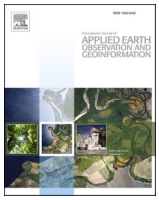
International Journal of Applied Earth Observation and Geoinformation 124 (2023) 103503

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Waveform classification and index sharing based-effective intensity data extraction of airborne laser bathymetry

Xue Ji [a,b,c], Zhen Dong [b], Lin Zhang [d], Mingchang Wang [a,*], Yi Ma [e], Jingyu Zhang [e], Xueping Lin [f], Jonathan Li [g]

[a] College of Geoexploration Science and Technology, Jilin University, Changchun, JL 130026, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, HB 430079, China
[c] Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources, Guangzhou, GD 510000, China
[d] School of Mechatronic Engineering, Changchun University of Technology, Changchun, JL 130012, China
[e] Technology Innovation Center for Ocean Telemetry, Ministry of Natural Resources, Qingdao, SD 266061, China
[f] Island Research Center, Ministry of Natural Resources, Pingtan, FJ 350400, China
[g] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Airborne laser bathymetry
Full-waveform
Intensity extraction
Waveform classification

ABSTRACT

Alongside widely accepted geometric information, airborne laser bathymetry (ALB) typically captures the temporal profile sample (waveforms) of both the emitted laser pulses and their echoes. These waveforms provide radiometric properties (backscattering intensity) of sensed targets and assist with precise strip registration, classification of fine ground cover (sediment), and advanced geometric modelling. However, effective intensity data extraction is essential because the intensity provided by the LAS file is mixed with invalid signals, multiple echo intensity, and intermittent intensity variations caused by improper interference of automatic gain control, making it unusable for direct analysis. To address this issue, a flexible index sharing mechanism between waveform, coordinate, and intensity data is constructed for interference intensity tracking. A novel waveform classification-based approach is proposed to efficiently extract ALB intensity by dividing waveform data into six categories using morphological differences and topographic data of waveforms. To ensure accurate analysis, duplicate and invalid waveforms are eliminated, leaving only genuine intensity readings. Additionally, a triple spline fit is employed to restore oversaturated waveform segments that were previously suppressed due to exceeding the device's maximum measurable limit. To address the problem of mixing multiple return intensities, waveforms are decomposed using various decomposition models. This approach ensures that different waveform categories retain their respective varying return intensities. The approach is then tested on an ALB dataset collected using Optech Aquarius around Yuanzhi Island in the South China Sea. The results demonstrate a considerable advancement in data quality when compared to LAS file intensity products with a reduction in maximum deviation of 883 [digital number, DN], standard deviation of 350 [DN], and mean absolute error of 310 [DN].

1. Introduction

Airborne laser bathymetry (ALB) has emerged as a highly effective tool for mapping coastal zones on islands due to its potential for rapid and efficient acquisition of high-quality data (Lee, 2003; Ji et al., 2022; Guo et al., 2022). In contrast to traditional bathymetric measurements utilizing shipboard sonar, ALB is performed from an aircraft, eliminating the need for a vessel and reducing the risks of danger and damage to crew and equipment, especially in areas of shallow water (Ji et al., 2020; Pereira et al., 2015). Additionally, ALB instruments are equipped with full-waveform systems that extend the complete echo waveform of the emitted pulse and the backscattered echoes (Wang & Philpot, 2007). Therefore, ALB provides not only range data, but also an additional physical property: the strength of the backscattered signal, typically referred to as intensity. Both the 3D geometry and intensity of the measured object are generated by the laser round-trip process. The recorded intensity of each echo, recorded as spectral information, has been proven to be highly relevant in various applications, such as strip

* Corresponding author.
E-mail address: wangmc@jlu.edu.cn (M. Wang).

https://doi.org/10.1016/j.jag.2023.103503
Received 2 July 2023; Received in revised form 13 September 2023; Accepted 19 September 2023

adjustment, feature extraction, coastal environment monitoring, sediment classification or segmentation (Eren et al., 2018), surface modeling, biological habitats mapping (Tuell & Park, 2004; Zavalas et al., 2014), water quality inversion (Richter et al., 2017) as well as natural hazard assessment (Kashani et al., 2015; Long et al., 2011; Yan et al., 2015).

LiDAR systems have been widely studied and applied for gathering geometric information. However, there has been increasing interest in exploring the potential applications of intensity and full waveform data in recent years. To date, a significant amount of research has been conducted to investigate laser intensity in Airborne Laser Scanning (ALS) (Ahokas et al., 2006; Sevara et al., 2019; Wu et al., 2021) and Terrestrial Laser Scanner (TLS) (Errington et al., 2015; Fang et al., 2015; Kaasalainen et al., 2011), leaving a relative lack of studies addressing the intensity data obtained by ALB. The initial models for effective intensity correction in ALB, pioneered by Guenther et al. (Guenther, 1985) and Wang et al. (Wang & Philpot, 2002; Wang & Philpot, 2007), were based on a simplification of the LiDAR equation using physical principles. Since then, further refinements have been made to these models by incorporating a variety of environmental influence factors (Feygels et al., 2003; Kopilevich et al., 2005; Tuell et al., 2005). Apart from the traditional intensity correction models, (Abdallah et al., 2012) proposed a different approach that employed a comprehensive full waveform energy composition perspective to decompose the received power into five distinct components, included power returned from the surface, column, and bottom, as well as background power returned from the air column and detector noise power. Corrected intensities have been demonstrated to be highly effective in enhancing the attribute characterization of sensed targets and in offering significant potential for benthic habitat classification (Collin et al., 2008; Narayanan et al., 2009; Tamondong et al., 2020). Despite the promising results of corrected intensities, the extraction of intensity data in ALB has received limited attention. Interference stemming from multiple echo intensities, especially surface return intensities, coupled with incorrect Automatic Gain Control (AGC) interventions. It results in a significant level of noise being introduced into the intensity products, as visualized in Fig. 1. This poses a significant challenge to the analysis and correction of intensity data, reducing its usefulness for various applications. Most existing signal detection models for full-waveform LiDAR focus only on peak localization or coordinate calculation, disregarding invalid and interfered intensity data, which makes it difficult to achieve high-quality intensity image production.(Wang et al., 2015; Guenther et al., 2000; Ji et al., 2022; Guo et al., 2022). Therefore, it is essential to develop effective methods for extracting intensity data in ALB systems to mitigate these challenges and enable more accurate underwater sensing and classification.

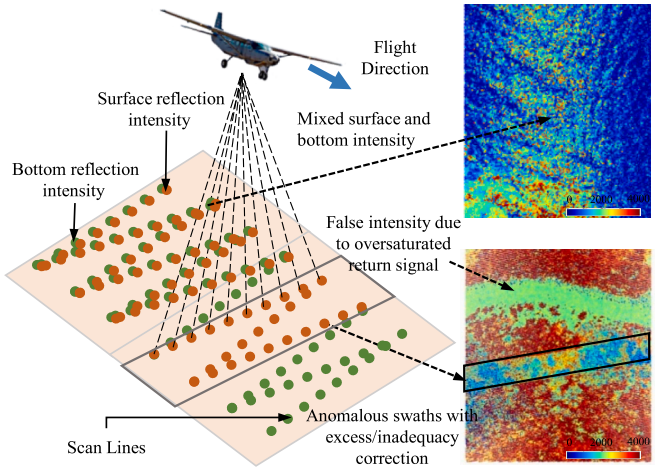


Fig. 1. A large amount of interference data is mixed into the intensity image.

