Sea Surface Signal Extraction for Photon-Counting LiDAR Data: A General Method by Dual-Signal Unmixing Parameters

Zhen Wen, Xinming Tang, Guoyuan Li, Bo Ai, Guanghui Wang, Jiaqi Yao, and Fan Mo

***Abstract*—The ice, cloud, and land elevation satellite-2 (ICESat-2) is the only satellite that produces photon-counting light detection and ranging data, and is equipped with the advanced topographic laser altimeter system. ICESat-2 provides sea surface height product; however, its approach of the product is unsuitable for areas with sub-surface signals. Conventional denoising methods applied to sea surface photon data of variable density involve the use of different empirical parameters. Considering the distribution of sea surface signal photons, we propose a general open-source method using a dual-signal unmixing parameter (DSUMP), which incorporates the Gaussian distribution of dual-signal peaks to determine the sea surface range. This method facilitates the direct extraction of sea surface photons under various observation conditions—day or night, strong or weak beams, and including or excluding seabed photons—without requiring any variable parameters. The elevation error by DSUMP within 0.1m accounts for more than 97%. The mean absolute error is within 0.01 m compared to sea surface photons obtained via manual extraction. Different model parameters show stable denoising accuracy, only affects operating efficiency. The proposed method introduces a novel denoising technique for extracting sea surface elevation from ICESat-2 altimetry data, and its applicability can be extended to various point cloud data with similar distributions.**

***Index Terms*—Gaussian mixture model, ICESat-2, photon-counting light detection and ranging (LiDAR), point cloud denoising, sea surface height.**

Manuscript received 22 April 2023; revised 25 August 2023 and 6 October 2023; accepted 31 October 2023. Date of publication 3 November 2023; date of current version 23 November 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 41971425, Grant 41871382 and Grant 62071279, and in part by the Special Funds for Creative Research under Grant 2022C61540. *(Corresponding author: Xinming Tang.)*

Zhen Wen is with the College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China, and also with the Land Satellite Remote Sensing Application Center, MNR, Beijing 100048, China (e-mail: iwzhen@163.com).

Xinming Tang is with the College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China, and also with the Land Satellite Remote Sensing Application Center, MNR, Beijing 100048, China (e-mail: tangxinming99@qq.com).

Bo Ai is with the College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China (e-mail: aibo@sdust.edu.cn).

Guoyuan Li, Guanghui Wang, and Fan Mo are with the Land Satellite Remote Sensing Application Center, MNR, Beijing 100048, China (e-mail: ligy@lasac.cn; wanggh@lasac.cn; surveymofan@163.com).

Jiaqi Yao is with the Academy of Ecological Civilization Development for JING-JIN-JI, Tianjin Normal University, Tianjin 300382, China (e-mail: yaojiaqi@tjnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3329962

I. Introduction

SEA surface height is a key oceanographic parameter, and is utilized in models of ocean circulation and tides as well as tsunami warnings. Long-term tide gauge measurements are a major source of data for studying relative sea level changes in coastal waters [1]; however, owing to the limited number of gauge stations, this type of data is unsuitable for large-scale and high-resolution observations. Passive reflectometry and interferometry systems have been proposed as an alternative [2], and global navigation satellite system reflection (GNSS-R) technology has been shown to provide reliable data for determining sea surface height [3]. This includes shore-based [4], space-based [5], and satellite-based [6], [7] data, which are mainly limited by the location of the receiver. Furthermore, satellite radar altimeters, such as the TOPEX/Poseidon [8], Jason-1/2 [9], and Envisat [10] altimeters, have been used in sea surface height acquisition. Although radar altimeters have high observation accuracy in terms of sea surface measurements, they operate at a large scale and provide low resolution data. The ice, cloud, and land elevation satellite-2 (ICESat-2) was successfully launched in 2018, and provides a new technical approach for sea surface height acquisition [11].

The objective of the ICESat-2 mission was to provide elevational data for ice sheets and vegetation canopies, as well as provide topographic measurements of cities, lakes, reservoirs, oceans, and land surfaces. The Level2 data product (ATL03) combines the photon time-of-flight with the satellite position and attitude to determine the photon location (i.e., latitude, longitude, and elevation) detected by ATLAS [12]. The Level3A data products were processed from Level2 data, and includes glacier and ice sheet, vegetation canopy, ocean surface, and inland water heights [11]. ATL12, the ocean surface height product [13], comprises specific length-scale surface height and estimates of height distribution, roughness, surface slope, and apparent reflectance. ICESat-2 is widely used in oceanography to detect and measure changes in sea surface height [14] and to calculate significant wave height [15], peak wavelength, and period of ocean waves [16], making it an important data source for oceanographic measurements [17]. ICESat-2 single track can recover the marine geoid at wavelengths >20 km which is similar to the best radar altimeter data [18].

The utilization of data from ATL03 to ATL12 has been relatively low for the following reasons.

1) *Weak Beam With a Low Signal*: ATL12 primarily uses three strong beams of level 2 products as input signals.
2) *Ocean Subsurface Signal*: Submarine signals create vertical double peaks owing to the high energy of ATLAS, and automatic separation of the sea surface and submarine signals is difficult.
3) *Strong Sunlight Background Noise*: Data acquired during the day has a low signal-to-noise ratio, making denoising difficult. In general, although a denoising method [19] can achieve good results when submarine signals are excluded, but it is difficult to set a common threshold.

There are researchers to select the most appropriate surface detection methods for different scenarios, but the method is not suitable for the above conditions [20]. The official ATL12 extraction of sea-surface signals mainly involves determining the photon distribution range of the sea surface using statistical histograms. However, this method becomes unsuitable when submarine signals are present. Consequently, the application scenarios of ATL12 data are limited, and reliable and general automatic separation models for surface and submarine signals are lacking. Currently, methods for separating submarine and sea surface signals rely on statistical histograms, which delineate the range of high-density photons based on experience [21], [22]. Typically, an automatic processing method first determines the peak of sea surface signals and then uses a buffer distance to determine the range; however, these distances are uncertain, and are 3–5 times the standard deviation [23], [24] or a fixed value [25], [26]. There is a lack of general methods for analyzing data with different densities and conditions.

Sea surface height exhibits periodic fluctuations and limited variations within a range; consequently, sea surface signal photons are highly concentrated within a certain elevation range compared with photons associated with land elevation. Accordingly, in this article, a general open-source signal extraction method was developed to extract sea surface signals with strong or weak noise. This method addresses the shortcomings traditional denoising methods, such as density-based clustering methods and Gaussian filtering, which rely on variable empirical parameters, and effectively extracts sea-surface photons even in the presence of submarine signals. The extraction results were verified using a high-confidence product from ATL03 and manual extraction, and evaluated quantitatively and qualitatively. Moreover, we analyzed the influence of horizontal and vertical bins on the accuracy and efficiency of range extraction. Notably, the dual-signal unmixing parameter (DSUMP) algorithm effectively extracted sea-surface photons under diverse observation conditions without any variable parameters, including day and night, strong and weak beams, and inclusion or exclusion of seabed photons. The greatest advantage of this method is its universal applicability and robustness compared with other algorithms. The results of our assessment indicated that the method is stable in regions with different photon densities and can be widely applied.

