Title Page

Measuring shallow-water bathymetric signal strength in lidar point attribute data using machine learning

Kim Lowell[1], Brian Calder, Anthony Lyons
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** – klowell@ccom.unh.edu

Biographical Notes

Kim Lowell is a Research Data Scientist and holds a PhD. In Forest Biometrics and MSc. In Data Analytics. He has considerable experience in the analysis of geospatial information and imagery to address land management issues while also accounting for uncertainties inherent in those data. His current focus the application of machine learning and deep learning to improve the accuracy of shallow-water bathymetric charts.

Brian Calder is a Research Professor at CCOM and holds a PhD. In Electrical and Electronic Engineering. He has worked on a number of signal processing problems, including real-time grain size analysis, seismic processing, and wave-field modeling for shallow seismic applications. His current research focu is on statistically robust automated data cleaning approaches and tracing uncertainty in hydrographic data.

Anthony Lyons is a Research Professor who holds a PhD in Oceanography. He conducts research in the field of underwater acoustics and acoustical oceanography. His current areas of interest include high-frequency acoustic propagation and scattering in the ocean environment, acoustic characterization of the seafloor, and quantitative studies using synthetic aperture sonar.

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

Disclosure: The authors have no potential competing nor financial interest in the work presented.

Data and codes availability statement: The codes that support the findings of this study are available at the link https://figshare.com/s/8c563c3032927af0b25d. 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 to demonstrate how codes function. 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/htdata/lidar2_z/geoid12b/data/6246 as compressed .laz files. These can be decompressed using the LASzip tool which can be downloaded from laszip.org.

Total word count: 7870 **Word count excluding title page and references**: 6297

Measuring shallow-water bathymetric signal strength in lidar point attribute data using machine learning

Abstract

The goal of this work was to evaluate if routinely collected but seldom used airborne lidar metadata – "point attribute data" (PAD) – analyzed using machine learning/artificial intelligence can improve extraction of shallow-water (less than 20 m) bathymetry from lidar point clouds. Extreme gradient boosting (XGB) models relating PAD to an existing bathymetry/not bathymetry classification were fitted and evaluated for four areas near the Florida Keys. The PAD examined include "pulse specific" information such as the return intensity and PAD describing flight path consistency. The R^2 values for the XGB models were between 0.34 and 0.74. Global classification accuracies were above 80% although this reflected a sometimes extreme *Bathy/NotBathy* imbalance that inflated global accuracy. This imbalance was mitigated by employing a probability decision threshold (PDT) that equalizes the true positive (*Bathy)* and true negative (NotBathy) rates. It was concluded that 1)the strength of the bathymetric signal in the PAD should be sufficient to increase accuracy of density-based lidar point cloud bathymetry extraction methods and 2)ML can successfully model the relationship between the PAD and the *Bathy/NotBathy* classification. A method is also presented to examine the spatial and feature-space distribution of errors that will facilitate quality assurance and continuous improvement.

Keywords : lidar metadata, hydrography, bathymetry, Florida Keys

Introduction

By at least 2005, lidar ("light detection and ranging") had been identified as a "proven technology" for a variety of marine information needs (Wozencraft and Millar 2005) including the production of nautical depth charts. Compared to more conventional acoustic sonar data, lidar data typically comprise noisier point clouds (Chust *et al.* 2008) and a major processing challenge is distinguishing between pulse returns that are unwanted noise and those that truly describe ocean depth. Nonetheless lidar waveform data -- the energy returned from each lidar pulse recorded in nanosecond intervals -- . have been shown to be able to distinguish among fine sand, coarse sand, and rocks on ocean floors in shallow waters (Eren *et al*. 2019). Waveform lidar data have also been used to estimate vertical uncertainty (Parrish *et al*. 2014), describe coastal vegetation (Rogers *et al*. 2015; Tulldahl and Wikström 2012), and characterize marine seafloor habitats (Collin *et al*. 2011; Tulldahl 2014). It has also been shown that depth and benthic habitat can be mapped by combining hyperspectral data and lidar waveform data (Tuell *et al*. 2005).

