

An Automatic Algorithm to Extract Nearshore Bathymetric Photons Using Pre-Pruning Quadtree Isolation for ICESat-2 Data

Guoping Zhang^{id}, Shuai Xing, Qing Xu^{id}, Fubing Zhang, Mofan Dai, and Dandi Wang^{id}

Abstract—The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) equips with a novel photon-counting LiDAR system, which can generate underwater reflections in nearshore environments. However, due to the water reflection, scattering, and absorption, the distribution of bathymetric photons in the nearshore data varies with depth. The existing bathymetric photon extraction algorithms need more adaptability to seafloor topography. The changing density of bathymetric photons and the fluctuation of underwater topography make the noise removal of nearshore data full of challenges. This study proposed a bathymetric photon extraction algorithm using pre-pruning quadtree isolation (PQI). First, the pre-pruning step judges whether to stop the growth of quadtree in advance during quadtree isolation (QI) to avoid excessive division of noise photons. Second, the maximum inter-class variance algorithm (also called the Otsu method) obtains the best threshold of isolation depth (ID) and extracts bathymetric photons. The algorithm was tested on the Florida coast. The results show that the PQI algorithm can wholly and accurately extract bathymetric photons with different acquisition times from the data. The F1-score of the extracted results is 93.96%. This study provides an intelligent solution to processing bathymetric data in nearshore environments worldwide.

Index Terms—Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2), nearshore environment, photon-counting LiDAR, pre-pruning quadtree isolation (PQI), signal extraction.

I. INTRODUCTION

INFLUENCED by natural and human activities, accurately measuring the underwater topography in a nearshore environment is challenging. Multi band images can retrieve the relative depth according to the radiation transmission equation but depend on the in situ measurement data to provide the absolute water depth, which is blank on many islands and reefs [1]. LiDAR uses a blue-green laser pulse, which can penetrate the water and map underwater spatial structures

directly with sub-meter precision [2]. Besides synthetic aperture radar (SAR) and sonar, it is an effective way to survey an area with a depth of less than 5 m using LiDAR [3]. However, the limitations of the airborne laser system (ALS) in spatial-temporal resolution restrict the further application of this technology.

In 2018, the National Aeronautics and Space Administration (NASA) launched the new generation LiDAR satellite, which is named the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) [4]. Its photon-counting LiDAR called Advanced Topographic Laser Altimeter System (ATLAS), equips with three pairs of laser beams (including one strong beam and one weak beam), which can emit laser pulses of 532 nm. The laser footprint obtained by ICESat-2 is 13 m in diameter, and the footprint interval is only 0.7 m [5], which is much higher than the previous LiDAR satellite missions. The researchers unexpectedly found that ICESat-2 can record the reflected signals from underwater up to 40 m [6]. This new data provides a rich opportunity for satellite-derived bathymetry (SDB) in global nearshore environments.

The high sensitivity of ATLAS cannot prevent noise photons from flooding the raw data. In the nearshore environment, the laser pulse is reflected, scattered, and absorbed by water during transmission, which makes the spatial characteristics of the photons different from other types [7]. Specifically, bathymetric photons are composed of sea surface photons and seafloor photons. The spatial density of seafloor photons decreases with depth and is also affected by water turbidity. At present, the bathymetric photon extraction algorithm fits the sea surface by using the wave spectrum [8] and filters underwater noises by clustering the density of seafloor photons through the neighborhood with adaptive size and density threshold [9], [10]. The algorithms have been tested, but the horizontally placed elliptical neighborhood makes them theoretically unable to adapt to the underwater terrain. Furthermore, with the increased water depth, the noise photons would lead to potentially wrong extraction.

Zhang et al. [11] proposed a noise removal algorithm based on quadtree isolation (QI), which isolated each photon through quad-spatial division and extracted the signal photon according to the isolation depth (ID). It works without input parameters and has strong adaptability to different topographies. Unfortunately, this algorithm is not designed for nearshore data and cannot identify noise photons near the seafloor.