In order to eliminate the mixed invalid and interfering intensity data in the ALB intensity images, an effective method for extracting intensity data from ALB is presented, which not only aids in the study of the transmission of laser radiation but also facilitates potential intensity data mining applications. The developed approach proposes a novel intensity censoring mechanism by integrating waveform classification and index sharing.

Our main contributions are summarized as follows:

- A flexible index sharing mechanism is designed to be implemented throughout the method. It enables one-to-one correspondence between waveform, terrain and intensity data, facilitating multiple echo intensity deletion, invalidation and interference intensity tracking.
- Fine classification of waveforms is proposed through topographic data assistance and multi-parameter thresholding to eliminate repetitive and invalid waveforms. Different processing modes are applied to each category to obtain the final intensity data of the target. This approach ensures that only the relevant echo data are considered in the intensity correction process.
- Intensity recovery based on waveform fitting and intensity offset statistics is performed for suppressed oversaturated waveform segments. The effectiveness of the new method is fully demonstrated by comparative experiments of test and setup on the ALB dataset collected by Optech Aquarius near zhiyuan island in the South China Sea.

The layout of this article is as follows. Section 2 introduces the ALB data and their distribution locations, while the effective intensity extraction method of ALB based on waveform classification is introduced in detail in Section 3. Section 4 evaluates the effectiveness of the proposed method with specific experiments and quantitative analysis, and Section 5 shows the technical and financial support received in the course of this study.

2. Study area and data

The data utilized in the study are acquired by an Optech Aquarius system on May 1, 2013 around Zhiyuan Island in the South China Sea, as show in Fig. 2. The airborne survey consisted of a total distance of approximately 4,800 m, covering an area of 3 Km^2 and generating about 2,440,000 sampling points. The Aquarius system, introduced in 2011, is a lightweight ALB system. It is compatible with the Optech's Airborne Laser Terrain Mapping (ALTM) systems and has the capability of conducting both terrestrial and underwater topographic surveys, with a maximum depth of 13 m (Guo et al., 2023). The predetermined operational altitude for the aircraft has been set at 300 m, along with a configured scanning nadir angle of 20° and a detection frequency of 550 kHz for the laser beam.

3. Method

3.1. Overview

After analyzing the sources of intensity noise in ALB systems, three broad types of interference are identified. The first type refers to invalid intensity data generated by invalid waveforms, the second includes suppressed or enhanced intensity data due to echo killer (handles oversaturated waveforms) and AGC intervention, and the last type pertains to interference intensity caused by multiple echoes, such as sea surface and plant cover area echoes. Further details on invalid waveforms and oversaturated waveforms will be presented in Subsections 3.3 and 3.4 as they contribute significantly to the ALB system interference noise and require special attention.

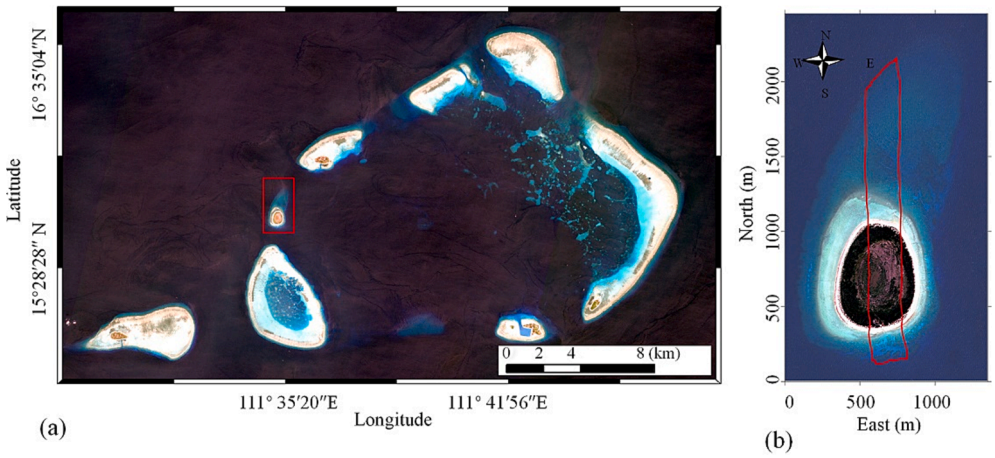


Fig. 2. Geographic location of data collection: (a) location of the sea area where the study area is located, and (b) enlarged view of the island where the data collection area is located.

This paper proposes a novel approach to intensity extraction based on the classification of waveforms, which has been developed through an analysis of the intensity noise distribution. The approach, depicted in Fig. 3, comprises multiple stages, including repeated waveform rejection, identification and rejection of invalid waveforms, identification and recovery of oversaturated waveforms, segmentation of bathymetric and land waveforms, waveform decomposition, and intensity detection. The proposed method is effective in mitigating the negative impact of interference noise from sources such as invalid and oversaturated waveforms, as well as multiple echoes, thereby improving the accuracy of intensity data obtained from the ALB system.

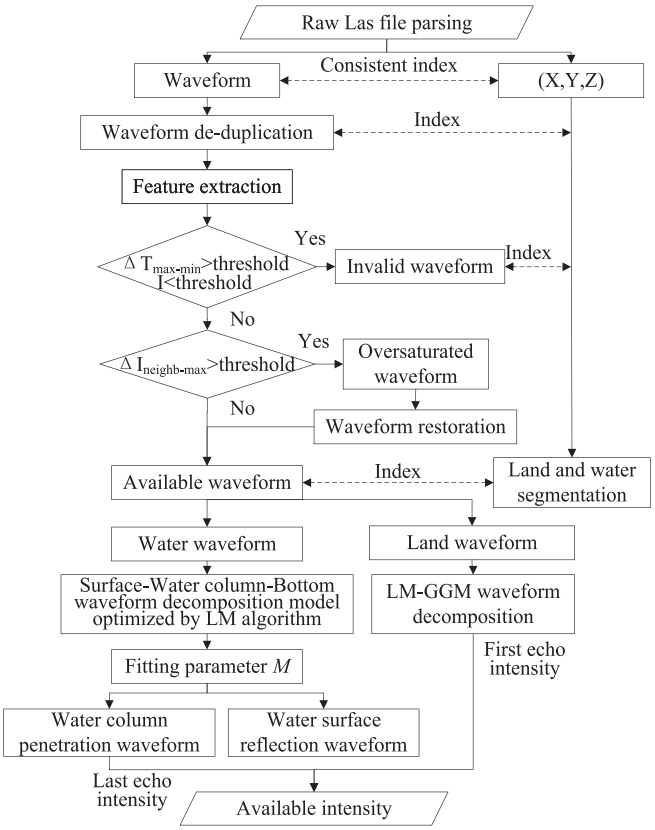


Fig. 3. Flowchart of effective intensity data extraction for ALB.

3.2. Repeated waveform rejection

The original LAS file contains coordinate, intensity, and full waveform data that correspond to one another. In situations where the laser pulse interacts with multiple targets during propagation, the sensor generates multiple coordinates and corresponding intensity data while repeatedly recording the full waveform data. In areas dense with vegetation or water, the proportion of duplicate waveforms can become as high as 50%. In response to this challenge, our proposed method initiates the identification and removal of repeated waveform data to enhance the accuracy and efficiency of data processing.

