Title Page

Extracting Shallow-water Bathymetry from Lidar Point Clouds Using Pulse Attribute Data: Merging Density-based and Machine Learning Approaches

Kim Lowell[1], Brian Calder
Center for Coastal and Ocean Mapping and Joint Hydrographic Center
University of New Hampshire
24 Colovos Road,
Durham, NH 03824 UNITED STATES

[1]**Corresponding Author**

Titles and email addresses:

Kim Lowell, Research Data Scientist, klowell@ccom.unh.edu (ORCID: 000-0002-8326-4022)
Brian Calder, Research Professor, brc@ccom.unh.edu

Funding: This work was supported by the National Oceanic and Atmospheric Administration (NOAA) Grant NA15NOS400020.

Disclosure: The authors have no potential competing financial nor non-financial interests in the work presented.

Data availability statement: Data that support the findings of this study are available at the link doi.org/10.6084/m9.figshare.12597404. SBET data in the required format are provided at the figshare link. Though the .las data used are available to the public, the authors are not authorized to make them directly available. A small sample of the data for a single data tile are provided at the figshare link. Complete data sets (2016_420500e_2728500n.laz, 2016_426000e_2708000n.laz, 2016_428000e_2719500n.laz, and 2016_430000e_2707500n.laz) can be downloaded from https://coast.noaa.gov/digitalcoast/data/ (Data set name: 2016 NGS Topobathy Lidar: Key West FL') as compressed .laz files. These can be decompressed using the LASzip tool which can be downloaded from laszip.org.

Total word count: 9600 **Word count excluding title page and references**: 7746

Extracting Shallow-water Bathymetry from Lidar Point Clouds Using Pulse Attribute Data: Merging Density-based and Machine Learning Approaches

Abstract

To automate extraction of bathymetric soundings from lidar point clouds, two machine learning (ML[1]) techniques were combined with a more conventional density-based algorithm. The study area was four data "tiles" near the Florida Keys. The density-based algorithm determined the most likely depth (MLD) for a grid of "estimation nodes" (ENs). Unsupervised *k*-means clustering determined which EN's MLD depth and associated soundings represented ocean depth rather than ocean surface or noise to produce a preliminary classification. An extreme gradient boosting (XGB) model was fitted to pulse return metadata – e.g., return intensity, incidence angle -- to produce a final *Bathy/NotBathy* classification. Compared to an operationally produced reference classification, the XGB model increased global accuracy and decreased the false negative rate (FNR) – i.e., undetected bathymetry – that are most important for nautical navigation for all but one tile. Agreement between the final XGB and operational reference classifications ranged from 0.84 to 0.999. Imbalance between *Bathy* and *NotBathy* was addressed using a probability decision threshold that equalizes the FNR and the true positive rate (TPR). Two methods are presented for visually evaluating differences between the two classifications spatially and in feature-space.

Keywords: shallow water bathymetry, airborne lidar, Florida Keys, extreme gradient boosting, k-means clustering

1. Introduction and Approach

It is generally accepted that Hickman and Hogg (1969) authored the first article published on the use of airborne lidar ('light detection and ranging') data for bathymetric mapping. They observed that due to limitations on the penetration of light through water, lidar is most appropriate for shallow water charting. Heritage and Hetherington (2007) noted that the initial focus of lidar research had been primarily on the sensor and data acquisition rather than data analysis or specific applications. It could be argued that this tendency has continued with Andersen et al. (2017), for example, stating that as of 2017 there was no standardized accepted methodology for extracting surface points from green lidar point data alone – although green and near-infrared lidar data are currently combined operationally to extract water surface

[1] A list of abbreviations is provided at the end of the article.

returns. Nonetheless, recent review articles (Kashani et al. 2015; Kutser et al. 2020) suggest that difficulties associated with extracting application-specific information from lidar point clouds are now continually being addressed by the scientific community. As lidar data processing/analysis research has increased in recent years, new concepts for improving lidar sensors have also continued (e.g., Kinzel et al. 2021; Mandlburger et al. 2020; Mitchell and Thayer 2014).

This increased interest in lidar data analysis has been driven by a desire to decrease data acquisition costs to better understand various phenomena. For example, lidar data are proving to be particularly useful to characterize benthic habitat. Various analytical approaches have been explored: characterizing lidar waveforms (Collin et al. 2008; Eren et al. 2018), using machine learning approaches (e.g., Pittman et al. 2009; Su et al. 2019), and classifying benthic habitat using variables that describe characteristics of lidar pulses (Tulldahl and Wikstrom 2012) – 'soundings' in marine parlance.

The focus of this article is the use of airborne lidar data for shallow water bathymetry charting – defined herein as water depths less than 20 m (although Jawak et al. 2015 noted that lidar can penetrate up to 60m under ideal conditions). Much bathymetric depth work has focused on analysing the full lidar waveform for a single spectral wavelength (see, for example, Pe'eri and Philpot 2007; Fernandez-Diaz et al. 2014; Wang et al. 2015; Xing et al. 2019.) Waveform soundings have the potential to identify the water surface and bottom due to increased reflectance from both. Single wavelength waveform data are operationally advantageous because they potentially decrease sensor complexity but are disadvantageous because of increased data volumes compared to multi-wavelength systems that collect point data. Approaches to using waveform soundings for bathymetric mapping are varied and examples include near-surface water modelling (Zhao et al. 2017), analysis of water column backscatter (Kinzel et al. 2012; Nagle and Wright 2016), and a 'surface-volume-bottom' approach that provides a time-saving closed-form solution (Schwarz et al. 2019). The analysis of lidar point– rather than lidar waveform – data has also received considerable attention (see, for example, Brzank et al. (2008), Yang et al. 2020) including its combination with data from passive sensors (e.g., Dietrich 2017, Agrifiotis et al. 2019a).

Numerous researchers have examined ways of extracting bathymetry from such waveform data algorithmically (e.g., Lyzenga et al. 2006, Pacheco et al. 2015, Li et al. 2019). In such work, lidar is sometimes used primarily as the reference data against which analytical methods are evaluated (e.g., Agrifiotis et al. 2019b). In recent years, many such studies have examined various machine learning techniques: neural networks (Liu et al. 2015), support vector machines (Misra et al. 2018; Wang et al.

2018), principal components analysis (Gholamalifard et al. 2013), partial least squares (Niroumand-Jadidi et al. 2018), and random forests (Kogut and Weistock 2019).

The present research is focused on extracting shallow water bathymetry from lidar point clouds – i.e., identifying which lidar soundings represent the ocean bottom. The approach adopted combines a density-based algorithm developed for multi-beam echo sounder (MBES) sonar data with machine learning (ML) techniques. This methodological fusion is explored to overcome two considerable challenges. First, lidar data are collected from airborne platforms, resulting in a substantial number of soundings that represent the ocean surface and near-surface. Second, no ground-truth data are available for training ML models. The latter difficulty was also recognized and addressed by Kerr and Purkis (2018), who developed a workflow for optical data. The former is suggestive of a weak bathymetric signal within a cloud of lidar soundings. The latter is particularly vexing because it creates a processing circularity: to determine which lidar soundings represent ocean depth one needs at the least an initial depth estimate or, more ideally, a *Bathy/NotBathy* designation for each sounding. This article describes a method that overcomes both of these difficulties and documents the results relative to a reference classification that is produced by operationally adopted procedures.

2. Study Area and Lidar Data

The airborne lidar data used for this work were captured by the United States National Oceanic and Atmospheric Administration (NOAA) between April 22 and 25, 2016, in the vicinity of Key West, Florida (24°33′ N, 81°46′ W). Data were acquired by collecting lidar soundings over multiple overlapping flight lines generally having a north-south orientation using a Riegel™ VQ-880-G sensor that employs a counter-clockwise circular scan and a 20^0 scan angle. The nominal flying altitude of 400 m above mean sea level results in an individual swath width of approximately 300 m and the pulse frequency of 45,000 pulses per second provides a spatial density of approximately 10 soundings sq m^{-1} for a single flight line. The lidar data were post-processed by NOAA by "cutting" the data from all flight lines into 500m-by-500m data "tiles" aligned north-south and east-west with the Universal Transverse Mercator (UTM) projection.

Data for four tiles (Figure 1) were provided by NOAA in the format of the LAS data standard Version 1.4-R13 (Point Data Record Format 6) (ASPRS 2013). These tiles were selected because they are representative of the range of sounding densities, depths, and ocean floor characteristics encountered in operational shallow water bathymetric mapping (Table 1). For convenience, the first five digits of each tile's northing are employed as its identifier as well as a depth indicator – Shallow, Deep, Deeper, or Deepest. The overlap in flight lines produces a combined average sounding density between 13 and 30 returns m^{-2}, although sounding density varies across each tile and is considerably higher where flight lines overlap.