Manuscript received 7 November 2022; revised 3 March 2023; accepted 28 March 2023. Date of publication 31 March 2023; date of current version 10 April 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 41876105 and Grant 41371436. (Corresponding author: Shuai Xing.)

Guoping Zhang is with the Institute of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China, and also with the Collaborative Innovation Center, Geo-Information Technology, Zhengzhou, Henan 450001, China (e-mail: zac Zhang1996@gmail.com).

Shuai Xing, Qing Xu, Fubing Zhang, Mofan Dai, and Dandi Wang are with the Institute of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China (e-mail: xing972403@163.com; 13937169139@139.com; zhangfubing@163.com; daidai_1115@163.com; wdd_93@163.com).

Digital Object Identifier 10.1109/LGRS.2023.3263469

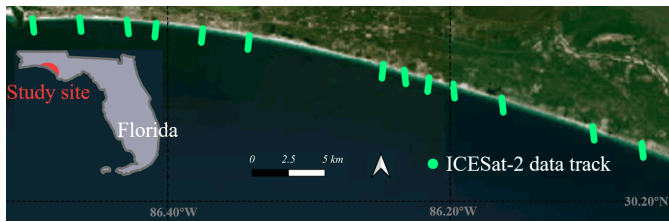


Fig. 1. Study site in Florida. The green lines show the ICESat-2 data tracks.

TABLE I
ACQUISITION DATE AND TIME OF ICESAT-2 ATL03 DATASETS

Date	Time	Latitude range (°N)
2019-01-28	Daytime	30.2720 - 30.2945
2019-04-29	Daytime	30.2720 - 30.3066
2019-06-16	Daytime	30.3310 - 30.3505
2019-09-29	Nighttime	30.3718 - 30.3835
2019-10-28	Nighttime	30.3165 - 30.3465
2019-11-16	Daytime	30.3608 - 30.3795

Considering the spatial characteristics of bathymetric photons, pre-pruning QI (PQI) was developed. Different from the traditional QI, pre-pruning decides whether to stop the growth of quadtree in advance by judging whether photons are isolated before and after each space division, thus enhancing the recognition ability of noise photons. The maximum inter-class variance algorithm (also called the Otsu method) is used to automatically determine the threshold of ID and extract bathymetric photons from the raw data as well as possible.

II. STUDY AREA AND DATA

A. Study Site

The study site is on the coast between Destin and Panama (86°01'W–86°31'W, 30°15'N–30°24'N), Northern West Florida. The water body is clear, which enables ICESat-2 to record the bathymetric topography with a depth of more than 15 m within several hundred meters along the track. The east-west rocky coastline makes the ICESat-2 data easy to collect, and the water's backscattering effect on laser pulses makes the data contain massive noise photons, which poses a challenge.

B. Test Data

The longitude, latitude, and elevation of photons acquired by ATLAS are recorded in ICESat-2 ATL03 data according to laser beams (numbered gt1r, gt1l, gt2r, gt2l, gt3r, and gt3l). In this study, six pieces of data that passed through the study site in 2019 were collected, in which the strong beams were used for verification. Two-thirds of the data were acquired in the daytime, and the rest were acquired at night. These data were cut into ~ 1 km segments in advance, as shown in Fig. 1, and the data details are shown in Table I.

III. ALGORITHMS

A. Pre-Pruning QI

The spatial distribution of signal photons is denser than noise photons due to the difference in reflectivity, so the

