A Comparison and Review of Surface Detection Methods Using MBL, MABEL, and ICESat-2 Photon-Counting Laser Altimetry Data

Huan Xie, *Senior Member, IEEE*, Qi Xu, Dan Ye, Jianghao Jia, Yuan Sun, Peiqi Huang, Ming Li, Shijie Liu, *Member, IEEE*, Feng Xie, Xiaolong Hao, and Xiaohua Tong, *Senior Member, IEEE*

***Abstract*—The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) was launched on September 15, 2018. It is the first photon-counting laser altimeter satellite, which is of great significance for the research into laser altimetry. ICESat-2 is, however, highly sensitive and susceptible to environmental influences. In addition to surface returns, a lot of nonsurface photons are found in the data. It is, therefore, necessary to study an effective method to separate the surface signal from background information. In this article, we review the existing surface detection methods for photon point cloud data and select seven methods for comparison. Four sources of photon-counting data were considered in the experiments: The Multiple Altimeter Beam Experimental Lidar (MABEL), the Chinese Multibeam LiDAR, The Advanced Topographic Laser Altimeter System (ICESat-2/ATLAS), and MATLAS (using MABEL data to simulate the expected ATLAS photon point cloud). Four scenarios of land, land ice, sea ice, and ocean were also considered. Each surface detection method was tested in 12 experiments, and the different methods were finally compared by qualitative and quantitative measures. We were, thus, able to establish the advantages and disadvantages of each method, which will be of great significance for scholars studying surface detection methods.**

***Index Terms*—Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2), laser altimeter, MABEL, photon-counting, surface detection.**

Manuscript received May 26, 2021; revised June 15, 2021; accepted June 20, 2021. Date of publication July 1, 2021; date of current version August 11, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 41822106 and 41631178, in part by the National Key Research and Development Program of China under Grant 2018YFB0505400, in part by the National High Resolution Ground Observation System of China under Grant 11-Y20A12-9001-17/18, in part by the Dawn Scholar of Shanghai Program under Grant 18SG22, in part by the State Key Laboratory of Disaster Reduction in Civil Engineering under Grant SLDRCE19-B-35, and in part by the Fundamental Research Funds for the Central Universities of China. (*Corresponding author: Huan Xie*.)

Huan Xie is with the College of Surveying and Geo-Informatics and the State Key Laboratory of Disaster Reduction in Civil Engineering, and also the Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China (e-mail: huanxie@tongji.edu.cn).

Qi Xu, Dan Ye, Jianghao Jia, Yuan Sun, Peiqi Huang, Shijie Liu, and Xiaohua Tong are with the Shanghai Key Laboratory of Space Mapping and Remote Sensing for Planetary Exploration, College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: smxj_xq@tongji.edu.cn; yedan516@tongji.edu.cn; 1751985@tongji.edu.cn; sunyuan@tongji.edu.cn; 1750693@tongji.edu.cn; liusjtj@tongji.edu.cn; xhtong@tongji.edu.cn).

Ming Li and Feng Xie are with the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Beijing 100049, China (e-mail: lpfilter@163.com; xf@mail.sitp.ac.cn).

Xiaolong Hao is with the Beijing Tracking and Communication Technology Research Institute, Beijing 100094, China (e-mail: xlong_hao@163.com).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2021.3094195

I. Introduction

LASER altimetry is one of the frontiers and core technologies for spatial information acquisition in earth observation and deep space exploration [1]. As a result, laser altimetry data are widely used in surveying, meteorological, geological, environmental, and other fields, in applications such as polar ice sheet measurement [2] and change detection [3], [4], topographic mapping [5], [6], biomass estimation [7], disaster monitoring and assessment [8], etc. Most of the existing earth observation laser altimeters, such as the Ice, Cloud and Land Elevation Satellite /the Geoscience Laser Altimeter System (ICESat/GLAS), the ZiYuan-302 (ZY-302) laser altimeter, and the Gaofen 7 (GF-7) laser altimeter, use linear detection systems with high-energy laser consumption. However, such laser altimeters have a limited repetition frequency when collecting large-scale and multitemporal three-dimensional (3-D) data on space platforms such as satellites, resulting in low data density [9]. In the past few years, quantum information technology has developed rapidly, and photon detection is now one of the key technologies [10]. The essence of photon detection is high-sensitivity probability detection. It can detect photons in harsh environments and complex target characteristics with a loss of resolution and a decrease in ground resolution. The combination of photon detection technology and laser radar technology has brought laser radar technology to a new level. This kind of LiDAR technology is also highly reliable [11], [12]. Photon detection reduces the energy consumption and the mass of the loading of laser radar, and can effectively reduce the weight of satellite loads [13]. At the same time, based on the photon-counting detection mode, the low energy consumption also makes LiDAR multibeam detection easier to achieve [14]. Therefore, photon-counting LiDAR (PCL) technology is likely to be the major approach used to carry out satellite laser altimetry and 3-D imaging in the future, and will be the main way to resolve the contradiction between LiDAR energy consumption and acquisition frequency. The ICESat-2 satellite was launched in September 2018, and is the first earth observation satellite using photon-counting technology. The LiDAR Surface Topography

mission, the Active Sensing of CO_2 Emissions over Nights, Days, and Seasons mission, and the Aerosols-Cloud-Ecosystem (ACE) mission are all planning to adopt photon-counting technology when launched [15]. According to the national civil space infrastructure medium and long-term development plan (2015–2025), China is also planning a photon-counting laser altimeter satellite [16], [17], and an airborne system has been implemented by the Shanghai Institute of Technical Physics of the Chinese Academy of Sciences with 51 beams.

With the photon-counting and multibeam mode, the sampling density of the laser footprint is greatly improved, but PCL data still have the problem that the data density in the transverse direction is far lower than in the along-track direction. PCL systems have high sensitivity and can detect weak signals, but they are easily affected by atmospheric scattering, solar radiation, and the instrument's own dark count. As a result, PCL data contain large numbers of nonsurface photons, which are widely distributed both in the air and beneath the surface, which is completely different from the large-footprint full-waveform data obtained by the traditional spaceborne laser altimetry systems, and the 3-D laser point cloud data acquired by ground and airborne LiDAR. It is therefore necessary to identify the surface photons for PCL data, to provide accurate terrain information for the subsequent scientific research and application.

In this article, we focus on the surface detection methods for PCL data. Interested readers can find information regarding the basic physical and mathematical relations of the different surface detection methods. We begin by presenting a brief history of the development of PCL, without repeating the basic physical principles behind the technology (which are discussed in detail in [14]), and the difference between PCL and large-footprint full-waveform laser altimetry data. In the following sections, we review the existing surface detection methods for photon-counting laser altimetry data. We then select several typical methods and compare their performances. The different methods were adopted to process the same PCL data. We chose data from four different sources: the Multiple Altimeter Beam Experimental LiDAR (MABEL) instrument, the ICESat-2 Advanced Topographic Laser Altimeter System (ATLAS), the Chinese Multibeam LiDAR (MBL) system, and MATLAS (using MABEL data to simulate the expected ATLAS photon point cloud). Finally, we provided a qualitative and quantitative evaluation of the experimental results.

II. Photon-Counting Lidar Data

In order to verify the advantages and disadvantages of the respective algorithms, we selected 12 datasets from four sources and different scenarios: three MABEL datasets (land ice, ocean, and land), seven ICESat-2/ATLAS datasets (land ice-flat, land ice-uneven, ocean data, land/low vegetation, land/high vegetation, land/vegetation-uneven, sea ice), one MBL dataset (land/urban area), and one MATLAS dataset (land/vegetation). The data we selected are very typical, basically covering data of different sensor types and various characteristics (including day and night, land area with vegetation, land area without vegetation, urban area, ocean area, sea ice area, land ice area, etc.). The specific selection of data is shown in Table I. The profiles of each type of experimental data are shown in Fig. 1, and the location distributions are shown in Fig. 2.

A. Multiple Altimeter Beam Experimental Lidar

The MABEL instrument is an airborne PCL system, which is used as a demonstrator and validation tool for the ICESat-2 photon-counting altimetry concept. The MABEL instrument is carried by the high-altitude ER-2 aircraft, which fly at an altitude of 20 km. The pulse repetition frequency of the MABEL instrument varies from 5 kHz to 20 kHz. The pulse is transmitted every ∼4 cm along track and produce 2 m footprint with 5 kHz repetition rate [18]. The MABEL instrument uses both green (532 nm) and near-infrared (NIR, 1064 nm) laser wavelengths to detect the surface. There are 24 beams on MABEL, containing eight near-infrared beams and 16 green beams [19].

The first demonstration flights of the MABEL instrument took place in December 2010, and the flight data were published on the ICESat-2 website (https://icesat-2.gsfc.nasa.gov/legacy-data/mabel/data/browse/index.html). In order to verify the performance of ICESat-2 on different surface types, the MABEL data include data on different surface types such as land, ocean, sea ice, land ice, and inland water. In this article, MABEL data of three different land-cover types are selected: 1) land ice data with fluctuations; 2) ocean data with slight fluctuations in the sea surface; and 3) land data covered by vegetation with uneven background photons and large terrain fluctuations. These data correspond to data Nos. 1–3 in Table I.

B. ICESat-2 ATLAS/ATL03

ICESat-2 operates at a speed of 7 km/s at a height of ∼500 km, with an inclination of 92°. Each orbit is referred to as a reference ground track (RGT), and there are 1387 RGTs in one cycle. It takes 91 days for the satellite to complete a cycle.

ATLAS is the only instrument carried by ICESat-2. It is a photon-counting detector with a 10-kHz 532-nm laser that is used to obtain high-resolution altimetry measurements. Based on the velocity of the satellite and its repetition rate, one laser pulse is transmitted every 0.7 m along the ground tracks. ATLAS has a 17-m diameter footprint and six beams, which are divided into three groups. Each group contains a weak beam and a strong beam; for instance, both GT1L and GT1R belong to group one. The distribution of strong and weak beams is not absolute, and it depends on the direction of flight of the satellite. When ATLAS is oriented forward, GTxL (where x represents the number of the group) is in front of GTxR by a distance of 2.5 km, and the GTxL beams are always the weak beams. In contrast, when ATLAS moves in the negative direction, the GTxR beams are ahead and become the weak beams. The centers of the two adjacent groups are 3.3 km apart [20], [21].

ICESat-2 product data release is handled by the National Snow and Ice Data Center, and the entire data system includes three levels and 21 data products (ATL00–ATL21). ATL01 and ATL02 are the Level 1 products, which contain raw data. ATL03 and ATL04 belong to the Level 2 products, and ATL06–ATL21 belong to the Level 3 products. Based on ATL02, Precise Orbit

TABLE I
EXPERIMENTAL DATA

Data source	Number	Type	Name	Beam	Time
MABEL	1	Land ice	mabel_l2_20140718t015400_010_1.h5	50	all
	2	Ocean	mabel_l2_20120412t105300_010_1.h5	50	all
	3	Land	mabel_l2_20140802t003000_010_1.h5	50	all
ICESat-2/ ATLAS (ATL03)	4	Land ice	ATL03_20181021200424_03540110_001_01.h5	1r	400-402 s
	5	Land ice	ATL03_20181109172947_06430103_001_01.h5	1r	202-204 s
	6	Ocean	ATL03_20190430100235_04910304_001_01.h5	1l	55-57 s
	7	Land	ATL03_20181124001516_08610105_002_01.h5	1l	277-279 s
	8	Land	ATL03_20190201032619_05300202_001_01.h5	1r	450-452 s
	9	Land	ATL03_20190205031800_05910202_001_01.h5	1l	460-462 s
	10	Sea ice	ATL03_20190412130642_02180305_002_01.h5	1l	20-22 s
MBL	11	Land		26	all
MATLAS	12	Land	matlas_l2_20120920t231600_010_2_8B_1VEG_02.h5		all