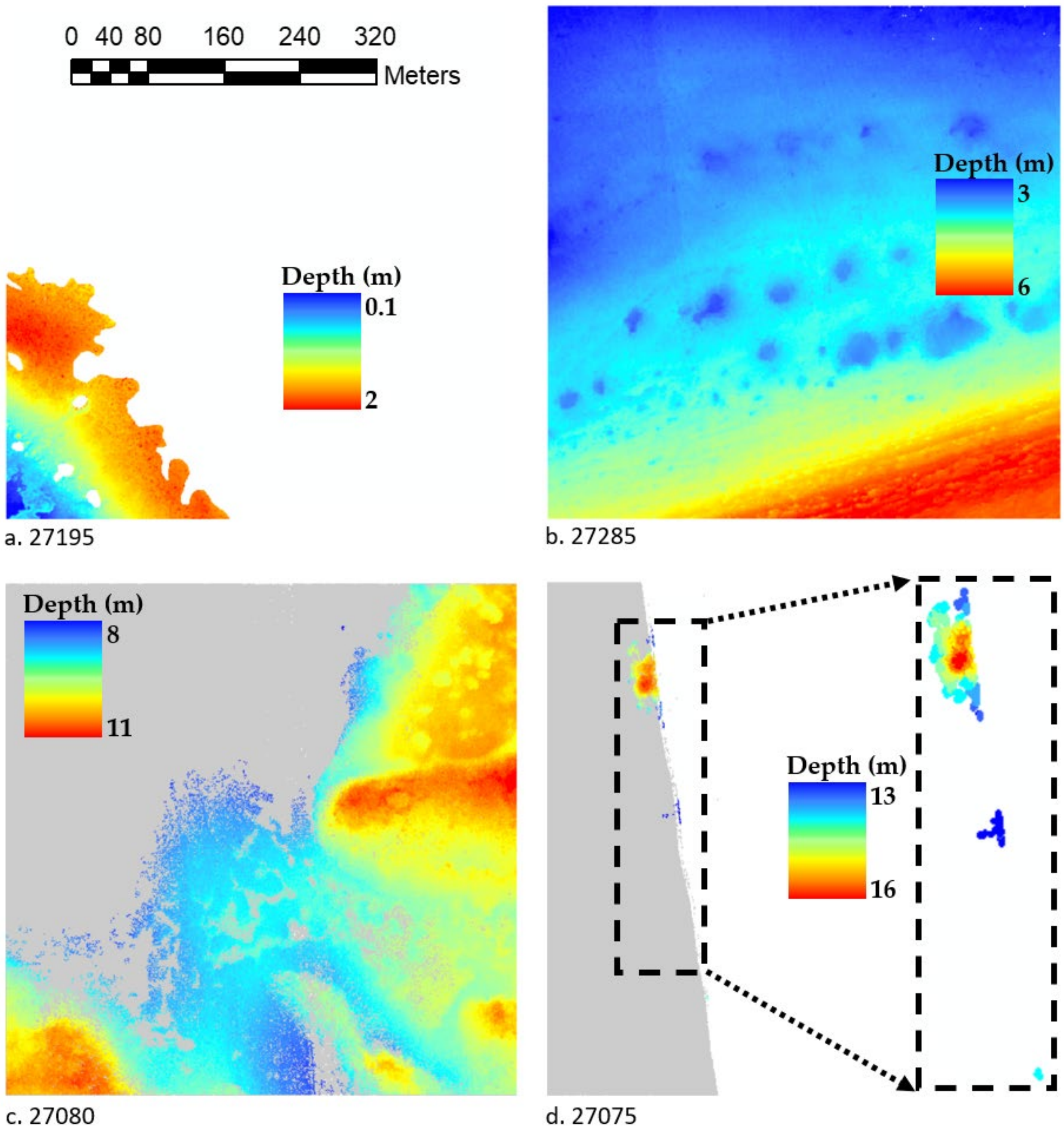

Figure 1. Depth maps (1m pixels) for the four tiles based on depth determined by NOAA. White areas have no usable data. Gray areas have usable data, but no soundings were identified as bathymetry by NOAA. Due to sparseness of NOAA-identified bathymetry on the deepest tile (27075; Fig. 1d), an enlargement of the area containing bathymetry is shown with bathymetric soundings accentuated and gray background removed.

Table 1. Descriptive information about the data tiles employed in this study.

Identifier (North-ing)	Relative Depth	Description	Area (km^2)	Approx. MSL depth range (m)	Total Soundings (million)	Mean return density (pts/m^2)	% Bathy-metry	Number of flight lines
27195	Shallow	Shallow area including some mangrove swamps	3	0 to 2	0.6	27.6	78	5
27285	Deep	Gradual slope with a few scattered mounds about 1 m tall	25	3 to 6	7.6	30.4	76	7
27080	Deeper	Gradual slope cut by relatively shallow channels; the northwest is poorly classified	25	8 to 11	3.7	14.8	21	7
27075	Deepest	Depth mostly beyond limit of lidar penetration except for mound in northeast and isolated points on eastern edge.	7.5	13 to 16	0.9	13.3	0.4	2

Attached to each sounding are its geographic (UTM) coordinates, depth, time of acquisition, and a variety of metadata that we term “sounding attribute data” (SAD; Table 2). Lidar depth is expressed in meters relative to mean sea level determined using NOAA’s VDATUM tool (https://vdatum.noaa.gov/). Sounding-based SAD are either acquired by the lidar instrument or were derived post-acquisition. Also provided were Smoothed Best Estimate of Trajectory (SBET) data. These are produced by the Applanix software by post-processing pulse return data from each flight line using a proprietary method based on a tightly coupled extended Kalman filter. SBET data have had noise removed to describe the most likely airplane position and orientation at 200 Hz. Flight path and orientation consistency are described in the SBET data by the standard deviations for the x, y, and z location of the plane and its yaw, pitch, and roll extracted from the Kalman filter’s post-observation covariance matrix. SBET values were assigned to individual soundings by matching time of acquisition.

An additional variable was created to characterize platform stability at the moment of data acquisition. It was observed that the crenularity – i.e., the deviation from a straight line -- of the margin of soundings of individual flight lines varied along the flight line (Figs. 2a and 2b). We hypothesized that this crenularity reflected local wind conditions that may in turn impact surface water conditions and lidar reflectance characteristics. To quantify this, sounding cloud ‘edge points’ were identified algorithmically along the length of the flight path (Fig. 2a). The two end soundings were considered ‘corner’ soundings and the equation of the straight line between them calculated (Fig. 2b). The orthogonal distance from each edge sounding to the straight line was determined (Fig. 2c) and the absolute value of this deviation was assigned to each sounding by matching time of acquisition. This variable is termed *abs_devia*.

Table 2. Depth and sounding attribute data (SAD) employed in this study for machine learning (ML) modelling. Variable names are *italicised* throughout the article.

SAD Type	**Nature**	**Variable (*Name*)**
Depth	***Depth***	Depth as provided via lidar post-processing (*depth* in m)
Sounding-based	***Pulse-specific***	• Intensity of sounding return (*intensity*; 16-bits – i.e., maximum value is 65536) • Number of soundings (*numreturns*) • Sounding number from a given lidar pulse (*return_no*) • First sounding of many (*first_of_many*; 0 or 1) • Last sounding of many (*last_of_many*; 0 or 1) • Last sounding (*last*; 0 or 1) • Scan direction (*scan_direct;* -1 (backwards) or +1 (forward)) • Azimuth from airplane to pulse (*azim2plse*; 0^0 to 360^0 in decimal degrees) • Incident scan angle corrected for yaw, pitch, and roll (*inciangle*; recorded in decimal degrees) • Difference between pulse direction and airplane heading (*plse_frm_hdng*; 0^0 to 90^0 decimal degrees)
Airplane stability	***SBET***	• Aircraft positional – sum of standard deviations of x, y, and z (*stdXYZ*) • Aircraft platform – sum of standard deviations of Yaw, Pitch, and Roll (*stdYwPtRl*) • Deviation from flight path – see Figure 2 and text for explanation (*abs_devia*)

Most of the SAD variables can be considered ‘direct features’ (Höfle and Rutzinger 2011) that are measured, although *abs_*devia would be considered an ‘indirect feature’ as it is derived post-acquisition. The SAD variables in Table 2 were retained for analysis because of their documented or hypothesized impact on light reflectance directly and bathymetric signal indirectly. Examples of their documented impact include *intensity* (Schmidt et al. 2012), *abs_devia* and SBET variables as surrogates for surface waves (Westfeld et al. 2017; Maas et al. 2019; Lowell et al. 2021), and *inciangle* (Birkeback et al. 2018; Okhrimenko and Hopkinson 2020).

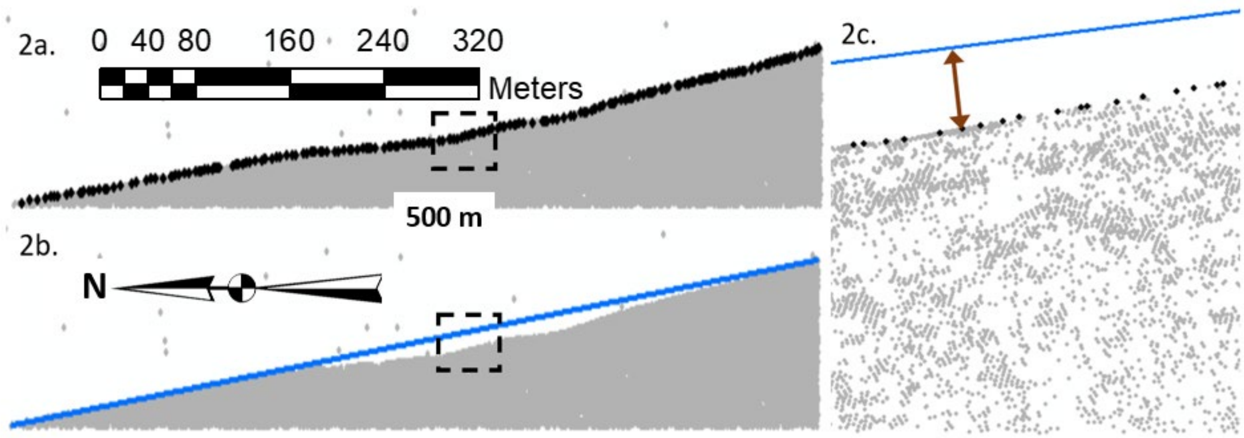

Figure 2. Extraction of orthogonal deviations from a single flight line (the western edge of the deepest tile 27075; Fig. 1d). To conserve space, North points to the left. a) Lidar point cloud (gray) and derived edgepoints (black). The dashed box is the area of enlargement in Fig. 2c. b) "Corner-to-corner" straight flight path (blue line). c) Orthogonal deviation of a single edge point from the corner-to-corner flight path edge of lidar point cloud.

Finally, also available for each sounding is the *Bathy/NotBathy* classification produced by NOAA using in-house methods. The LAS data standard classes of interest herein are 'Bth', 'Unc', and 'LP-Nz' that generally represent bathymetry, water surface, and water column noise, respectively. For the current work, these were condensed into two classes – *Bathy* ('Bth' only) and *NotBathy* ('Unc' and 'LP-Nz'). This *Bathy/NotBathy* classification is used only as the reference classification against which the results of the method developed are compared – i.e., it is not used in the method developed. Moreover, although the NOAA *Bathy/NotBathy* classification is the most authoritative available and is an appropriate standard for comparison since it is used operationally, it is not 'ground truth' produced via direct measurement or observation.

For analysis, notable in the data tiles employed is the spatial distribution of *Bathy* – i.e., the not-gray points in Figure 1. Tile 27195/Shallow (Fig. 1a) is the shallowest tile, is located in an area of mangrove swamps, and all of it except the southwest area has been classified by NOAA as being above sea level. For such areas, NOAA creates a data exclusion mask so that only soundings from aquatic areas are classified as *Bathy/NotBathy*. For consistency, the same practice is adopted herein and only data from the colored area shown in Fig. 1a are employed in subsequent analyses. Also notable – and representative of real-world conditions – are the incomplete *Bathy* coverages of Tiles 27080/Deeper (Fig. 1c) and 27075/Deepest (Fig. 1d). This results from increasing depths that ultimately exceed the depth limit of lidar penetration. It is most pronounced for Tile 27075/Deepest on which only 0.4% of soundings are *Bathy* (Table 1).