3.3. Invalid waveform identification

Invalid waveforms are a type of waveform with extremely similar morphologies, posing a significant obstacle to the accurate interpretation of intensity data, as shown in Fig. 4(a). Typically, these waveforms are generated by the sensor and are characterized by low intensity, with values ranging primarily between 30 and 50, as illustrated in Fig. 4(b). Due to their distinct decreasing trend over the recording time, the maximum and minimum intensity values of invalid waveforms typically appear at the very beginning and end of the waveform, respectively. To identify and reject invalid waveforms and corresponding intensity values, the proposed method deploys a criterion based on the time deviation between the minimum and maximum intensity values, denoted as $\Delta T_{\max-\min}$.

$$\Delta T_{\max-\min} = T_{\min} - T_{\max} \tag{1}$$

where $T_{\min}$ and $T_{\max}$ are the times corresponding to the minimum and maximum intensities, waveforms with relatively small intensity and large $\Delta T_{\max-\min}$ values that satisfy the predefined threshold are considered as invalid waveforms and are consequently removed from the dataset. This step is crucial in eliminating unwanted signals and improving data accuracy for subsequent analysis.

3.4. Oversaturated waveform identification and correction

Oversaturated waveforms occur when the intensity level of a waveform exceeds the maximum reception range of the system. As a result, the waveform is clipped or compressed by the device, as illustrated in Fig. 5. To prevent excessive or low-intensity, many systems incorporate an AGC program that continually adjusts the gain or reduction of the signal to maintain its intensity level within the dynamic range of the system (Korpela et al., 2010). Depending on the device's specification, the intensity data may be recorded using different bit depths and ranges,

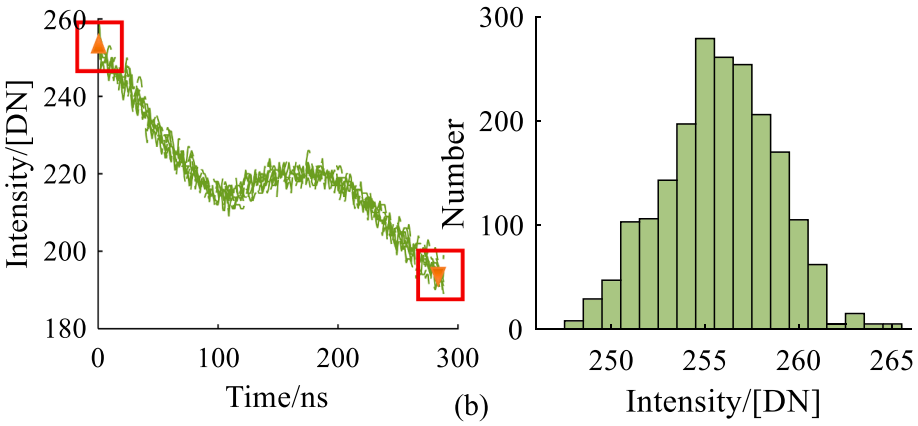


Fig. 4. Invalid waveforms and statistical histograms of recorded intensity data: (a) spatial pattern of invalid waveforms, (b) statistical histogram of the intensity corresponding to the invalid waveforms recorded.

such as floating-point intensity (ranging from 0 to 1), 8-bit intensity data (ranging from 0 to 255) or 16-bit intensity data (ranging from 0 to 65535). A higher bit-depth provides more precise intensity resolution (Vain & Kaasalainen, 2010). The phenomenon of oversaturation caused by strong signals has been identified as a significant contributor to inaccuracies in distance measurements. To this end, AGC implementations have been instrumental in mitigating this effect. Specifically, AGC are designed to enhance the number of received echoes while simultaneously maintaining oversaturation below an acceptable threshold by adjusting the gain value within a defined range. However, the inappropriate configuration of AGC algorithms may lead to the generation of irregular and inconsistent intensity data. As a result, the utility of such data for analysis and interpretation may be limited, especially in complex environments.

Fig. 5 shows the waveforms affected by oversaturation frequently present with truncated shapes, resulting in sudden increases and decreases in intensity. Such a type of waveform can be identified by the deviation of sampled intensities at adjacent times. Specifically, $\Delta I_{\text{neighb}-\max}$ is defined as the maximum value of the intensity deviation between the sampled data at time $t+1$ and t, and $\Delta I_{\text{neighb}-\min}$ is the minimum value of the deviation.

$$\Delta I_{\text{neighb}-\max} = \max(I_{t+1} - I_t) \tag{2}$$

$$\Delta I_{\text{neighb}-\min} = \min(I_{t+1} - I_t) \tag{3}$$

where $t = 1, 2, 3, \cdots N-1$, and N is the maximum value of time sampling. Once a threshold value is set for either $\Delta I_{\text{neighb}-\max}$ or $\Delta I_{\text{neighb}-\min}$, the oversaturated waveform can be identified, and the start and end positions of the truncated waveform segment can then be determined based on both deviation values.

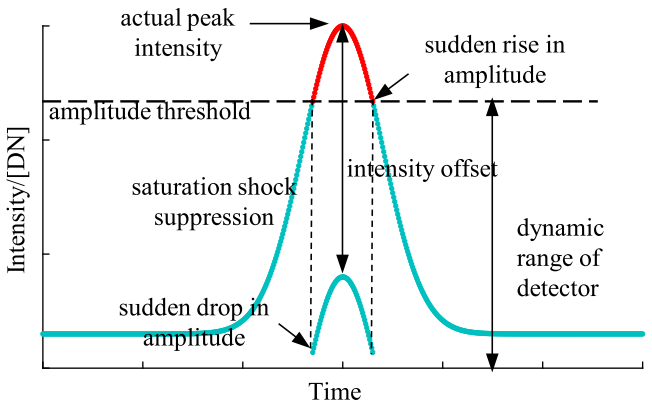


Fig. 5. Display of oversaturated waveform morphology.

Once the oversaturated waveform segment has been identified and flagged, the next step is to recover the truncated waveform fragment. This is achieved using the triple spline interpolation technique (Kung & Rota, 1979), which involves generating a smooth curve through the oversaturated waveform points to fill in the missing data points of the truncated waveform segment. Triple spline interpolation is a widely used method for reconstructing functions from approximate values or sampling points. It involves generating a smooth curve through the oversaturated waveform points to complement the truncated waveform segment. Calculation of the longitudinal displacement of truncated waveform segments becomes possible once the high-return waveform segment is estimated via the three-sample interpolation method. The recovered oversaturated waveform can then be merged back into the waveform file for segmentation purposes.

3.5. *Bathymetric and land waveform segmentation and waveform decomposition*

Adapting the effective intensity extraction process to the scanning environment and signal reflection characteristics is essential. In general, the laser signals aimed at land areas record the reflected signal of the target, while signals aimed at water areas contain both surface and bottom reflected signals. Therefore, to determine the effective intensity for land waveforms, the first echo must be considered, as it represents the top of the nearest target or surface. Conversely, for bathymetric waveforms, the effective intensity is defined as the seabed echo which occurs as the second or third echo, and reflects the reflectivity of the seafloor. A threshold value is set on the elevation data to split the recovered oversaturated waveform and remaining waveform into separate water and land regions. Once this has been accomplished, the bathymetric waveform is further separated into two distinct categories based on the laser's behavior as it interacts with the water surface. The first category is the surface reflection waveform, which refers to the laser beam reflecting directly back to the sensor from the water surface and should be remove in intensity image, and the second category is the water column penetration waveform, which corresponds to the laser pulse penetrating the water column, reaching the bottom and then reflecting back to the sensor. Due to the complex interactions between the laser beam and the target surface, the retrieved waveform usually contains multiple echo components, each with a different intensity and duration. To handle this complexity and eliminate the intensity mixing that is attributable to multiple echoes, different waveform decomposition models are used for both bathymetric and land waveforms. Additionally, only the intensity values related to valid targets are retained. A generalized Gaussian model (GGM) optimized by the Levenberg-Marquardt (LM) algorithm (LM-GGM) is constructed to decompose the