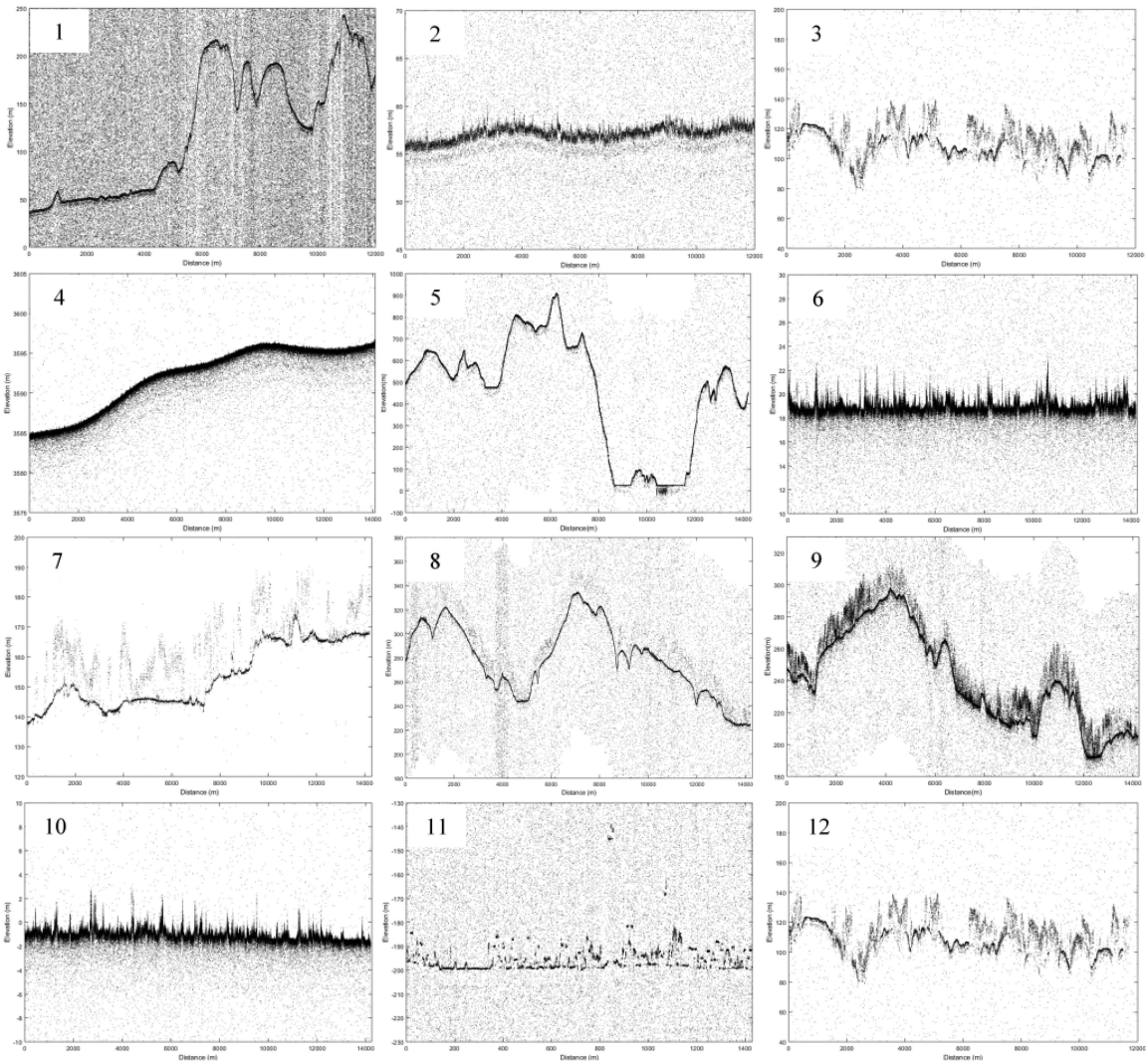

Fig. 1. Experimental data profiles (X: orbital distance, Y: elevation).

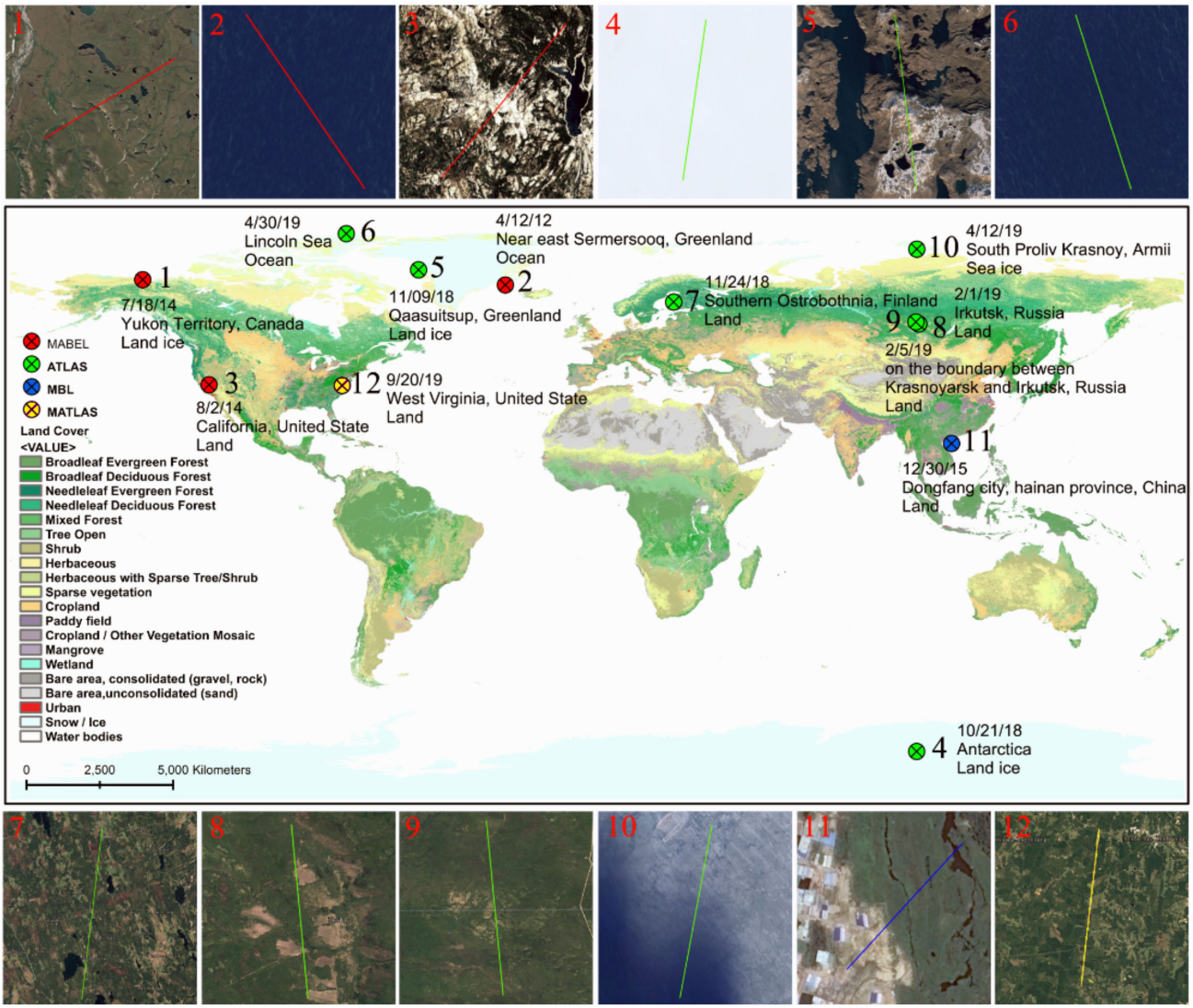

Fig. 2. Data distribution map, with the dots indicating the data locations. The MABEL (red), ATLAS03 (green), MBL (blue), and MATLAS (yellow) data are labeled with the time, location, and type of data. The base map is the Global Land Cover by National Mapping Organizations map with a 15-s resolution.

Determination, Precision Pointing Determination, and other ancillary data, ATL03 generates the data of each photon's time, latitude, longitude, and height, which is significant for the higher-level products. The ATL03 product is described as a bridge between the lower-level products and the higher-level products. In addition, based on other data resources, ATL03 classifies each photon into five surface types (land, ocean, sea ice, land ice, inland water) and provides a preliminary assessment of each photon. In addition, for each photon event, ATL03 provides corrected heights for multiple geophysical corrections (i.e., geoid, ocean tides, dynamic atmospheric correction) and all the instrument information for the higher-level data algorithms.

In this article, seven ATL03 datasets were used in the experiments to verify the performance of the seven surface detection algorithms. The datasets comprised two land ice datasets, one ocean dataset, three vegetation-covered land dataset, and one sea ice dataset, which correspond to Nos. 4–10 in Table I. Among the different datasets, the first land ice dataset was made up of data with less undulation, and the second dataset was made up of undulating data, and included a segment of sea. For the ocean dataset, this was almost flat on the sea surface. There were large differences in the three terrestrial datasets. The first land dataset was covered by a small amount of vegetation; the second dataset contained background photons that were unevenly distributed, with less vegetation coverage; and the third dataset contained many background photons, which were distributed relatively evenly, with dense vegetation on the surface.

C. *Chinese Multibeam LiDAR*

MBL system was developed by the Shanghai Institute of Technical Physics of the Chinese Academy of Sciences and is the first multiple-beam photon-counting LiDAR system of China. The airborne test of the MBL system was carried out in December 2015 in Hainan province, China. The test flight speed was 200 km/h. The MBL system collects data using the

1064 nm wavelength with 51 beams, and the maximum detection distance across the orbit is 3 km. Each beam transmits pulses at 1.5 μJ, with a 3-kHz repetition rate. The data are presented in geographic coordinates of each photon point [22]. In this article, we used the MBL data of the middle beam (channel 26). The data were obtained in a residential area, with some undulating houses on the surface.

D. MATLAS

MATLAS data simulate the expected performance of the ICESat-2 ATLAS instrument. The simulation is undertaken as per the ATLAS 10-kHz laser fire, with a 0.7 m sample spacing. MATLAS signal photons are subsampled from MABEL signal photons. The numbers of signal and noise photons are then adjusted according to the predicted using the ATLAS instrument model design cases [23]. MATLAS data were used for the quantitative evaluation as this dataset is the only dataset, which has the correct surface type for each photon. In this article, land data with vegetation cover were selected to quantitatively describe the performance of the surface detection algorithms.

III. Comprehensive Review of the Surface Detection Methods for Photon-Counting LiDAR Data

Surface detection is the first step and the key step to accurately extracting surface information for photon-counting LiDAR in many applications. At present, the surface detection algorithms for photon point cloud data are basically based on the distribution of the photons. Due to the influence of the environment and the instrument itself, there are many background photons in both the air and beneath the surface, and they are randomly distributed in space. The density of background photons is smaller than that of surface photons [24]. Many scholars have researched photon detection algorithms and proposed many excellent algorithms. We studied these and found that the algorithms have similar steps.

1) *Preprocessing:* Determine the signal range based on the original data or other ways to reduce the processing amount of photons.
2) *Surface detection:* The core step for algorithms. The purpose of signal detection is accomplished in this step. There are many ways to achieve the goal in different algorithms. According our summary, there are three main ways of surface detection: 1) clustering; 2) feature calculation and separation; and 3) waveform statistics and processing. Different ways cause different results for different surface types.
3) *Subsequent processing*. For the better results, part of methods carried out subsequent processing, for example, Iterative processing, histogram statistics, and so on.

The specific processing flows are summarized and shown in Fig. 3. In the following sections, we analyze the existing methods introduced previously in detail.

A. Preprocessing

Photon detectors placed on satellites usually only record photons within a certain range gate. However, the data still contain a large number of background photons, which need to be removed before surface detection can take place. The main purpose of the preprocessing is to decrease the number of background photons and to reduce the amount of calculation needed in the subsequent steps. Image processing[25], [28], [29], envelope curves [26], and histogram statistics [26], [30]–[33] are the commonly used methods in the preprocessing step.

1) *Image processing*: The first preprocessing method involves rasterizing the profile point cloud along the orbit into a 2-D image according to the density of the photons. The image processing algorithm is then adopted to detect the surface. In the earlier data processing approaches, image processing was not only a method of data preprocessing, but was also used as a method of surface detection. Magruder *et al.* [28] proposed the Canny edge detector to extract surface photons. In this method, the profile data are rasterized into a 2-D (along-track direction and elevation) profile image, where the pixel values correspond to the number of photons within each cell. A Gaussian filter is then applied to broaden the edge of the high-density pixels, and to remove the background pixels. The gradient value and direction are then calculated, and a threshold of the gradient is determined using the resulting statistics. Awadallah *et al*. [29] used a median filter to remove a large amount of background photons, and then used a size filter to remove the remaining isolated regions. Since rasterization reduces the accuracy, and this method cannot filter out the photons among the surface and ground objects, such as the photons among a vegetation canopy, this results in low-accuracy processing results. This method has subsequently been used only for image preprocessing. Li *et al*. [25] segmented the imagery using the image processing technique, retaining the surface photons, and filtering out most of the nonsurface photons.
2) *Envelope curves:* The second preprocessing method relies on ancillary data to preserve possible surface photons. The ancillary datasets are usually a digital elevation model (DEM) or a canopy height map (CHM). In 2018, Popescu *et al*. [26] sought to reduce the range of the possible surface photons that are actually returned from the vegetation canopy and the ground. In this method, the minimum and maximum elevation thresholds are set based on the advanced spaceborne thermal emission and reflection radiometer global DEM, and global canopy height map, and photons within the two thresholds are extracted as possible surface photons.
3) *Histogram statistics:* The last preprocessing method involves dividing the distance/time segments into small elevation bins, and then counting the number of photons within each bin. Photons in a certain range above and below the bin containing the maximum number of photons, or photons in bins whose number are greater than several standard deviations, are considered possible surface photons. Moussavi *et al*. [31] established a histogram of the elevation frequency distribution, and calculated the standard deviation (sigma) based on the histogram. The cut-off threshold was set as 3 times sigma. Zhu *et al*. [30] and Nie *et al*. [32] used Moussavi's method to extract