3. Procedures

Figure 3 is a schematic showing the procedural flow of the work undertaken. This is explained below and was applied to each data tile individually.

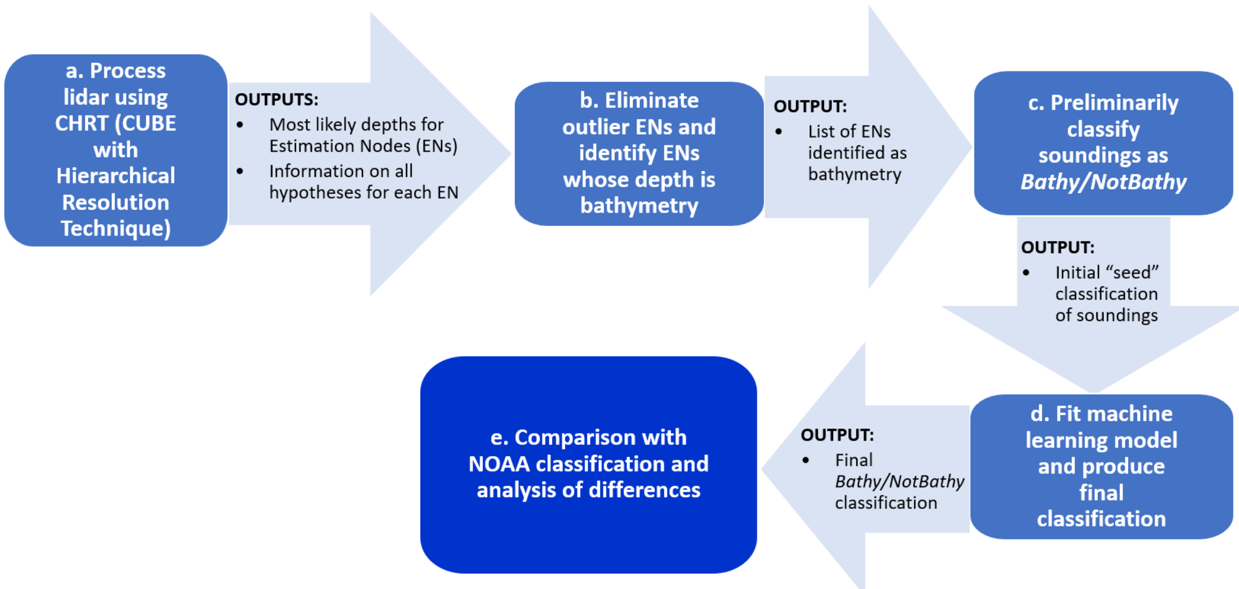

Figure 3. Schematic of data processing flow. Letters refer to parts of section 3 of the paper; the final step ("e.") is not further described in section 3.

3a. Process lidar using density-based/CHRT algorithm

This process identifies the most likely depth (MLD) for a north-south/east-west grid of 'estimation nodes' (ENs) established over each lidar data tile. This information is produced by processing the lidar point cloud data through a density-based algorithm as described below.

The algorithm employed in this study is CHRT (CUBE with Hierarchical Resolution Technique; Calder and Rice 2017) which is a modification of CUBE (Combined Uncertainty and Bathymetry Estimator; Calder and Mayer 2003). CHRT is incorporated into many software packages that are widely used operationally and scientifically for processing MBES sonar data (Lecours et al. 2016). Its scientific use has also been extended into other applications such as benthic habitat mapping (Calvert et al. 2015).

CHRT establishes a grid of ENs across an area of interest with the spacing of the grid determined by the density of the soundings. Given the non-rectangular nature of the spatial coverage for some tiles in this study (see Figure 1), some ENs on a grid have no associated lidar soundings and are removed from further analysis. Similarly, as will be explained subsequently, others have aberrant MLD values due to data anomalies and sparseness; these are also removed using outlier analysis.

To estimate the MLD for a single EN, the ‘neighboring’ soundings for the EN are identified. An EN’s ‘neighboring’ soundings are those within a geographic radius defined by the grid spacing for a tile. The radius used is the Euclidean distance to the ‘pixel corner’ defined by a point equidistant from four adjacent ENs (Figure 4). Thus 50-60% of soundings are neighbors of two ENs depending on local sounding density.

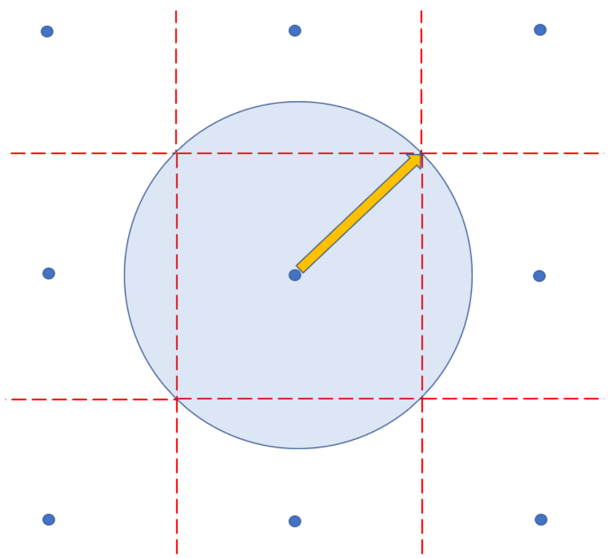

Figure 4. Search radius to determine “neighboring” pulse returns. Blue points are estimation nodes (ENs). The orange arrow defines the neighbor radius.

For each EN, its neighboring soundings are progressively ingested with the first sounding defining an initial depth ‘hypothesis.’ A variety of hyperparameters or ‘tuning’ parameters establish initial thresholds for determining if two soundings represent different depths; default parameters are based on user experience with the location and variability of depth frequency distributions for various depth conditions. The depth of the second sounding ingested is evaluated against the hyperparameters to determine if it also ‘belongs to’ the first depth hypothesis, or if its depth is ‘different enough’ to be considered a new hypothesis. This process continues until all neighboring soundings have been ingested and one or more depth hypotheses have been developed and characterized. As the process progresses, frequency distributions for each hypothesis are produced and their characteristics – rather than the initial hyperparameters -- increasingly control the assignment of newly ingested soundings to existing hypotheses, or to the creation of a new hypothesis. After all soundings have been ingested, disambiguation rules determine which hypothesis represents the MLD for the EN. A simplistic example is that the hypothesis with the deepest mean depth is considered the MLD although such a rule ignores factors such as turbidity in the water column or the number of soundings in the deepest hypothesis; current disambiguation rules generally identify the hypothesis having the greatest number of soundings as most likely.

Table 3 provides summary information about the ENs for each tile. One notable point is that the variability (standard deviation) of MLDs for tile 27080/Deeper is considerably larger than for other tiles. This reflects both a greater variability in geomorphometry for this tile and the large area in the northwest beyond the range of lidar penetration (Fig. 1c). Note that because each tile is processed individually (here and operationally) and the density of the EN grid on each tile depends on its sounding density, EN density varies across tiles. This causes edge artifacts when combining adjacent tiles into a seamless map; procedures for doing this are beyond the scope of this study.

Table 3. Estimation node (EN) information after removal of outliers. (See text for explanation.)

Identifier (Northing)	Relative Depth	Grid Spacing (m)	EN Grid: Rows*Cols	ENs used for analysis[1]	Mean hypotheses per EN	Mean soundings per hypothesis	MLD[2] range (m)	Mean MLD (m)	MLD standard deviation (m)
27195	Shallow	12.4	20*18	188	4.8	3827	0 to 1	0.7	0.2
27285	Deep	1.6	308*308	94853	4.6	125	2 to 7	4.5	0.7
27080	Deeper	3.0	167*167	27823	6.3	32	1 to 10	3.5	2.3
27075	Deepest	1.9	267*120	18446	5.2	82	1 to 18	1.5	0.5

[1]Estimation Nodes after removal of no-data and outlier estimation nodes.
[2]Most likely depth (MLD).

3b Eliminate outliers and identify ENs whose MLD is bathymetry

A two-phase outlier screening process is employed; Table 4 provides information on the results of this screening.

Table 4. Information about estimation node (EN) outlier screening.

Identifier (Northing)	Relative Depth	Total Grid ENs	ENs w/o Soundings	MD[1] Outliers	Beyond Lidar Penetration MLDs[2]	ENs Analysed
27195	Shallow	360	136	2	37	185
27285	Deep	94864	11	1232	0	93621
27080	Deeper	27889	0	318	14	27557
27075	Deepest	32040	13514	288	34	18204

[1]Mahalanobis Distance.
[2]Most Likely Depth (MLD).

First, 12 variables associated with each EN's hypotheses are used to calculate Mahalanobis distances (MDs; Mahalanobis 1936) for each EN. Examples of such variables are the number of hypotheses, total sounding, the number of soundings associated with the MLD hypothesis and non-MLD hypotheses, and the standard deviation of the depth of soundings associated with the MLD and non-MLD hypotheses. Prior to calculating the MDs, variables are normalized between 0 and 100 using max-min normalization. ENs are eliminated from subsequent analysis if their MD is in the outer 0.1% of the frequency distribution – i.e., their MD is more than approximately 3.3 standard deviations from the mean MD.

Second, airborne lidar cannot penetrate below certain ocean depths. Examination of depth frequency distributions across all tiles suggested that for the area studied, lidar could not penetrate below a depth of 20 m. Hence ENs whose MLD depth was greater than 20 m and that had not already been removed by the MD outlier analysis are eliminated as ''Beyond Lidar Penetration' MLDs.

MLD frequency distributions for the four tiles were highly irregular – e.g., not clearly normal or bi-modal – generally reflecting a separation of ocean surface and ocean bottom. Hence *k*-means clustering (Steinhaus 1957, McQueen 1967) is applied to the MLD of the ENs retained for analysis to separate them into two classes. Because a single variable – MLD -- is used in this clustering, this is equivalent to separating a frequency distribution along a single axis. The cluster having the greatest difference between its MLD and the average depth of all other hypotheses – i.e., the 'non-MLD' hypotheses -- is assumed to contain some EN hypotheses that are 'definitely' bathymetry. The other cluster has a smaller difference between the mean MLD and the mean depth of non-MLD hypotheses suggesting that both represent the ocean surface. Note that not all ENs will represent ocean floor or surface, but some will represent the water column. This is most likely to be problematic where water column soundings are more prevalent than ocean floor soundings – i.e., in highly turbid waters or beyond the limits of lidar depth penetration.

