	Sensitivity (%)
Classical DBSCAN clustering algorithm	Number of signal points (29965)	23387	6578	82.93	86.50
	Number of noise points (29963)	3650	26313		
Improved DBSCAN clustering algorithm	Number of signal points (28206)	26090	2116	94.89	96.50
	Number of noise points (31722)	947	30775		

et al. 2022). Additionally, we will optimize preprocessing steps for point cloud data to better eliminate near-ground noise points. This optimization may involve more sophisticated denoising techniques, consideration of spatiotemporal features, and a more accurate modeling of laser beam propagation paths (Kui et al. 2023). We plan to integrate these optimization methods with the previously proposed improved DBSCAN clustering algorithm, forming a more comprehensive and efficient denoising framework (You et al. 2023). Furthermore, we will focus on addressing challenges in point cloud data processing in complex terrains and densely vegetated areas. Through an in-depth study of point cloud characteristics in these complex scenes, we will design targeted algorithms and strategies to effectively differentiate vegetation, ground surfaces, and other features, thereby enhancing classification accuracy and reliability (Liu et al. 2021). Finally, emphasis will be placed on integrating and comparing the proposed classification and denoising methods with ground validation data. In this paper, the results obtained from visual judgement reading are used as a validation data comparison, which still has some limitations. Through comparison with field observations, we can comprehensively evaluate the algorithm's performance and provide robust guidance for future improvements.

5 Conclusions

The improved DBSCAN algorithm proposed in this paper is aimed at the problem that the manual single threshold selection of DBSCAN clustering algorithm, which causes the algorithm to struggle with difficult terrain changes. The algorithm improves a grouping method based on elevation and distance statistics and an automatic radius search method. The improved algorithm reduces the number of manual attempts, enhances the automation of the denoising framework, and significantly enhances the efficiency of the clustering process. The improved algorithm is used to denoise the data in the experimental area, and good results with great improvements over the classical DBSCAN are achieved. The accuracy of the improved DBSCAN clustering algorithm can reach 95.49%, 94.22% and

94.89% in three experimental areas, respectively. These are 34.76%, 6.39% and 11.96% higher than the classical DBSCAN clustering algorithm, respectively. This indicates that the improved DBSCAN clustering algorithm has higher denoising accuracy. The sensitivity of the improved DBSCAN clustering algorithm can reach more than 96% in those experimental areas, indicating that the method is better for positive classification. For the problem of confusing the signals of near-surface vegetation points and ground points found in the experiments, considering the feasibility of different algorithms under different conditions and the accuracy of the results, various denoising algorithms can be used in combination to denoise various terrains. Multi-source data fusion processing can be considered as part of the denoising framework to improve the classification of point cloud signals, providing more valuable data support for the scientific application of ICESat-2 data.

Abbreviations

ICESat-2	Ice cloud and land elevation satellite-2
NASA	National aeronautics and space administration
ATLAS	Advanced topographic laser altimeter system
ATL03	Advanced topographic laser altimeter system level 03
ATL08	Advanced topographic laser altimeter system level 08
DBSCAN	Density-based spatial clustering of applications with noise
OPTICS	Ordering points to identify the clustering structure
NTEMS	National terrestrial ecosystem monitoring system
Eps	Epsilon
MinPts	Minimum points

Author contributions

Conceptualization, writing—original draft preparation and methodology were performed by WD and YJC. Writing—review and editing were performed by YJC, LFY and LQH; software and validation were performed by YJC. Data curation was provided by LFY. Formal analysis was performed by WD and LQH. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials

Publicly available datasets were analyzed in this study. ICESat-2 data were downloaded from the NASA National Snow and Ice Data Center (NSIDC) (<https://nsidc.org/data/icesat-2>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China. ²National Demonstration Center for Experimental Surveying and Mapping Education (Shandong University of Science and Technology), Qingdao 266590, China.