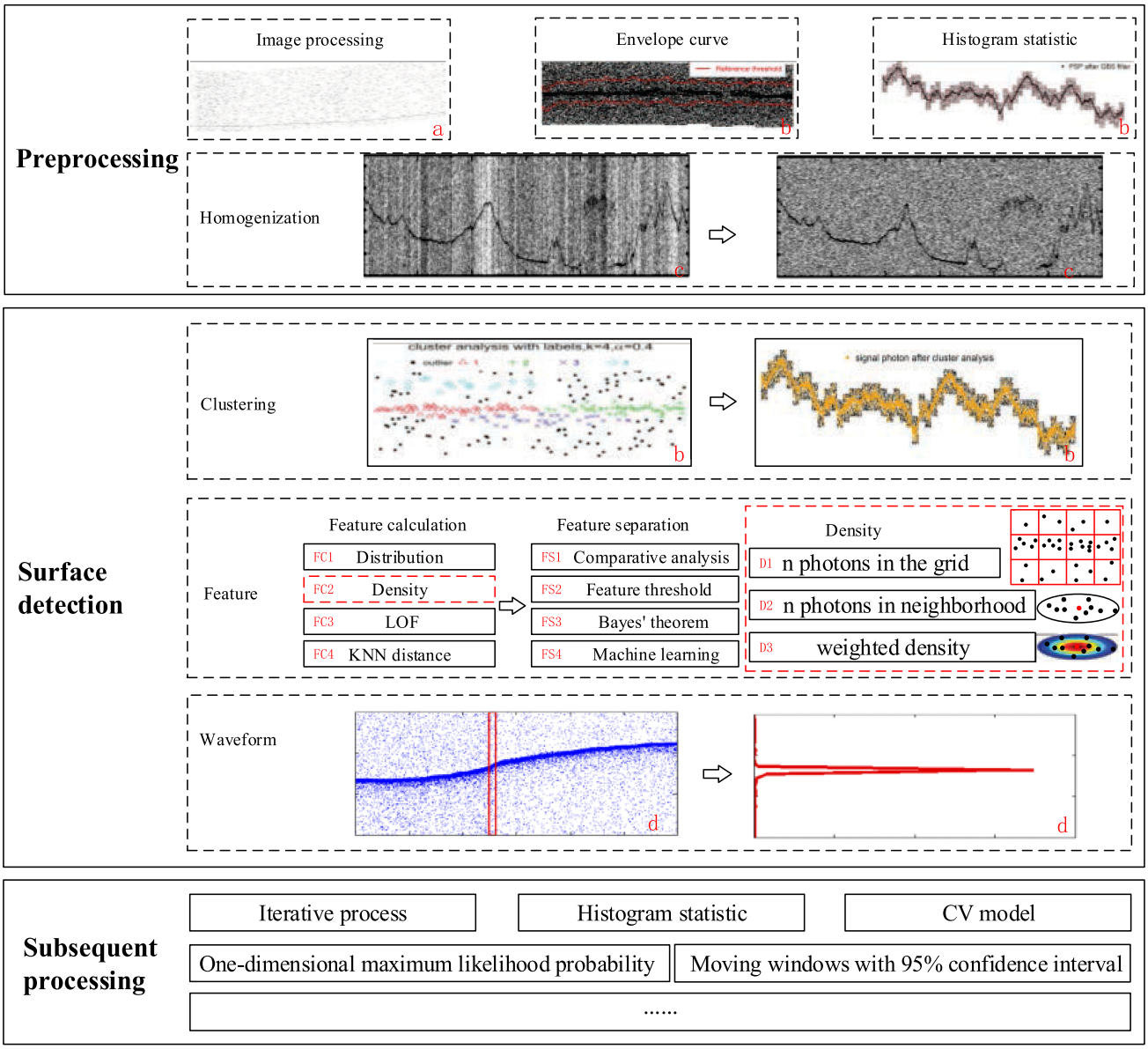

Fig. 3. Classification chart of the existing PCL data surface detection algorithms, according to the processing flow. Image (a) is from Li *et al.* (2017) [25]; image (b) is from Popescu *et al.* (2018) [26]; image (c) is from Wang *et al.* (2016) [24]; and image (d) is from Markus *et al.* (2017) [27].

possible surface photons. Popescu *et al.* [26] also proposed a preprocessing method based on histogram statistics. This method retains the photons within a certain elevation width of the peak of the statistical histogram for each distance segment along the orbit. In some studies, this method has also been directly used to detect the surface, without other subsequent steps. Horan *et al.* [33] divided the photons into along-track distance segments containing the same number of photons, and counted the number of photons in each specific elevation range bin. In this method, the photons within an elevation threshold are considered as surface photons, and the elevation threshold is set to a certain elevation range above and below the elevation corresponding to the maximum number of photons. For different terrains, the width of the bin and threshold must be reset, and according to Horan *et al.* [33], the results are optimal when the along-track distance segments contain 25 photons and the elevation range is 50 cm.

Due to the differences in surface spectral reflectance, atmospheric conditions, and aircraft speed, the density of photons is inconsistent, and the background photons are unevenly distributed in the direction along the track. The inconsistent photon density greatly limits the performance of some surface detection algorithms. Specifically, background photons may be incorrectly classified as surface photons, while surface photons may not be correctly detected. Therefore, to address this issue, Wang *et al.* [24] homogenized the background photons along the orbital direction according to the 0.02 s instantaneous noise_rate (channelxxx/altimetry/noise_rate) provided in the MABEL HDF5 data file.

B. Surface Detection

There are three main ways to achieve surface detection. The first is distinguishing surface photons and background photons by clustering. The second method is waveform statistics. This method involves converting PCL data into a waveform by statistics, and the surface is then detected by a waveform processing method. The third method involves calculating the feature values and then separating the background and surface photons based on their feature differences.

1) *Clustering:* The first surface detection method can be used to measure the similarity between different data sources and to classify the data sources into different clusters. The clustering method can also be used to detect the surface of PCL data. Popescu *et al*. [26] applied the R package named tclust to eliminate a certain proportion of photons. The principle of the method is to limit the covariance matrix to achieve maximum likelihood estimation in different ways, and the data are clustered into k classes and out-of-group photons. Narine *et al*. [34], [35] adopted this method for surface detection.
2) *Waveform statistics:* The second surface detection method is based on the principle of the photon point cloud can be converted into a waveform by histogram statistics. This method involves processing the photon point cloud data using full-waveform processing, which allows us to quantitatively estimate the surface and canopy heights. The photons are merged to a waveform in a certain length of time/distance along the track, and then the height is directly acquired according to waveform processing technology [15], [27], [36]–[38].
3) *Feature calculation:* The third surface detection method is a key step in surface detection. There are four main features that have been calculated in the existing studies, namely, the data distribution feature [28], [39], the photon density [25], [30]–[32], [40]–[49], the k-nearest neighbor distance [24], [50], and the local outlier factor [51], [52]. Among the different features, the photon density is the most commonly adopted feature.

Step 1: Feature value calculation of each photon

① *Distribution feature:* The distributions of the surface photons and background photons of PCL data are usually different. The surface photons are typically more closely distributed, with the background photons being randomly distributed in space. The surface photons can, thus, be separated from the background photons based on the distribution differences. Magruder *et al*. [28] selected some of the background photons, usually more than 100 m above or below the surface. They then calculated the average distance of each photon to its neighborhood, and established a background photon distribution. Gwenzi *et al*. [39] assumed that the background photons obey a Poisson distribution. In this method, the expected number of background photons is determined according to the average background statistics of the bins and the Poisson distribution. If the number of photons in a bin exceeds the expected value, the excess amount is removed by random sampling, and the remaining photons are classified as background photons.

② *Density feature:* The surface photons of PCL data are typically denser than the background photons. Separating surface photons from background photons based on density is, therefore, an effective method. There are several existing methods [53], [54] used for calculating the density, which involve calculating the number of photons in a grid and then calculating the number of photons in its neighborhood. The neighborhood can be a circle or ellipse centered on the photon. Alternatively, the weighted density is calculated in its neighborhood for each photon.

1) Number of photons in a grid. Brunt *et al*. [40] detected surface photons through a multiple statistical method. In this method, the data are first divided into bins with a width and height of 25 m and 10 m, and the mean and standard deviation of all the bins are calculated. Photons in the bins with a photon number greater than the mean plus variance are considered to be potential surface photons, and the standard deviation is recalculated for the remaining bins. Photons in the bins with more than 15 times the standard deviation are also considered as potential surface photons, and photons within ± 10 m of potential surface photons in the elevation direction are also considered as surface photons. Finally, photons within 0.5 m above or below the median height of these surface photons are defined as surface photons. Chen *et al*. [43], Glenn *et al*. [44], and Farrell *et al.* [45] all used Brunt's method to detect the surface for PCL data. *Tang et al*. [46] considered the number of photons in a 3-D cell as the density. Ma *et al*. [47] proposed an ocean surface detection method based on the Joint North Sea Wave Project (JONSWAP) spectra and Levenberg–Marquardt (LM) nonlinear least-squares fitting. The JONSWAP wave spectra are widely used to express surface profiles. In this method, the histogram of the number of photons in each 20 m vertical segment is calculated, and the segments are reserved, which have 1.2 times the density of the average density of all the segments. The sea level is fitted every 500 m segment using the LM method, and the initial parameters can be calculated based on the JONSWAP spectra.
2) Number of photons in a circle or ellipse. Zhang *et al*. [48], [49] modified the density based spatial clustering of applications with noise (DBSCAN) algorithm. Since the surface photons are linearly and tightly distributed in the direction along the orbit, the shape of the search area in DBSCAN is changed from a circle to an ellipse, and the density is the number of photons in the elliptical area centered on the photon. Photons in the horizontal direction have a greater weight relative to the center of the search area than photons in the vertical direction. Therefore, continuous photons in the horizontal direction are more likely to be classified as surface photons. Moussavi *et al*. [31] extracted photons within a 40-m elevation buffer of the inflection point of the statistical histogram, and calculated the probability matrix of each photon, which is also called the normalized density, and is the number of photons in a rectangular region divided by the maximum density of the

data. Nie *et al*. [32] defined density as being the number of photons in the neighborhood of each photon, where the neighborhood is an elliptical area centered on the photon. In high relief areas of terrain, the density is calculated by the horizontal elliptical nucleus, which does not reflect the density well. According to the direction of the slope of the terrain, more ground photons can be included in the elliptical nucleus. The method of calculating the density of the elliptical nucleus in the same direction as the slope can better reflect the density. Zhu *et al*. [30] counted the number of photons in the elliptical filter kernel, where the direction of the elliptical kernel changes according to the direction of the ground. This method of changing the direction of the elliptical nucleus was first proposed by Xie *et al*. [41].

3) Since the surface photons in PCL data are more closely distributed in the orbital direction than the vertical direction, many algorithms give different weights to each photon in the neighborhood, and the weight usually differs in the along-track direction and the transverse direction. Li *et al*. [25] used the anisotropic norm to calculate the density value of the remaining photons after extracting the coarse surface using an image processing technique. In this method, histogram statistics are calculated on the density values, and a low-pass filter is applied, which is implemented by a five-point moving average algorithm. Xie *et al.* [41] considered the slope factor and changed the direction of the elliptical filter kernel to be consistent with the slope. In this way, for an inclined surface, it is possible to obtain better density results. Sparse data on the slope surface can, therefore, achieve a greater density, which is beneficial to the separation of background photons. By traversing the various directions, it is possible to acquire the direction with the maximum density, which is the direction of the best filter kernel. Photons in the filter kernel are given a linear weight in time and are Gaussian in height, and the maximum density is considered as the final density. Herzfeld *et al*. [42] calculated the density weight using the radial basis function for an ice surface. Since the ice surface extends more in the horizontal direction than the vertical direction, the anisotropic norm is combined to make the horizontal direction weight greater than the vertical direction weight.

③ *K-nearest neighbor distance feature*: The *k*-nearest neighbor distance is another way to judge discrete levels. It can be separated according to the discrete differences between surface photons and background photons. That is, the *k*-nearest neighbor distance of the surface photons is smaller than that of the background photons. Xia *et al*. [50] proposed a method based on local distance statistics, where they calculated the mean distance of each photon to its *k*-nearest photons and performed histogram statistics on the calculated results. Wang *et al*. [24] obtained the probability distribution function of the distance between each point to its *k*-nearest neighbor, and then, according to Bayesian decision theory, calculated the probability of each point being a background or surface photon.

④ *Local outlier factor (LOF) feature*: the LOF is an unsupervised outlier detection method that computes a score for a photon, which indicates the local density around the given photon to its nearest neighbors. Chen *et al*. [51], [52] utilized the unbalanced distribution using range search and a multiwindow size histogram filter to distinguish the surface and background photons. The LOF value is calculated using the ellipse search area for the photons.

Chen *et al*. [55] calculated 12 features, including the *k*-nearest neighbor distance, the difference between each photon and the height quantile of all the photons in a certain size window. The feature selection technique was then applied by sorting the importance of the variables using a random forest algorithm. The three most representative features were then selected based on the importance level of the variables. Finally, a random forest classifier was established.

Step 2: Surface detection by separating features

After calculating the features, it is necessary to separate the surface photons and the background photons, according to the feature differences. There are different feature separation methods for different features. For the distribution features, the separation is usually performed by comparing the difference between the distributions [28], [39]. For the density, LOF, and *k*-nearest neighbor distances, a threshold for the features is usually set [25], [30]–[32], [40]–[42], [46]–[48], [50]–[52]. In the case of prior knowledge, Bayesian theory can also be used for the separation [24]. Some other methods achieve separation by machine learning [55].

① *Comparative analysis:* In 2012, after establishing the background photon distribution, Magruder *et al*. [28] divided all the data into smaller cells, which were processed with the same method, and they compared the result with the background photon distribution. The photons were then gradually removed until the probability distribution function completely matched the background photons, and all the removed photons were surface photons. Gwenzi *et al*. [39] removed the photons beyond the expected number of background photons by random sampling, and the remaining photons were considered as background photons.