II. Data and Methods

The DSUMP algorithm was used to separate the sea surface from noise photons and other weaker signals. Parameters were calculated by finding the maximum value of the statistical histogram area associated with the Gaussian curve and adding an attenuation parameter to the area, which decays exponentially with the distance from the Gaussian peak. The dual-signals refer to the seabed signals and the sea surface signals in the case where the data including these two signals, the noise signals and the sea surface signals in the case where there are only sea surface signals. Therefore, the method is applicable in both cases, including and excluding submarine signals. The framework of this article, including the data, model, validation, and application, is shown in Fig. 1.

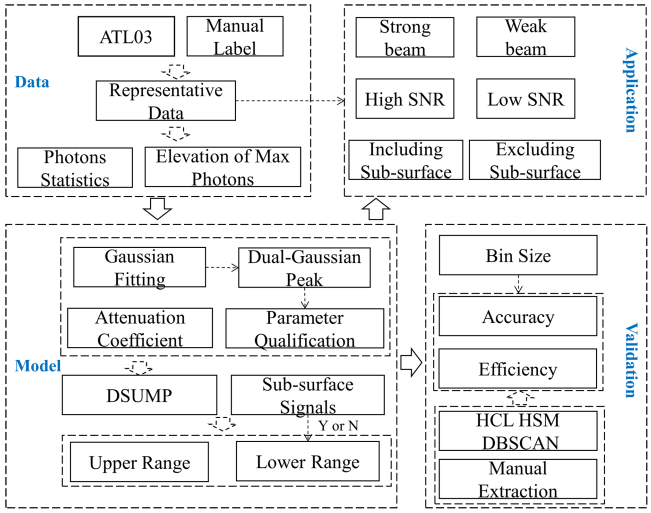


Fig. 1. Data analysis framework.

The ICESat-2 data are publicly available from the National Snow and Ice Data Center. Geolocated photon data (ATL03) were obtained online (https://nsidc.org/data/atl03). The core code and the results of this article are available at https://github.com/wen-zhen/DSUMP.

A. ICESat-2 Data

The ICESat-2 and ATLAS instruments use photon-counting light detection and ranging (LiDAR) and ancillary systems to perform three main measurements: time-of-flight of photons from ATLAS to Earth and back; pointing vectors of photons as they are transmitted by ATLAS; and position in space when the photon is recorded. ATLAS emits green (532-nm) 10-kHz laser pulses approximately every 0.7 m along the track. Each emitted laser pulse is split by a diffractive optical element in ATLAS, producing six individual beams arranged in three pairs. Each pair of beams is emitted at different energies (the energy ratio between the weak and strong beams is approximately 1:4). The distance between strong-strong beams or weak-weak beams in the cross-track direction is approximately 3.3 km, and the distance between strong and weak beams in the along-track direction is approximately 2.5 km [11].

ATL03 is a Level2 product generated by the data products of Precision Orbit Determination, Precision Pointing Determination, and ATL02, and which contains the geolocation ellipsoidal heights of photon events downlinked from ATLAS. These

TABLE I
CHARACTERISTICS OF THE SELECTED DATA

NO.	ATL03	D/N	Start of latitude	End of latitude	Beam	Bottom
1	ATL03_20210214031837_07961007_005_01.h5	D	16.81	16.84	1R	I
2		D	16.43	16.46	3R	E
3	ATL03_20210717075633_03621201_005_01.h5	D	17.08	17.12	1R	I
4		D	16.58	16.61	2R	E
5	ATL03_20210417002225_03541107_005_01.h5	D	16.75	16.78	2L	E
6		D	16.97	17.00	3L	I
7	ATL03_20200419053723_03620701_005_01.h5	D	16.13	16.15	2L	E
8		D	16.16	16.19	1L	I
9	ATL03_20210818183003_08571207_005_01.h5	N	16.16	16.18	1R	I
10		N	16.53	16.55	2L	I
11	ATL03_20210716200216_03541207_005_01.h5	N	16.37	16.4	3R	E
12		N	16.76	16.79	2L	E
13	ATL03_20210617212611_12991107_005_01.h5	N	16.32	16.34	1R	I
14		N	17.12	17.15	2R	E
15	ATL03_20210916170607_12991207_005_01.h5	N	16.20	16.22	3L	E
16		N	16.05	16.07	2L	I

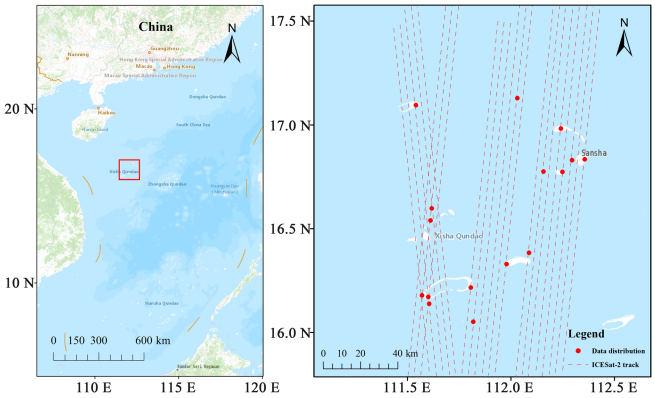


Fig. 2. Distribution of the data over longitude 111.5°E to 112.5°E and latitude 15.5°N to 17.5°N. The red line is the data track of ATL03 and the red photons the approximate location of the data.

TABLE II
DATA DENSITY

NO.	Density pts/m	Signal pts/m	Noise pts/m	SNR
1	3.68	1.56	2.12	0.74
2	2.41	1.49	0.92	1.62
3	5.80	2.88	2.92	0.99
4	2.22	1.44	0.78	1.85
5	1.25	0.49	0.76	0.64
6	2.15	0.56	1.59	0.35
7	1.62	0.56	1.07	0.52
8	2.26	0.58	1.68	0.35
9	3.80	2.46	1.34	1.84
10	1.18	0.57	0.61	0.93
11	3.77	3.64	0.13	28
12	0.69	0.66	0.03	22
13	3.26	1.24	2.02	0.61
14	0.89	0.84	0.05	16.8
15	2.29	2.24	0.05	44.8
16	5.41	4.76	0.65	7.32

heights are corrected for certain geophysical phenomena (e.g., atmospheric refraction and tides) and are classified as likely signal photon events or possible background photon events with four levels of confidence [12]. The experimental data are mainly distributed in low-latitude areas, without ice sheets, in the South China Sea, and in areas near some islands and reefs. The data selected to test the efficacy of our method included day or night and strong or weak beams, as well as data that included or excluded submarine photons. The data are presented in Table I and Fig. 2. We analyzed the density of the selected data, and separated signal and noise by manually labeling the data. The values of total density, signal density, noise density, and signal-noise-ratio (SNR) are given in Table II. The unit of density is the number of photons per distance along the track (pts/m). The density of the selected data ranged from 0.69 to 5.8 pts/m, and the SNR was between 0.35 and 44.8, indicating that the selected data can be considered representative.

The data were mainly from the Xisha Islands (see Fig. 2), one of the four major archipelagos in the South China Sea. The main body of the Xisha Islands comprises the Crescent and Amphitrite Groups. The island area is approximately 10 km^2, the total coastline is 518 km long, and the sea area exceeds 500 000 km^2. The water was clear, and the diffusion attenuation coefficient was relatively small.