Received: 21 September 2023 Accepted: 18 September 2024

Published online: 08 October 2024

References

- Chen T, Song C, Luo S, Ke L, Liu K, Zhu J (2022) Monitoring global reservoirs using ICESat-2: assessment on spatial coverage and application potential. *J Hydrol* 604:127257. <https://doi.org/10.1016/j.jhydrol.2021.127257>
- Ester M, Kriegel HP, Sander J, Xu X, Simoudis E, Han J, Fayyad U (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD'96: Proc Second Int Conf Knowledge Disc Data Mining*. <https://doi.org/10.5555/3001460.3001507>
- Hao T, Cui H, Hai G, Qiao G, Li H, He Y, Li R (2022) Impact of Slopes on ICESat-2 elevation accuracy along the CHINARE route in East Antarctica. *IEEE J Sel Top Appl Earth Obs Remote Sens* 15:5636–5643. <https://doi.org/10.1109/JSTARS.2022.3189042>
- Herzfeld UC, Lawson M, Trantow T, Nylen T (2022) Airborne validation of ICESat-2 ATLAS data over crevassed surfaces and other complex glacial environments: Results from experiments of laser altimeter and kinematic GPS data collection from a helicopter over a surging arctic glacier (Negribreen, Svalbard). *Remote Sens* 14(5):1185. <https://doi.org/10.3390/rs14051185>
- Kacimi S, Kwok R (2022) Arctic snow depth, ice thickness, and volume from ICESat-2 and CryoSat-2: 2018–2021. *Geophys Res Lett*. <https://doi.org/10.1029/2021GL097448>
- Kui M, Xu Y, Wang J, Cheng F (2023) Research on the adaptability of typical denoising algorithms based on ICESat-2 data. *Remote Sens* 15(15):3884. <https://doi.org/10.3390/RS15153884>
- Li W, Niu Z, Shang R, Qin Y, Wang L, Chen H (2020) High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with sentinel-1, sentinel-2 and landsat-8 data. *Int J Appl Earth Obs Geoinf* 92:102163. <https://doi.org/10.1016/j.jag.2020.102163>
- Li X, Gao Z, Chen X, Sun S, Liu J (2021) Research on estimation method of geometric features of structured negative obstacle based on single-frame 3D laser point cloud. *Inf* 12(6):235. <https://doi.org/10.3390/info12060235>
- Liu A, Cheng X, Chen Z (2021) Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sens Environ* 264:112571. <https://doi.org/10.1016/j.rse.2021.112571>
- Malambo L, Popescu SC (2021) Assessing the agreement of ICESat-2 terrain and canopy height with airborne lidar over US ecozones. *Remote Sens Environ* 266:112711. <https://doi.org/10.1016/j.rse.2021.112711>
- McGarry JF, Carabajal CC, Saba JL, Reese AR, Holland ST, Palm SP, Swinski JPA, Golder JE, Liiva PM (2021) ICESat-2/ATLAS onboard flight science receiver algorithms: purpose, process, and performance. *Earth Sp Sci*. <https://doi.org/10.1029/2020EA001235>
- Meng W, Li J, Tang Q, Xu W, Dong Z (2022) ICESat-2 laser data denoising algorithm based on a back propagation neural network. *Appl Opt* 61(28):8395–8404. <https://doi.org/10.1364/ao.469584>
- Michaelides RJ, Bryant MB, Siegfried MR, Borsa AA (2021) Quantifying surface-height change over a periglacial environment with ICESat-2 laser altimetry. *Earth Sp Sci*. <https://doi.org/10.1029/2020EA001538>
- Mulverhill C, Coops NC, Hermosilla T, White JC, Wulder MA (2022) Evaluating ICESat-2 for monitoring, modeling, and update of large area forest canopy height products. *Remote Sens Environ* 271:112919. <https://doi.org/10.1016/j.rse.2022.112919>
- Neuenschwander A, Guenther E, White JC, Duncanson L, Montesano P (2020) Validation of ICESat-2 terrain and canopy heights in boreal forests. *Remote Sens Environ* 251:112110. <https://doi.org/10.1016/j.rse.2020.112110>