The mean and standard deviation for the MLDs of the cluster identified as 'definitely' containing bathymetry are used to define the bathymetry MLD confidence interval for the MLDs for all ENs. The shallower MLD limit of the interval is the one-sided 99.9% confidence limit whereas the deeper MLD limit is the one-sided 95% confidence limit. These 'imbalanced limits' were found to address the irregularly shaped bathymetry frequency distributions across all tiles better than equal 'shallower/deeper' MLD confidence limits.

The MLD bathymetry confidence interval is used to classify the MLDs of all ENs as *Bathy* or *NotBathy*. That is, the MLD hypotheses of all ENs contained within the bathymetry confidence interval are classified as

Bathy. All other hypotheses, even including those whose mean depth falls in the bathymetry depth interval but are not the MLD for their EN, are classified as *NotBathy*.

Alternatives to this classification rule were evaluated, including clustering on all hypotheses rather than on MLD hypotheses only, classifying as *Bathy* all hypotheses – MLD and non-MLD -- whose mean depth fell in the bathymetry depth interval, and using the range instead of the standard deviation to define the bathymetry confidence interval. None performed as well as the clustering approach and classification rule adopted.

3c Preliminarily classify pulse returns as *Bathy* or *NotBathy*

The neighboring soundings for each EN are classified as *Bathy* if they are associated with the MLD and the MLD has been classified as *Bathy*; otherwise they are classified as *NotBathy* (Table 5). Soundings that are neighbors of two ENs whose *Bathy* or *NotBathy* classification agrees are assigned to the agreed class. Two-neighbor soundings whose classifications do not agree are termed 'mixed' and are assigned to *Bathy*. Tile 27195/Shallow with the least separation between ocean surface and ocean depth had the largest percentage of mixed soundings and Tile 27075/Deepest had the smallest (Table 5). This assignment scheme for mixed soundings has the effect of decreasing the number of false negatives (FNs) – undetected bathymetry – which is a more serious error than a false positive (FP) – erroneously labelled bathymetry – in nautical chart production. Assigning mixed soundings to *Bathy* also was found to improve the skill of the subsequently fitted machine learning model.

Table 5. *Bathy/NotBathy* classification information for soundings that were neighbors of two estimation nodes (ENs).

Identifier (Northing)	Relative Depth	Soundings Neighboring Two ENs	Pure Bathy Soundings	Pure NotBathy Soundings	"Mixed" Soundings	% "Mixed"
27195	Shallow	307500[1]	138500	85700	83300	27
27285	Deep	4246400	2307400	1547700	391300	9
27080	Deeper	2081400	600900	1343100	137400	7
27075	Deepest	542100	3100	538400	600	<1

[1]All values rounded to nearest 100.

3d Fit a machine learning model to produce a final *Bathy/NotBathy* classification for all soundings

At this point a preliminary or 'seed' classification has assigned each sounding to the *Bathy* or *NotBathy* class. However, only soundings associated with the MLD hypothesis will have been classified as *Bathy*. This is problematic in areas where all hypotheses – MLD and others – are representative of bathymetry.

This occurs on tiles where bathymetry is commonplace (27195/Shallow and 27285/Deep; Table 1), as well as in areas where bathymetry is locally concentrated such as the northeastern area of 27075/Deepest (Fig. 1d). Hence this assignment is a 'preliminary' or 'seed' classification that can be considered conservative – i.e., soundings classified as *Bathy* have an extremely high 'true probability' of being *Bathy,* but the classification likely contains a high number of FN errors.

Therefore, to produce a final classification, extreme gradient boosting (XGB) (Friedman 2001) is used to fit a model using the seed classification. XGB is a decision-tree machine-learning technique that progressively fits numerous simple or 'shallow' models/trees. Each successive tree is fitted with a focus on the worst-predicted observations of the previous tree. Once statistical convergence or the maximum number of trees is achieved, a composite XGB model is produced using a 'majority vote' approach. Lowell et al. 2021) have demonstrated that the SAD employed (Table 2) contain a substantial amount of bathymetric signal. Specifically, machine learning models that used the variables in Table 2 as independent variables and NOAA's *Bathy/NotBathy* as the dependent variable produced R^2 values between 0.61 and 0.99 and global classification accuracies between 90% and 99.9%.

Hence, an XGB model that uses the seed CHRT/clustering-based classification as the dependent variable and the variables in Table 5 as predictor variables is fitted for each tile. This model is then used to estimate *p(Bathy)* – i.e., the probability of each pulse return being *Bathy* for each sounding. Soundings are classified as *Bathy/NotBathy* by applying to the *p(Bathy)* values a probability decision threshold (PDT) that equalizes the true positive (*Bathy*) rate (TPR) and true negative (*NotBathy*) rate (TNR) rather than the conventional PDT of 0.50; we term this alternative PDT the "optimal decision threshold" or "ODT." Lowell et al. (2021) demonstrated that the use of the ODT mitigates the impacts of the accuracy of a class – *Bathy* or *NotBathy* – comprising a strong majority of soundings being maximized at the expense of the accuracy of the minority class. Problems associated with applying a conventional PDT of 0.50 were notable for all tiles, but especially for 27075/Deepest on which only 0.4% of pulse returns were identified by NOAA as being *Bathy*.

4. Analysis, Results, and Discussion

Initially assessed are 1) how well the 'preliminary/seed' *Bathy/NotBathy* classification derived from clustering the EN MLDs performs relative to the NOAA reference classification and 2) if the final XGB model-based classification improves the preliminary seed classification relative to the NOAA reference classification. Methods that can be used for continuous improvement of classification methodology are subsequently presented.

4a. Classification accuracies

Recall that the preliminary seed classification produced by clustering EN MLDs from the CHRT algorithm is considered a conservative classification because only the most certain soundings are classified as *Bathy*. Nonetheless, because it will be subsequently refined using an XGB model, it only needs to be 'sufficiently accurate' that the XGB model will be able to detect and describe underlying relationships between bathymetry and the SAD variables (Table 2). If this occurs, the XGB model should be able to identify soundings not classified as *Bathy* initially, but that have a high *p(Bathy)* nonetheless. Subsequently reclassifying all soundings based on the *p(Bathy)* values should thus expand the number of soundings correctly classified as *Bathy* or *NotBathy*. A truly ideal outcome would be that the preliminary seed classification is identical to the NOAA reference classification and thus does not require additional processing.

Figure 5a suggests that the preliminary seed clustering classification relates strongly to the NOAA reference classification. Global accuracy – or more precisely 'agreement between the two' – is the percentage of all soundings that are correctly classified as *Bathy* or *NotBathy*; it is at least 85% for all tiles. Moreover, the percent of correctly classified *Bathy/NotBathy* soundings – true positives (TPs) and true negatives (TNs), respectively – is 81% or more for all tiles. Similarly, the percentage of false positives (FPs; *NotBathy* soundings incorrectly labelled *Bathy*) and false negatives (FNs; *Bathy* soundings incorrectly classified as *NotBathy*) is 20% or lower for all tiles.

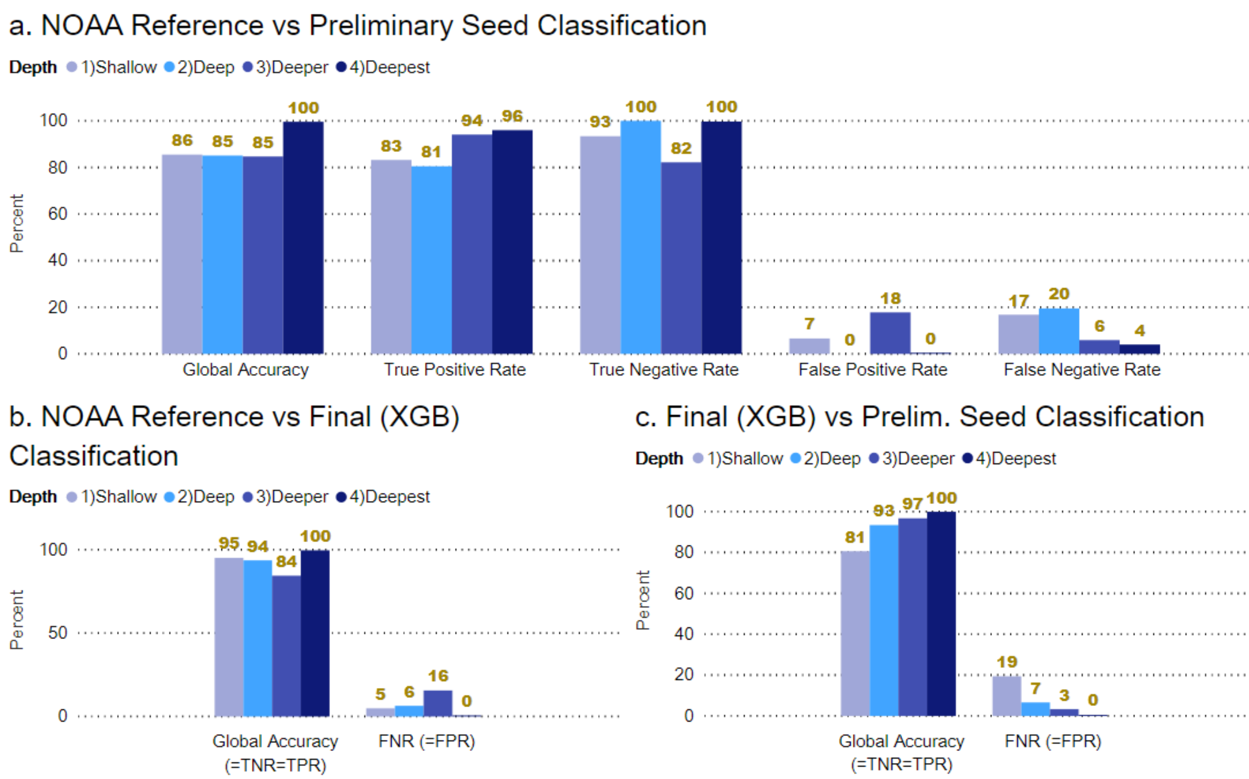

Figure 5. Classification accuracy/comparisons for various classification schemes.

The ability of an XGB model to improve on the preliminary classification – i.e., to harmonize it with the NOAA classification -- can be assessed by comparing Figs. 5a and 5b. Readers are reminded that to classify *p(Bathy)* as *Bathy/NotBathy*, the ODT that equalizes the TNR and TPR was employed throughout. Since the FNR is 100 minus the TPR, the FPR is 100 minus the TNR, and the TPR and TNR are equal, the FNR and FPR are equal. The use of the ODT also has the impact of making the global accuracy equal to the TNR and TPR – e.g., if the classification of *NotBathy* soundings is 80% correct and the classification of *Bathy* soundings is 80% correct, global accuracy for all soundings must also be 80%.