Despite these demonstrated uses of lidar waveform data, the large data volumes make them difficult to use operationally for the production of nautical charts. Hence it is of interest to be able to alternatively employ what Tulldahl (2014) termed "lidar point data" – sometimes also referred to as "discrete lidar returns" – that do not

record the entire temporal waveform of each lidar pulse return. One approach to mapping ocean depth that has proved reasonably successful for lidar point data and (more conventional multi-beam echo sounder (MBES) (sonar) point data) is based on the use of localised pulse return density. One example described by Nayegandhi (2009) and Nagle and Wright (2016) is underpinned by a random consensus filter (Fischler and Bolles 1981). Another that is broadly representative of density-based approaches and that also has been implemented operationally is the CUBE (Combined Uncertainty and Bathymetry Estimator) algorithm. First described in 2003 (Calder and Mayer 2003), it proved successful for producing nautical charts from MBES data while decreasing manual editing effort and maintaining computational efficiency. CUBE has since been enhanced using variable data density information (Calder and Rice 2017); the updated algorithm is known as CHRT (CUBE with Hierarchical Resolution Techniques).

CHRT establishes a variable resolution grid of points across an area being mapped. For each grid point, "neighboring" pulse returns are identified. The frequency distribution of the depths of these neighbors – which is generally multi-modal and resembles an individual-pulse waveform -- is analyzed to identify individual modes or "depth hypotheses" using a process comparable to unsupervised clustering in one-dimensional space. An algorithmic decision is then made about which hypothesis is the most likely depth of the grid point using relatively simplistic "disambiguation rules" that rely largely on local context.

Density-based approaches employ only the depth estimated for each pulse return. Ignored are pulse return metadata such as the intensity of each return and the incidence angle of the return-generating pulse; we term these "point attribute data" (PAD). If the PAD collectively contain information about which pulse returns ("soundings" in marine parlance) represent bathymetry, the PAD may enable improved bathymetric mapping. Evidence of the successful use of discrete lidar return PAD exists for other applications such as building detection, for example (Matikainen *et al*. 2009). Similarly, "multi-return" PAD such as rugosity – essentially local variability of return intensity -- have been used for landcover classification and object detection in a ML approach (Zhao *et al*. 2018). Furthermore sounding PAD may be useful as surrogates for factors such as ocean bottom characteristics that are difficult to measure directly over large areas but that are known to impact return reflectance (see for example Chust *et al*. (2008)).

Hence one objective of this work is to evaluate this premise for shallow-water bathymetry. The work is necessarily constrained to shallow water – defined as depths less than 20 m – because of the limit of lidar penetration in water. We note that though shallow waters comprise a relatively small percentage of the ocean, because of human use and technical limitations, Salameh *et al.* (2019) recognized shallow water as a bathymetric priority for ocean mapping.

That PAD contain a strong bathymetric signal independent of the depth recorded for each sounding is a necessary but not sufficient condition for its operational use in improving nautical charts using density-based approaches like CHRT. It must also be possible to model the bathymetric signal despite the nature of the relationships of individual PAD variables and their interaction(s) to bathymetry being unknown. In the introduction article to a marine-themed issue of the ICES Journal of Marine Science, Beyan and Browman (2020) noted the potential of machine learning and artificial intelligence (ML/AI) to address issues of unknown relationships and interactions. In the same issue Malde *et* al. (2020) discussed a number of ML/AI techniques and their potential for increasing efficiency in processing expanding volumes of marine data.

ML/AI techniques have been applied to idar data for a variety of purposes. Examples include:

- Convolutional neural networks for landcover/vegetation) and objects (Zhao *et al*. 2018), and buildings (Nahhas *et al*. 2018).
- Boosted trees and random forests for built infrastructure such as powerlines and buildings (Kim and Sohn 2011; Guo *et al*. 2015).
- AdaBoost tree-based classifiers for landscape objects such as buildings, trees, and cars (Zhang *et al*. 2016).
- Bayesian networks for land cover and objects (Kang *et al*. 2016).
- Neural networks and boosted regression trees for marine corals (Collin *et al.* 2018; Pittman *et al*. 2009) and benthic habitat (Narayanan and Sohn 2009).
- Support vector machines (SVM) for urban classification (Mallet *et al*. 2011).

ML/AI has also been explored for bathymetric mapping. For example, Collin *et al*. (2008) used principal component analysis (PCA) and cluster analysis to extract a shallow water digital depth model (DDM) from waveform lidar data. To produce DDMs, ML/AI has also been applied to optical imagery alone (Mohamed *et al*. 2016), a combination of optical imagery and sonar (Vojinovic *et al*. 2013), and a combination of optical imagery and lidar (Macon *et al*. 2008). The use of ML/AI to extract bathymetry from the PAD of lidar soundings, however, does not appear to have been explored.