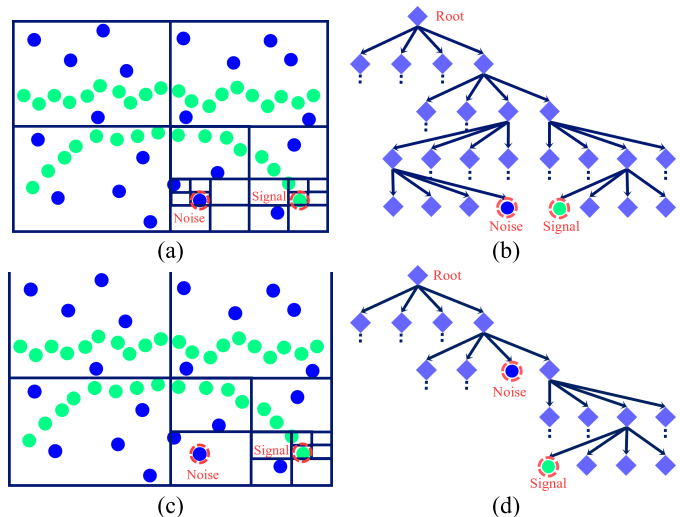


Fig. 2. Examples illustrating bathymetric scenario of ICESat-2 data: (a) error caused by traditional QI, (b) quadtree generated by traditional QI where the purple diamonds represent the tree nodes, (c) result generated by PQI, and (d) quadtree generated by PQI.

noises can be removed by describing the spatial distribution of photons. Unlike the algorithm based on density clustering, traditional QI regards the noise removal process as photon isolation. The photons are separated from the surrounding photons by recursive spatial divisions. The method of photon isolation corresponds to a quadtree, and each isolated photon is in a different layer of the quadtree. The layer of quadtree where each photon locates is called ID. Since isolating signal photons needs more times of space division than separating noise photons, the ID of signal photons is also greater. Therefore, signal photons can be extracted from the raw data by setting the threshold of ID through manual testing of binary classification algorithms.

Although traditional QI has achieved ideal results on different land surfaces [12], it may also lead to potential errors during the processing of nearshore data. In the coastal environment, bathymetric photons comprise sea surface and seafloor photons. Due to the reflection of the sea surface, the spatial distribution of sea surface photons is often more compact than that of the seafloor photons. The signal-to-noise ratio (SNR) underwater is lower than that of the above-water environment due to water scattering and the absorption of laser pulses. With the increased water depth, the SNR will be further reduced, and the noise photons will become challenging to distinguish, resulting in the errors of traditional QI.

A typical example is shown in Fig. 2(a), where the blue dots represent noise photons, green dots represent signal photons, and the dotted pink lines highlight the analyzed photons. Although the noise photon in the figure is far from the seafloor photon, to isolate it from the nearest photon (which is also a noise photon), both the noise photon and the sea floor photon were isolated after four times division. As shown in Fig. 2(b), they are all at the fourth level of the quadtree and cannot be distinguished by ID. The traditional QI performed two times of weak division (the second and third division) because these two times of division failed to isolate the noise photon from its

nearest neighbor. The traditional algorithm only separates the noise photon from the nearest photon in the fourth division leading to the overestimation of the noise photon. With the increase of water depth, the noise photons caused by water scattering and the increasingly reduced seafloor photons make this kind of situation not uncommon.

The traditional QI divides photons according to the position of the nearest neighbor photon. Space division ignores the position of noise photons in the whole nearshore environment, thus overestimating the spatial density of underwater noise photons. Therefore, pre-pruning is introduced to improve this situation [13]. The core idea of pre-pruning is to judge whether there are noise photons near bathymetric photons before and after the spatial division is further divided before each time of quadtree growth. The pre-pruning step will be considered if the photons are not further divided into sub-regions. To avoid excessive division of noise photons, the growth of the isolated quadtree will be stopped in advance, and no spatial division will be carried out. Therefore, when using a pre-pruning quadtree to extract bathymetric photons, the conditions for stopping spatial division are as follows:

1) The number of photons contained in the space before the division is the same as that in a subspace after division, which means the photons are not further divided.

2) The number of photons included in the divided subspace is one or zero, which means the photons have been isolated.

When PQI isolates the same noise photon [see Fig. 2(c)], it is found that the next time division cannot separate the photons, so after the second space division, the quadtree is cut off and stops growing. In the pre-pruning quadtree, the noise photon is at the second layer while the signal photon is still at the fourth layer. The pre-pruning step amplified the difference between signal and noise.