② *Feature threshold:* Brunt *et al*. [40] applied multiple statistical methods, and as mentioned earlier, multiple fixed thresholds were set. Moussavi *et al*. [31] set the lowermost high density as the threshold, and the spline curve was used to fit the lowest density photons. The photons within one to three standard deviations were considered as background photons, and were removed. Xia *et al*. [50] set the mean plus *t* times standard deviation according to the *k*-nearest neighbor distance statistical result as the threshold. Zhang *et al*. [48] estimated the number of photons in the elliptical kernel of the surface photons based on the average density. If the average density was greater than this estimated value, the photon in the center of the ellipse was classified as a surface photon. Tang *et al*. [46] set the threshold to the average number of photons of 27 individual elements around the volume element. Li *et al*. [25] set the density corresponding to the maximum value of the histogram as a threshold, and filtered out photons with

a smaller density. They then recalculated the density for the remaining photons, and filtered out the high-density background photons that may exist. Xie *et al*. [41] removed the background photons that were far away from the ground according to the density threshold *T*. Due to the ground signal photons being closer to the density center of the photon cloud, the maximum density difference between each photon and its neighboring area is then calculated. If the difference is greater than 3*T*, it is considered to be a background photon; otherwise, it is considered to be a surface photon. Herzfeld *et al*. [42] calculated two adaptive density thresholds. The first threshold was the maximum density of the background photon area plus the offset value, and the second threshold was the percentile that met the first threshold. Nie *et al*. [32] set the density threshold based on a density frequency histogram, and fine surface detection was achieved. Zhu *et al*. [30] improved the method of determining density thresholds. They considered that there are two peaks in the density distribution histogram, with the smaller being the peak of the background photons, and the larger being the peak of the ground signal photons. The Gaussian curves are fitted according to the two peaks, and the intersection of the two Gaussian curves is considered to be the density threshold. Ma *et al*. [47] calculated the fit error and the root-mean-square error. When the absolute fit error of the remaining photons is greater than twice the fit error, these residual photons are classified as background photons and discarded. Subsequently, an additional fitting process is used to further discard the background photons. Chen *et al*. [51], [52] calculated the histogram of the LOF value. In this method, the threshold is set as twice the difference between the LOF minimum and the peak value. Photons less than the threshold are considered to be surface photons, and the other are considered to be background photons.

③ *Bayes' theorem:* Wang *et al*. [24] obtained the probability distribution function of the distance between each point to its *k*-nearest neighbors, and then, according to Bayesian decision theory, calculated the posterior probability of each point being background or surface photon. In this method, if the surface photon probability is greater than the background probability, it is considered to be a surface photon; otherwise, it is considered to be a background photon.

④ *Machine learning:* Chen *et al*. [55] applied a machine learning method to classify photons after selecting a few important features using a random forest classifier, and the established model was then extended to the whole study area.

C. Subsequent Processing

After the abovementioned processing steps, some background photons near the surface will still be classified as surface, which will affect the data accuracy. For example, there can be a large error in extracting the photons at the top of the canopy and surface photons. In general, there are several methods for addressing the problem, which are listed as follows.

1) *Confidence interval filter:* Popescu *et al*. [3] selected a series of 95% confidence interval filters within each horizontal interval window to keep photons within the 95% confidence interval until the relative change of the total number of signal points between two adjacent iterations is $<5\%$.
2) *Chan–Vese model:* Chen *et al*. [43] used the Chan–Vese (CV) segmentation algorithm [56] to obtain potential surface photons after using Brunt's method to obtain the initial surface contour.
3) *Maximum likelihood probability filter:* In order to further reduce more background points, Farrell *et al*. [45] adopted a 1-D maximum likelihood probability filter along the track, after adopting Brunt's method.
4) *Iterative process:* To eliminate the outliers and reduce the effect of misclassified noise photons, Moussavi *et al.* [31] used an iterative process based on cubic spline interpolation. In this method, if the photons are within three standard deviations of the spline fit, they are classified as surface return; otherwise, they are removed. This process is repeated three times to ensure that as many background photons are removed as possible, while still retaining enough data.
5) *Histogram statistic:* Nie *et al*. [32] established a localized elevation frequency histogram. In this method, a localized elevation frequency histogram is first constructed by setting the elevation bin size and the distance bin size. The elevation threshold is then set according to the local elevation frequency histogram to further eliminate isolated background photons below the surface and above the canopy.

IV. Results and Discussions

In the experiments, we used seven methods (see Table II) to perform surface detection and processing on 12 PCL datasets. The 12 PCL datasets cover four sensor types (MABEL, ATLAS, MATLAS, and MBL) and almost all surface types. The methods we selected are fundamental methods, other methods are improved from those. We reckoned that these methods could reflect their processing performance. Intuitively, the seven methods can all extract the surface photons, but there are still differences between the methods. At the end of this section, we will discuss variations or improving versions of these methods.

In the following, we introduce seven algorithms and analyze the effect of each method. In Section IV-A, we describe the flow diagrams of algorithm. In Section IV-B, we analyze the processing effects of the different methods on different surface types and data sources, In Section IV-C, we analyze the advantages and disadvantages of each method. Table II summarizes the steps of the seven methods. And details of the quantitative evaluation of different methods are provided in Sections III–VI. The detailed results are provided in the supplementary file.

A. Method Description

In this section, we describe the seven typical algorithms in detail. In addition, we summarize the algorithm and show it in Table II

1) *Histogram based method [40]* (*HBM*): The method is based on the histograms of the photon data to distinguish

TABLE II
BRIEF INTRODUCTION TO THE MAIN STEPS OF THE SEVEN METHODS ACCORDING TO THE CLASSIFICATION SCHEME DESCRIBED IN THIS ARTICLE

Method	Preprocessing	Feature calculation (FC)	Feature separation (FS)
1. The histogram based method (HBM) [40]		Histogram statistics and calculation of the mean and standard deviation	The first threshold is set Photons are excluded that are more than the threshold The second threshold is set Photons are added that are more than the threshold Photons are supplemented around the selected photons Photons around the median photon are selected as surface photons
2. The Local distance method (LDM) [50]		Calculation of the mean KNN distance Histogram statistics	The threshold is set according to the histogram
3. The modified DBSCAN method (MDBSCANM) [48]		Calculation of the number of photons in the ellipse	The clustering parameters are supplemented The threshold is set according to the parameters
4. The novel filtering model (NFM) [24]	Homogenized density of non-ground photons	Derivation of the distribution of distance to KNN	The probability is calculated according to Bayesian decision theory
5. The adaptive directional filter method (ADFM) [41]		Setting of the weights of the horizontal and vertical directions Calculation of the density in the ellipse in all directions Identification of the maximum density	The first threshold T is set The maximum density difference in a circular neighborhood is calculated The second threshold is set to three times T
6. The density-dimension method (DDM)[42]	Selection of surface and non-ground slab	Calculation of the density using an anisotropic kernel and radial basis function	The first threshold is set according to the density of background photons The photons satisfying the threshold are selected
7. The improved localized statistics method (ILSM) [30]	Histogram statistics and selection of photons in the buffer range	Calculation of the number of photons in the ellipse as the density Identification of the maximum density	The histogram statistics of the density are calculated The threshold is set according to the histogram

rough signal photons and noise photons. Assuming that the background photons are randomly distributed and at the same time the returned pulse is symmetric. First, generate photon data of 0.125 s (∼25 m; 625 shots) and histogram the data at 10-m vertical resolution. Second, calculate the mean and standard deviation of total bin counts and select bins with more counts than the mean plus the standard deviation as potential signal photons. Third, calculate the mean and standard deviation of the remaining photons and add bins with more than 15 times standard deviation of background counts to the potential signal bins. Finally, for the remaining photons, add any photon within ±10 m of the signal photons obtained in steps 2 and 3 as the signal photons. In addition, all photons within ±0.5 m of the median photon elevation are also considered as signal photons.

2) *Local distance method [50]* (*LDM*): In this method, the noise points in the data are eliminated by the local distance statistics method. First, calculate the total distance dist_i from each point to the nearest k points around it in the point cloud. Second, count the frequency histogram of the local total distance. Finally, set threshold to eliminate noise points. Since the spatial distribution of the four research objects in this article is relatively uniform, it can be assumed that the frequency histogram of the local total distance dist_i of the noiseless point cloud is a Gaussian distribution, and its mean and variance are estimated through the local statistics of the point cloud. If dist_i is greater than the sum of its overall mean and t times the standard deviation, then point i is a noise point.
3) *Modified DBSCAN method [48]* (*MDBSCANM*): This approach makes some improvements on DBSCAN clustering model and then modifies the shape of the search area. In the first stage, since the density of most clusters (surface returns) in the horizontal direction is greater than that in the vertical direction and the photon distribution in the horizontal direction is linearly and tightly, it is reasonable to modify the search area from a circle to an ellipse shape. In the second stage, estimate the clustering parameters, including a Eps-neighborhood of a point and the minimum number of points in that Eps-neighborhood. First, extract some sample data from the test dataset and calculate the area. Then, calculate the area and the average point density of the Eps-neighborhood of a point. If the average point density for clusters is higher than the average density of the whole data set, the photon in the center of the ellipse is classified as a surface photon.
4) *Novel filtering model [24]* (*NFM*): Under the assumption that the noise point is isolated and the neighbors of the noise are less than that of the topographic point, the novel filtering model improves the defects of the previous method in dealing with the background photon inhomogeneity. First, because of the inconsistency of noise

density, the effectiveness of density-based noise detection is limited. Therefore, it is necessary to adaptively scale the noise density of the original photon counting point cloud. Second, we apply the method introduced in [1] to derive the distance distribution between a point and its *k*-nearest neighbor (KNN) in 2-D space, so as to distinguish noisy photons after adaptive noise scaling. Finally, in order to divide the data into background and surface photon, we apply a Bayesian decision theory after determining the probability of KNN. If the probability of a surface photon is greater than that of a background photon, it is considered to be a surface photon. Otherwise, it is considered to be a background photon.

5) *Adaptive directional filter method [41]* (*ADFM*): On the basis of MDBSCANM [48] modifying the shape of the search area, ADFM [41] takes the influence of terrain factor into consideration and traverses each filtering direction to determine the optimal filtering direction of the filtering kernel. First, define an elliptical region centered on point p, and determine whether any point q is within the ellipse. Since the distance between the noise point and the ground point is very close, in order to make the density value obtained by statistics more distinguishable, the filter kernel set the weight according to the distance between two points. Second, traverse each filtering direction, and use the direction with the largest density statistics result (i.e., when the direction of the filtering kernel is consistent with the direction of the terrain) as the optimal filtering direction of the filtering kernel. Finally, after calculating the density value of each point, a threshold T is given to determine whether each point is a noise point. Due to the high coherence between the signal and noise of photon-counting lidar, it is necessary to calculate the maximum density value of the point cloud within the range of radius c with any point as the center. Subsequently, the difference between the maximum density value and the point density value is calculated. If the difference is greater than the given threshold value, the point is removed as a noise point.
6) *Density-dimension method [42]* (*DDM*): Density-dimension method utilizes the anisotropic kernel and radial basis function to calculate the weighted density. One significant advantage of *DDM* [42] is its autoadaptive capability, which enables it to distinguish background and surface photons in complex environmental conditions. First, before the density-dimension algorithm is executed, perform a large-scale signal and noise slabs separation to define a noise data set and determine the height range of the ice surface. Second, since the ice surface generally extends more horizontally than vertically, an anisotropic kernel is introduced and combined with the radial basis function to calculate the density field. Finally, determine autoadaptive threshold to separate noise versus signal photons and associate signal photons to the ice surface.
7) *Improved localized statistics method* (*ILSM*): The last method is proposed to detect surface photon based on an improved localized statistics-based algorithm. It involves two main steps to implement this algorithm. First, divide the raw data into partial bins and build the elevation frequency histogram. A buffer zone is then created within 150 m above and below the average elevation value of the bin to remove the obvious noise photons and contain the potential signal photons. Finally, in order to further remove the remaining noise photons, calculate the density of photons in each direction to get the maximum density of each photon. Then, generate the local photon density distribution histogram and detect all potential peaks of the histogram by local maximum. After that, Gaussian curve is used to fit the noise peak and signal peak and the intersection of the two Gaussian curves is considered to be the density threshold.

B. Evaluation Method

For fair comparison, we have made the ground truth data. For MABEL data and ATLAS data, confidence evaluation is officially given in the data. Confidence value [57] is divided into five levels (0, noise; 1, padding point; 2, low-confidence point; 3, medium-confidence point; 4, high-confidence point). On the basis of the official label, we interpret visually the low-confidence points mainly, and modify these to obtain a more reliable label. For MATLAS data, the official label contains clear noise point and signal point, so we directly used it as ground truth data. For MBL data, the photon label is not announced. Based on satellite images from Google Earth, the ground truth data are obtained by visual interpretation completely. The results of ground truth could be viewed in Fig. 4.