Figure 5b suggests that the XGB model improved the initial cluster-based 'seed' classification. Through the use of an XGB model, global accuracy improved for all tiles except 27080/Deeper for which it decreased by a single percent (85% to 84%). The TPR follows a similar pattern, although the TNR did decrease for Tile 27285/Deep (from 100% to 94%). Of greatest real-world interest is that the FNR has dropped considerably for all tiles except 27080/Deeper. In operational practice, a decrease in FNs not only means improved navigational safety but also considerable cost savings. That is, because FNs are the most serious error for nautical navigation, considerable human-time is spent verifying FNs. Hence reducing the number of FNs decreases the time spent on manual editing. Thus except for the 27080/Deeper tile, the ML model has improved classification in a way that benefits operational workflows and improves navigational safety. We note a considerable portion of the 27080/Deeper tile (northwest of Figure 1c) exceeds the depth of lidar penetration which undoubtedly impacts the accuracy of identifying *Bathy* soundings.

Results can potentially be further improved by better understanding the XGB model fitted to the binary *Bathy/NotBathy* cluster-based "seed" classification using the variables in Table 2 as the independent variables. The classification accuracy of the model on this preliminary classification– i.e., the classification used to fit the model rather than NOAA's reference classification -- is at least 81% for all tiles (Fig. 5c). The goodness-of-fit/explanatory power of the model can also be evaluated by calculating an R^2 for binary dependent variables (McFadden1974) that is conceptually equivalent to, and interpreted in the same manner as, the more familiar R^2 value associated with linear regression. The R^2 values for individual models are relatively high (Table 6) – particularly considering the large number of soundings (at least 600,000; Table 1) used to fit the models. The number of variables with an 'importance value' (a measures of a variable's contribution to an XGB model) greater than zero (0) was at least 11 for all tiles and the five most important variables contained at least 95% of the total importance for all tiles except 27195/Shallow. Coupled with the fact that other than *depth*, an inconsistent variety of SAD variables were important, the information that provides discrimination between *Bathy* and *NotBathy* soundings appears to be distributed among a suite of variables specific to each tile; XGB as a model development technique is able to accommodate this variability.

These findings are potentially most relevant for the 27195/Shallow tile. Its relatively low cumulative importance of the five most important variables (0.85) suggests that for shallow areas where the distance between the ocean surface and ocean floor is less than the noise in the lidar sounding cloud, *depth* contains a smaller proportion of the information the provides discrimination between *Bathy* and *NotBathy* soundings than it does for deeper areas. Finally, SBET variables were among the five most important variables in two of the models and *abs*_devia was present in one suggesting that both SAD associated with individual soundings and SAD that describe flight path and airplane stability contain information that provides discrimination between *Bathy* and *NotBathy* soundings.

Table 6. Information about XGB models.

Identifier (Northing)	**Relative Depth**	**R-squared[1]**	**Number of Important Variables**	**Five most important variables[2]**	**Cumulative importance of the five most important variables**
27195	Shallow	0.48	12	*depth, last, first_of_many, stdYwPtRl, inciangle*	0.85

27285	Deep	0.79	14	*depth, return_no, last_of_many, intensity, plse_frm_hdng*	0.95
27080	Deeper	0.87	12	*depth, num_returns, return_no, last_of_many plse_frm_hdng*	0.99
27075	Deepest	0.97	11	*depth, plse_frm_hdng, abs_devia, scan_direct, stdXYZ*	0.99

[1]McFadden's (McFadden 1974) pseudo R^2 which cannot be tested for statistical significance.
[2]In descending order of importance.

4b. Continuous Improvement.

Because neither the NOAA classification nor the XGB classification developed can be considered ground 'truth' that results from direct measurement, being able to characterize differences between the two is useful for improving the NOAA classification, the final XGB classification, or both. Two methods were developed to provide such information.

The first method focusses on 'feature' or 'statistical' space and entails comparing *p(Bathy)* values from the XGB model with the NOAA *Bathy/NotBathy* classification using logistic regression. The approach is comparable to binning the soundings by *p(Bathy)* values and then determining if the proportion of soundings classified as *Bathy* by NOAA in each bin equals the bin mid-point class value. The logistic regression approach employed, however, provides information along the continuum of *p(Bathy)* values without requiring an arbitrary number of bins. In this approach, NOAA's *Bathy/NotBathy* classification is used as the dependent variable and the following logistic equation is fitted:

$$p' = \left(1 + e^{\left(-(b_0 + b_1 L)\right)}\right)^{-1} \tag{1}$$

where L is:

$$L = ln\left(\frac{p}{(1-p)}\right) \tag{2}$$

and *p* is the *p(Bathy)* estimated by the XGB model. For each tile, if the NOAA classification and *p(Bathy)* values from the XGB model are identical over the entire *p(Bathy)* range of 0.0 to 1.0, b_0 and b_1 in Equation (1) will be 0.0 and 1.0, respectively. Furthermore, R^2 for Equation 1 will be 1.0 with an associated log-likelihood *p* that is infinitesimally small. Such a 'logistic agreement model' was fitted for each tile (Table 7).

Table 7. Information on logistic agreement models.

Identifier (Northing)	Relative Depth	R-squared[1]	log-likelihood *p*	b_0[2]	b_1[3]	n
27195	Shallow	0.85	<0.001	1.1*	2.05*	576,000
27285	Deep	0.79	<0.001	19.3*	2.83*	7,599,000
27080	Deeper	0.52	<0.001	-1.3*	0.60*	3,706,000
27075	Deepest	0.77	<0.001	-2.3*	0.70*	983,000

[1]McFadden's pseudo R^2 that cannot be tested for statistical significance.
[2] * signifies the intercept value is significantly different from 0.0 at α=0.001.
[3] * signifies the slope value is significantly different from 1.0 at α=0.001.

The relatively high R^2 values and low log-likelihood *p* values for the logistic models (Table 7) suggest a strong and significant relationship between the *p(Bathy)* produced by the XGB model fitted on the preliminary classification and the NOAA *Bathy*/*NotBathy* classification. However, for all models the intercepts (b_0) and slopes (b_1) are significantly different from 0.0 and 1.0, respectively.

To assess (dis)agreement over the entire probability range, the logistic agreement models of Table 7 can be displayed graphically by plotting *p′* vs. *p* over the interval {0,1}. This was done using the ODT that is specific to each tile – i.e., by stretching *p(Bathy)* values in the range {0, ODT} to the interval {0.0, 0.5}, and stretching *p(Bathy)* values in the range {ODT, 1.0) .0 to the interval {0.5, 1.0}. Note that this segmenting of the probability range is the cause of the graphical discontinuities present at a *p(Bathy)* value of 0.5 for some tiles (Figure 6).

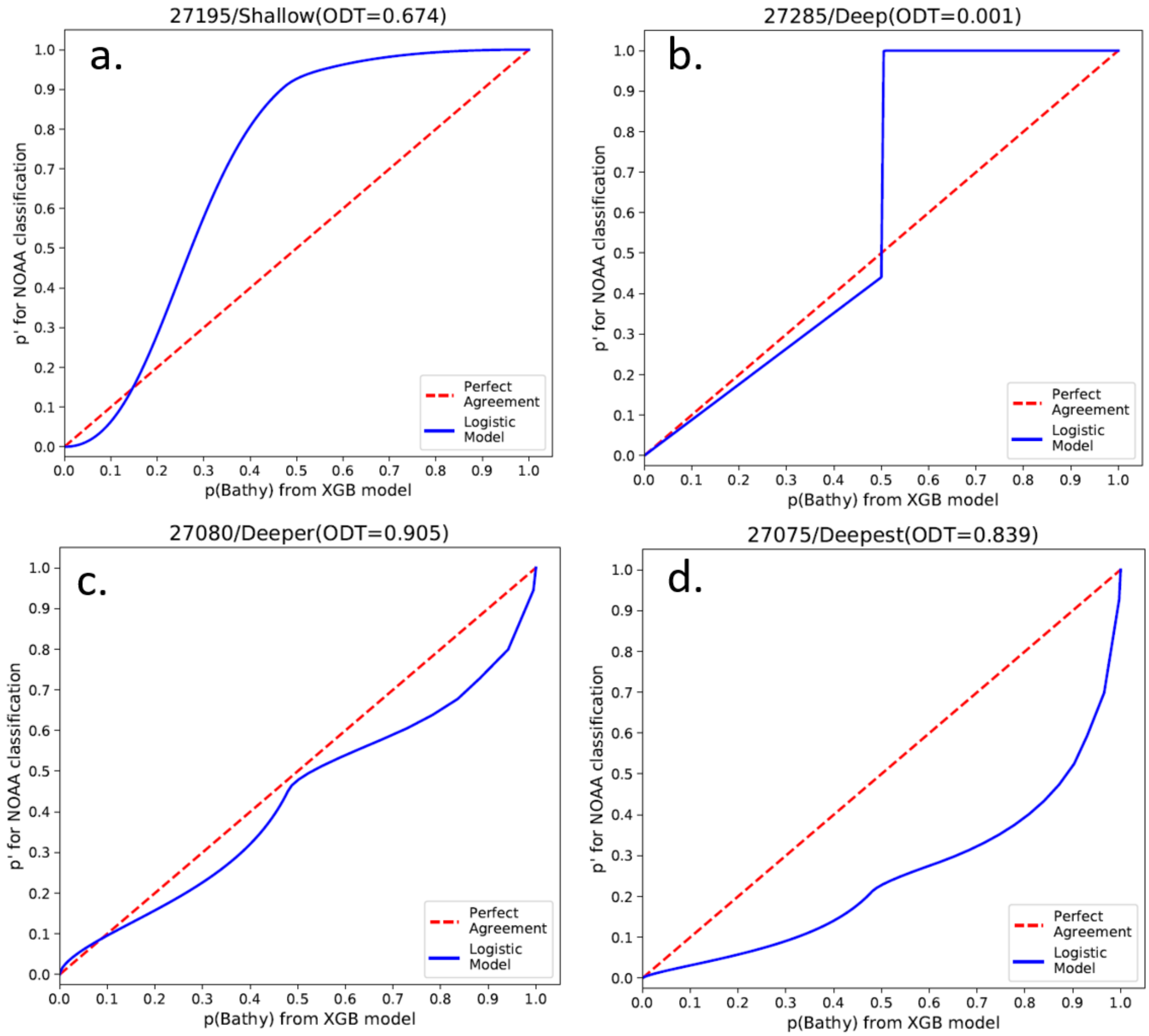

Figure 6. Agreement between p(Bathy) from NOAA classification and CHRT-based XGB model using logistic agreement models. ("ODT" is the tile-specific optimal decision threshold.)