B. Bathymetric Photon Extraction Using Otsu Method

Since the ID of bathymetric photons is always greater than that of noise photons, the Otsu method can measure the variance of ID [13], and the ID with the maximum variance can be selected as the threshold value for extracting bathymetric photons. Assuming that n is the photon number and t is the number of potential signal photons, the Otsu method is used to calculate according to the following:

$$\sigma^2 = \omega_b(t)(\mu_b(t) - \mu(t))^2 + \omega_n(t)(\mu_n(t) - \mu(t))^2 \quad (1)$$

where σ^2 is the interclass variance, $\omega_b(t)$ is the proportion of potential signal photons, $\omega_n(t)$ is the proportion of potential noise photons, $\mu_b(t)$ represents the average ID of signal photons, $\mu_n(t)$ represents the average ID of potential noise photons, and $\mu(t)$ represents the average ID of all photons, as follows:

$$\omega_b(t) = \frac{t}{n} \quad (2)$$

$$\omega_n(t) = 1 - \frac{t}{n} \quad (3)$$

$$\mu_b(t) = \frac{\sum_{i=1}^t ID_i}{t} \quad (4)$$

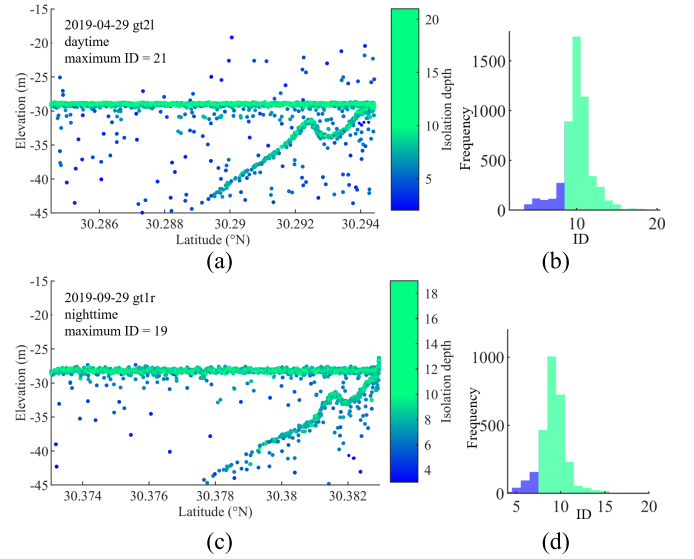


Fig. 3. Results of ID. (a) and (b) are ID and its histogram of daytime data. (c) and (d) are ID and its histogram of nighttime data.

$$\mu_n(t) = \frac{\sum_{i=t+1}^n ID_i}{n-t} \quad (5)$$

$$\mu(t) = \omega_b(t)\mu_b(t) + \omega_n(t)\mu_n(t). \quad (6)$$

By changing t , $\mu_b(t)$, and $\mu_n(t)$ are recalculated, and σ^2 is renewed. When σ^2 becomes the largest, the ID is selected as the threshold, and the photons with an ID greater than the threshold are extracted as bathymetric photons.

IV. RESULTS AND DISCUSSION

A. Verification of ID

When a pre-pruning quadtree is used to process the raw data, photons are continuously divided until the number of photons in the window before and after the division does not change. Record the ID of each photon, which corresponds to the layer in the quadtree. Due to the different spatial distribution, signal photons' ID is larger than noise photons. Fig. 3(a) and (c) show the IDs of daytime and nighttime data, respectively, in which the date, track number, and other details are marked in the upper left corner. The greater the ID, the greener the photons, the smaller the ID, and the bluer the photons.

From the results, it can be found that the bathymetric photons are greener than noise photons. Specifically, the sea surface photons are the greenest, indicating that the spatial distribution of these photons is the closest. With the increase of water depth, the color of seafloor photons gradually changes from green to blue. Due to the absorption of laser energy and the influence of water scattering, the SNR of nearshore data is decreasing. In this way, when the elevation is less than -40 m (at this time, the water depth is more than 10 m), the photon distribution on the seafloor is sparse, and the ID of seafloor photons at this time can hardly be distinguished from that of noise photons. However, most seafloor photons were accurately extracted. The IDs calculated by the PQI algorithm can effectively distinguish bathymetric photons from noise photons.