In this article, the qualitative and quantitative analysis are all conducted. We mainly complete qualitative analysis through visual observation. Additionally, *recall (R)*, *precision (P),* and *F-Measure (F)* [32] are used as quantitative evaluation indicators. *Recall* represents how many positive samples are predicted correctly. *Precision (P)* indicates how many of the samples predicted to be positive are correct. *F-Measure (F)* is the weighted harmonic average of Precision and Recall. The calculation equation of the indicators can be referred to as follows:

$$R = \frac{TP}{TP + FN}$$

$$P = \frac{TP}{TP + FP}$$

$$F = \frac{2P \cdot R}{P + R}$$

where *TP* represents the signal photons that are detected correctly; *FP* represents the noise photons that are detected incorrectly by algorithm; *FN* represents the signal photons are detected as noise photons.

C. Different Methods on Different Surface Types

In this section, we describe how we analyzed the different surface detection methods both qualitatively and quantitatively. The result of each dataset detected by each algorithm has list in Table III–Table VI.

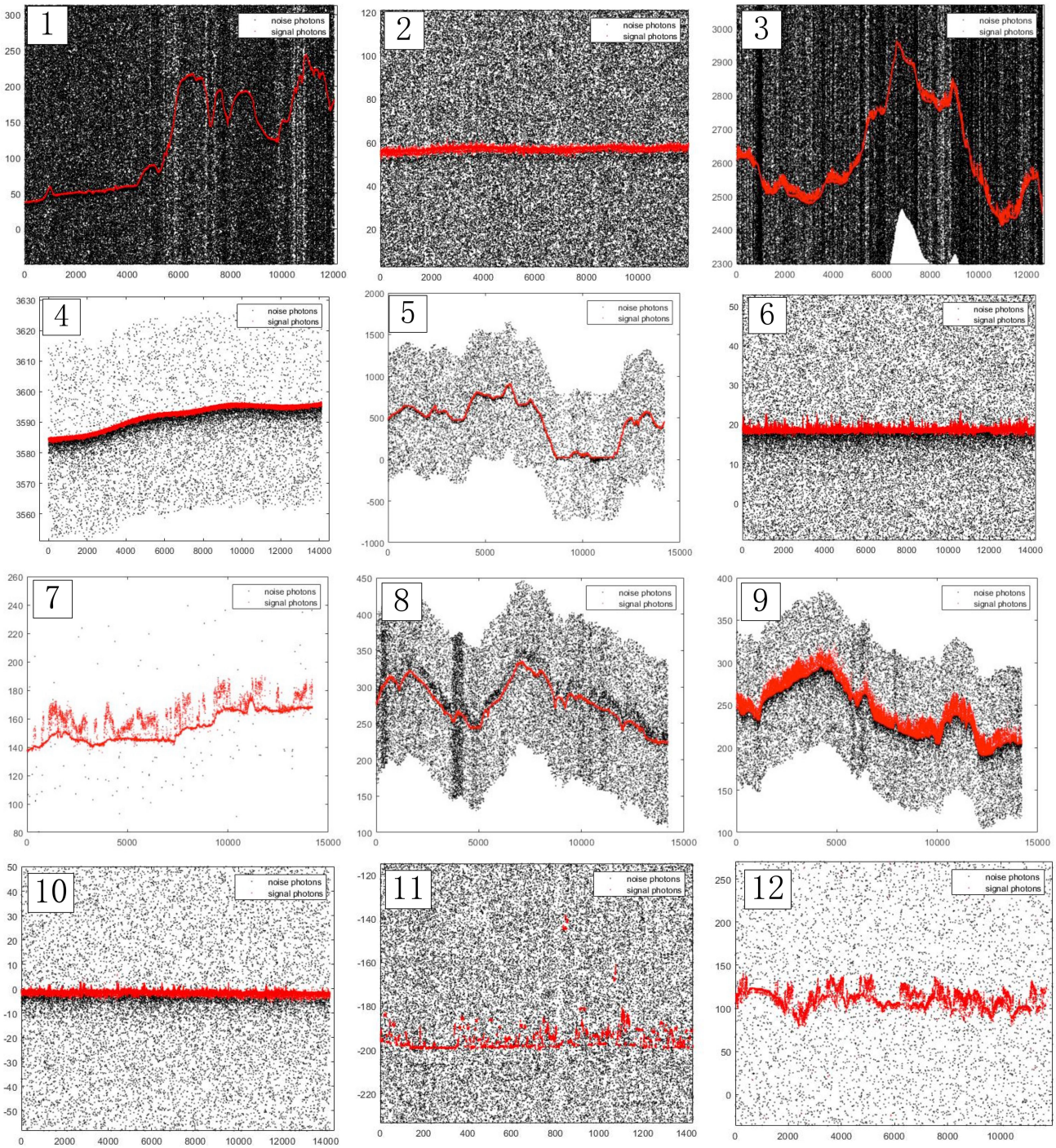

Fig. 4. Ground truth data of each dataset. The red points represent signal points, while the black points represent noise points. The pictures labeled with 1–3 are MABEL data, labeled with 4–10 are ATLAS data. The 11th picture is MBL data and the12th picture is MATLAS data. More details of datasets could refer to Table I.

1) For the surface of the ocean, the terrain is relatively gentle. We selected a set of MABEL ocean data and a set of ATLAS ocean data. Ocean data, due to the flat terrain, is less difficult, so all methods have achieved good results. The three methods HBM [40], ILSM [30], LDM [50], and ADFM [41] perform better, they perform well in Recall, Precision, and F-Measure, which means these methods could detect as many signal photons as possible while ensuring the correct rate. Other methods, MDBSCANM [48], NFM [24], and DDM [42], also perform well in Precision, but unsatisfactory in Recall, which means that there are a lot of signals that have not been detected. We noted in the previous section that HBM [40] has a good segmentation ability on flat terrain. The results of Dataset6

TABLE III
SEVEN METHODS FOR THE INDICATORS RESULTS OF OCEAN DATA

DATASETS Number	Land types	Data description	Methods	indicators		
				Recall(R)	Precision(P)	F1-Measure
2	Ocean	MABEL Data of Ocean	HBM [40]	0.8469	0.9960	0.9154
			LDM [50]	0.9617	0.9507	0.9562
			MDBSCANM [48]	0.4732	1	0.6425
			NFM [24]	0.0828	0.9739	0.1526
			ADFM [41]	0.9876	0.9132	0.9489
			DDM [42]	0.5538	0.9967	0.7120
			ILSM [30]	0.9582	0.9841	0.9710
6	Ocean	ATLAS Data of Ocean	HBM [40]	0.9727	0.9957	0.9841
			LDM [50]	0.9124	0.9988	0.9536
			MDBSCANM [48]	0.9810	0.9952	0.9880
			NFM [24]	1	0.9505	0.9746
			ADFM [41]	0.9999	0.9534	0.9762
			DDM [42]	0.4425	0.9973	0.6130
			ILSM [30]	0.9999	0.9590	0.9791

TABLE IV
SEVEN METHODS FOR THE INDICATORS RESULTS OF SEA ICE DATA

DATASETS Number	Land types	Data description	Methods	indicators		
				Recall(R)	Precision(P)	F1-Measure
10	Sea Ice	ATLAS Data of Sea Ice	HBM [40]	0.9207	0.9999	0.9587
			LDM [50]	0.9007	0.9999	0.9477
			MDBSCANM [48]	0.9755	0.9997	0.9874
			NFM [24]	0.9999	0.9881	0.9940
			ADFM [41]	0.9999	0.9774	0.9885
			DDM [42]	0.4166	0.9983	0.5879
			ILSM [30]	1	0.9823	0.9911

TABLE V
SEVEN METHODS FOR THE INDICATORS RESULTS OF LAND ICE DATA

DATASETS Number	Land types	Data description	Methods	indicators		
				Recall(R)	Precision(P)	F1-Measure
1	Land Ice	MABEL Data of north of the Canada; Observation area has obvious slope influence	HBM [40]	0.9121	0.9877	0.9484
			LDM [50]	0.9371	0.9714	0.9539
			MDBSCANM [48]	0.6508	0.9954	0.7870
			NFM [24]	0.8147	0.8664	0.8397
			ADFM [41]	0.9882	0.9127	0.9489
			DDM [42]	0.5211	0.9749	0.6792
			ILSM [30]	0.9893	0.9089	0.9474
4	Land Ice	ATLAS Data of Antarctica; The slope of the observation area is relatively gentle	HBM [40]	0.9514	0.9984	0.9744
			LDM [50]	0.9478	0.9994	0.9539
			MDBSCANM [48]	0.9988	0.9925	0.9956
			NFM [24]	1	0.9552	0.9771
			ADFM [41]	0.9518	0.9993	0.9750
			DDM [42]	0.4219	0.9986	0.5932
			ILSM [30]	1	0.9835	0.9917
5	Land Ice	ATLAS Data of Greenland; The observation area has large elevation fluctuations and large slope changes	HBM [40]	0.5573	0.9998	0.7156
			LDM [50]	0.9885	0.9992	0.9938
			MDBSCANM [48]	0.4970	0.9974	0.6634
			NFM [24]	0.9994	0.9941	0.9967
			ADFM [41]	0.9905	0.9978	0.9941
			DDM [42]	0.3655	0.9949	0.5346
			ILSM [30]	0.9999	0.9898	0.9948

TABLE VI
SEVEN METHODS FOR THE INDICATORS RESULTS OF LAND DATA

DATASETS Number	Land types	Data description	Methods	indicators		
				Recall(R)	Precision(P)	F1-Measure
3	Land	MABEL Data of north America; Observation area is the mountain with large slope, and there are a lot of vegetation on land	HBM [40]	0.4312	0.9842	0.5997
			LDM [50]	0.7328	0.9593	0.8309
			MDBSCANM [48]	0.2676	0.9862	0.4209
			NFM [24]	0.5442	0.9357	0.6882
			ADFM [41]	0.8009	0.9432	0.8662
			DDM [42]	0.4576	0.9420	0.6160
			ILSM [30]	0.9880	0.8358	0.9056
7	Land	ATLAS Data of Europe. The data is observed in night with low background noise rate.	HBM [40]	0.5321	0.9998	0.6946
			LDM [50]	0.9099	0.9988	0.9523
			MDBSCANM [48]	0.9051	0.9998	0.9501
			NFM [24]	0.7554	0.9988	0.8602
			ADFM [41]	0.9816	0.9981	0.9898
			DDM [42]	0.3550	0.9998	0.5240
			ILSM [30]	0.9522	0.9979	0.9745
8	Land	ATLAS Data of Siberia. The data is observed in the daytime with little vegetation..	HBM [40]	0.8824	0.9974	0.9364
			LDM [50]	0.9002	0.9848	0.9406
			MDBSCANM [48]	0.9115	0.9449	0.9279
			NFM [24]	0.9934	0.9053	0.9473
			ADFM [41]	0.9926	0.9503	0.9712
			DDM [42]	0.4865	0.9897	0.6524
			ILSM [30]	0.9992	0.9365	0.9668
9	Land	ATLAS Data of Siberia. The data is observed in the daytime with many vegetation..	HBM [40]	0.7459	0.9962	0.8531
			LDM [50]	0.6567	0.9993	0.7926
			MDBSCANM [48]	0.8618	0.9670	0.9114
			NFM [24]	0.9265	0.9466	0.9365
			ADFM [41]	0.8599	0.9589	0.9068
			DDM [42]	0.4267	0.9967	0.5975
			ILSM [30]	0.9265	0.9466	0.9365
11	Land	MBL Data of Sanya of China. The data is observed in the daytime. The data contains many building in the city.	HBM [40]	0.8646	0.9890	0.9226
			LDM [50]	0.8425	0.9985	0.9139
			MDBSCANM [48]	0.9363	0.9864	0.9607
			NFM [24]	0.9999	0.9245	0.9607
			ADFM [41]	0.9944	0.9219	0.9568
			DDM [42]	0.4684	0.9181	0.6203
			ILSM [30]	0.9381	0.9216	0.9298
12	Land	MATLAS Data of West Virginia of US. The data contains a little vegetation with gentle ground	HBM [40]	0.4288	0.9961	0.5995
			LDM [50]	0.4166	0.9793	0.5846
			MDBSCANM [48]	0.9650	0.9672	0.9661
			NFM [24]	0.9907	0.9652	0.9778
			ADFM [41]	0.9878	0.9710	0.9793
			DDM [42]	0.4166	0.9793	0.5846
			ILSM [30]	0.9899	0.9630	0.9762

processed by HBM[40] and DDM[42] are shown in Fig. 5. LSM sets the vertical and horizontal weights. ILSM [30] and ADFM [41] add a slope search step, this is conducive to the detection of the upper and lower edge signals so its results are good in the case of processing after pulse noise.

2) For the surface of sea ice, the data characteristics are very similar to those of the ocean. But the sea ice area has a more complex area that is usually a mixture of ocean and ice, so the data dispersion of sea ice is higher. The indicators show that LDM [50], ILSM [30], and ADFM [41] perform better. The Precision of DDM [42] and LSM is higher than Recall, which means that the ability of the finding the accuracy of the signal is better than the ability to find the number of signals. However, the ADFM [41] tends to find more signal photons, resulting in a decrease in the accuracy of the detected signals.
3) For land ice data, the fluctuation is greater than for ocean and sea ice data, and there are no other features on the surface, such as vegetation, buildings, etc. We selected three sets of typical data, Dataset1, Dataset4, and Dataset5.