B. DSUMP Denoising Method

Traditional methods based on data density [19], [27], [28], images [29], and histograms [13], [23] cannot split sea surface and bottom signals. The method used here overcomes this limitation by adopting an unmixing parameter as the central basis and utilizes the Gaussian fitting model [30], [31], [32]. Relevant applications of photon-counting LiDAR data in the field of sea surface detection have developed rapidly and achieved remarkable results. Sea surface data extraction generally uses empirical parameters, such as filtering methods. The following processes are involved in the DSUMP method and its verification.

1) The vertical distance of the maximum photon number distribution was determined using histogram statistics. Sea surface extraction was divided into two parts for processing, and unmixing was performed separately.
2) The Gaussian fitting model was used as the kernel for the DSUMP. Dual Gaussian peaks are considered to have subparameter limiting conditions: the relevant parameters

decay exponentially with increasing distance, and the weak Gaussian peak is smaller than the strong Gaussian peak.

3) The extraction results were verified and analyzed. The model was verified using ATL03 high-confidence photons, sea surface signals, and manual extraction. The influence of the horizontal and vertical bins on the extraction of the upper and lower boundaries was assessed to verify the robustness of the model.
4) An application test was conducted, in which the model was applied to different situations, such as strong or weak beams and areas with different SNR, including or excluding seabed photons.

Photons within the spatial segment (a photon distribution perpendicular to the track) form a double-peak histogram, with the PDP near the sea surface and the SDP near the seafloor [22]. However, DSUMP remains effective even when the photons are regarded as having a Gaussian distribution (secondary Gaussian peak) that excluding the seabed signal, which confirms that the unmixing parameters are general.

The sea surface and sea bottom are separated by double-signal unmixing parameters, and the sea surface and sea bottom (or noise) are regarded as the primary Gaussian peak (PGP) and secondary Gaussian peak (SGP), respectively, which correspond to strong and weak signals of the photon double-layer, where SGP is the elevation corresponding to the peak value of the Gaussian formed on the sea surface, and PGP is the elevation corresponding to the peak value of the Gaussian distribution formed by underwater photons. The graphical meaning of the parameter is the sum of the areas of the histogram generated by the Gaussian curve covering the statistical histogram, considering the distance to the Gaussian peak

$$\mathrm{UM}_{yj} = \sum_{i=0}^{n} \frac{CG_{ij} \times \mathrm{bin}_y}{e^{dtp}}, j = 1, 2 \tag{1}$$

$$y = z_{\min}, z_{\min} + \mathrm{bin}_y, +z_{\min} + 2 \times \mathrm{bin}_y, \cdots, z_{\max}. \tag{2}$$

The index of an unmixing parameter is calculated in (1). e^{dtp} is an attenuation parameter that is greater than 0. It is a weight parameter, and its purpose is to weaken the related statistical histogram far from the Gaussian peak. The closer to the Gaussian peak, the smaller the e^{dtp} and the larger the e^{-dtp}; that is, the smaller the attenuation coefficient, the greater the weight. Thus, the elevation obtained by the maximum value of the statistical histogram area is correct when the Gaussian peak is misclassified. If there is no attenuation parameter—for example, when the division value divides the photons of the sea surface—the elevation obtained by argmax will decrease and an incorrect segmentation result will be obtained. UM_{yj} is the parameter of the correlation area covered by a single Gaussian peak, where j is the primary or SGP and represents the PGP or SGP when j is 1 or 2; y is the elevation value (i.e., the assumed segmentation elevation in the calculation process) used to calculate UM_{yj}; bin_y is the vertical split distance parameter, and 0.1 m is the interval used to obtain unmixing parameter more accurately [21]. It is difficult to describe the distribution characteristics of photons if the interval is too small, and noise signals may be introduced if the interval is too large. Only a common interval (0.1 m) was used in the algorithm to display the experimental results [21]. In principle, the algorithm is universally applicable across intervals. i is the bin number for all the ranges when bin_y is determined. The calculation of y is shown in (2), where $z_{\max}$ and $z_{\min}$ are the maximum and minimum photon elevation values, respectively. e is a natural constant in physical optics, reflecting the relationship between reflectivity and water depth [33], [34]. CG_{ij} is the Gaussian peak correlation area parameter after the Gaussian fitting of a single bin_y, which includes some limiting rules because the PGP contains a strong signal and is much larger than the SGP, as shown in (3). dtp is the distance from each bin_y to the PGP or SGP, which is the distance from the photon to the two peaks when y is determined. The calculation of these parameters is shown in (4) and (5)

$$\mathrm{CG}_{ij} = \begin{cases} j = 1 \begin{cases} C_{ij}, C_{ij} < gs_{i_j} \\ gs_{i_j}, C_{ij} > gsp_j \end{cases} \\ j = 2 \begin{cases} C_{ij}, C_{ij} < gsp_j \\ gs_{i_j}, C_{ij} > gsp_j \text{ and } C_{ij} < gsp_1 \end{cases} \end{cases} \tag{3}$$

where gs_{ij} is the Gaussian fitting value for a bin_y, gsp_j is the fitted primary and SGP (the PGP when j is 1), and C_{ij} is the height of a single statistical histogram. This method considers C_{ij} as CG_{ij} when $C_{ij} < gs_{ij}$, and $\mathrm{gs}_{\mathrm{ij}}$ as CG_{ij} when $C_{ij} > gsp_j$. The constraint condition was added ($C_{ij} < gsp_1$) when calculating the sub-Gaussian peak time, which indicated that SGP was smaller than PGP

$$\mathrm{dtp} = \begin{cases} \left| z_{\max} - z_j - (\mathrm{len}\,(C_{ij}) - i) \times \mathrm{bin}_y \right|, a \text{ or } b \\ \left| z_j - z_{\min} - i \times \mathrm{bin}_y \right|, c \text{ or } d \end{cases} \tag{4}$$

$$\begin{cases} j = 1 \text{ and } bgs = bgs_1, \ a \\ j = 2 \text{ and } bgs = bgs_2, \mathrm{b} \\ j = 1 \text{ and } bgs = bgs_2, \mathrm{c} \\ j = 2 \text{ and } bgs = bgs_1, \mathrm{d} \end{cases}. \tag{5}$$

CG_{ij}, as calculated in (3), reflects the true area of the related histogram covered by the two Gaussian peaks; however, the greater the distance from the Gaussian peak, the greater the probability of noise. Therefore, an exponential attenuation is added to the parameter. The area of the histogram involved in the summation decays exponentially as dtp increases. z_j is the water depth at the Gaussian peak, and len is a parameter for calculating the length of C_{ij}. The conditions were set as shown in (5) to calculate dtp considering the relative positions of the PGP and SGP. The values of the upper and lower ranges of the sea surface require two-parameter calculations (with the maximum value obtained from the statistical value as the boundary) because the positions of the main Gaussian peaks are different. The upper surface was obtained by calculating the unmixing parameter above the sea surface (denoted bgs_1), and the lower surface was obtained by calculating the unmixing parameter below the sea surface (denoted bgs_2).