- Neumann TA, Martino AJ, Markus T, Bae S, Bock MR, Brenner AC, Brunt KM, Cavanaugh J, Fernandes ST, Hancock DW, Harbeck K, Lee J, Kurtz NT, Luers PJ, Luthcke SB, Magruder L, Pennington TA, Ramos-Izquierdo L, Rebold T, Skoog J, Thomas TC (2019) The ice, cloud, and land elevation satellite-2 mission: a global geolocated photon product derived from the advanced topographic laser altimeter system. *Remote Sens Environ*. <https://doi.org/10.1016/j.rse.2019.111325>
- Petty AA, Bagnardi M, Kurtz NT, Tilling R, Fons S, Armitage T, Horvat C, Kwok R (2021) Assessment of ICESat-2 sea ice surface classification with Sentinel-2 imagery: Implications for freeboard and new estimates of lead and floe geometry. *Earth Sp Sci*. <https://doi.org/10.1029/2020EA001491>
- Tian X, Shan J (2021) Comprehensive evaluation of the ICESat-2 ATL08(advanced topographic laser altimeter system level 08) terrain product. *IEEE Trans Geosci Remote Sens* 59(10):8195–8209. <https://doi.org/10.1109/TGRS.2021.3051086>
- Vernimmen R, Hooijer A, Pronk M (2020) New ICESat-2 satellite LiDAR data allow first global lowland DTM suitable for accurate coastal flood risk assessment. *Remote Sens* 12(17):2827. <https://doi.org/10.3390/rs12172827>
- Wang X, Pan Z, Glennie C (2016) A novel noise filtering model for photon-counting laser altimeter data. *IEEE Geosci Remote S* 13(7):947–951. <https://doi.org/10.1109/LGRS.2016.2555308>
- Xie H, Ye D, Hai G, Tong X (2020) An adaptive filter based on local slope by adjusting elliptic filter kernel for ICESat-2 photon counting laser altimetric data. *EPJ Web Conf* 237:01003. <https://doi.org/10.1051/epjconf/20203701003>
- Xie H, Xu Q, Ye D, Jia J, Sun Y, Huang P, Li M, Liu S, Xie F, Hao X, Tong X (2021) A comparison and review of surface detection methods using MBL, MABEL, and ICESat-2 photon-counting laser altimetry data. *IEEE J. Sel Top Appl Earth Obs Remote Sens* 14:7604–7623. <https://doi.org/10.1109/JSTARS.2021.3094195>
- Xing Y, Huang J, Gruen A, Qin L (2020) Assessing the performance of ICESat-2/ATLAS multi-channel photon data for estimating ground topography in forested terrain. *Remote Sens* 12(13):2084. <https://doi.org/10.3390/rs12132084>
- You H, Li Y, Qin Z, Lei P, Chen J, Shi X (2023) Research on multilevel filtering algorithm used for denoising strong and weak beams of daytime photon cloud data with high background noise. *Remote Sens* 15(17):4260. <https://doi.org/10.3390/RS15174260>
- Zhang Z, Liu X, Ma Y, Xu N, Zhang W, Li S (2021) Signal photon extraction method for weak beam data of ICESat-2 using information provided by strong beam data in mountainous areas. *Remote Sens* 13(5):863. <https://doi.org/10.3390/rs13050863>
- Zhang G, Xu Q, Xing S, Li P, Zhang X, Wang D, Dai M (2022) A noise-removal algorithm without input parameters based on quadtree isolation for photon counting LiDAR. *IEEE Geosci Remote Sens Lett* 19:1–5. <https://doi.org/10.1109/LGRS.2021.3081721>
- Zhu X, Nie S, Wang C, Xi X, Hu Z (2018) A ground elevation and vegetation height retrieval algorithm using micro-pulse photon-counting lidar data. *Remote Sens* 10:1962. <https://doi.org/10.3390/rs10121962>
- Zhu X, Wang C, Xi X, Nie S, Yang X, Li D (2020) Research progress of ICESat-2/ATLAS data processing and applications. *Infrared Laser Eng* 49(11):20200259. <https://doi.org/10.3788/IRLA20200259>
- Zhu X, Nie S, Wang C, Xi X, Wang J, Li D, Zhou H (2021) A noise removal algorithm based on OPTICS for photon-counting LiDAR data. *IEEE Geosci Remote Sens Lett* 18:1471–1475. <https://doi.org/10.1109/LGRS.2020.3003191>
- Zhu J, Yang PF, Li Y, Xie YZ, Fu HQ (2022) Accuracy assessment of ICESat-2 ATL08 terrain estimates: a case study in Spain. *J Cent South Univ* 29:226–238. <https://doi.org/10.1007/s11771-022-4896-x>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.