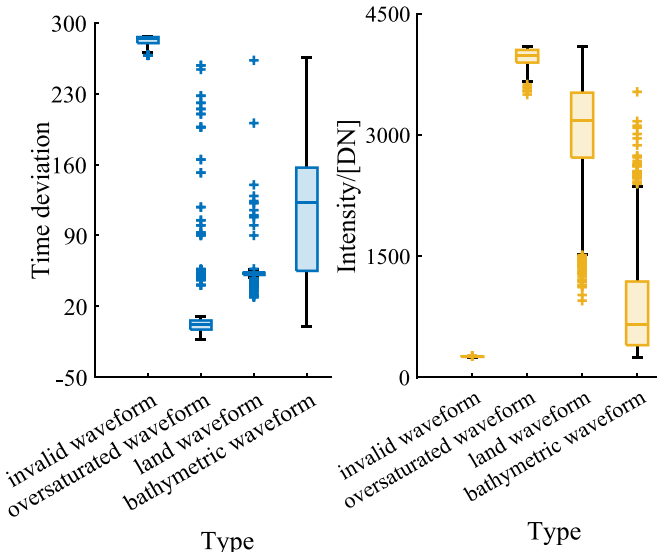


Fig. 6. Box plots corresponding to different categories of waveform characteristics.

land waveform (Ji et al., 2021). The land waveform $y(t)$ is defined as the sum of M gaussian components and the integration of the average noise $\overline{\text{noise}}$:

$$y(t) = \sum_{j=1}^{M} A_j e^{-\frac{(t-\mu_j)^{\gamma_j}^2}{2\omega_j^2}} + \int \overline{\text{noise}} \tag{4}$$

where A_j, μ_j, ω_j and γ_j are the amplitude, peak position, half-wave width and waveform shape parameters of the j-th gaussian component, respectively. The LM algorithm is highly effective in optimizing the GGM model due to its ability to handle noisy data and to find solutions to ill-posed problems. It is also highly robust and converges fast, rendering it an appropriate tool for large-scale remote sensing applications. The objective function, which is defined as the summation of the differences between the modeled and measured waveforms, is minimized by updating the parameters at each iteration.

$$g = \min\|f(t) - y(t)\| \tag{5}$$

To stop the iteration, the threshold condition ε, and the maximum number of iterations Max_N is set as the termination condition. Let $y(t) = F(t,P)$ be a function of the unknown coefficients $P = \left\{A_j, \mu_j, \omega_j, \gamma_j, \overline{\text{noise}}\right\} = \{p_1, p_2, p_3, \cdots, p_m\}$, where $A_j, \mu_j, \omega_j, \gamma_j$ are the parameters of the j th echo component, and m is the total number of parameters. The vector of parameters to be estimated is initialized as $P^0 = \{p_1{}^0, p_2{}^0, p_3{}^0, \cdots, p_m{}^0\}$. To obtain a more accurate estimation, we can expand $F(t,P)$ by Taylor series, neglecting quadratic and higher-order terms for computational simplicity.

$$F(t,P) \approx F\left(t,P^0\right) + \frac{\partial F(t,P)}{\partial p_1}|_{P=P^0}\left(p_1 - p_1{}^0\right) + \frac{\partial F(t,P)}{\partial p_2}|_{P=P^0}\left(p_2 - p_2{}^0\right) + \frac{\partial F(t,P)}{\partial p_3}|_{P=P^0}\left(p_3 - p_3{}^0\right) + \cdots + \frac{\partial F(t,P)}{\partial p_m}|_{P=P^0}\left(p_m - p_m{}^0\right) \tag{6}$$

Then, the LM algorithm is driven to estimate the parameter vector, specifically including the following steps:

Step 1: Initialize the damping coefficient ζ and calculate the Jacobi matrix $J(t)$. Assume that the function is Π.

$$\Pi = \sum_{i=1}^{n}\left\{f(t_i) - \left[F(t_i,P) + \sum_{j=1}^{m}\frac{\partial F(t_i,P)}{\partial p_j}|_{P=P^0} \bullet \left(p_j - p_j{}^0\right)\right]^2\right\} + \zeta\sum_{j=1}^{m}\left(p_j - p_j{}^0\right)^2 \tag{7}$$

where ζ (≥ 0) is the damping coefficient. In order to make the first-order partial derivative of the function Π with respect to P equal to 0, i.e., to minimize Π, and $J(t)$ is introduced as:

$$J(t) = \begin{bmatrix} \frac{\partial F(t_1,P)}{\partial p_1} & \frac{\partial F(t_1,P)}{\partial p_2} & \cdots & \frac{\partial F(t_1,P)}{\partial p_m} \\ \frac{\partial F(t_2,P)}{\partial p_1} & \frac{\partial F(t_2,P)}{\partial p_2} & \cdots & \frac{\partial F(t_2,P)}{\partial p_m} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial F(t_n,P)}{\partial p_1} & \frac{\partial F(t_n,P)}{\partial p_2} & & \frac{\partial F(t_n,P)}{\partial p_m} \end{bmatrix} \tag{8}$$

Step 2: Calculate the Hesse matrix from the Jacobi matrix.

$$\Gamma = J(t)^T J(t) \tag{9}$$

Step 3: Calculate the residuals ξ^N of the original waveform and the fitted waveform corresponding to the N th iteration

$$\xi^N = \left|f(t) - F(t,P)^N\right|^2 \tag{10}$$

Step 4: Calculate the step size ΔP.

$$\Delta P = -[\Gamma + \zeta D]J(t)\xi^N \tag{11}$$

where D is the unit matrix.

If ξ^N is greater than ε, it means that the algorithm has not yet converged to the desired level of accuracy. In this case, the algorithm continues to update the parameters using a step size ΔP and re-executes step 2. Once the algorithm has converged, the final parameter vector P is output as the optimal solution. This continues until the algorithm reaches the maximum number of iterations Max_N or until the convergence criterion is met.

The bathymetric waveform, a fundamental data used for seafloor mapping, is defined as a combination of multiple echo components, specifically, surface, water column, bottom, and noise. Among these components, the surface and bottom echoes are frequently modeled as Gaussian functions, while the water column component, $f_c(t)$, is typically modeled as a triangle.

$$y(t) = \sum_{j=1}^{M} A_j e^{-\frac{(t-\mu_j)^{\gamma_j}^2}{2\omega_j^2}} + f_c(t) + \int \overline{\text{noise}} \tag{12}$$

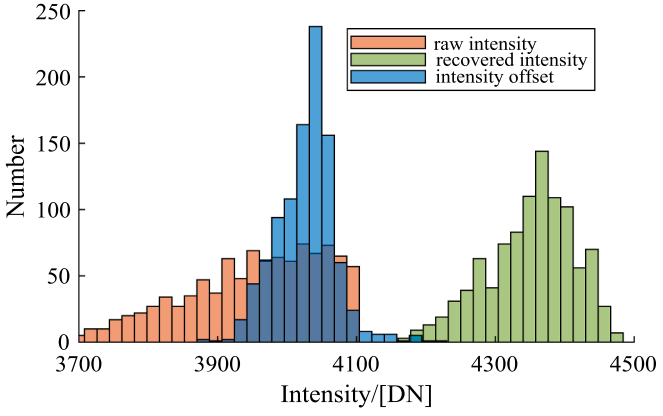


Fig. 7. Intensity statistics before and after intensity recovery of oversaturated waveforms.