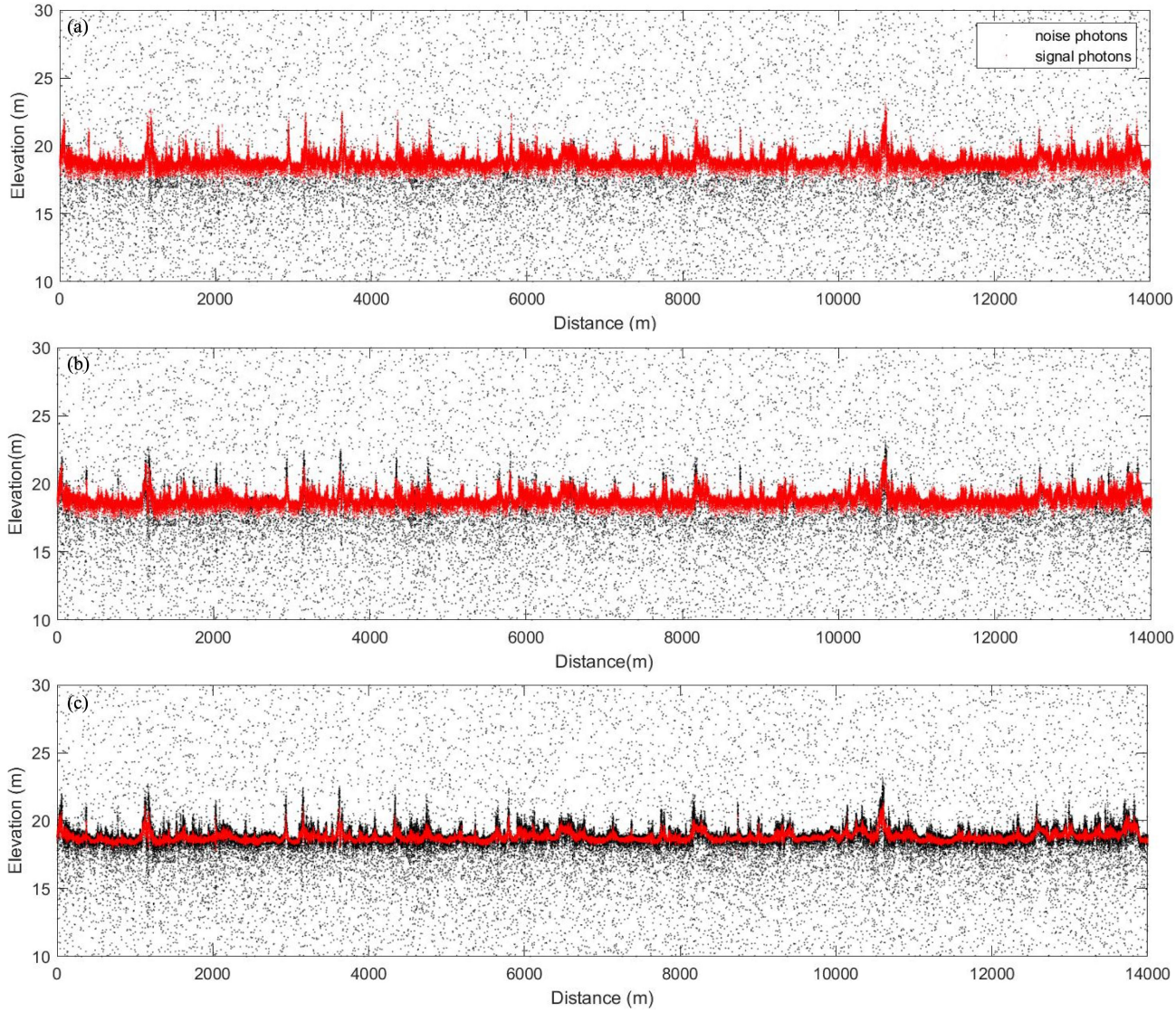

Fig. 5. (a) *Ground truth data* of the *Dataset6*. (b) Result of HBM. (c) Result of DDM. Both processing results preserve the shape of the surface, however, HBM contains more signals, and DDM lost many edge signals.

Dataset1 is MABEL data from north of Canada with obvious steep slope. Dataset4 is ATLAS data from Antarctica with flat slope. Dataset5 is ATLAS data from Greenland with steep slope. These three datasets cover all possible features of land ice.

The indicators of results show that every algorithm could extract signal photons completely, but three algorithms, LDM [50], ADFM [41], ILSM [30], perform very well and stable, and F-Measure value of these methods are all above 0.94. The HBM [40] and MDBSCANM [48] method have shown a good ability to find ice signals, but it is unsatisfactory in slope processing, which is obvious in Dataset5. NFM [24] performs better in land ice, it can show a relatively stable performance for slope areas. The detail result of land ice dataset processed by NFM is shown in Fig. 6. DDM [42] pays more attentions to the accuracy of the detected signal photons.

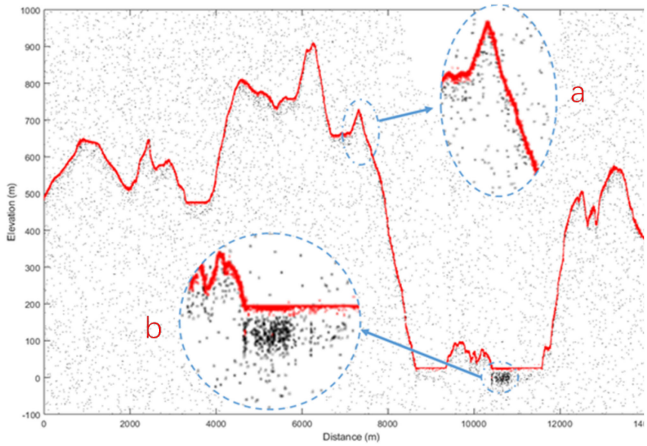

Fig. 6. Results of NFM processing of the ATL03 land ice dataset. Area a shows that the method can avoid breakpoints in areas with large slopes. Area b shows that the method can process after pulse noise.

4) For land areas, they are characterized by more complex types of ground features. In this experiment, we selected six datasets, including various conditions such as day and night, whether vegetation is covered, etc. These groups of data are divided into four groups.

   The first group contains Dataset7, which is observed in night by ATLAS. The characteristic of this group is that the noise rate is relatively low. Due to lack of noise photons, it is easier for the algorithm to mistake the ground object photon as a noise photon. Among all the methods, ADFM [41] and ILSM [30] performs best, and these methods score high in *Recall, Precision,* and *F-Measure*. LDM [50], MDBSCANM [48], and NFM [24] lose some vegetation signals with low signal rates. HBM [40] and DDM [42] could only detect ground signals, and lose a lot of vegetation signals.

   The second group contains *Dataset8*, which is observed in daytime with little vegetation. Six methods score greater than 0.9 of F-Measure. However, DDM [42] does not perform well of Recall, which lead to a low F-Measure.

   The third group contains Dataset3, Dataset9, Dataset12, which are observed in daytime with a lot of vegetation by different sensors. Among all the results, ILSM [30] and ADFM [41] perform very well in all three datasets. MDBSCANM [48] and NFM [24] perform well in last two datasets, but lost many signals in Dataset3, which means that these two methods are more sensitive to the parameters of different sensors. LDM [50] could find out ground signals and few vegetation signals, but lost signals of the top of the canopy. HBM [40] and DDM [42] could only detect ground truth, and lost most of the vegetation signals.

   The fourth group contains Dataset11, which is observed by MBL in the seaside city called Sanya. From the profile of the data, many building signals can be observed, and below the building signals, there will be gaps of the ground signals. This data examines the ability of algorithm to process discontinuous signals. The experimental results are similar to the second group. Only DDM [42] does not perform well in Recall. However, the signals lost by DDM [42] are the edge signals, and from a qualitative perspective, DDM [42] have a good result.

D. Method Analysis

In this section, combining the process efficiency (Fig. 7) of each algorithm and performance on different surface types in the previous parts, we will deeply analyze the characteristics of each algorithm.

1) HBM [40] involves selecting a very small interval as the processing area for each step in the distance along the track. The vertical histogram statistics are performed on this interval, and the signal of this area is found by setting the threshold twice. Because the interval selected by the method in the direction along the orbital distance is very small, the method has good adaptability when it encounters background photon unevenness. From the perspective of processing efficiency, HBM's efficiency is extremely high and only one-hundredth to one-thousandth of other methods. High efficiency and good processing capacity for flat areas are the characteristics of HBM. However, at the same time, because it only searches for possible surface photons based on the histogram of the vertical direction of fixed ground distance, the algorithm does not make good use of the information of vegetation or buildings. In addition, when the area slope is large, the vertical histogram will show a waveform with small peaks and large variances. This makes the surface signal area wider in the vertical direction and results in more background photons being classified as surface photons.
2) LDM [50] involves calculating the sum distance of KNN of each photon, and surface photon detection is performed at the photon scale. Compared with HBM [40], the photon boundary found by this method is clearer. Although there is no design step for slope correction, the method is still quite satisfactory in the slope area. However, it also has some defects: ① in the daytime, the photon density in the vertical direction is greater than the photon density in the orbital direction, so the nearest N photons that the algorithm looks for are mostly vertical photons; and ② after counting the KNN features of all the photons, the algorithm adopts the global histogram method, but for uneven background noise, it cannot achieve a high processing accuracy.
3) MDBSCANM [48] involves calculating the number of photons in a certain range of photons as the characteristic value. Compared with LDM, the search area of the algorithm changes to an ellipse, which solves the problem of uneven photon distribution along the orbital and vertical directions, to a certain extent.

However, this type of method has a unified problem: when classifying photons, the global histogram method is still adopted. When the background photons are uneven, the classification accuracy is greatly reduced. What is more, this method is sensitive to parameters and has poor slope processing capabilities.

4) NFM [24] improves the defects of the previous methods in the treatment of uneven background photons. In the preprocessing stage, the homogenized density method is used to uniformize the area with uneven background photon. This algorithm retains a relatively wide height range of signal called padding the signals in ATBD03. In the feature calculation, the KNN value of the LDM [50] is still used, but in the feature segmentation section, Bayesian theory is used, which is innovative. From the perspective of processing results, this method shows excellent performance when processing datasets with vegetation on the ground. However, the algorithm needs the prior information, and the accuracy of this information is related to the accuracy of the segmentation result, which means it is sensitive to the parameters. In addition, NFM [24] can still retain few noises far away from the main body of the signals.
5) ADFM [41] can be seen as an improvement of MDBSCANM [48]. Based on MDBSCANM [48], the vertical

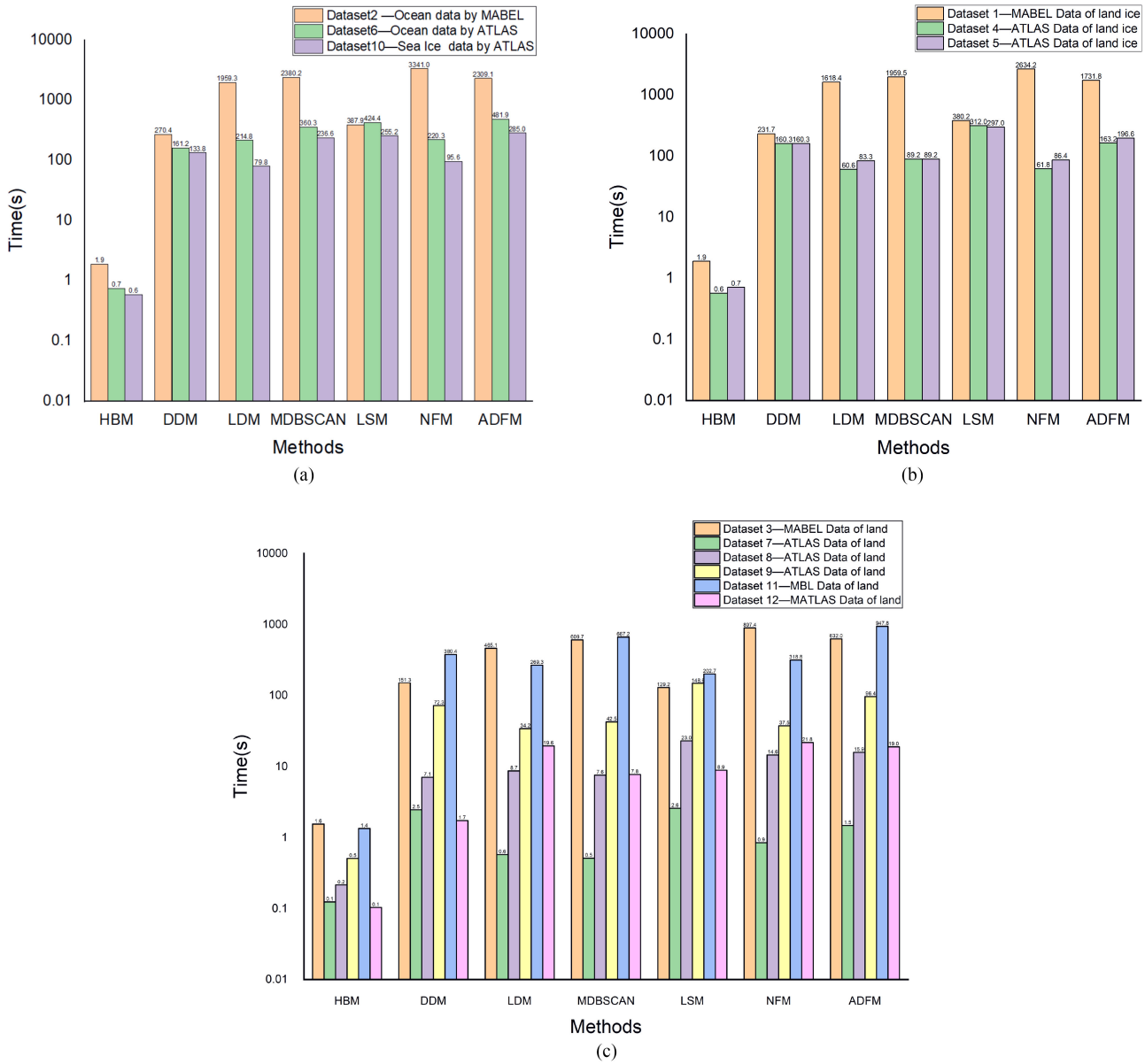

Fig. 7. Calculation time of each method for different data. (a) Situation of sea ice data and ocean data, (b) situation of land ice, and (c) situation of land. All the methods is programmed using MATLAB, and running using a computer with Intel(R) Core(TM) i9-9900 CPU. It can be seen from the figures that HBM's computing power is relatively strong, and the time-consuming is only ∼1/100 of other methods. The calculation time of other methods will vary with the number of photons in Datasets.

and horizontal weights are set when calculating the feature value, which better solves the problem of uneven density distribution in different directions. At the same time, the step of searching the surface slope is added, so that the algorithm can better detect areas with steep terrain. And this method has a good performance in all surface types. However, searching for the slope in a fixed step size greatly increases the complexity of the algorithm. Furthermore, the effect of the improvement is not obvious for the case of uneven background photon.