The graphs of *p'* vs. *p(Bathy)* suggest bias along the range of *p(Bathy)* values whose magnitude varies by tile. For Tiles 27080/Deeper (Fig. 6c) and 27075/Deepest (Fig. 6d), relative to the NOAA *Bathy/NotBathy* classification XGB *p(Bathy)* values are overestimated over the entire range, resulting in a relatively large number of false positives (FPs). Given that these are the two tiles having depths that exceed lidar's penetration capability, we hypothesize that the FPs are spatially concentrated on the deeper edges of areas that NOAA identified as bathymetry; this will be examined explicitly. The XGB model for Tile 27285/Deep (Fig. 6b) performs reasonably well below the ODT but severely underestimates *p(Bathy)* above the ODT. For practical purposes, this may not be problematic. This indicates that, according to the XGB model, any sounding whose *p(Bathy)* is above the ODT is 'definitely' *Bathy*. Accordingly, all pulse returns having a *p(Bathy)* value greater than the ODT will be classified as *Bathy* – which is likely to be the correct classification for the vast majority of such soundings. Tile 27195/Shallow shows that below a *p(Bathy)* value of about 0.15 the XGB model overestimates *p(Bathy)* thereby producing FP errors, but higher *p(Bathy)* values are underestimates thereby resulting in FN errors. That Fig. 5b does not indicate a large number of FNs for Tile 27195/Shallow suggests that examination of the spatial distribution of FNs (and FPs) might be particularly useful for this tile.

Examination of the geographic distribution of the differences between the ML-based and reference classifications is the second method of characterizing misclassification errors. To examine the spatial distribution of the FNs and FPs, each tile was divided into 20 m pixels. If there is no spatial bias, the errors will be distributed across each tile as the lidar pulse returns are – i.e., areas having a high density of pulse returns should have a comparably high density of FNs and FPs. To determine if the densities of pulse returns and errors were similar and therefore not spatially biased, the differences between the percent of total lidar pulse returns and percent of FNs and FPs in each pixel can be calculated with negative values indicating an 'excess' of FNs or FPs. These differences can then be displayed spatially (Figures 7, 8, 9, and 10) such that negative/brown values indicate 'too many' FNs or FPs, while positive/green values represent 'too few.'

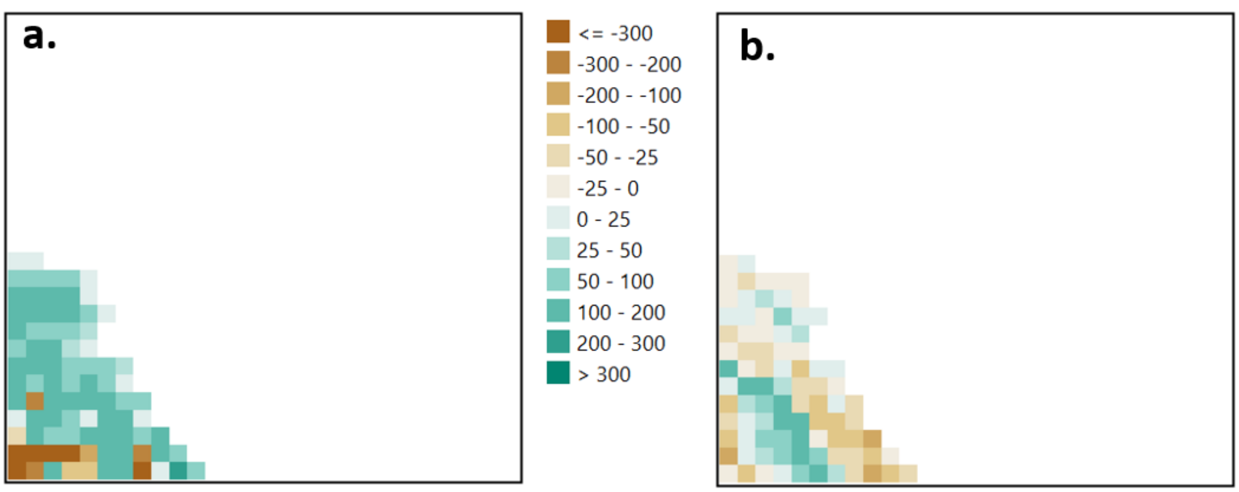

Figure 7. Tile 27195/Shallow. a. Difference between percent of *Bathy* points and False Negatives (FNs) in each pixel (times 100). b. Difference between percent of *NotBathy* points and False Positives (FPs) in each pixel (times 100). (Negative values indicate an "excess" of FNs or FPs.)

The pattern for FNs (undetected *Bathy*) for Tile 27195/Shallow (Fig. 7a) is unexpected: *Bathy* is fairly accurately detected in the shallower northeast edge of data (relatively few FNs) where there are also about the expected number of FPs (Fig. 7b), but there are 'too many' FNs in the deeper southwestern portion. Also of interest is that there is an area (the green northwest-to-southeast band in Fig. 7b) between the shallow northeast and deeper southwest where there are ‘too few’ FPs. It is also notable that the magnitude of differences for FPs is less than for FNs as indicated by the more muted colors in Fig. 7b.

The information for the other tiles can be interpreted similarly:

Tile 27285/Deep (Figure 8): There is an ‘excess’ of FNs (undetected *Bathy*) in the southeastern area (Fig. 8a) which is the shallowest area of the tile, and an excess of FPs (erroneously detected *Bathy*) in the northwest. (The dark green north-south bands in Fig. 8a correspond to areas of flight line overlap where sounding density is abnormally high.) The magnitude of differences is greater for FNs than for FPs as indicated by the more muted colors for the latter (Fig. 8b). Noting that in practical terms FNs are potentially more serious than FPs, the southeast of this tile may be an area where continuous improvement efforts should be concentrated, although verifying the FPs in the northwest would also lead to better accuracy for that portion of Tile 27285/Deep.

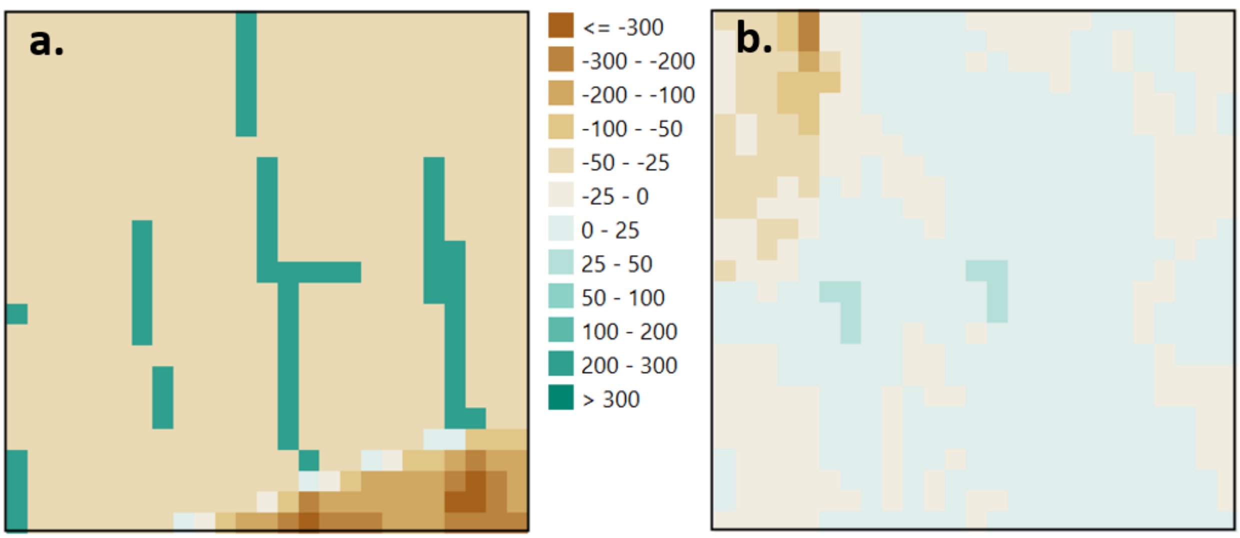

Figure 8. Tile 27285/Deep. a. Difference between percent of *Bathy* points and False Negatives (FNs) in each pixel (times 100). b. Difference between percent of *NotBathy* points and False Positives (FPs) in each pixel (times 100). (Negative values indicate an “excess” of FNs or FPs.)

Tile 27080/Deeper (Figure 9): There is an ‘excess’ of FNs in the northeastern and southwestern quadrants of this tile (Fig. 9a). These are the shallowest area of this tile (see Fig. 1c) and are generally surrounded by areas of relatively low differences (muted greens and browns). Figure 9b indicates that ‘too many’ FPs are present on the eastern edge of this tile suggesting that either the XGB classification identifies too many *Bathy* soundings in this area, that the NOAA classification does not identify enough, or both. The ‘too few’ FNs on the northwestern edge of *Bathy* pulse returns (Fig. 1c) coupled with the low level of FPs in this area suggest that for this tile, the overall error rate is low at the limits of lidar light penetration. The ‘too high’ number of FPs on the shallowest eastern edge of this tile may suggest that NOAA’s classification procedures, the XGB model, or both could be improved by further studying this area and the depths it represents.