Fig. 3 also shows the ID histogram and the thresholds obtained by the Otsu method are marked (the ID threshold of daytime data is 8, while that of nighttime data is 7). For different acquisition times, the number of photons in daytime data is more than that in nighttime data, which shows that under the action of sunlight, more background noises and backscattering noises are recorded, and the reflection effect of water surface and underwater topography is more substantial. Compare the ID value corresponding to the maximum frequency in the histogram. When the ID value of daytime data is 10, the maximum frequency is more than 1500, while the maximum frequency of the nighttime histogram is less than 1000 when the ID value is 9. Because the ID value is only related to the number of times, the photon is divided. The spatial photon distribution of daytime data is dense, so it must go through more quad-space division when segmenting it.

After calculating each photon's ID, the ID threshold is computed using by Otsu method. The extracted results are shown in Fig. 4. The results show that the sea surface photons are entirely and correctly extracted. Additionally, the Otsu method extracted most seafloor photons in shallow water, but there are some omissions while extracting the seafloor photons in the deep water. In Fig. 3(c), the seafloor photons at 30.379°N are not extracted because their spatial distribution is sparse and their IDs are small, so they cannot be distinguished from the noise photons in the data from the perspective of ID. Since dozens of seafloor photons only account for a small part of the nearshore environment, this does not mean that using the Otsu method to calculate the threshold is invalid. The results show that the isolated depth calculated by the Otsu method can extract bathymetric photons from the raw data. However, the spatial distribution of seafloor photons in deep water areas challenges marking bathymetric photons.

Besides, because the isolated depth obtained by spatial division measures the photon distribution and avoids the elliptical density neighborhood used in [9] and [10], the PQI algorithm is adaptable to seafloor topography. The results in Figs. 3 and 4 show that the changes in seafloor topography do not affect the ID of photons, and the bathymetric photons under different topography are extracted.

Compare the results obtained by PQI and QI in Fig. 4. QI also removes noise from the air, but when processing underwater data, some noise photons near the sea surface and underwater terrain photons are not recognized. Although the proportion of these noise photons in all photons is low, their randomness in spatial position makes it challenging to retrieve water depth from ATLO3 data. These noise photons are effectively identified and removed from the PQI results, which shows that introducing the pre-pruning step gives quad space segmentation the ability to actively identify small noise clusters and ensure the accuracy of the extracted sounding photons.

Quantitative verification is also made. Because ATL 03 does not provide signal labels in the nearshore environment, some studies choose in situ data to test the performance. However, this cannot explain whether the photons are fully extracted, so the reference bathymetric photons are

TABLE II
PERFORMANCE EVALUATION INDEXES FOR DATA WITH DIFFERENT ACQUISITION TIME

Method	Time	P (%)	R (%)	$F1$ (%)
PQI	Daytime	95.87	90.64	93.18
	Nighttime	96.70	92.55	94.58
	All	96.12	91.88	93.96
QI	Daytime	93.70	87.64	90.56
	Nighttime	92.87	89.55	91.18
	All	93.16	88.89	90.96

available by visual interpretation (uploaded to the website, https://drive.google.com/drive/folders/19WfWCsxYuRL5Jp2hAqDkH1NbnCOF_eBd?usp=sharing). The results are compared with the reference data, and precision P , recall rate R , and F1-score $F1$ are calculated

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (7)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (8)$$

$$F1 = \frac{2P \cdot R}{P + R} \quad (9)$$

where N_{TP} indicates the number of correctly extracted photons, N_{FP} indicates the number of incorrectly extracted photons, and N_{FN} indicates the number of bathymetric photons that have not been extracted. Therefore, P measures the reliability of the extraction algorithm, R represents the completeness of the results, and $F1$ measures the comprehensive performance of the extraction algorithm.

Table II shows the evaluation indexes of the results. The P is higher than 96%, indicating that the extraction algorithm can extract most bathymetric photons. The R is close to 92%, slightly lower than the P , showing that some of the bathymetric photons are undetected. This finding is consistent with the above qualitative analysis, and the ability to detect bathymetric photons is slightly weaker than the ability to extract photons correctly. The $F1$ is close to 94%, higher than the $F1$ of the QI result, which indicates that the proposed algorithm can extract wholly and correctly, and the experimental results are ideal.