In the process of calculating the upper and lower surfaces of the photons on the sea surface, the PGP is always part of the

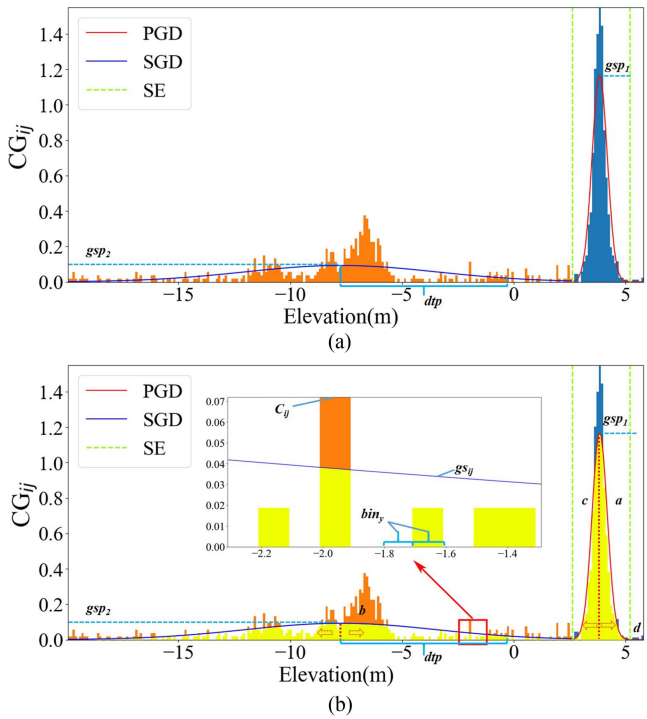


Fig. 3. Data are regarded as two Gaussian distributions including bottom signals; the bottom noises are regarded as a distribution when there are no bottom signals. PGD: And SGD are the primary and secondary Gaussian distribution fitting curves, respectively. The yellow area involved in the calculation is the area covered by Gaussian curve, which will decay exponentially from the Gaussian peak to both sides. Variables are defined in the equations in the text. (a) Logic of the parameter. (b) Result of the logic.

photons on the sea surface. However, the relative positions of the SGPs are different. When calculating the upper value of the sea surface, the SGP is on the right side of the PGP, but when calculating the lower value of the sea surface, it is on the left side. However, the number of bins ranged from $z_{\min}$ to $z_{\max}$; therefore, we adopted different strategies when calculating the distance. For the Gaussian peak on the left (i.e., cases a and b), a is the PGP when calculating the upper value of photons, and b is the SGP when calculating the lower value), using $z_{\min}$ as the benchmark. For the Gaussian peak on the right (i.e., cases c and d), c is the PGP when calculating the lower limit of the photons, and d is the SGP when calculating the upper limit of the photons), using $z_{\max}$ as the benchmark. Thus, (4) and (5) help calculate the distance between each bin of the statistical histogram and the Gaussian peak

$$\mathrm{SE} = \arg\max\left(UM_{y1} + UM_{y2}\right). \tag{6}$$

Finally, the optimal segmented elevation (SE) value was determined from the area of the correlation histogram covered by two exponentially attenuated Gaussian peaks, as shown in (6). We hypothesized that an elevation value must exist that separates the noise from the sea surface and maximizes the obtained UM_{yi}; the argmax function obtains the elevation when UM_{yi} is at its maximum. The logical results of this method are shown in Fig. 3.

In applying this model, errors will arise when using the same elevation value to divide photons in a large range, owing to the inclination of the sea surface. Therefore, a finite bin (0.005 in latitude, approximately 500 m) was used to divide the data in the latitude direction to further extract the sea surface photon data. The results of a demonstration of the method are shown in Fig. 4. The bin division only affects the operating efficiency of the program and does not reduce the efficiency of the sea surface extraction results, which are discussed in detail in the Discussion.

A particular bin (the demonstration in Fig. 4. is 0.005 in latitude, approximately 500 m) was used to divide the data in the latitudinal direction to further extract the sea surface photon data, as shown in Fig. 4(a). We took area A as an example to illustrate these steps. First, we calculated the largest density of photons with respect to elevation, as shown in Fig. 4(b), and the water depth at which the largest density of photons was found to be 3.727 m. We then calculated the upper and lower unmixing parameters. The sea surface-submarine dual-Gaussian results are shown in Fig. 4(c) (left), with a magnified view (right) showing the elevation value; the upper unmixing parameter value was 5.428 m. Similarly, the lower unmixing parameter value was 2.227 m [see Fig. 4(d)]. The above steps were repeated for the six regions, a–f, shown in Fig. 4(a), and the results are shown in Fig. 4(e).

III. Results

The following techniques were used to verify the DSUMP method qualitatively and quantitatively: Density-Based Spatial Clustering of Applications with Noise (DBSCAN, a density-based clustering method) [35], histogram statistical method (HSM, the method used to produce ATL12) [13], ATL03 high-confidence labels (HCLs) [12] and Gaussian fitting (GSF) [21]. The DBSCAN method mainly relies on two empirical parameters: the scanning radius (eps) and the minimum number of included points (minPts). Starting with an unvisited point, it finds all nearby points within an eps distance, and if the number of nearby points is greater than minPts, the point is marked as a signal point. In our evaluation of DSUMP, the scanning radius was 1 m, and the minPts of different data were determined manually for verification. The HSM was used in the process of ATL12 production at a certain bin in the vertical direction to extract the sea surface, and the value of the signal in the histogram was greater than the median of the histogram. As the median is typically zero, zero was moved from the histogram, and we subsequently extracted the signal range using a median multiple (MM) value of the histograms, where the histograms are greater than the MM. The vertical bin was 0.1 m in the experiment, and the MMs of the different data varied. The HCL method classifies each photon event as either a likely signal photon event or a background photon event, and provides a confidence assessment of these classifications (0: noise, 1: added to pad likely signal photon events, 2: low confidence signal, 3: medium confidence signal, 4: high confidence signal), the specific value range should be adjusted empirically according to different data. The GSF regards the data as having a Gaussian distribution in the vertical direction, and the range of sea surface photons is 3 $\times$ standard deviation.

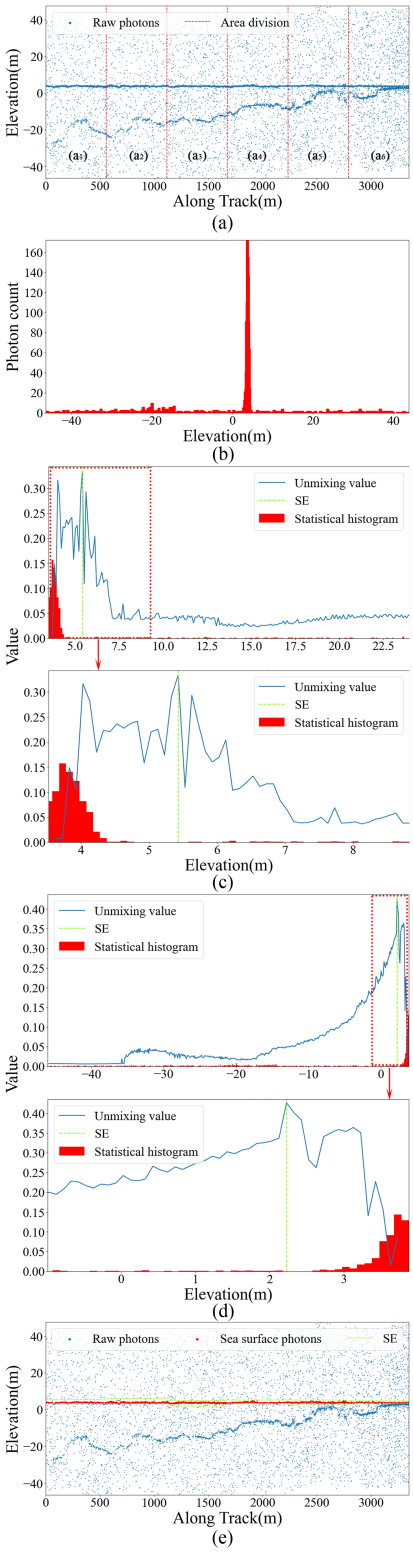


Fig. 4. Demonstration of DSUMP. (a) Green line is the data segment, approximately 500 m. (b) Vertical axis is the number of photons, and the horizontal axis the number of bins, the width of each bin is 0.1 m. (c)–(e) Blue dots are photon events for all confidence levels, red dots are photon events from DSUMP, green line is the unmixing elevation, which is $UM_{y1}+UM_{y2}$ in (6). (a) Photon data divided horizontally. (b) Elevation of the maximum photon distribution. (c) Upper surface unmixing value. (d) Lower surface unmixing value. (e) Result of the demo.