Hence the second objective of this article is to evaluate if ML/AI techniques provide a means to detect and model the bathymetric signal in lidar PAD such that soundings representing shallow-water bathymetry can be better identified and providing a means to improve shallow-water bathymetric mapping.

Study Area and Data

Airborne lidar data collected between April 22 and April 25, 2016 were available for the area surrounding Key West, Florida, United States – approximately 24°35' N latitude and 81°45' W longitude. Data were acquired by the National Oceanic and Atmospheric Administration (NOAA) Remote Sensing Division using a RIEGL™ VQ-880-G lidar flown at a nominal altitude of 400 m above mean sea level at a speed of 190 km per hour. During operation, near infrared (1064 nm) and green (532 nm) pulses are generated using a counterclockwise circular scan. The maximum effective measurement rate at 550 kHz is 177,200 measurements per second for the infrared and 550,000 measurements per second for green. (NOAA operates at 145 kHz.) The scan angle of 20^0 results in a swath approximately 300m wide. Data were acquired continuously along pre-determined overlapping flightpaths, and then post-processed into 500 m-by-500 m square tiles. Tiles were available as "LAS" files in accordance with Version 1.4-R13 (Point Data Record Format 6) of the LAS data standard (ASPRS 2013).

Provided with the LAS files was the NOAA *Bathy/NotBathy* classification of each return/sounding. This is produced using a combination of semi-automated procedures and human intervention. In these processes, the infrared data are used to eliminate above-water soundings and to produce a water surface model that determines to which green pulses refraction corrections are applied. For this study, this NOAA *Bathy/NotBathy* classification was considered to be "ground truth" although in bathymetric mapping it is in reality a "best estimate" of ocean depth.

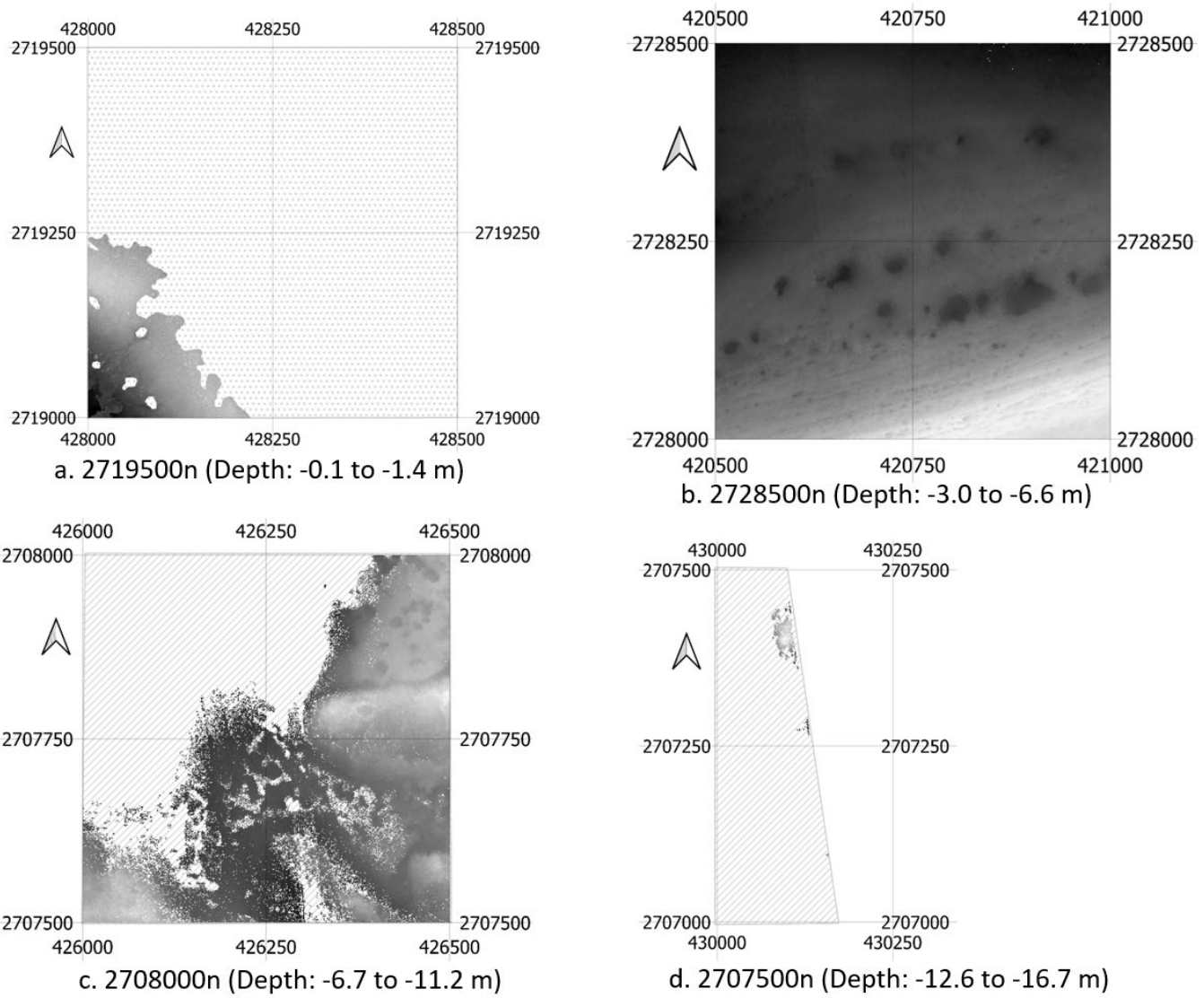
Four data tiles located in UTM Zone 17N (NAD83) were employed (Table 1; Figure 1). The NOAA naming convention for these files includes the UTM coordinates of the northeast corner of each tile; for brevity in this paper the northing for each tile is used as its identifier. These tiles were selected because they represent a range of physical and data conditions and a variety of real-world issues – e.g., a range of depths and ocean bottom characteristics (Table 1), potentially weak and strong bathymetric signals (e.g., 2707500n in which only 0.4% of returns are bathymetry versus 2728500n in which 76% of points are bathymetry), a range of sounding densities and total number of soundings. Tile 2707500n covers the deepest area, is almost entirely at the depth limit of airborne lidar in the area studied (approximately 20 m), and was only partially covered due to the location of available flight paths. Tile 2719500n is largely covered by land – a characteristic that NOAA addresses by creating a polygonal mask to eliminate land areas; for this reason, in this study only the unmasked area was analyzed.