B. Effect of Data Acquisition Times

The data acquired at different times have different SNRs. We distinguish the data according to the acquisition time to understand the effect of SNR on bathymetric photon extraction. Fig. 3 shows the daytime data and nighttime data IDs. The maximum ID of data in the daytime is greater than that of data in the nighttime. This result is because there are more noise photons in daytime data, which requires more space division to isolate the photons. Solar background noises and water scattering noises caused by sunlight are all over the daytime data, so it is more difficult to accurately extract the bathymetric photons from the daytime data than from the nighttime data.

The indexes confirm this result in Table II. The extraction results of nighttime data are better than those of daytime data

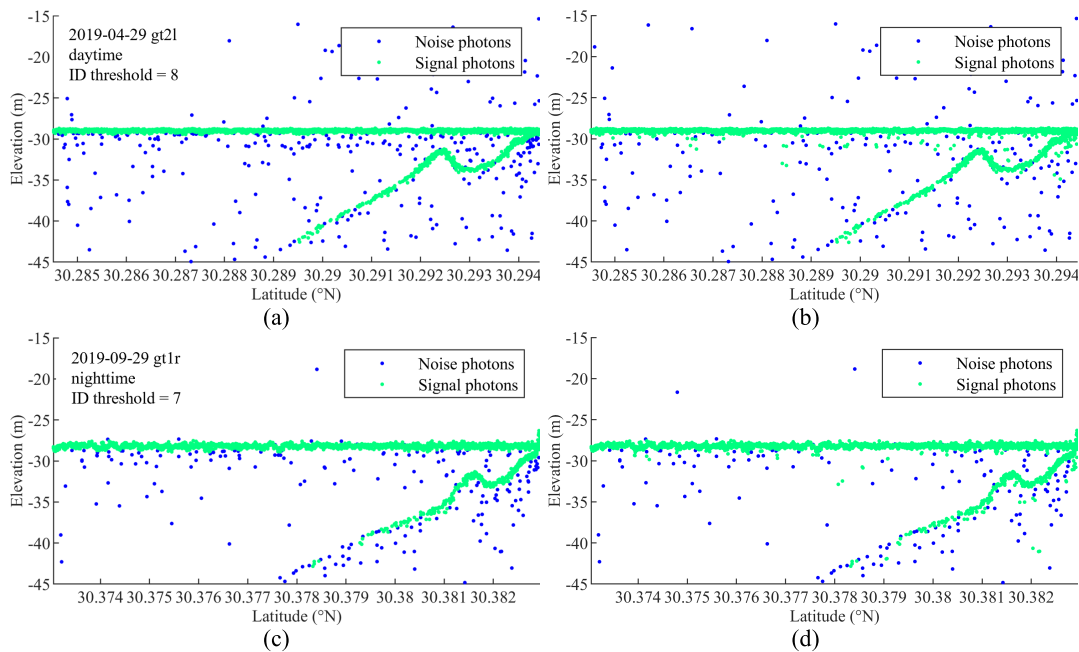


Fig. 4. Results of bathymetric photon extraction: (a) PQI result of daytime data, (b) QI result of daytime data, (c) PQI result of nighttime data, and (d) QI result of nighttime data.

in all indexes, which indicates that when the SNR of data is lower, the bathymetric photon extraction results are better. Further comparing the results of different acquisition times, the difference of three indexes between daytime and nighttime results is less than 2%, and the SNR has limited effect during extraction. The proposed PQI is insensitive to the SNR.

V. CONCLUSION

ICESat-2 has excellent potential in global bathymetry and has become an important data source for nearshore research. To automatically extract the bathymetric photons in ICESat-2 bathymetric data, PQI is proposed. The spatial distribution of photons is transformed into a pre-pruning quadtree, and the position of photons corresponds to the IDs. The Otsu method obtains the threshold and extracts the photons. Through qualitative and quantitative analysis, it is found that the PQI algorithm can wholly and accurately extract bathymetric photons. The influence of the acquisition times is minimal.