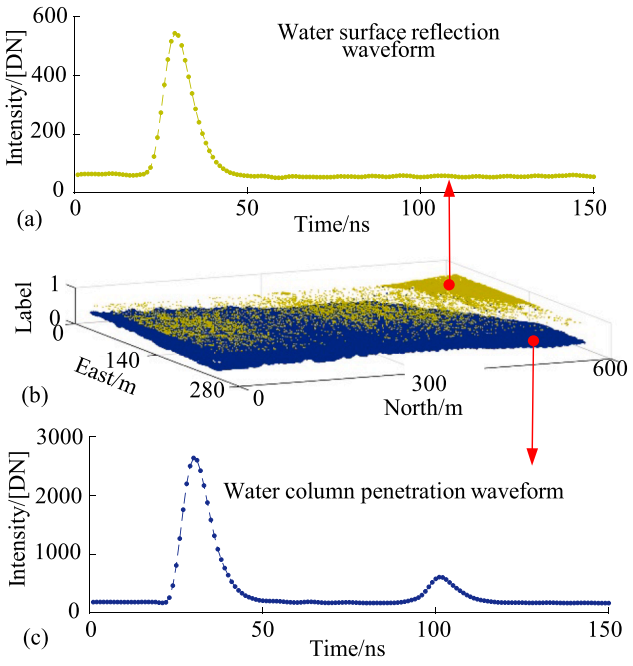


Fig. 8. Water surface reflection waveform and water column penetration waveform segmentation results: (a) the surface reflection waveform, corresponding to Label 1 in (b), (b) Lables 0 and 1, and (c) the water column penetration waveform, corresponding to Label 0 in (b).

$$f_c(t) = \begin{cases} 0, t < a \\ c\dfrac{x-a}{b-a}, a \le t \le b \\ c\dfrac{x-d}{b-d}, b < t \le d \\ 0, t > d \end{cases} \tag{13}$$

where, (a,0), (b,c) and (d,0) are the coordinates of the triangle vertices. Moreover, to accurately determine the model parameters for the bathymetric waveform, the LM algorithm is frequently introduced. When M is equal to 1 the waveform corresponds to the water surface reflection waveform. A waveform that penetrates the water column typically has an M value of 2, but if it encounters phytoplankton, the transmission may be strong enough to produce an M value of 3. The intensity of the last echo component in a waveform over water represents the seafloor properties, while for land waveforms, the first echo component is typically used to determine the properties of the target.

3.6. Accuracy assessment

Upon performing waveform classification and effective intensity extraction in ALB data processing, it is vital to evaluate the quality and precision of the extracted intensity data to ensure its reliability. Accordingly, various performance indices are employed to assess the extracted data, including intensity deviation ΔY, standard deviation *STD*, mean squared error *MSE*, mean absolute error *MAE*, and mean absolute percentage error *MAPE*:

$$\Delta Y = y_{max} - y_{min} \tag{14}$$

$$STD = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-1}} \tag{15}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \tag{16}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y} - y_i| \tag{17}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y} - y_i}{y_i}\right| \tag{18}$$

where, $Y = \{y_1, y_2, y_3, \cdots y_n\}$ is the set of intensities, n is the number of intensities, and $\hat{y}$ is the mean value of Y.

4. Results and discussion

4.1. Identification feature analysis of invalid waveforms

Identifying invalid waveforms in ALB data processing is a critical task that requires the use of appropriate identifying features. In this study, the maximum intensity time deviation $\Delta T_{\max-\min}$ and intensity I are utilized as identifying features for this purpose. Statistical box plots are then obtained for these features corresponding to four waveform types, including invalid waveform, oversaturated waveform, land waveform, and bathymetric waveform, as shown in Fig. 6. The statistical box plots indicate that the invalid waveform type possesses a distinct feature profile compared to the other waveform types, where it exhibits a higher value of $\Delta T_{\max-\min}$ and a lower value of I.

To identify invalid waveforms, a segmentation threshold of 90% of the total recording time is utilized for $\Delta T_{\max-\min}$, while a threshold of 55 [DN] (the smallest integer that exceeds the intensity distribution interval) is employed for I. Waveforms that exceed the threshold for

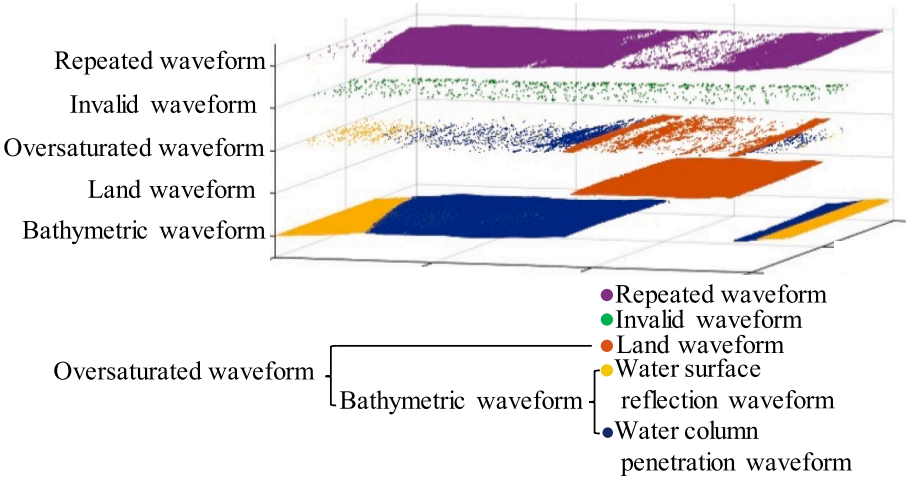


Fig. 9. Waveform classification results.