6) DDM [42] uses another way of calculating the feature value. Computationally, this method uses an anisotropic kernel and radial basis function to calculate the feature value. The improvement lies in giving weight to the distance from the center position, and the closer to the center position, the higher the weight. In terms of threshold setting, this method uses a local threshold, which also solves the problem of uneven background noise. Compared with MDBSCANM [48], this method retains the core surface photons and discards the edge signal photons (we refer to the middle photons of the vertical direction as the "core surface photons", and the photons above and below the core signal photons as the "edge signal photons"), because DDM [42] is more demanding when performing feature segmentation. This feature also leads to the lower Recall value of this method in all datasets, especially on datasets

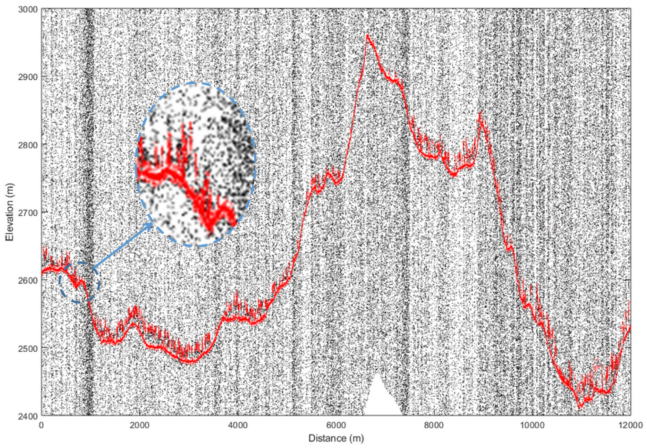

Fig. 8. Results of NFM processing of the land dataset of MABEL. From the enlarged view, some signal photons above the surface are better preserved.

with vegetation. However, when dealing with simple data types such as sea ice datasets, land ice datasets, etc., its qualitative effect is still well with signal continuity and signal accuracy.

7) ILSM [30] can be seen as a simplification and improvement of ADFM [41]. In terms of preprocessing, it uses local histogram statistics to select regions that may be signals. Compared with the preprocessing method of DDM [42], this method can select data regions without resorting to external data. For the calculation of the feature value, ILSM [30] is used to calculate the number of photons in the fixed ellipse range of each photon, and to set the signal slope search step. A global histogram is then built and a threshold set to distinguish the surface photons from the background photons.

The seven algorithms can all extract surface photons well, but they perform differently on different aspects, which means that they perform differently on different types of surface. HBM [40] shows a better performance when dealing with relatively gentle surface data such as ocean and sea ice data. DDM [42] can extract the surface subjects well and has a strong advantage in surface detection, especially land ice data. NFM [24] has advantages in land data with vegetation. MDBSCANM [48] and LDM [50] show better performance in land data. ADFM [41] and ILSM [30] show a comprehensive performance.

Improved versions or the combination of these methods were also developed to achieve better results [40], [48], [53], [54], [57], [58].For instance, *The modified histogram based method* (MHBM)[57], the official method of ATL03, could be seen the improvement of the HBM [40]. It adopted variable bin size according different surface types, and slope strategy is used in method. MHBM has the ability to adapt to different surfaces to a certain extent, and at the same time solves part of the slope problem. *The differential, regressive, and Gaussian adaptive nearest neighbor* (DRAGANN)[53], [54], the official method of ATL08, also uses the number of points around a point as feature value like MDBSCAN [48]. This method sets the ratio between the distance along the track and the elevations, and to a certain extent solves the problem of the uneven density of photons in the direction along the track and the elevation, which is an improvement of MDBSCAN [48]. *The modified OPTICS method* (MOPTICSM) [58] combines the advantages of NFM, MDBSCANN, and OPTICS, which continues homogenizing noise photons of NFM, adopts the distance processing method between two points of MDBSCANM [48], and use OPTICS clustering method to complete the denoising. The step of homogenizing noise photons solves the problem of uneven noise distribution, the step of adopting a horizontal ellipse searching area solves the problem of uneven distribution in different directions of photon density, and OPTICS clustering method performs better than DBSCAN method in parameter sensitivity. This method performs well when dealing with areas with vegetation. From the perspective of the development trend of the method, compared with the original method, the improved method shows stronger applicability to specific surface types and research areas.

V. Conclusion

In this article, we have reviewed the existing surface detection methods for photon point cloud data, and summarize the methods. The full process of all methods could be roughly divided into three steps as follows.

1) *Preprocessing:* Image processing, envelope curves, and histogram statistics are the commonly used methods to reduce the number of background photons.
2) *Surface detection:* We summarize three types of methods: the clustering methods, the feature calculation methods, and waveforms methods.
3) *Subsequent processing:* Iterative process, histogram statistics, and CV model are always subsequent processing methods.

In order to explore the characteristics of different types of methods, we selected 7 methods and 12 data for experiments. The seven methods include: HBM [40], LDM [50], MDBSCANM [48], NFM [24], ADFM [41], DDM [42], ILSM [30]. In total 12 datasets we selected contain many typical surface types, including land ice, ocean, land, and sea ice, from four types of photon-counting sensors, i.e., MABEL, ICESat-2/ATLAS, MBL (China), and MATLAS.

We have carried out a quantitative and qualitative evaluation of the experimental results. After analysis of the result, HBM [40] showed a better performance for gentle surface data such as ocean and sea ice. DDM [42] performed well in surface and surface features detection. NFM [24] has advantages in land data with vegetation. MDBSCANM [48] and LDM [50] show better performance in land data. ADFM [41] and ILSM [30] show a comprehensive performance. If there are no high requirements for surface detection, ADFM [41] and ILSM [30] are recommended in various surface type. Although the NFM [24] method has not been widely promoted, we still believe that the innovation of the method is very novel. With the improvement of the stability of this method, we think it will still have more outstanding performance.

The findings of this review article will help researchers to select the most appropriate surface detection methods for different scenarios.

REFERENCES

[1] C. Wang, X. LI, and H. Peng, "Research on simulation and analysis of spaceborne full waveform laser altimeter," *Laser Optoelectron. Prog.,* vol. 52, no. 10, pp. 276–282, 2015.

[2] H. Zwally *et al.*, "ICESat's laser measurements of polar ice, atmosphere, ocean, and land," *J. Geodynamics,* vol. 34, no. 3/4, pp. 405–445, 2002.

[3] T. Schenk and B. Csatho, "A new methodology for detecting ice sheet surface elevation changes from laser altimetry data," *IEEE Trans. Geosci. Remote Sens.,* vol. 50, no. 9, pp. 3302–3316, Sep. 2012.

[4] H. Xie *et al.*, "A comparative study of changes in the lambert glacier/amery ice shelf system, east Antarctica, during 2004–2008 using gravity and surface elevation observations," *J. Glaciology,* vol. -1, no. 235, pp. 1–17, 2016.

[5] J. B. Blair, D. L. Rabine, and M. A. Hofton, "The laser vegetation imaging sensor: A medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography," *ISPRS J. Photogrammetry Remote Sens.,* vol. 54, no. 2/3, pp. 115–122, 1999.

[6] H. Araki *et al.*, "Lunar global shape and polar topography derived from Kaguya-LALT laser altimetry," *Science,* vol. 323, no. 5916, pp. 897–900, 2009.

[7] M. A. Lefsky *et al.*, "Estimates of forest canopy height and aboveground biomass using ICESat," *Geophysical Res. Lett.,* vol. 32, no. 22, pp. 1–4, 2005.

[8] G. Metternicht, L. Hurni, and R. Gogu, "Remote sensing of landslides: An analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments," *Remote Sens. Environ.,* vol. 98, no. 2/3, pp. 284–303, 2005.

[9] A. W. Yu *et al.*, "Space laser transmitter development for ICESat-2 mission," in *Solid State Lasers Xix: Technology and Devices*, W. A. Clarkson, N. Hodgson, and R. K. Shori eds., vol. 7578. Bellingham, WA, USA: SPIE, 2010.

[10] R. H. Hadfield, "Single-photon detectors for optical quantum information applications," *Nature Photon.,* vol. 3, no. 12, p. 696–705, 2009.

[11] J. Degnan, D. Wells, R. Machan, and E. Leventhal, *Second Generation Airborne 3D Imaging Lidars Based on Photon Counting (Optics East)*. Bellingham, WA, USA: SPIE, 2007.

[12] L. Hou, *Research On Key Technologies of Photon Countin Imaging Lidar in Moving Conditions*, Shanghai, China: Shanghai Inst. Tech. Physics, 2013.

[13] J. Degnan *et al.*, "Design and performance of an airborne multikilohertz photon-counting microlaser altimeter," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.,* vol. 34, no. 3/W4, pp. 9–16, 2001.

[14] J. J. Degnan, "Photon-counting multikilohertz microlaser altimeters for airborne and spaceborne topographic measurements," *J. Geodynamics,* vol. 34, no. 3, pp. 503–549, 2002.

[15] M. F. Jasinski, J. D. Stoll, W. B. Cook, M. Ondrusek, E. Stengel, and K. Brunt, "Inland and near-shore water profiles derived from the high-altitude multiple altimeter beam experimental lidar (MABEL)," *J. Coastal Res.,* vol. 76, no. sp1, pp. 44–55, 2016.

[16] Ministry of Land and Resources of the People's Republic of China, "The planning and development observation satellite on marine and land (2011–2020)," R 2012.

[17] The National Development and Reform Commission of the People's Republic of China, The Ministry of Finance of the People's Republic of China, and State Administration of Science, M. o. Finance, and N. D. T. a. I. Bureau, "National civil space infrastructure long and medium term development plan (2015-2025)," R 2015.

[18] L. A. Magruder and K. M. Brunt, "Performance analysis of airborne Photon- Counting Lidar data in preparation for the ICESat-2 mission," *IEEE Trans. Geosci. Remote Sens.,* vol. 56, no. 5, pp. 2911–2918, May 2018, doi: 10.1109/tgrs.2017.2786659.

[19] M. McGill, T. Markus, V. S. Scott, and T. Neumann, "The multiple altimeter beam experimental lidar (MABEL): An airborne simulator for theICESat-2Mission," *J. Atmospheric Ocean. Technol.,* vol. 30, no. 2, pp. 345–352, 2013, doi: 10.1175/jtech-d-12-00076.1.

[20] W. Abdalati *et al.*, "The ICESat-2 laser altimetry mission," *Proc. IEEE,* vol. 98, no. 5, pp. 735–751, May 2010, doi: 10.1109/jproc.2009.2034765.

[21] R. Kwok *et al.*, "Surface height and sea ice freeboard of the arctic ocean from ICESat-2: Characteristics and early results," *J. Geophysical Res., Oceans,* vol. 124, no. 10, pp. 6942–6959, 2019.

[22] M. Li, *Research On Technologies of Photon Counting Lidar based On Fiber Optics*. Shanghai, China: Shanghai Inst. Tech. Phys., 2017.

[23] D. Hancock. "MATLAS selected slides from David Harding SDT presentation." [Online]. Available: https://icesat-2.gsfc.nasa.gov/data/matlas/docs/Slides20141204.pdf

[24] X. Wang, Z. Pan, and C. Glennie, "A novel noise filtering model for photon-counting laser altimeter data," *IEEE Geosci. Remote Sens. Lett.,* vol. 13, no. 7, pp. 947–951, Jul. 2016.

[25] M. Li, Y. Guo, G. Yang, and R. Shu, "A noise filter method for the Push-broom photon counting lidar and airborne cloud data verification," *Sci. Technol. Eng., J.*, vol. 17, no. 9, pp. 53–58, 2017.