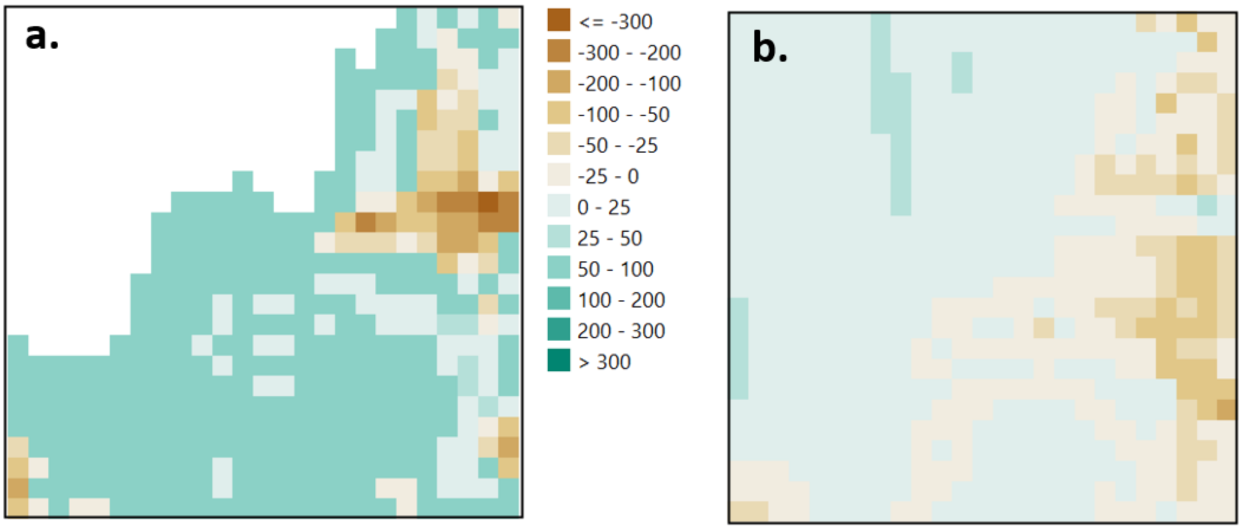

Figure 9. Tile 27280/Deeper. a. Difference between percent of *Bathy* points and False Negatives (FNs) in each pixel (times 100). b. Difference between percent of *NotBathy* points and False Positives (FPs) in each pixel (times 100). (Negative values indicate an “excess” of FNs or FPs.)

Tile 27075/Deepest (Figure 10): There is an ‘excess’ of FNs – undetected *Bathy* -- in the northeastern area (Fig. 10a) which is the shallowest area and the area with the highest density of *Bathy* lidar pulse returns (see Figure 1d). There are ‘too many’ FPs in the same area (Fig. 10b) further suggesting that *Bathy* may be under-detected in this area (and/or that the XGB model performs poorly in this area). Interestingly, however, FPs are fairly widely distributed spatially. This distribution of FPs strongly suggests that undetected *Bathy* soundings might be present throughout the tile. It is possible that the widely distributed FPs result from NOAA adopting a conservative approach to extracting bathymetry from areas near the limit of lidar ocean penetration. Regardless of the reason, these results demonstrate how this type of analysis could help to guide continuous improvement.

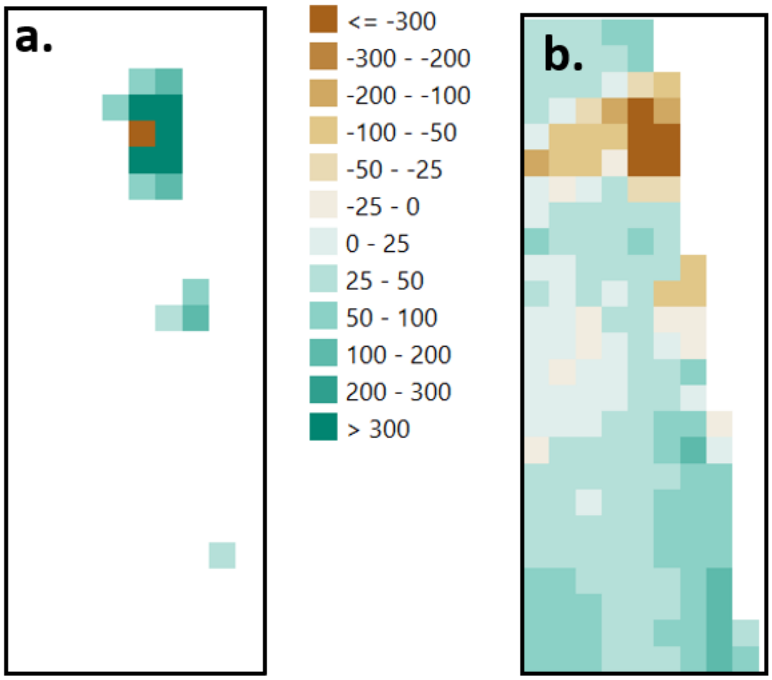

Figure 10. Tile 27075/Deepest. a. Difference between percent of *Bathy* points and False Negatives (FNs) in each pixel (times 100). b. Difference between percent of *NotBathy* points and False Positives (FPs) in each pixel (times 100). (Negative values indicate an "excess" of FNs or FPs.)

5. Summary and Conclusions

Lidar soundings that identify bathymetry could be extracted from lidar point clouds without the need for an *a priori* estimate of depth with average global accuracies, true positive rates, and true negative rates of 93% compared to a reference classification for four 500 m-by-500m lidar data tiles located near Key West, Florida. These 'accuracies' are achieved relative to a reference classification that is used operationally but that also has an unknown level of uncertainty. The accuracy of a preliminary *Bathy/NotBathy* classification derived solely from a density-based algorithm coupled with unsupervised clustering was improved for three of the tiles by fitting and applying a machine learning extreme gradient boosting model to produce a final *Bathy/NotBathy* classification. Models for each tile were fit using the preliminary classification as the dependent variable and 14 SAD variables such as the intensity and incidence angle of each pulse return as independent variables. Though depth was consistently the most important SAD variable, the models for all four tiles contained at least 11 SAD variables indicating that the information that distinguishes between *Bathy* and *NotBathy* soundings is dispersed among numerous SAD variables with no consistency across all tiles. Moreover, the information that distinguishes between *Bathy* and *NotBathy* soundings is spread among SAD that quantify pulse reflectance characteristics and airplane stability with the importance of individual SAD.

Two methods were employed to characterize differences between the reference and the ML-based classification in feature/statistical and geographic space. These are exemplified by feature-space information in Table 7 and Figure 6, and spatial information presented in Figures 7 through 10. One use of this information would be refining the XGB final classification. However, a potentially more valuable application would be continuous improvement of NOAA processing methodology. Because the true level of accuracy in the XGB and NOAA classifications is unknown, the results in Table 7, Figure 6, and Figures 7 through 10 may be viewed as identifying differences between two independent classifications rather than differences against 'truth.' The differences might be due to weaknesses in either or both classification(s), or the large-difference areas may be where bathymetry is simply difficult to extract. Knowing this could lead to revisions in XGB and/or NOAA classification procedures depending on confidence in a classification, the severity or potential practical consequences of the observed differences, and a variety of other factors.

In closing, the dual objectives of this work are recalled: to diminish the impacts of extracting bathymetry from lidar sounding clouds for shallow water using machine learning, and to achieve this without the circularity of needing a pre-existing classification or even depth estimate. If one accepts that the NOAA classification used for evaluation is an authoritative – but not error-free – reference, we argue that these objectives have been achieved across a range of depth and data conditions. While refinement of methods and a better understanding of the nature of errors can certainly provide improvements, this work at least provides a workflow to decrease time-consuming manual effort in extracting bathymetry from lidar sounding clouds.

List of Abbreviations

CHRT – CUBE (Calder and Mayer 2003) with Hierarchical Resolution Technique (Calder and Rice 2017)

EN – Estimation Node for the density-based algorithm

FN, FNR – False Negative Rate (*Bathy* soundings erroneously identified as *NotBathy*)

FP, FPR – False Positive Rate (*NotBathy* soundings erroneously identified as *Bathy)*

MD – Mahalanobis Distance (used for outlier analysis)

ML – Machine Learning

MLD – Most Likely Depth

NOAA – National Oceanic and Atmospheric Administration

PDT – Probability Decision Threshold

SAD – Sounding Attribute Data

TN – True Negative Rate (*NotBathy* soundings correctly identified as *NotBathy*)

TP – True Positive Rate (*Bathy* soundings correctly identified as *Bathy*)

UTM – Universal Transverse Mercator

Funding

This work was supported by the National Oceanic and Atmospheric Administration (NOAA) Grant NA15NOS400020.

Bibliography

Agrifiotis, P., D., Skarlatos, A., Georgopoulos, and K., Karantzalos. 2019a. Depthlearn: learning to correct the refraction on point clouds derived from aerial imagery for accurate dense shallow water bathymetry based on SVMs-fusion with LiDAR point clouds. *Remote Sensing* 11: 1225, 31 pp. DOI: 10.3390/rs11192225.

Agrifiotis, P., D., Skarlatos, A., Georgopoulos, and K. Karantzalos. 2019b. Shallow water bathymetry mapping from UAV imagery based on machine learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W10, pp. 9–16. DOI: https://doi.org/10.5194/isprs-archives-XLII-2-W10-9-2019.

American Society for Photogrammetry and Remote Sensing. 2013. LAS Specification Version 1.3-R13 (15 July 2013).

Andersen, M., A. Gergely, Z. Al-Hamdani, F. Steinbacher, L. Larsen, and V., Ernstsen. 2017. Processing and performance of topobathymetric lidar data for geomorphometric and morphological classification in a high-energy tidal environment. *Hydrology and Earth System Sciences* 21: 43-63. DOI: 10.5194/hess-21-43-2017.

Birkeback, M., F. Eren, S. Pe'eri, N. Weston. 2018. The effect of surface waves on airborne lidar bathymetry (ALB) measurement uncertainties. *Remote Sensing* 10: 453, 19 pp. DOI: 10.3390/rs10030453.

Brzank, A., C. Heipke, J., Goepfert, and U. Soergel. 2008. Aspects of generating precise digital terrain models in the Wadden Sea from lidar-water classification and structure line extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 63: 510-528.

Calder, B., and L. Mayer. 2003. Automatic processing of high-rate, high-density multibeam echosounder data. *Geochemistry, Geophysics, Geosystems*, 4(6), 1048 (22 pp.). DOI: 10.1029/2002GC000486.

Calder, B., and G. Rice. 2017. Computationally efficient variable resolution depth estimation. *Computers & Geosciences* 106: 49-59. DOI: dx.doi.org/10.1016/j.cageeo.2017.05.013.

Calvert, J., J. String, M., Service, C., McGonigle, and R. Quinn. 2015. An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data, *ICES Journal of Marine Science* 72(5): 1498-1513. DOI: doi.org/10.1093/icesjms/fsu223.

Collin, A., P. Archambault, and B., Long. 2008. Mapping shallow water seabed habitat with the SHOALS. *IEEE Transactions on Geoscience and Remote Sensing* 46(10): 2947-2955. DOI: 10.1109/TGRS.2008.920020.

Dietrich, J., 2017. Bathymetric structure-from-motion: extracting shallow stream bathymetry from multi-view stereo photogrammetry. *Earth Surface Processes and Landforms* 42(2): 355-364.

Eren, F., S. Pe’eri, Y. Rzhanov, and L. Ward. 2018. Bottom characterization by using airborne lidar bathymetry (ALB) waveform feature obtained from bottom return residual analysis. *Remote Sensing of Environment* 206: 260-274. DOI: doi.org/10.1016/j.rse.2017.12.035.