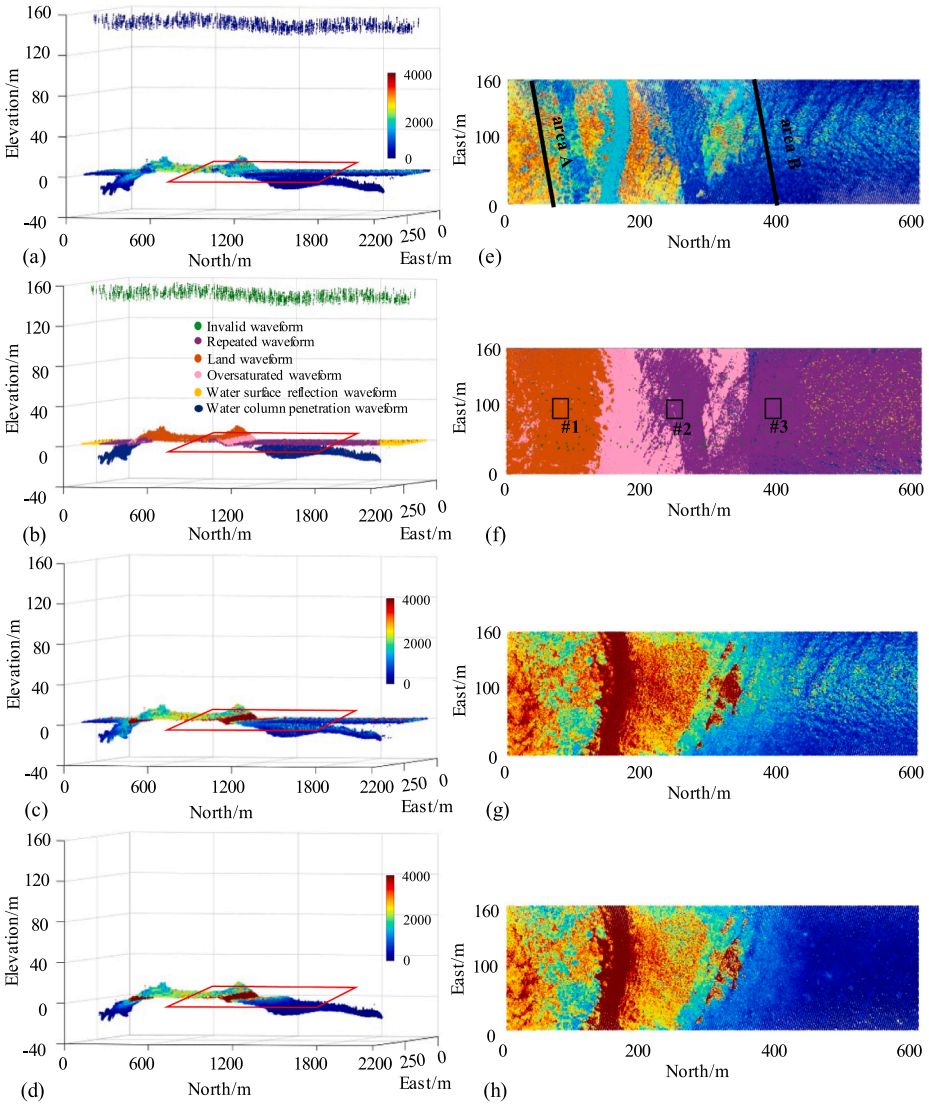


Fig. 10. Demonstration of effective intensity data extraction process and comparison of intensity data obtained through different routes: (a) Original point cloud intensity, (b) waveform classification result, (c) intensity data obtained by the RLD algorithm, (d) intensity value after effective intensity extraction by proposed approach, (e), (f), (g) and (h) corresponding to the enlarged top view of the red boxed area in (a), (b), (c) and (d), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$\Delta T_{\max-\min}$ and have a corresponding I value below the established threshold are considered as invalid waveform. Using this method, approximately 0.2% of the total waveforms recorded in the study area are identified as invalid waveforms.

4.2. Intensity correction analysis of oversaturated waveforms

The identification and adjustment of oversaturated waveforms are crucial in the processing of ALB data. At times, the intensity data may be modified due to limitations in the system detector dynamic range and AGC intervention, resulting in an erroneous representation of the original signal. To mitigate this issue, it is necessary to adjust the intensity data to achieve accurate results. The maximum and minimum values of adjacent sampled intensity deviation, defined as $\Delta I_{\text{neighb}-\max}$ and $\Delta I_{\text{neighb}-\min}$, respectively, are calculated to identify oversaturated waveforms. If the value of $\Delta I_{\text{neighb}-\max}$ exceeds 2500 [DN] (60% of the maximum intensity value) or if $\Delta I_{\text{neighb}-\min}$ drops below −2500 [DN], then the waveform is classified as oversaturated waveform.

This study utilizes a random selection of 2000 oversaturated waveforms to investigate the impact of employing triple spline fit on the intensity data. For each waveform, triple spline fit is employed to determine the intensity offsets of truncated waveform segments. As shown in Fig. 7, the mean value of intensity offset is 4027 [DN]. Subsequently, the truncated waveform segments are recovered, and the intensity data is re-extracted. Notably, the intensity data obtained post-recovery exhibits higher values compares to the system recorded data product (raw intensity) due to the absence of dynamic control range constraints during the extraction process.

4.3. Separation of water surface reflection waveform and water column penetration waveform

In the signal capture process of the bathymetric waveform, there is an admixture of the water surface reflection waveform and the water column penetration waveform. The former waveform only records surface return intensity and must be identified and rejected. To achieve this goal, an LM algorithm-optimized surface-water column-bottom-noise waveform decomposition model is driven, which is capable of decompose bathymetric waveform and detecting the number of Gaussian components present in them.

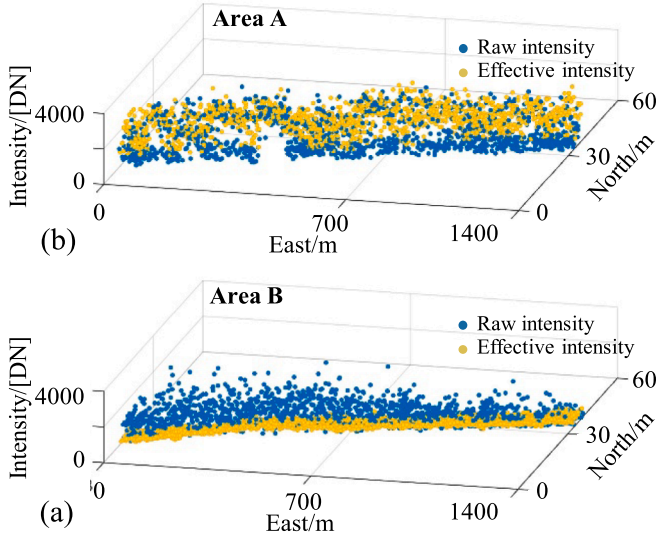


Fig. 11. Effective intensity before and after extraction: (a) effect before and after effective intensity extraction in terrestrial areas, and (b) corresponds to the results in water.

When the model detects a single Gaussian component within a waveform, it indicates the presence of a surface reflection signal exclusively. Conversely, detection of more than a single Gaussian component demonstrates the existence of a seafloor return signal. To demonstrate the effectiveness of our approach, a piece of water within the study area was randomly chosen for waveform decomposition. Fig. 8 shows the segmentation results of the surface reflection and bottom penetration waveforms, where (a) shows the surface reflection waveform, corresponding to Label 1 in (b), and (c) shows the water column penetration waveform, corresponding to Label 0 in (b).

4.4. Waveform classification

The process of waveform classification is carried out for all waveforms within the study area, and the classification results are obtained as shown in Fig. 9. The highest purple point on the diagram represents the repeated waveforms, which constitute one-third of the overall waveforms observed in the study area. The green dots on the diagram represent invalid waveforms that are present throughout the acquisition area, and thus should be excluded from further analysis. The intermediate layer of the diagram illustrates oversaturated waveforms that primarily found on beaches, roads, and mirrored sea surface. Those waveforms are subsequently classified into three subcategories: land, water surface reflection, and water column penetration waveforms. Within the diagram, the orange dots correspond to the land waveforms, while the bottommost layer represents bathymetric waveforms that are further classified into water column penetration waveforms (depicted in blue) and water surface reflection waveforms (depicted in yellow).

Table 1
Accuracy statistics of effective intensity and raw intensity.

Parameters		#1	#2	#3
Raw intensity	**ΔY**	3731	3617	3017
	STD	914.16	1038.28	470.22
	MSE	22.92	27.60	11.62
	MAE	780.32	913.08	343.56
	MAPE	1.49	0.81	0.97
Effective intensity	**ΔY**	3252	2632	1832
	STD	634.45	493.18	245.20
	MSE	21.18	17.17	8.59
	MAE	525.57	393.96	186.56
	MAPE	0.27	0.13	0.25

Within the study area, water surface reflection waveforms constituted approximately 20.3% of the entire number of bathymetric waveforms, whilst water column penetration waveforms accounted for 79.7%.