Table 1. Description of tiles employed in this study from shallowest to deepest. (See Figure 1 for a spatial representation of each tile.)

Identifier (Northing)	Relative Depth	Area with data (m^2)	Area with Bathy-metry (m^2)	Approx. MSL depth range (m)	Total Returns area with data (million)	Return density (pts/m^2)	% Bathy-metry	Flight-paths
2719500n	Shallow	29,500	29,500	-0.1 to -1.4	0.6	20.	78	4
2728500n	Deep	250,000	250,000	-3.0 to -6.6	7.6	30.	76	7

2708000n	Deeper	250,000	131,600	-6.7 to -11.2	3.7	15.	21	7
2707500n	Deepest	65,000	1,200	-12.6 to -16.7	1.0	13.	0.4	2

158



159

160 Figure 1. Depth maps (1m pixels) for the four tiles based on depth determined by NOAA. Striped areas contain no
161 soundings identified as bathymetry by NOAA. Dotted areas are above sea level and have been removed from
162 analysis. Lighter colored areas are shallowest; dark areas are deepest. Coordinates are UTM Zone 17N (NAD 83).

163 Three types of sounding PAD were provided by NOAA and/or derived by the authors (Table 2). "Return-based"
164 PAD are associated with each pulse whereas "SBET" (Smoothed Best Estimate of Trajectory) and "lidar-edge" PAD
165 are more macro and reflect the flightpath and aircraft characteristics when data were collected.

166 "Return-based" PAD describe the green pulse returns and were largely provided by NOAA although *rel_return_no*
167 that normalizes *return_no* between 0 and 1 was derived. Determining the actual scan angle (*abs_scanangl*)
168 required correcting the sensor scan angle of 20^0. for aircraft pitch and roll.