[26] S. C. Popescu *et al.*, "Photon counting lidar: An adaptive ground and canopy height retrieval algorithm for ICESat-2 data," *Remote Sens. Environ.,* vol. 208, pp. 154–170, 2018, doi: 10.1016/j.rse.2018.02.019.

[27] T. Markus *et al.*, "The ice, cloud, and land elevation satellite-2 (ICESat-2): Science requirements, concept, and implementation," *Remote Sens. Environ.,* vol. 190, pp. 260–273, 2017, doi: 10.1016/j.rse.2016.12.029.

[28] L. A. Magruder, M. E. Wharton, K. D. Stout, and A. L. Neuenschwander, "Noise filtering techniques for photon-counting ladar data," Laser Radar Technology and Applications XVII, 2012.

[29] M. Awadallah, S. Ghannam, L. Abbott, and A. Ghanem, "Active contour models for extracting ground and forest canopy curves from discrete laser altimeter data," in *Proc. 13th Int. Conf. Lidar Appl. Assess. Forest Ecosyst.*, 2013, pp. 129–136.

[30] X. Zhu, S. Nie, C. Wang, X. Xi, and Z. Hu, "A ground elevation and vegetation height retrieval algorithm using micro-pulse photon-counting lidar data," *Remote Sens.*, vol. 10, no. 12, pp. 1–23, 2018, doi: 10.3390/rs10121962.

[31] M. S. Moussavi, W. Abdalati, T. Scambos, and A. Neuenschwander, "Applicability of an automatic surface detection approach to micro-pulse photon-counting lidar altimetry data: Implications for canopy height retrieval from future ICESat-2 data," *Int. J. Remote Sens.,* vol. 35, no. 13, pp. 5263–5279, 2014, doi: 10.1080/01431161.2014.939780.

[32] S. Nie *et al.*, "Estimating the vegetation canopy height using micro-pulse photon-counting LiDAR data," *Opt. Exp.*, vol. 26, no. 10, pp. A520–A540, May 14, 2018, doi: 10.1364/OE.26.00A520.

[33] K. H. Horan and J. P. Kerekes, "AN automated statistical analysis approach to noise reduction for photon-counting lidar systems," in *Proc. IEEE Int. Geosci. Remote Sens. Symp., IEEE Int. Symp. Geosci. Remote Sens.*, 2013, pp. 4336–4339.

[34] L. L. Narine, S. Popescu, A. Neuenschwander, T. Zhou, S. Srinivasan, and K. Harbeck, "Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data," (in English), *Remote Sens. Environ.*, vol. 224, pp. 1–11, Apr. 2019, doi: 10.1016/j.rse.2019.01.037.

[35] L. L. Narine, S. Popescu, T. Zhou, S. Srinivasan, and K. Harbeck, "Mapping forest aboveground biomass with a simulated ICESat-2 vegetation canopy product and landsat data," *Ann. Forest Res.,* vol. 62, no. 2, pp. 69–86, 2019.

[36] K. M. Brunt *et al.*, "MABEL photon-counting laser altimetry data in alaska for ICESat-2 simulations and development," *Cryosphere,* vol. 10, no. 4, pp. 1707–1719, 2016, doi: 10.5194/tc-10-1707-2016.

[37] R. Kwok *et al.*, "Profiling sea ice with a multiple altimeter beam experimental lidar (MABEL)," *J. Atmospheric Ocean. Technol.,* vol. 31, no. 5, pp. 1151–1168, 2014, doi: 10.1175/jtech-d-13-00120.1.

[38] D. Gwenzi and M. A. Lefsky, "Prospects of photon counting lidar for savanna ecosystem structural studies," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.,* vol. XL-1, pp. 141–147, 2014, doi: 10.5194/isprsarchives-XL-1-141-2014.

[39] D. Gwenzi, M. A. Lefsky, V. P. Suchdeo, and D. J. Harding, "Prospects of the ICESat-2 laser altimetry mission for savanna ecosystem structural studies based on airborne simulation data," *ISPRS J. Photogrammetry Remote Sens.,* vol. 118, pp. 68–82, 2016, doi: 10.1016/j.isprsjprs.2016.04.009.

[40] K. M. Brunt, T. A. Neumann, K. M. Walsh, and T. Markus, "Determination of local slope on the Greenland ice sheet using a multibeam photon-counting lidar in preparation for the ICESat-2 mission," *IEEE Geosci. Remote Sens. Lett.,* vol. 11, no. 5, pp. 935–939, May 2014, doi: 10.1109/lgrs.2013.2282217.

[41] F. Xie, G. Yang, R. Shu, and M. Li, "An adaptive directional filter for photon counting lidar point cloud data," *J. Infrared Millimeter Waves, J.*, vol. 36, no. 1, pp. 107–113, 2017, doi: 10.11972 /j.

[42] U. C. Herzfeld, T. M. Trantow, D. Harding, and P. W. Dabney, "Surface-Height determination of crevassed glaciers–mathematical principles of an autoadaptive density-dimension algorithm and validation using ICESat-2 simulator (SIMPL) data," *IEEE Trans. Geosci. Remote Sens.,* vol. 55, no. 4, pp. 1874–1896, Apr. 2017.

[43] S. Jiang, L. Wang, C. Tang, Y. Cheng, B. Chen, and Y. Pang, "A denoising approach for detection of canopy and ground from ICESat-2's airborne simulator data in Maryland, USA," in *Proc. Adv. Laser Technol. Appl.*, 2015, vol. 9671, Paper 96711S.
[44] N. F. Glenn *et al.*, "Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass," *Remote Sens. Environ.,* vol. 185, pp. 233–242, 2016, doi: 10.1016/j.rse.2016.02.039.
[45] S. L. Farrell, K. M. Brunt, J. M. Ruth, J. M. Kuhn, L. N. Connor, and K. M. Walsh, "Sea-ice freeboard retrieval using digital photon-counting laser altimetry," *Ann. Glaciology,* vol. 56, no. 69, pp. 167–174, 2017, doi: 10.3189/2015AoG69A686.
[46] H. Tang, A. Swatantran, T. Barrett, P. DeCola, and R. Dubayah, "Voxel-Based spatial filtering method for canopy height retrieval from airborne single-photon lidar," *Remote Sens.,* vol. 8, no. 9, pp. 1–13, 2016, doi: 10.3390/rs8090771.
[47] Y. Ma, R. Liu, S. Li, W. Zhang, F. Yang, and D. Su, "Detecting the ocean surface from the raw data of the MABEL photon-counting lidar," *Opt. Exp.,* vol. 26, no. 19, pp. 24752–24762, Sep. 17, 2018, doi: 10.1364/OE.26.024752.
[48] J. Zhang, J. Kerekes, B. Csatho, T. Schenk, and R. Wheelwright, "A clustering approach for detection of ground in micropulse photon-counting LiDAR altimeter data," in *Proc. Geosci. Remote Sens. Symp.*, 2014, pp. 177–180.
[49] Z. Jiashu and J. Kerekes, "An adaptive density-based model for extracting surface returns from photon-counting laser altimeter data," *IEEE Geosci. Remote Sens. Lett.,* vol. 12, no. 4, pp. 726–730, Apr. 2015, doi: 10.1109/lgrs.2014.2360367.
[50] S. Xia, C. Wang, X. Xi, S. Luo, and H. Zeng, "Point cloud filtering and tree height estimation using airborne experiment data of ICESat-2," *J. Remote Sens.*, vol. 18, no. 4, pp. 1199–1207, 2014, doi: 10.11834 /jrs.20144029.
[51] B. Chen *et al.*, "Ground and top of canopy extraction from photon-counting LiDAR data using local outlier factor with ellipse searching area," *IEEE Geosci. Remote Sens. Lett.,* vol. 16, no. 9, pp. 1447–1451, Sep. 2019, doi: 10.1109/lgrs.2019.2899011.
[52] B. Chen *et al.*, "Potential of forest parameter estimation using metrics from photon counting LiDAR data in howland research forest," *Remote Sens.,* vol. 11, no. 7, pp. 1–21, 2019, doi: 10.3390/rs11070856.
[53] A. Neuenschwander *et al.*, "Ice, cloud, and land elevation satellite 2 (ICESat-2) algorithm theoretical basis document (ATBD) for land - Vegetation Along-Track products (ATL08)."
[54] A. Neuenschwander and K. Pitts, "The ATL08 land and vegetation product for the ICESat-2 mission," (in English), *Remote Sens. Environ.,* vol. 221, pp. 247–259, Feb. 2019, doi: 10.1016/j.rse.2018.11.005.
[55] B. Chen, Y. Pang, Z. Li, H. Lu, and X. Liang, "Photoncounting LiDAR point cloud data filtering based on random forest algorithm," *J. Geo-Inf. Sci.,* vol. 21, no. 6, pp. 898–906, 2019, doi: 10.12082/dqxxkx.2019.190013.
[56] T. F. Chan, B. Y. Sandberg, and L. A. Vese, "Active contours without edges for vector-valued images," *J. Vis. Commun. Image Representation,* vol. 11, no. 2, pp. 130–141, 2000, doi: 10.1006/jvci.1999.0442.
[57] T. Neumann, A. Brenner, D. Hancock, J. Robbins, J. Saba, and K. Harbeck, "Ice, cloud, and land elevation satellite–2 (ICESat-2) project: Algorithm theoretical basis document (ATBD) for global geolocated photons (ATL03)," *Nat. Aeronaut. Space Admin., Goddard Space Flight Center*, 2021. [Online]. Available: https://nsidc.org/sites/nsidc.org/files/technical-references/ICESat2_ATL03_ATBD_r004.pdf
[58] X. Zhu *et al.*, "A noise removal algorithm based on OPTICS for photon-counting LiDAR data," *IEEE Geosci. Remote Sens. Lett.,* vol. 18, no. 8, pp. 1471–1475, Aug. 2021.

Huan Xie (Senior Member, IEEE) received the B.S. degree in surveying engineering, and the M.S. and Ph.D. degrees in cartography and geo-infomation from Tongji University, Shanghai, China, in 2003, 2006, and 2009, respectively.

From 2007 to 2008, she was with the Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany, funded by the China Scholarship Council, as a visiting scholar. Her research interests include satellite laser altimetry and hyperspectral remote sensing.

Qi Xu was born in Shandong Province, China in 1997. He received the B.S. degree in geographic information science from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2015. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing from Tongji University, Shanghai, China.

His research interests include satellite single-photon data processing and applications.

Dan Ye was born in Jiangxi Province, China in 1995. She received the B.S. degree in surveying and mapping engineering from East China Jiaotong University, Nanchang, China, in 2017, and the M.S. degree in surveying and mapping engineering from Tongji University, Shanghai, China, in 2020.

Her research interests include the extraction of ground information from single-photon laser altimetry data.

Jianghao Jia received the B.Eng. degree in surveying and geo-informatics from Tongji University, Shanghai, China, in 2021.

His research interests include remote sensing and satellite laser data processing.

Yuan Sun received the B.S. degree in surveying and mapping engineering from Nanjing University of Information Science and Technology, Nanjing, China, in 2020. She is currently working toward the Ph.D. degree in surveying and mapping science and technology with Tongji University, Shanghai, China.

Her research interests include satellite laser altimetry data processing and applications.

Peiqi Huang was born in Guangxi, China in 2000. She received the B.S. degree in surveying and mapping from Tongji University, Shanghai, China, in 2021.

Ming Li was born in 1984. He received the B.S. and M.S. degrees in electronic engineering from Shanghai JiaoTong University, Shanghai, China, in 2010 and the Ph.D. degree in physical electronics from the University of Chinese Academy of Sciences, Shanghai, in 2017.

He is currently an Associate Researcher with the Shanghai Institute of Technical Physics of the Chinese Academy of Sciences, a master's supervisor, a member of the Youth Innovation Promotion Association of the Chinese Academy of Sciences, a member of the Shanghai Astronautical Association. His main research interests include laser active photoelectric detection technology and system engineering, aerospace remote sensing technology, etc.

Dr. Li was the electronics designer of the laser three-dimensional imaging payloads of the Chinese lunar exploration projects "Chang'e-3", and the chief designer of electronics for the quantum science satellite "Mozi"; won the first prize of Shanghai Science and Technology Progress Award One, one paper each in Science, Physical Review Letters and other journals, multiple papers in SCI journals such as Applied Optics, and the 2018 Cleveland Prize winner of the American Association for the Advancement of Science.

Shijie Liu (Member, IEEE) received the B.S. degree in surveying engineering, and the M.S. and Ph.D. degrees in cartography and geo-information from Tongji University, Shanghai, China, in 2005, 2008, and 2012, respectively.

He is currently an Associate Professor of photogrammetry and remote sensing with the College of Surveying and Geoinformatics, Tongji University. His research interests include geometric exploitation of high-resolution remote sensing and its applications.

Xiaohua Tong (Senior Member, IEEE) received the Ph.D. degree in traffic engineering from Tongji University, Shanghai, China, in 1999.

From 2001 to 2003, he was a Postdoctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. He was a Research Fellow with Hong Kong Polytechnic University, Hong Kong, in 2006, and a Visiting Scholar with the University of California, Santa Barbara, CA, USA, from 2008 to 2009. His research interests include photogrammetry and remote sensing, trust in spatial data, and image processing for high-resolution satellite images.