4.5. Effective intensity data extraction and accuracy evaluation

The proposed approach utilizes the intensity of the bottom echo component for water column penetration waveform and the intensity of the first echo for the land waveform to reconstruct a new intensity image. To quantitatively evaluate the extracted effective intensity, the extracted intensity data is compared with the intensity value recorded in the LAS file and the intensity image obtained using the Richardson-Lucy deconvolution (RLD) algorithm, which has been shown to outperform bathymetric waveforms as compared to various waveform signal detection methods.(Wang et al. 2015). Fig. 10 shows the results of the proposed method, where Fig. 10(a) displays the intensity data recorded in the system's LAS file, while Fig. 10(b) shows the waveform classification results. Fig. 10(c) is the intensity image obtained by RLD algorithm and Fig. 10(d) presents the effective intensity values based on the waveform classification results. In addition to the results presented in Fig. 10(a-d), Fig. 10(e-h) provides a more detailed representation of the effective intensity data extracted by the proposed method in both the land and water areas. Fig. 10(e), located in the top right corner, clearly shows significant noise from interference and continuous anomalies in scan line intensity in the original intensity image due to the water surface echo intensity and inappropriate AGC intervention. In the case of low-reflective objects such as water, when the sensor detects a continuous sequence of laser pulses that return intensities below the minimum value set by AGC, the sensor may initiate the AGC augmentation mechanism. Conversely, the AGC reduction function is activated to reduce the gain value and avoid saturation if the sensor detects a continuous sequence of laser pulses with intensities exceeding the AGC maximum threshold. However, continuous adjustment of the gain value without proper discernment between high and low reflectivity targets may result in certain undesirable effects, compromising the quality of the extracted data. This could lead to delays in the adjustment processes and unjustifiable compensation of the intensity data within the threshold range. Moreover, the duration of the signal may not be sufficient to either amplify or attenuate the intensity data points above the threshold, leading to aberrant regions of the adjusted intensity data and unexpected variations in the measured intensity readings over time.

As shown in Fig. 10(h), it is clear that effective intensity extraction addresses the aforementioned hindrances, thereby producing a more natural rendering of features. Compared with Fig. 10(g) and (h), it is evident that the signal detection model fails to produce the desired intensity image without accounting for the root cause of interference intensity generation. Compared to the method proposed in this paper, the RLD algorithm increases the intensity extraction effort by 33% because it does not pre-judge the class of the returned signal, which also leads to the problem of intensity mixing of multiple targets produced by a single emitted light source. Additionally, the intensity images obtained by the RLD algorithm are filled with significant noise originating from invalid intensities and high return signals from the water surface. The inability to process oversaturated signals results in the intensity of high-return signals deviating from their true values. Fig. 10(e) shows the locations of two randomly selected areas, namely Area A (land area) and Area B (bathymetry area), within the study area. Fig. 11 presents the intensity results with and without effective intensity extraction for both areas. The analysis of the results indicates that the effective intensity for the land area mainly clusters in regions with high intensity values, primarily because the initial echo of the laser pulse assimilates the most potent energy, thereby leaving the subsequent echoes, particularly the second or multiple signals, with inadequate energy. Therefore, these echoes can be ignored. Conversely, the echoes from the bathymetric waveforms that have relatively low intensity values, such as the bottom echoes, are the most essential and informative. These echoes provide critical

information about the terrain and structure of the seafloor. However, the surface echo value in the sounding waveform belongs to the interference data and must be eliminated.

To evaluate the effectiveness of our method, three test sites with a fixed area of 20 m × 20 m were randomly selected and evaluated before and after applying the proposed method, as illustrated in Fig. 10 (f). The intensity deviation ΔY, standard deviation *STD*, mean squared error *MSE*, mean absolute error *MAE*, and mean absolute percentage error *MAPE* were assessed for each test sites. Table 1 shows the outcomes obtained using the proposed method. As compared to the intensity data products recorded by the system, our method is effective in extracting intensities with considerably lower values for all the assessed indices. This validation was attributed to the removal of multiple return intensities, resulting in greater regional consistency in the remaining data.

5. Conclusions

This paper has presented an effective intensity extraction method based on waveform classification to eliminate interference and obtain accurate intensity data that reflects the properties of underwater features. The novel intensity extraction method incorporates an index sharing mechanism that integrates intensity data, full waveform data, and terrain data to track and eliminate interference intensities originating from multiple echoes, suppressed high return signals, and invalid signals recorded by the acquisition system. In addition, the refined waveform classification framework is a crucial step in the processing of ALB data, as it enables the removal of unwanted waveforms and interference signals that impede the extraction of accurate intensity data for both land and bathymetric regions.

The findings can be summarized as:

- For land waveforms, the first target intensity is retained, while for water column penetration waveforms, the bottom return intensity is kept.
- Accurate intensity estimation is often hindered by the presence of oversaturated waveforms, which occur when the signal return intensity surpasses the system's dynamic range, resulting in waveform clipping and consequently problematic estimates of true backscatter intensity. To address this issue, the proposed ALB data processing workflow incorporates a fitting methodology for oversaturated waveforms to estimate the offset of the clipped waveform segment. By utilizing this offset, the problematic estimates of the true backscatter intensity resulting from oversaturated waveforms can be rectified.
- By comparing the intensity values registered by the system, the efficacy of the novel method can be evaluated in terms of providing smoother transitions and increasing regional consistency in the intensity date. A traditional full-waveform signal detection model that fails to comprehensively account for the origins and distribution patterns of interference intensity distribution is inadequate for generating high-quality intensity images.

Recent studies have revealed promising potential for utilizing ALB intensity data in various areas, including seafloor habitat mapping, substrate classification, and water column environment inversion. By combining the intensity data with elevation data, a more comprehensive and precise portrayal of the target objects and their features can be achieved. Consequently, subsequent research endeavors will focus on developing and evaluating ALB intensity correction schemes to improve the accuracy and uniformity of intensity information.

CRediT authorship contribution statement

Xue Ji: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Zhen Dong:** Data curation, Writing – original draft. **Lin Zhang:** Conceptualization, Supervision, Methodology, Writing – review & editing. **Mingchang Wang:** Conceptualization. **Yi Ma:** Supervision. **Jingyu Zhang:** Supervision. **Xueping Lin:** Software, Validation, Writing – review & editing. **Jonathan Li:** Conceptualization, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

We extend our sincerest appreciation to the esteemed Prof. William Philpot at the Cornell University and Dr. Wai-yeung Yan at the Hong Kong Polytechnic University for their invaluable guidance and stimulating perspectives, which played a crucial role in shaping the outcome of this scholarly investigation. The present study is supported by the National Natural Science Foundation of China (No. 22FAA01871), the open research fund program of the State Key Laboratory of information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University (No. 22S02), the open fund of Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources (No. MESTA-2022-B005) and Technology Innovation Center for Ocean Telemetry, Ministry of Natural Resources (No. 2022002), the National Natural Science Foundation of China (No. 42301505, No.52305445, No. 41871381, No. 42171407).

References

Abdallah, H., Baghdadi, N.N., Bailly, J.S., Pastol, Y., Fabre, F., 2012. Wa-LiD: A new LiDAR simulator for waters. IEEE Geosci. Remote Sens. Lett. 9, 744–748.

Ahokas, E., Kaasalainen, S., Hyyppä, J., & Suomlainen, J. (2006). Calibration of the optech ALTM-3100 laser scanner intensity data using brightness targets. In.

Collin, A., Archambault, P., Long, B., 2008. Mapping the shallow water seabed habitat with the SHOALS. IEEE Trans. Geosci. Remote Sensing 46, 2947–2955.

Eren, F., Pe Eri, S., Rzhanov, Y.A., Ward, L.G., 2018. Bottom characterization by using airborne lidar bathymetry (ALB) waveform features obtained from bottom return residual analysis. Remote Sens. Environ. 206, 260–274.