TABLE III
RESULTS OF DSUMP–PHOTONS

NO.	Precision (%)	Recall (%)	F1 (%)	IOU (%)
1	97.8	100	98.9	97.8
2	97.8	100	98.9	97.8
3	99.8	95.3	97.5	95.2
4	99.2	100	99.6	99.2
5	97.4	100	98.7	97.4
6	99	97.9	98.5	97
7	98.8	97.8	98.3	96.6
8	98.2	97.1	97.7	95.4
9	100	96.8	98.3	96.7
10	98.4	100	99.2	98.4
11	99.6	99	99.3	98.6
12	99	99.3	99.1	98.3
13	98.7	99.1	98.9	97.8
14	99.6	100	99.8	99.6
15	99.2	100	99.6	99.2
16	99.4	99.2	99.3	98.6

The reason for selecting the DBSCAN method was its representation of point-based algorithms. The intention was to demonstrate that the point-density-based algorithm simultaneously extracts surface and seabed photons, although further determination of the range is necessary. Empirical parameters such as the radius and photon number vary with photon density. Although the official confidence label of ATL03 does not explicitly extract sea surface photons (due to the ambiguity and division of the ground photon into several confidence levels), sea surface photons could still be determined (excluding seabed signals) based on the confidence level. However, when submarine signals were present, such as data whose NO is 1, 3, or 6 in Table I even if the confidence level is 4, the sea surface and seabed photons cannot be directly separated. Furthermore, the HCL cannot distinguish between sea-surface and sea-bottom photons in the presence of a seabed photon. The empirical parameters of the GSF method are necessary, and include a range of sea surface photons (e.g., 5 m) both above and below the maximum photon elevation value as the calculation area. However, this approach is not suitable if the sea surface elevation difference exceeds a certain range. Thus, these methods cannot be applied when seabed signals are present.

We used Precision, Recall, F1, and IOU to test the algorithm; however, owing to the high photon density on the sea surface, the parameters will be very high even if misclassified photons are generated. As given in Table III, we propose to use other parameters to enable practical application.

To verify the accuracy of the DSUMP method, we compared it with a traditional method that involves manual labeling. The manual range of signal photons was marked at a distance of 500 m. To verify model stability, the ATL03 data were divided into eight types according to strong or weak beam (R/L), day or night (D/N), and including or excluding submarine photons (I/E) (L-D-I, L-D-E, R-D-I, R-D-E, L-N-I, L-N-E, R-N-I, R-D-E).Two examples are provided for each type of data, and all data are given in Table I. Because the laser spot size of ATL03 was approximately 10 m, the latitude bin was set to

TABLE IV
RESULTS OF DSUMP-QUANTITATIVE

NO.	DBSCAN				HSM				HCL				GSF				DSUMP			
	MAE m	ME m	RLT 001	RLT 01	MAE m	ME m	RLT 001	RL T01	MAE m	ME m	RLT 001	RL T01	MAE m	ME m	RLT0 01	RLT0 1	MAE m	ME m	RLT0 01	RL T01
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.42	87	99
2	0.05	0.67	29	82	0.002	0.08	94	100	0.02	0.48	75	96	0.008	0.204	79	99	0.006	0.16	85	99
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.006	0.05	79	100
4	0.03	0.31	34	94	0.002	0.08	93	100	0.004	0.09	87	100	0.005	0.08	83	100	0.002	0.08	96	100
5	0.03	0.28	22	79	0.003	0.18	95	99	0.01	0.35	86	96	0.03	1.56	66	91	0.006	0.20	92	98
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.44	88	98
7	0.1	12	20	75	0.005	0.31	94	99	0.02	0.31	73	95	0.08	1.9	44	86	0.01	0.3	86	98
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.29	82	97
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.006	0.14	82	99
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.004	0.13	93	99
11	0.003	0.07	89	100	0.001	0.03	98	100	0.004	0.04	87	100	0.001	0.03	98	100	0.003	0.13	93	99
12	0.03	0.24	42	95	0.001	0.04	98	100	0.002	0.18	95	99	0.001	0.06	97	100	0.003	0.21	93	99
13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.92	87	98
14	0.06	0.96	23	75	0.003	0.31	96	99	0.006	0.35	91	97	0.001	0.07	98	100	0.002	0.19	97	99
15	0.009	0.09	76	100	0.001	0.02	97	100	0.002	0.03	96	100	0.001	0.025	97	100	0.001	0.02	97	100
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.002	0.01	93	100

TABLE V
EMPIRICAL PARAMETERS OF TRADITIONAL METHOD (WITHOUT UNDERSEA SIGNALS)

NO.	DBSCAN	HSM	HCL
2	3	3	3, 4
4	3	3	3, 4
5	2	4	3, 4
7	3	5	3, 4
11	3	10	4
12	1	3	4
14	2	10	4
15	3	10	4

0.0001 (approximately 10 m) to count the median elevation in the vertical direction as the elevation. The maximum error (ME), mean absolute error (MAE), and the photon ratio less than 0.1 m (RLT01), and that less than 0.01 ms (RLT001) were used to evaluate the accuracies of different methods, including DBSCAN, HSM, HCL, GSF, and DSUMP. The performance of the model is given in Table IV and Fig. 6. Three of the five methods we compared (HSM, HCL, and DBSCAN) require manually entered parameters.

The results given in Table IV indicate that HSM, DSUMP, and HCL yield good results when submarine photons were excluded; however, the DBSCAN algorithm ignores many signal photons, resulting in low accuracy. The three methods (DBSCAN, HSM, and HCL) must be adjusted manually to achieve optimal results. The empirical parameters are given in Table V. The parameters of DBSCAN, HSM, and HCL are the number of photons within a distance of 1 m, MM, and confidence level, respectively. However, in DSUMP, the empirical parameters did not need to be adjusted, and the same parameters are used for all data. The MAE was within 0.01 m, and the RLT01 was over 97% across all conditions. The three methods remove noise while retaining signals, and two types of signals, including submarine photons, were retained in the data (see Fig. 5). Although automatic separation of the two Gaussian peaks, that is, separating the sea surface signal from the sea bottom, is difficult, DSUMP effectively solved this problem even when the two signals were closely spaced (see Fig. 6). This shows that submarine signals do not induce prominent errors in sea-surface signal extraction. The results obtained from the 16 datasets indicate that the DSUMP method effectively extracts sea-surface signals with high precision and stability and without the need for manual parameters.