169

170 Table 2. Lidar sounding point attribute data (PAD) employed in this study.

Meta-data Type	Variable Name (*Abbreviation*)	Description	Range
Return-based	Absolute value of scan angle (*abs_scnangl*)	Absolute value of signed angle (as recorded negative is the left side of airplane; positive is the right side) at which a pulse was generated including the roll of the aircraft.	16 - 23
	Intensity (*intensity*)	Integer value representing the magnitude of the pulse return.	7302 -- 37448
	First of many (*first_of_many*)	Flag indicating if a return was the first of a pulse that generated multiple returns.	0 (No) – 1 (Yes)
	Last (*last*)	Flag indicating if a return was the final return from a pulse.	0 (No) – 1 (Yes)
	Number of returns (*numreturns*)	Number of returns from a single pulse	1 - 6
	Relative Return Number (*rel_return_num*)	(Return Number-1)/(Number of Returns-1)	0.0 – 1.0 (0.0 for a *last* return)
	Return number (*return_no*)	Return number	1 - 6
	Scan direction (*scan_direct*)	Flag indicating if pulses were generated in the forward direction of the flightpath	0 (No) – 1 (Yes)
	Single (*single*)	Flag indicating if a pulse generated a single return.)	0 (No) – 1 (Yes)
	Time (*gpstime*)	GPS-standardized time in seconds when a pulse was captured.	NA (Used for data preparation but not analysis)
SBET[1]	Aircraft positional uncertainty (*stdXYZ*)	Simple sum of individual standard deviations for each positional attribute (*x, y,* and *z*) from proprietary Smoothed Best Estimate of Trajectory (SBET) process.	0.022 – 0.030
	Aircraft stability uncertainty (*stdYwPtRl*)	Simple sum of individual standard deviations for each aircraft-related attribute – Yaw, Pitch, and Roll -- from SBET process.	0.027 – 0.031
Lidar-edge	Deviation absolute value (*absdevia*)	Absolute value of orthogonal distance from lidar point cloud "corner-to-corner" flight path[2].	0 – 7.7
	Maximum abs. deviation (*maxabsdev*)	Maximum *absdevia* over a given flight path.	0.6 – 8.3

171 [1]Smoothed Best Estimate of Trajectory. [2]See text for explanation.

172

Notably absent from the return-based PAD is *z* (depth). Though clearly related to a sounding's probability of being bathymetry – *p(Bathy)* – reader's are reminded that one of the main goals of this work was to evaluate the strength of the bathymetric signal in the PAD only. Therefore the *z/depth* value was intentionally not included in the PAD used to develop ML/AI models.

Additional return-based PAD not considered are instrument-dependent PAD such as amplitude and reflectance collected by the RIEGL sensor employed (RIEGL 2017). These are embedded in the "extrabytes" of the RIEGL data structure, that is not part of the 1.4-R13 LAS data standard. The use of such proprietary extrabytes would limit the broader applicability of this work. Moreover, extrabyte data are not used in NOAA's lidar processing workflows and are routinely eliminated from the LAS files by NOAA in large part due to data volumes

SBET PAD are produced by the Applanix Corporation POSPac Mobile Mapping Suite (MMS) using POS AV -- an integrated hardware and software system. It includes a motion sensor that is generally a GNSS[1]-aided INS[2] with PPK[3] corrections from a local base station. The SBET PAD are based on a tightly coupled extended Kalman filter and its post-observation covariance matrix. SBET PAD are provided for each flight path at 0.005 second intervals (i.e., 200 observations per second) and describe the location and uncertainty of the aircraft position in three dimensions as well as the aircraft's roll, pitch, and yaw. SBET uncertainty PAD were employed because they may indicate -- among other factors such as surge, sway, acceleration, etc. -- wind-related turbulence that might in turn affect sea surface characteristics that impact pulse reflectance.

It was similarly surmised that deviation(s) of a flightpath from a straight line might be indicative of temporally local weather effects that impact pulse reflectance. Figures 2a and 2b show two-dimensional examples of lidar point clouds with the SBET flight line (thick black line) from which the lidar data were collected. The deviation of the actual flight from a straight line such as the SBET flight line is apparent in the edge of the gray lidar point cloud. To quantify the crenulation of the actual flight path, a northeastern and southeastern "corner" of a flightpath's point cloud was identified and the equation of this "corner-to-corner" line calculated (the thin gray line in Figures 2c and 2d). Edge points of the lidar point cloud were identified and their orthogonal deviation from the corner-to-corner line calculated. These were summarized as described for the "Lidar edge" PAD in Table 2.

[1] Global Navigation Satellite System.
[2] Inertial Navigation System.
[3] Post-processing Kinematic.