Errington, A.F.C., Daku, B.L.F., Prugger, A., 2015. Reflectance modelling using terrestrial LiDAR intensity data. IEEE International Conference on Imaging Systems and Techniques (IST) 2015, 1–6.

Fang, W., Huang, X., Zhang, F., Li, D., 2015. Intensity correction of terrestrial laser scanning data by estimating laser transmission function. IEEE Trans. Geosci. Remote Sensing 53, 942–951.

Feygels, V., Kopilevich, Y.I., Surkov, A.I., Yungel, J., Behrenfeld, M.J., 2003. Airborne lidar system with variable-field-of-view receiver for water optical properties measurement. In, SPIE Optics + Photonics.

Guenther, G.C., Cunningham, A.G., LaRocque, P.E., & Reid, D.J. (2000). Meeting the Accuracy Challenge in Airborne Bathymetry. In.

Guenther, G.C. (1985). Airborne Laser Hydrography: System Design and Performance Factors. In.

Guo, Y., Feng, C., Xu, W., Liu, Y., Su, D., Qi, C., Dong, Z., 2023. Water-land classification for single-wavelength airborne LiDAR bathymetry based on waveform feature statistics and point cloud neighborhood analysis. Int. J. Appl. Earth Obs. Geoinf. 118, 103268.

Ji, X., Yang, B., Tang, Q., 2020. Seabed sediment classification using multibeam backscatter data based on the selecting optimal random forest model. Appl. Acoust. 167, 107387.

Ji, X., Tang, Q., Xu, W., Li, J., 2021. Island features Classification for single-wavelength airborne LiDAR bathymetry based on full-waveform parameters. Appl. Opt.

Kaasalainen, S., Jaakkola, A., Kaasalainen, M., Krooks, A., & Kukko, A. (2011). Analysis of Incidence Angle and Distance Effects on Terrestrial Laser Scanner Intensity: Search for Correction Methods. *Remote Sensing , 3*, 2207-2221.

Kashani, A.G., Olsen, M.J., Parrish, C.E., Wilson, N., 2015. A review of LIDAR radiometric processing: from Ad Hoc intensity correction to rigorous radiometric calibration. Sensors (Basel, Switzerland) 15, 28099–28128.

Kopilevich, Y.I., Feygels, V., Tuell, G., Surkov, A.I., 2005. Measurement of ocean water optical properties and seafloor reflectance with scanning hydrographic operational airborne lidar survey (SHOALS): I. In, SPIE Optics + Photonics, Theoretical background.

Korpela, I., Ørka, H.O., Hyyppä, J., Heikkinen, V., Tokola, T., 2010. Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies. ISPRS J. Photogramm. Remote Sens. 65, 369–379.
Kung, J., Rota, G.C., 1979. A practical guide to splines. Adv. Math. 32, 81.
Lee, M., 2003. Benthic mapping of coastal waters using data fusion of hyperspectral imagery and airborne laser bathymetry. In, *Environmental*. Science.
Long, B., Aucoin, F., Montreuil, S., Robitaille, V., & Xharde, R. (2011). Airborne lidar bathymetry applied to coastal hydrodynamic processes. In.
Narayanan, R., Kim, H.B., Sohn, G., 2009. Classification of SHOALS 3000 bathymetric LiDAR signals using decision tree and ensemble techniques. In: In, 2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH), pp. 462–467.
Pereira, T.R.B., Vital, H., Barbosa, N.F., & Silva, A.A.D. (2015). Seafloor morphology and coastal erosion assessment using multibeam bathymetric analysis. In, 2015 IEEE/OES Acoustics in Underwater Geosciences Symposium (RIO Acoustics) (pp. 1-2).
Richter, K., Maas, H., Westfeld, P., Weiß, R., 2017. An approach to determining turbidity and correcting for signal attenuation in airborne lidar bathymetry. PFG – J. Photogramm. Remote Sens. Geoinformat. Sci. 85, 31–40.
Sevara, C., Wieser, M., Doneus, M., Pfeifer, N., 2019. Relative radiometric calibration of airborne LiDAR data for archaeological applications. Remote. Sens. 11, 945.
Tamondong, A.M., Tamondong, A.M., Cadalzo, I.E., Estabillo, M.S., Go, G.A., Cruz, C., Blanco, A.C., 2020. Evaluation of object-based classification methods for mapping benthic habitats using bathymetric LiDAR derivatives. Philippine Eng. J. 1–18.
Tuell, G., Feygels, V., Kopilevich, Y.I., Weidemann, A.D., Cunningham, A.G., Mani, R., Podoba, V.I., Ramnath, V., Park, J.Y., & Aitken, J. (2005). Measurement of ocean water optical properties and seafloor reflectance with scanning hydrographic operational airborne lidar survey (SHOALS): II. Practical results and comparison with independent data. In, SPIE Optics + Photonics.
Tuell, G., Park, J.Y., 2004. Use of SHOALS bottom reflectance images to constrain the inversion of a hyperspectral radiative transfer model. SPIE Defense + Commercial Sensing 5412.
Vain, A., Kaasalainen, S., 2010. Correcting airborne laser scanning intensity data for automatic gain control effect. IEEE Geosci. Remote Sens. Lett. 7, 511–514.
Wang, C., Li, Q., Liu, Y., Wu, G., Liu, P., Ding, X., 2015. A comparison of waveform processing algorithms for single-wavelength LiDAR bathymetry. ISPRS J. Photogramm. Remote Sens. 101, 22–35.
Wang, C., Philpot, W., 2002. Using SHOALS LIDAR system to detect bottom material change. In: In, IEEE International Geoscience and Remote Sensing Symposium, pp. 2690–2692.
Wang, C., Philpot, W.D., 2007. Using airborne bathymetric lidar to detect bottom type variation in shallow waters. Remote Sens. Environ. 106, 123–135.
Wu, Q., Zhong, R., Dong, P., Mo, Y., Jin, Y., 2021. Airborne LiDAR intensity correction based on a new method for incidence angle correction for improving land-cover classification. Remote Sens. (Basel) 13, 511.
Yan, W.Y., Shaker, A., El-Ashmawy, N., 2015. Urban land cover classification using airborne LiDAR data: A review. Remote Sens. Environ. 158, 295–310.
Zavalas, R., Ierodiaconou, D., Ryan, D., Rattray, A., Monk, J., 2014. Habitat classification of temperate marine Macroalgal communities using bathymetric LiDAR. Remote Sens. (Basel) 6, 2154–2175.

Further reading

Guo, K., Li, Q., Wang, C., Mao, Q., Liu, Y., Zhu, J., Wu, A., 2022a. Development of a single-wavelength airborne bathymetric LiDAR: System design and data processing. ISPRS J. Photogramm. Remote Sens. 185, 62–84.
Guo, K., Li, Q., Wang, C., Mao, Q., Liu, Y., Ouyang, Y., 2022b. Target echo detection based on the signal conditional random field model for full-waveform airborne laser bathymetry. IEEE Trans. Geosci. Remote Sensing 60, 1–21.
Ji, X., Yang, B., Tang, Q., Xu, W., Li, J., 2022a. Feature fusion-based registration of satellite images to airborne LiDAR bathymetry in island area. Int. J. Appl. Earth Obs. Geoinf. 109, 102778.
Ji, X., Yang, B., Wang, Y., Tang, Q., Xu, W., 2022b. Full-waveform classification and segmentation-based signal detection of single-wavelength bathymetric LiDAR. IEEE Trans. Geosci. Remote Sensing 60, 1–14.