IV. DISCUSSION

A. Special Case

The presence of double signals is the default condition for this method; however, signal photons above the sea surface were not present in some nighttime data (such as No. 16). In such cases, the unmixing operation above the sea surface can divide the original signal into two parts (see Fig. 7). The first test was conducted to address this special situation. All photons within a certain buffer zone, such as 2 m (the signal width generally does not exceed 2 m) were considered signal photons if no photons were observed outside the buffer.

B. Parameters of DSUMP

The method uses the same bin of elevation and latitude for all data; the elevation bin is 0.1 m, and the latitude bin is 0.5 (approximately 0.5 km). According to the analysis of the DSUMP principle that the sea-surface signal consistently clustered within a certain distance regardless of the setting value, the unmixing parameters were obtained based on elevation, with the maximum correlation area of the fitting curve being unaffected by the bin size. As shown in the histogram, the maximum and minimum values and width of the histogram vary, and differences persist in the statistical value of the histogram, ensuring the accuracy of the results. The efficiency and accuracy of different parameters were calculated by taking the no. 1 data as an example. Different parameters yielded good results, although the efficiency fluctuated (see Table VI). The efficiency of the method decreased as the bin size decreased for a piece of data.

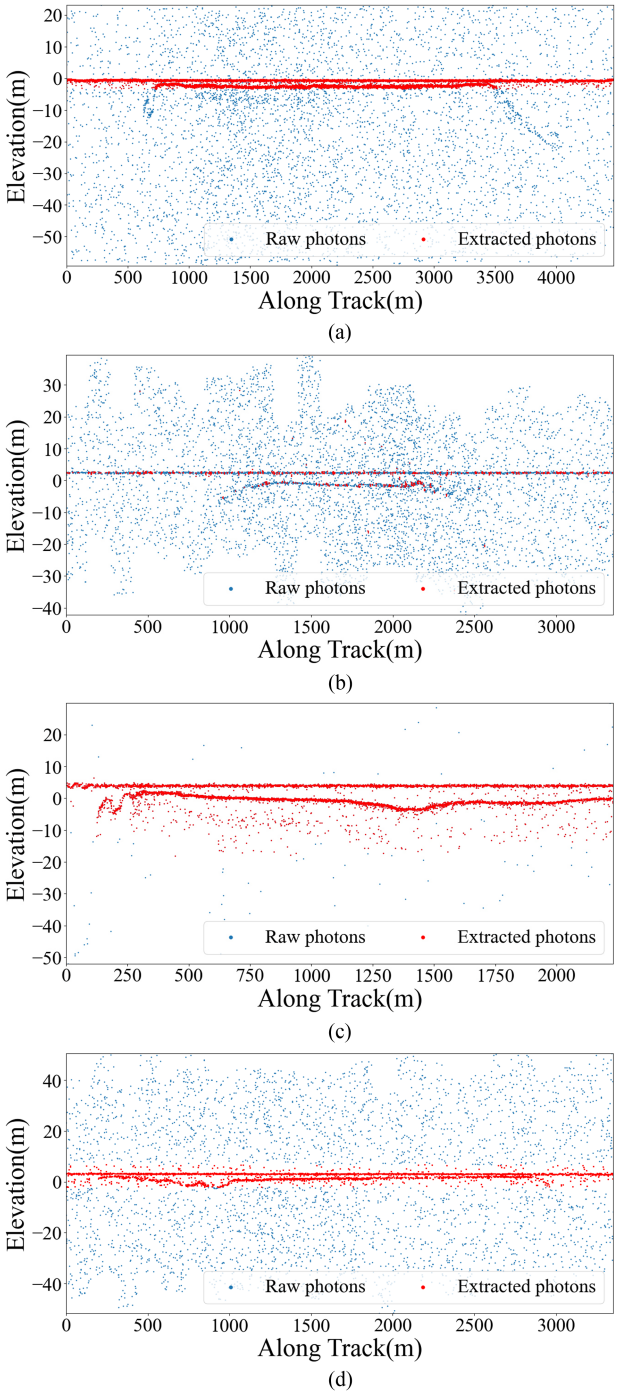


Fig. 5. Results of four methods (including bottom signals). Blue dots are photon events for all confidence levels; red dots are photon events from HSM, HCL, DBSCAN, and GSF. (a) HSM (data: NO 3, the parameter is 10). (b) DBSCAN (Data: NO 8, the parameter is 2). (c) HCL (data: No 13, the confidence is 4). (d) GSF (data: no 6, $\sigma = 3$).

TABLE VI
EFFECT OF SPACE ON EFFICIENCY AND ACCURACY

Elevation (m)	Latitude (km)	Time (min)	MAE (m)	ME (m)	RLT 001	RLT0 1
0.1	0.5	3	0.01	0.42	87	99
0.1	1	2	0.01	0.53	87	99
0.1	2	1.5	0.01	0.23	83	99
0.15	0.5	1.5	0.004	0.14	92	99
0.2	0.5	1	0.02	0.82	82	96

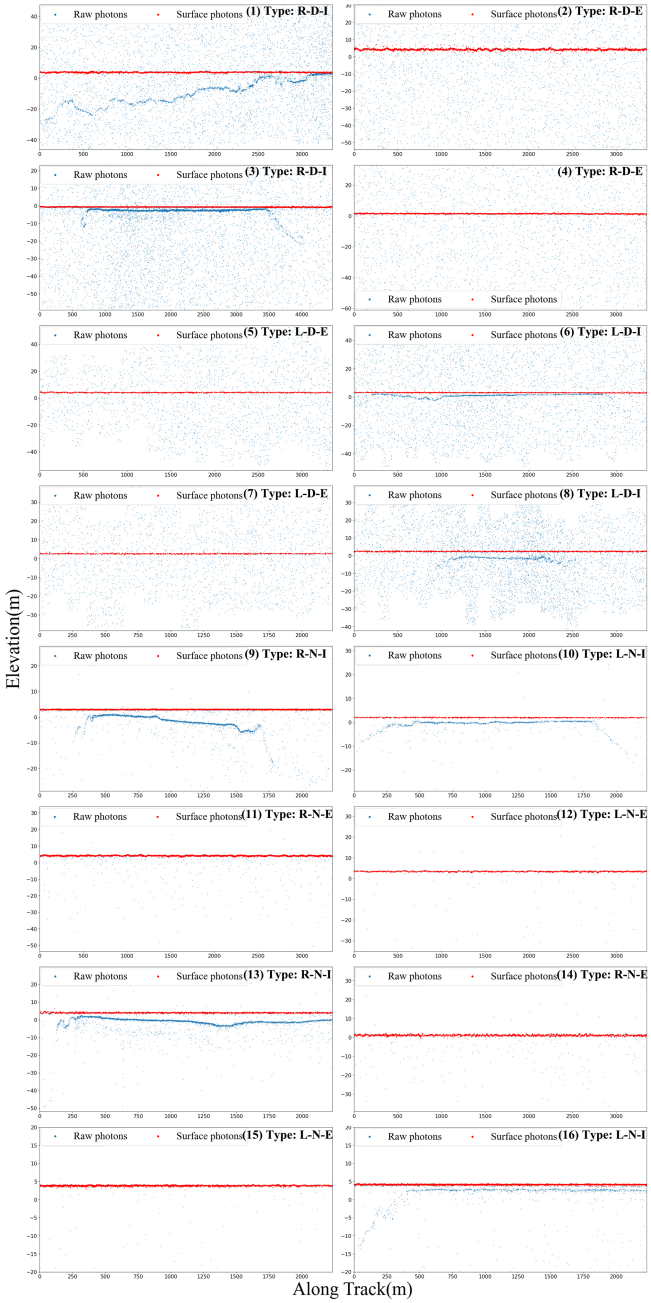


Fig. 6. Results of DSUMP for all types. Blue dots are photon events for all confidence levels and red dots are photon events from DSUMP. The indicated types are given in Table I.