Fernandez-Diaz, F., C. Glennie, W. Carter, R. Shrestha, M. Sartori, A. Singhania, C. Legleiter, and B. Overstreet. 2014. Early results of simultaneous terrain and shallow water bathymetry mapping using a single-wavelength airborne LiDAR sensor. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(2): 623-635. DOI: 10.1109/JSTARS.2013.2265255.

Friedman, J. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29(5): 1189–1232.

Gholamalifard, M., T. Kutser, A. Esmaili-Sari, A., Abkar, and B. Naimi. 2013. Remotely sensed empirical modelling of bathymetry in the Southeastern Caspian Sea. *Remote Sensing* 5(6): 2746-2762. DOI: 10.3390/rs5062746.

Heritage, G., and D. HetheringtonD. 2007. Towards a protocol for laser scanning in fluvial geomorphology. *Earth Surface Processes and* Landforms 32(1): 66-74.

Hickman, G., and J. Hogg. 1969. Application of an airborne pulsed laser for near shore bathymetric measurements. *Remote Sensing of Environment* 1: 47-58.

Höfle B., and M. Rutzinger. 2011. Topographic airborne LiDAR in geomorphology: a technological perspective. *Zeitschrift für Geomorphologie* 55(2): 1-29 (in English).

Jawak, S., S. Vadlamani, and A. Luis. 2015. A synoptic review on deriving bathymetry information using remote sensing technologies: models, methods, comparisons. *Advances in Remote Sensing* 4(2):147-162. DOI: 10.4236/ars.2015.42013.

Kashani, A., M. Olsen, C. Parrish, and N. Wilson. 2015. A review of LIDAR radiometric processing: from *Ad Hoc* intensity correction to rigorous radiometric calibration *Sensors* 15(11): 28099-28128. DOI: doi.org/10.3390/s151128099.

Kerr, J., and S. Purkis. 2018. An algorithm for optically deriving water depth from multispectral imagery in coral reef landscapes in the absence of ground-truth data. *Remote Sensing of Environment* 210: 307-324. DOI: doi.org/10.1016/j.rse.2018.03.024.

Kinzel, P., C. Legleiter, and P. Grams. 2021. Field evaluation of a compact, polarizing topo-bathymetric lidar across a range of river conditions. *River Research and Applications* 13 pp. DOI: doi.org/10.1002/rra.3771.

Kinzel, P., C. Legleiter, and J. Nelson. 2013. Mapping river bathymetry with a small footprint green LiDAR: applications and challenges. *Journal of the American Water Resources Association* (JAWRA) 49(1): 183-204. DOI: 10.1111/jawr.12008.

Kogut, T., and M. Weistock. 2019. Classifying airborne bathymetry data using the Random Forest algorithm. *Remote Sensing Letters* 10(9): 874-882. DOI: doi.org/10.1080/2150704x.2019.1629710.

Kutser, T., J. Hedley, C. Giardino, C., Roelfsema, and V. Brando. 2020. Remote sensing of shallow waters – a 50-year retrospective and future directions. *Remote Sensing of Environment* 240: 111619, 18 pp. DOI: doi.org/10.106/j.rse.2019.111619.

Lecours, V., M. Dolan, A., Micallef, and V. Lucieer. 2016. A review of marine geomorphometry, the quantitative study of the sea floor. *Hydrology and Earth System Sciences* 20: 3207-3244, DOI: https://doi.org/10.5194/hess-20-3207-2016.

Li, J., D. Knapp, S., Schill, C., Roelfsema, S., Phinn, M., Silman, J., Mascaro, and G. Asner. 2019. Adaptive bathymetry estimation for shallow coastal waters using Planet Dove satellite. *Remote Sensing of Environment* 232: 111302 (14 pp.). DOI: doi.org/10.1016/j.rse.2019.111302

Liu, S., Y. Gao, W. Zheng, and X. Li. 2015. Performance of two neural network models in bathymetry. *Remote Sensing Letters* 6(4): 321–330. DOI:10.1080/2150704X.2015.1034885.

Lowell, K., B. Calder, and A. Lyons. 2021. Measuring shallow-water bathymetric signal strength in lidar point attribute data using machine learning. *International Journal of Geographical Information Science* (in press). Published on-line. DOI:10.1080/13658816.2020.1867147.

Lyzenga, D., N. Malinas, and F. Tanis. 2006. Multispectral bathymetry using a simple physically based algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 44(8): 2251-2259.

Maas, H.-G., D. Mader, K., Richter, and P. Westfeld. 2019. Improvements in lidar bathymetry data analysis. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W10: 113-117–16. DOI: doi.org/10.5194/isprs-archives-XLII-2-W10-113-2019.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, Academic Press, (P. Zarembka, ed.), pp. 105-142.

McQueen, J. 1967. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press. pp. 281–297.

Mahalanobis, P. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India* 2(1): 49–55.

Mandlburger, G., Pfennigbauer, M., Schwarz, R., Flory, S., and L. Nussbaumer. 2020. Concept and performance evaluation of a novel UAV-borne topo-bathymetric LiDAR sensor. *Remote Sensing* 12(6):986. DOI: doi.org/10.3390/rs12060986.

Misra, A., Z. Vojinovic, B., Ramakrishnan, A., Luijendijk, and R. Ranasinghe. 2018. Shallow water bathymetry mapping using support vector machine (SVM) technique and multispectral imagery. *International Journal of Remote Sensing* 39(13): 4431-4450. DOI: doi.org/10.1080/01431161.2017.1421796.

Mitchell, S., and J. Thayer. 2014. Ranging through shallow semitransparent media with polarization lidar. *Journal of Atmospheric Ocean Technology* 31: 681-697. DOI: doi.org/10.1175/jtech-d-13-00014.1.

Nagle, D., and C. Wright. 2016. *Algorithms used in the Airborne Lidar Processing System (ALPS)*. United States Dept. of the Interior/ United States Geological Survey, Open File Report 2016-1046, 45 pp.

Niroumand-Jadidi, M., A. Vitti, and D. Lyzenga. 2018. Multiple optimal depth predictors analysis (MODPA) for river bathymetry: findings from spectroradiometry, simulations, and satellite imagery. *Remote Sensing of Environment* 218: 132-147. DOI: //doi.org/10.1016/j.rse.2018.09.022.

Okhrimenko, M., and C. Hopkinson. 2020. A simplified end-user approach to lidar very shallow water bathymetric correction. *IEEE Geoscience and Remote Sensing Letters* 17(1): 3-7.

Pacheco, A., J. Horta, C., Loureiro, and O. Ferreira. 2015. Retrieval of nearshore bathymetry from Landsat 8 images: a tool for coastal monitoring in shallow waters. *Remote Sensing of Environment* 159:102–116. DOI: 10.1016/j.rse.2014.12.004.

Pe’eri, S., and W. Philpot. 2007. Increasing the existence of very shallow LIDAR measurements using the red-channel waveforms. *IEEE Transactions on Geoscience and Remote Sensing* 45(5): 1217-1223.

Pittman, S., B. Costa, and T. Battista. 2009. Using lidar bathymetry and boosted regression trees to predict the diversity and abundance of fish and corals. *Journal of Coastal Research* 53: 27-38.

Schmidt, A., F. Rottensteiner, and U. Soergel. 2012. Classification of airborne laser scanning data in Wadden sea areas using conditional random fields. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXIX-B3: 161-166. DOI: https://doi.org/10.5194/isprs-archives-XLII-2-W10-9-2019.

Schwarz, R., G. Mandlburger, M. Pfennigbauer, and N. Pfeifer. 2019. Design and evaluation of a full-wave surface and bottom-detection algorithm for LiDAR bathymetry of very shallow waters. *ISPRS Journal of Photogrammetry and Remote Sensing* 150: 1-10. DOI: dx.doi.org/10.1016/j.isprsjprs.2019.02.002.

Steinhaus, H. 1957. Sur la division des corps matériels en parties. (English: “On the division of body materials in parts.”) Bulletin de l’Académie Polonaise des Sciences. (English: “Bulletin of the Polish Academy of Sciences.”) 4(12): 801–804. (In French.)

Su, D., F. Yang, Y. Ma, K. Zhang, J. Huang, and M. Wang. 2019. Classification of coral reefs in the South China Sea by combining airborne LiDAR bathymetry bottom waveforms and bathymetric features. *IEEE Transactions on Geoscience and Remote Sensing* 57(2): 815-828. DOI: 10.1109/TGRS.2018.2860931.

Tulldahl, H., and S. Wikström. 2012. Classification of aquatic macrovegetation and substrates with airborne lidar. *Remote Sensing of Environment* 121: 347-357. DOI: //doi.org/10.1016/j.rse.2012.02.004.

Wang, C., Q. Li, Y. Liu, G. Wu, P. Liu, and X. Ding. 2015. A comparison of waveform processing algorithms for single-wavelength LiDAR bathymetry. *ISPRS Journal of Photogrammetry and Remote Sensing*. 101**:**22-35. DOI: dx.doi.org/10.1016/j.isprsjprs.2014.11.005.

Wang, L., H. Liu, H. Su, and J. Wang. 2018. Bathymetry retrieval from optical images with spatially distributed support vector machines. *GIScience & Remote Sensing*. 56(3); 323-337. DOI: 10.1080/15481603.2018.1538620.

Westfeld, P., H. Maas, K. Richter, K., and R. Weiss. 2017. Analysis and correction of ocean wave pattern induced systematic coordinate errors in airborne LiDAR bathymetry. *ISPRS Journal of Photogrammetry and Remote Sensing* 128:314-325.

Xing, S., D. Wang, Q. Xu, Y. Lin, P. Li, L. Jiao, Z. Zhang, and C. Liu. 2019. A depth-adaptive waveform decomposition method for airborne LiDAR bathymetry. *Sensors* 19: 5065, 28 pp. DOI: 10.3390/s19235065.

Yang, A., Z. Wu, F. Yang, D. Su, Y. Ma, D. Zhao, and C. Qi. 2020. Filtering of airborne LiDAR bathymetry based on bidirectional cloth simulation. *ISPRS Journal of Photogrammetry and Remote Sensing* 163:49-61. DOI: doi.org/10.1016/j.isprsjprs.2020.03.004.

Zhao, J., X. Zhao, H. Zhang, and F. Zhou. 2017. Shallow water measurements using a single green laser corrected by building a near water surface penetration model. *Remote Sensing* 9**:** 426; 18 pp., DOI:10.3390/rs9050426.