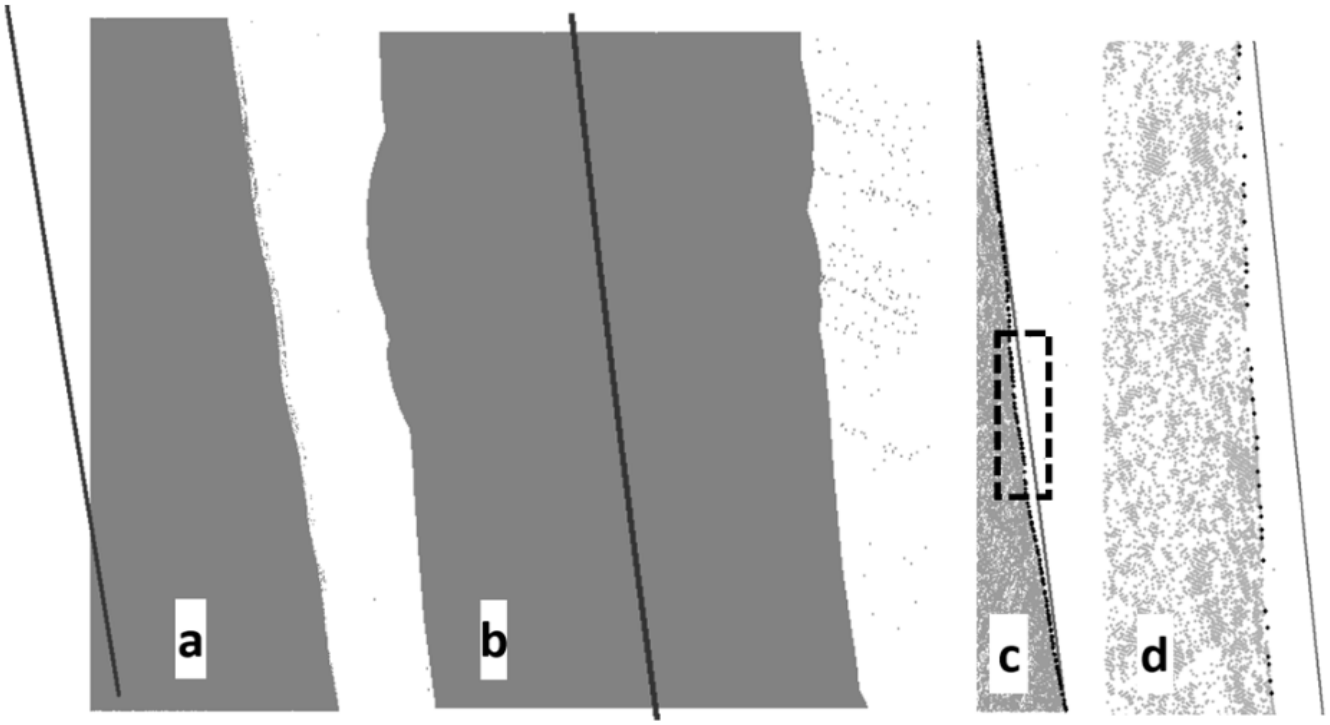


Figure 2. Gray points are lidar pulse returns. a. Flightpath with a relatively smooth easternscan edge. (The solid black line indicates the SBET aircraft location.) b. Flightpath with more crenulated (western and eastern scan edge. c. Scan edge and "corner-to-corner" straight flight path (thin gray line) with edge points (black); the dotted line indicates the area of enlargement in Figure 2d. d. Enlargement of a portion of Figure 2c.

The SBET variables *stdXYZ* and *stdYwPtRl,* and the lidar-edge variable *absdevia* were affixed to individual returns based on closest *gpstime*. For the lidar-edge flight path-based variable *maxabsdev* a single value was assigned to all returns on the relevant flight path.

Methods

The over-arching goals of this work are to assess the strength of the bathymetric signal in the PAD and to determine if ML is able to model this relationship to improve identification of discrete lidar returns that represent bathymetry. Achieving this translates into an operational goal of being able to assign to each lidar sounding a PAD-based probability of being bathymetry -- *p(Bathy)* -- that can improve the disambiguation rules used in density-based lidar point processing. The methodological outline to achieve these goals is to fit a ML/AI model for each tile to NOAA's *Bathy/NotBathy* classification using the PAD, estimate *p(Bathy)* for each sounding using the fitted model, classify each sounding as *Bathy* or *NotBathy*, and compare the ML/AI classification with NOAA's. High classification accuracy would indicate a strong relationship of the PAD and bathymetry and evidence that ML is suitable for such modelling; low classification accuracy suggests the opposite.