The optimal parameters can be obtained using the GridSearchCV autotuning method [36] within a certain range of parameters. However, the verification of the data (see Table VI) demonstrate that the parameters of the results have little effect on accuracy. Here, MAE describes the true level of error, which was within 0.02 m. Therefore, the parameter has little effect on practical applications. ME is an accidental event that may vary with the interval and involves the junction of the sea surface and seabed signals, which generates random errors and does not affect the overall application.

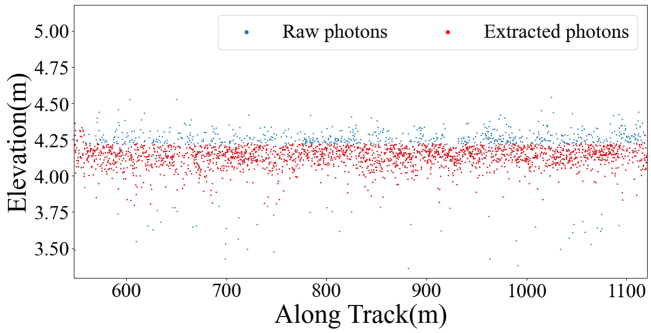


Fig. 7. Misclassification when there are no photons above the sea surface. Blue dots are photon events for all confidence levels, red dots are photon events under this condition.

V. Conclusion

The sea surface signal extraction method for photon-counting LiDAR with DSUMP is applicable to data with different densities and SNR, and achieves automatic extraction without experience parameters. The DSUMP method combines a Gaussian fitting curve and histogram statistics to construct an unmixing parameter related to the Gaussian coverage area to separate the noise and signals. The ME, MAE, and RLT series indicators quantitatively evaluated the sea-surface signal extraction results by calculating the median signal photon elevation. The RLT01 was as high as 97%, and the MAE was within 0.01 m in all conditions, demonstrating that the method has high precision and strong stability. This method can accurately extract sea surface signal photons and provides a new denoising technique for data with such characteristics. Considering its stability, the DSUMP method can potentially find widespread application.

Although the accuracy and stability of the method meet the application requirements, the efficiency is not high because the method fits Gaussian curves at all bins, and the program runs at single-core and single-process levels. Technical means such as multicore and multiprocess will be adopted in the future to improve efficiency. Eight categories and 16 datasets were used to verify the accuracy of the proposed method. Although the range of the SNR of the data involved is wide and the categories include all types of conditions, further application of more data is still necessary to prove the applicability of this method. The method will provide a new process for sea surface photons of ICESat-2 ATL03 and other data with similar distribution.

Acknowledgment

The authors thank the NASA ICESat-2 team for providing the data used in this article.

References

[1] R. G. Dean and J. R. Houston, "Recent sea level trends and accelerations: Comparison of tide gauge and satellite results," *Coastal Eng.*, vol. 75, pp. 4–9, 2013.

[2] M. Martin-Neira, M. Caparrini, J. Font-Rossello, S. Lannelongue, and C. S. Vallmitjana, "The PARIS concept: An experimental demonstration of sea surface altimetry using GPS reflected signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 142–150, Jan. 2001, doi: 10.1109/36.898676.

[3] W. Li, E. Cardellach, F. Fabra, S. Ribo, and A. Rius, "Assessment of spaceborne GNSS-R ocean altimetry performance using CYGNSS mission raw data," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 238–250, Jan. 2020, doi: 10.1109/TGRS.2019.2936108.

[4] Y. Zhang, L. Tian, W. Meng, Q. Gu, Y. Han, and Z. Hong, "Feasibility of code-level altimetry using coastal BeiDou reflection (BeiDou-R) setups," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 4130–4140, Aug. 2015, doi: 10 1109/JSTARS.2015.2446684.

[5] M. P. Clarizia and C. S. Ruf, "Wind speed retrieval algorithm for the cyclone global navigation satellite system (CYGNSS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4419–4432, Aug. 2016, doi: 10.1109/TGRS.2016.2541343.

[6] M. Martin-Neira, S. D'Addio, C. Buck, N. Floury, and R. Prieto-Cerdeira, "The PARIS ocean altimeter in-orbit demonstrator," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2209–2237, Jun. 2011, doi: 10.1109/TGRS.2010.2092431.

[7] C. Ruf et al., "CYG-NSS:Enabling the future of hurricane prediction remote sensing satellites," *IEEE Geosci. Remote Sens.*, vol. 1, no. 2, pp. 52–67, Jun. 2013, doi: 10.1109/MGRS.2013.2260911.

[8] R. S. Nerem, B. D. Beckley, J. T. Fasullo, B. D. Hamlington, D. Masters, and G. T. Mitchum, "Climate-change-driven accelerated sea-level rise detected in the altimeter era," in *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 9, pp. 2022–2025, 2018.

[9] S. Chander and T. J. Majumdar, "Comparison of SARAL and Jason-1/2 altimetry-derived geoids for geophysical exploration over the Indian offshore," *Geocarto Int.*, vol. 31, no. 2, pp. 158–175, 2016.

[10] K. Guerreiro, S. Fleury, E. Zakharova, A. Kouraev, F. Remy, and P. Maisongrande, "Comparison of CryoSat-2 and ENVISAT radar freeboard over Arctic sea ice: Toward an improved Envisat freeboard retrieval," *Cryosphere*, vol. 11, no. 5, pp. 2059–2073, 2017.

[11] T. A. Neumann et al., "The Ice, Cloud, and Land Elevation Satellite-2 Mission: A global geolocated photon product derived from the advanced topographic laser altimeter system," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111325.

[12] T. A. Neumann et al., "ATLAS/ICESat-2 L2A global geolocated photon data, version 5," 2021.

[13] J. H. Morison et al., "ATLAS/ICESat-2 L3A ocean surface height, version 5," 2021.

[14] M. Bagnardi, N. T. Kurtz, A. A. Petty, and R. Kwok, "Sea surface height anomalies of the arctic ocean from ICESat-2: A first examination and comparisons with CryoSat-2," *Geophys. Res. Lett.*, vol. 48, no. 14, 2021, Art. no. e2021GL093155.

[15] B. Nilsson, O. B. Andersen, H. Ranndal, and M. L. Rasmussen, "Consolidating ICESat-2 ocean wave characteristics with CryoSat-2 during the CRYO2ICE campaign," *Remote Sens.*, vol. 14, no. 6, 2022, Art. no. 1300.

[16] B. W. Klotz, A. Neuenschwander, and L. A. Magruder, "High-resolution ocean wave and wind characteristics determined by the ICESat-2 land surface algorithm," *Geophys. Res. Lett.*, vol. 47, no. 1, 2020, Art. no. e2019GL085907.

[17] X. M. Lu et al., "Ocean subsurface study from ICESat-2 mission," in *Proc. Photon. Electromagn. Res. Symp.-Fall*, 2019, pp. 910–918.

[18] Y. Yao et al., "Assessment of ICESat-2 for the recovery of ocean topography," *Geophys. J. Int.*, vol. 226, no. 1, pp. 456–467, 2021.