The PQI can be used to automatically process nearshore data, thus significantly reducing the demand for human resources. In the future, we will conduct studies in more challenging areas and introduce rough noise removal and post-processing steps into bathymetric photon extraction.

ACKNOWLEDGMENT

The authors are very grateful for the ICESat-2 data provided by NASA Earthdata.

REFERENCES

- [1] M. Niroumand-Jadidi, F. Bovolo, and L. Bruzzone, "SMART-SDB: Sample-specific multiple band ratio technique for satellite-derived bathymetry," *Remote Sens. Environ.*, vol. 251, Dec. 2020, Art. no. 112091, doi: [10.1016/j.rse.2020.112091](https://doi.org/10.1016/j.rse.2020.112091).
- [2] C. S. Gardner, "Ranging performance of satellite laser altimeters," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 1061–1072, Sep. 1992, doi: [10.1109/36.175341](https://doi.org/10.1109/36.175341).
- [3] N. A. Forfinski-Sarkozi and C. E. Parrish, "Active-passive spaceborne data fusion for mapping nearshore bathymetry," *Photogramm. Eng. Remote Sens.*, vol. 85, no. 4, pp. 281–295, Apr. 2019, doi: [10.14358/pers.85.4.281](https://doi.org/10.14358/pers.85.4.281).
- [4] B. Smith et al., "Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111352, doi: [10.1016/j.rse.2019.111352](https://doi.org/10.1016/j.rse.2019.111352).
- [5] 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, Nov. 2019, Art. no. 111325, doi: [10.1016/j.rse.2019.111325](https://doi.org/10.1016/j.rse.2019.111325).
- [6] P. Voosen, "Ice monitor delivers a bonus: Seafloor maps," *Science*, vol. 368, no. 6488, p. 224, Apr. 2020, doi: [10.1126/science.368.6488.224](https://doi.org/10.1126/science.368.6488.224).
- [7] N. Forfinski-Sarkozi and C. Parrish, "Analysis of MABEL bathymetry in Keweenaw Bay and implications for ICESat-2 ATLAS," *Remote Sens.*, vol. 8, no. 9, p. 772, Sep. 2016, doi: [10.3390/rs8090772](https://doi.org/10.3390/rs8090772).
- [8] 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](https://doi.org/10.1364/oe.26.024752).
- [9] Y. Ma et al., "Satellite-derived bathymetry using the ICESat-2 LiDAR and Sentinel-2 imagery datasets," *Remote Sens. Environ.*, vol. 250, Dec. 2020, Art. no. 112047, doi: [10.1016/j.rse.2020.112047](https://doi.org/10.1016/j.rse.2020.112047).
- [10] Y. Chen, Y. Le, D. Zhang, Y. Wang, Z. Qiu, and L. Wang, "A photon-counting LiDAR bathymetric method based on adaptive variable ellipse filtering," *Remote Sens. Environ.*, vol. 256, Apr. 2021, Art. no. 112326, doi: [10.1016/j.rse.2021.112326](https://doi.org/10.1016/j.rse.2021.112326).
- [11] G. Zhang et al., "A noise-removal algorithm without input parameters based on quadtree isolation for photon-counting LiDAR," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021, doi: [10.1109/LGRS.2021.3081721](https://doi.org/10.1109/LGRS.2021.3081721).
- [12] H. Xie et al., "Converting along-track photons into a point-region quadtree to assist with ICESat-2-based canopy cover and ground photon detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102872, doi: [10.1016/j.jag.2022.102872](https://doi.org/10.1016/j.jag.2022.102872).
- [13] M. Last and O. Maimon, "A compact and accurate model for classification," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 2, pp. 203–215, Feb. 2004, doi: [10.1109/TKDE.2004.1269598](https://doi.org/10.1109/TKDE.2004.1269598).
- [14] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979, doi: [10.1109/tsmc.1979.4310076](https://doi.org/10.1109/tsmc.1979.4310076).