Exploratory analysis was conducted to compare the merits of three different ML/AI techniques – regularized logistic regression (RLR), multi-layer perceptrons (MLPs; a type of neural network), and extreme gradient boosting (XGB). RLR was rejected because model fitting times were up to seven times longer compared to MLP and XGB. MLP was also rejected because of the difficulty of evaluating the contribution of individual PAD variables to each model – a well-recognized problem with ML/AI techniques. Hence XGB was the ML/AI technique retained for use.XGB is a decision-tree-based approach that progressively develops a sequence of "weak" or "shallow" models/trees and then combines them into a single model/tree using a weighted majority vote. Successively fit trees focus on improving the classification of the observations having the biggest errors. Information about the "importance" – but not statistical significance -- of each variable is produced with the "importance" of all variables summing to 1.0. XGB models were fitted using lasso regularization ("l1" regularization) that applies a numerical penalty to the model fitting cost function; this drives the contribution of "unimportant" variables to zero. Individual XGB models were fitted for each tile using NOAA's *Bathy*/*NotBathy* classification as the dependent variable and all of the PAD variables described in Table 2 and selected interactions as predictors. Prior to model fitting, all independent variables were normalised between 0 and 1 using max-min normalisation.

Models were evaluated using a variety of metrics. McFadden's adjusted (pseudo) R^2 (McFadden 1974) provides an assessment of goodness-of-fit for binary classification models. The other metrics employed are based on two different confusion matrices.

The first is a "conventional" confusion matrix that assigns observations to *Bathy* or *NotBathy* based on a probability decision threshold (PDT) of 0.50. However, an undesirable latent characteristic of ML/AI model fitting is that the accuracy of the most prevalent class –is increased at the expense of the accuracy of the minorityclass. Alternately stated, deemphasizing the true positive rate (TPR) of the minority class better achieves the model fitting goal of maximizing global accuracy. . While not a problem if all classes are equally present, class imbalance was present in all tiles used in this study (see Table 1). The impact on the goals of this study is that the true bathymetric signal in the PAD may be obscured.

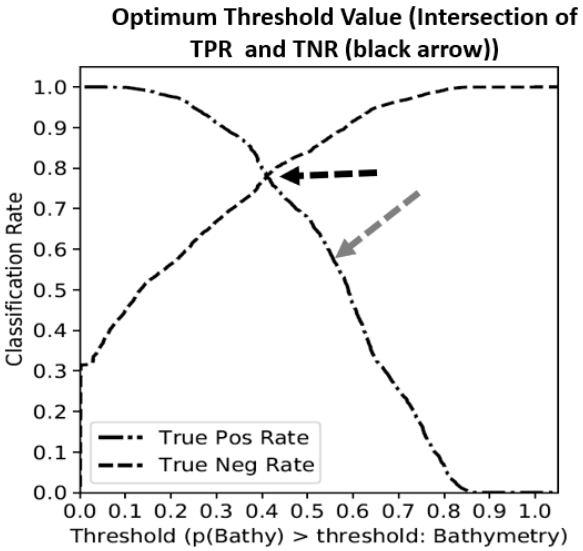
Lowell and Calder (2019) addressed this using an alternate PDT that equalizes the true negative rate (TNR) and TPR. Figure 3 provides an example. In the upper left is a confusion matrix resulting from a conventional PDT of 0.50. Its receiver operating characteristics (ROC) curve that indicates the trade-off between the TPR and false positive rate (FPR) at different PDTs is shown in the lower right. The gray arrow indicates the FPR (x axis) and TPR (y axis) produced by a PDT of 0.50. The threshold at which TPR and TNR are equal -- 0.41 in this example – is the intersection of these curves as shown in the graph in the upper right of Figure 4; this threshold is termed the

"optimal decision threshold" (ODT). The resulting confusion matrix is presented in the lower left and shows how the ODT equalizes TNR and TPR; it also makes false negative rate (FNR) and FPR equal.

Most important for evaluating the strength of the bathymetric signal is that the TPR – correct identification of bathymetry – has increased from 37% to 79%. And the FNR – undetected bathymetry that is the most important nautical navigation -- has decreased from 63% to 21%. In this example and relative to the goals of this study, comparison of the results for the PDT and the ODT suggests there is a reasonably strong bathymetric signal in the sounding PAD.