[19] J. S. Zhang 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.

[20] H. Xie et al., "A comparison and review of surface detection methods using MBL, MABEL, and ICESat-2 photon-counting laser altimetry data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7604–7623, 2021, doi: 10.1109/JSTARS.2021.3094195.

[21] Y. F. Chen, Y. Le, D. F. Zhang, Y. Wang, Z. G. Qiu, and L. Z. Wang, "A photon-counting LiDAR bathymetric method based on adaptive variable ellipse filtering," *Remote Sens. Environ.*, vol. 256, 2021, Art. no. 112326.

[22] H. J. Hsu et al., "A semi-empirical scheme for bathymetric mapping in shallow water by ICESat-2 and Sentinel-2: A case study in the South China Sea," *Int. Soc. Photogramm. Remote Sens. J. Photogramm. Remote Sens.*, vol. 178, pp. 1–19, 2021.

[23] Y. Ma et al., "Satellite-derived bathymetry using the ICESat-2 LiDAR and Sentinel-2 imagery datasets," *Remote Sens. Environ.*, vol. 250, 2020, Art. no. 112047.

[24] C. D. Liu et al., "Accurate refraction correction-assisted bathymetric inversion using ICESat-2 and multispectral data," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4355.

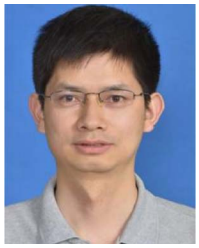
[25] H. Ranndal, P. S. Christiansen, P. Kliving, O. B. Andersen, and K. Nielsen, "Evaluation of a statistical approach for extracting shallow water bathymetry signals from ICESat-2 ATL03 photon data," *Remote Sens.*, vol. 13, no. 17, 2021, Art. no. 3548.
[26] C. S. Xie, P. Chen, D. L. Pan, C. Y. Zhong, and Z. H. Zhang, "Improved filtering of ICESat-2 LiDAR data for nearshore bathymetry estimation using Sentinel-2 imagery," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4303.
[27] B. Chen and P. Yong, "A denoising approach for detection of canopy and ground from ICESat-2's airborne simulator data in Maryland, USA," in *Proc. Appl. Opt. Photon.*, 2015, pp. 383–387.
[28] M. Ankerst et al., "OPTICS: Ordering points to identify the clustering structure," in *Proc. Assoc. Comput. Machinery SIGMOD, Int. Conf. Manage. Data*, 1999, pp. 49–60.
[29] M. Awadallah, S. Ghannam, A. 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. Assessing Forest Ecosyst.*, 2013.
[30] P. M. Atkinson, C. Jeganathan, J. Dash, and C. Atzberger, "Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology," *Remote Sens. Environ.*, vol. 250, 2012, Art. no. 112047.
[31] P. J. Diggle, P. Moraga, B. Rowlingson, and B. M. Taylor, "Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm," *Stat. Sci.*, vol. 28, no. 4, pp. 542–563, 2013.
[32] C. Proust-Lima, V. Philipps, and B. Liquet, "Estimation of extended mixed models using latent classes and latent processes: The R package LCMM," *J. Stat. Softw.*, vol. 78, no. 2, pp. 1–56, 2017.
[33] Z. Lee et al., "Hyperspectral remote sensing for shallow waters I. A semianalytical model," *Appl. Opt.*, vol. 37, no. 27, pp. 6329–6338, 1998.
[34] Z. Lee et al., "Hyperspectral remote sensing for shallow waters: 2 deriving bottom depths and water properties by optimization," *Appl. Opt.*, vol. 38, no. 18, pp. 3831–3843, 1999.
[35] M. Hahsler, M. Piekenbrock, and D. Doran, "DBSCAN: Fast density-based clustering with R," *J. Stat. Softw.*, vol. 91, no. 1, pp. 1–30, 2019.
[36] F. Pedregosa et al., "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

Zhen Wen received the B.S. degree in geography information systems in 2018 from the Shandong University of Science and echnology, Qingdao, China, where he is currently working toward the Ph.D. degree in photogrammetry and remote sensing.

He has been studying in the field of remote sensing and geographic information since 2018. His research interests include satellite laser altimetry data processing, optical image processing, and computer vision.

Xinming Tang received the B.S. degree from Nanjing University, Nanjing, China, in 1987, and the M.S. and Ph.D. degrees in geo-information science and computer application from the ITC Faculty of Twente University, Enschede, The Netherlands, in 1998 and 2004, respectively.

He is currently the Chief Engineer of Land Satellite Remote Sensing Application Center, Ministry of Natural Resources of China. He is the Chief Designer of the application system of satellite ZY-3 and the Deputy Chief Designer of the application system of China High-resolution Earth Observation System. He was the President of Commission I of the International Society for Photogrammetry and Remote Sensing and the Deputy Director-General of Satellite Surveying and Mapping Application Center, National Administration of Surveying, Mapping, and Geo-information of China. He is currently the Deputy Director-General of the GEO China Secretariat and a member of the GEO Program Board.



Guoyuan Li received the B.Sc. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2007, the M.Sc. degree in photogrammetry and remote sensing from the Graduate School of Chinese Academy of Sciences, Beijing, China, in 2010, and the Ph.D. degree in cartography and geographic information engineering from Wuhan University, Wuhan, China, in 2017.

He is currently an Associate Researcher with the Land Satellite Remote Sensing Application Center, Ministry of Natural Resources, China. His research interests include satellite laser altimetry and spaceborne photogrammetry and mapping.

Bo Ai received the B.S. and M.S. degrees in geography information system from Wuhan University, Wuhan, China, in 2001 and 2005, respectively, and the Ph.D. degree in geography information systems from the Shandong University of Science and Technology, Qingdao, China, in 2011.

He is currently the Director of the Geographic Information Department, Shandong University of Science and Technology. His research interests include ocean spatial-temporal modeling, maritime search, and rescue decision analysis.

Guanghui Wang received the B.S. degree in surveying and mapping engineering from Shandong Jianzhu University, Jinan, China, in 2006, and the M.S. degree in surveying and mapping engineering from China University of Mining and Technology, Xuzhou, China, in 2009.

He is currently an Assistant Research Fellow of the Land Satellite Remote Sensing Application Center, Ministry of Natural Resources, China. His research interests include image information extraction theory and method, remote sensing image information identification and extraction, spatial data mining.

Jiaqi Yao received the B.Sc. degree in surveying and mapping engineering from North China University of Science and Technology, Tangshan, China, in 2017, the M.Sc. degree in photogrammetry and remote sensing from Liaoning Technical University, in 2019, and the Ph.D. Degree in photogrammetry and remote sensing from College of Geomatics, Shandong University of Science and Technology, Qingdao, China, in 2022.

He is currently a Research Assistant with the Tianjin Normal University, Tianjin, China. His research interests include the application of deep learning in satellite LiDAR.

Fan Mo received the B.S. degree in surveying and mapping engineering from Shandong University of Technology, Zibo, China, in 2009, and the M.S. degree in surveying and mapping engineering from the Information Engineer University, China, in 2013.

He is currently an Assistant Research Fellow of LASAC, and employed in the Calibration Department. He has authored the laser altimeter calibration software. His research interest focuses on spaceborne laser altimetry. He also does some research in satellite attitude processing, including jitter frequency detection by different methods and attitude data correction.