Confusion Matrix: Threshold = 0.50

		Truth		
		NotBath	***Bath***	**Total**
Predicted	***NotBath***	2,751,500	497,500	**3,249,000**
	Bath	169,500	287,500	**457,000**
	Total	**2,921,000**	**785,000**	**3,706,000**
Global Accuracy (%): 82				
Individual Accuracies		**True Positive or Negative Rate(%)**		**False Positive or Negative Rate(%)**
NotBath		**94 (TNR)**		**63 (FNR)**
Bath		**37 (TPR)**		**6 (FPR)**



Confusion Matrix: Threshold = 0.41

		Truth		
		NotBath	***Bath***	**Total**
Predicted	***NotBath***	2,307,000	165,000	**2,472,000**
	Bath	614,000	620,000	**1,234,000**
	Total	**2,921,000**	**785,000**	**3,706,000**
Global Accuracy (%): 79				
Other Accuracies (%)		**True Negative or Positive Rate(%)**		**False Negative or Positive Rate(%)**
Not Bathymetry		**79 (TNR)**		**21 (FNR)**
Bathymetry		**79 (TPR)**		**21 (FPR)**

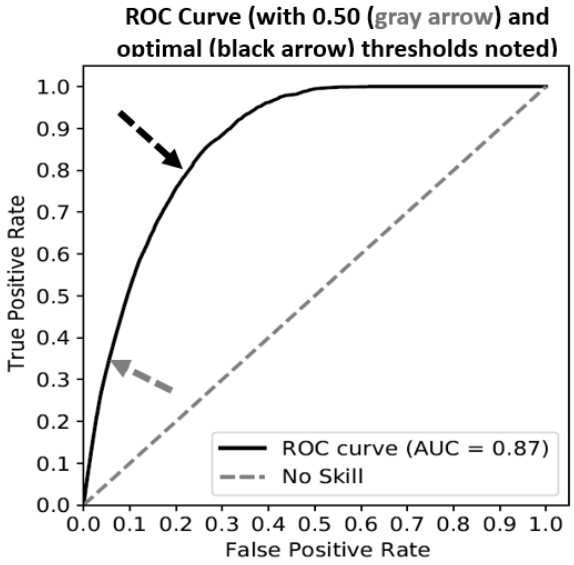


Figure 3. Example of the conventional (gray arrows) and optimal (black arrows) probability decision thresholds(PDTs) for tile 2708000n.

To enable evaluation of which PAD most relate to bathymetry, a rank-based variable scoring system was developed. For each model, the "contributing" variables – defined as those whose importance was greater than 0.0 – were ordered by decreasing importance. The variable in each position was assigned a score for that position:

$$\text{Position score} = (n+1)-r$$

where n is the total number of contributing variables and r is the rank position. If a position was occupied by a two-way interaction, each variable in the position received half of the score for that position. Scores were summed for each variable over all positions and the variables ranked in importance based on their total score.

As a final step, a method to examine the distribution of the classification errors in geographic and feature space was developed. Such analysis can provide more insight into where the relationship between the PAD and bathymetry is stronger/weaker, and may further be of interest for operational quality assurance and continuous improvement. Several spatial statistics and methods exist to test for complete spatial randomness in point patterns: nearest neighbor analysis (Boots and Getis 1988), quadrat analysis (Boots and Getis 1988), Ripley's *K* (Dixon 2002), and the join-count statistic for multi-nominal categorical data (Odland 1988). Herein, however, it is not expected that classification errors will be randomly distributed, but instead that areas having "high" sounding densities (due to flightpath overlap, for example) will have equally/proportionately high error rates.

To examine the expected versus actual spatial and feature-space distributions of classification errors, each tile was sub-divided into 20m (0.04 ha) pixels. This size was determined to be a useful trade-off between having "too few" or "too many" pulse returns in each pixel. The number of returns that were *Bathy*, *NotBathy*, TPs, FNs, TNs, and FPs were tabulated for each pixel and the percentage of the total number of returns of each type within each cell calculated. A pixel-based map was produced of the difference between percent of total *Bathy* minus percent of total FNs (i.e., negative numbers indicate an "excess" of FNs; positive numbers indicate fewer FNs than expected); the same was done for the percent of total *NotBathy* minus the percent of total FPs. This was explored visually.

To examine classification errors in feature space, a regression line was also fitted to the individual pixel values for each of these comparisons. Evaluation of this line relative to the characteristics of a 1:1 line – i.e., the percent of returns equals the percent of errors – provides a means to further evaluate classification error characteristics.

Results and discussion

Global information about model goodness-of-fit and accuracy is presented in Figure 4. (Pseudo) R^2 values suggest that the bathymetric signal in the PAD is reasonably high[4]. This is reinforced by relatively high global accuracies for all tiles, although as noted that these are achieved for the PDT of 0.50 through an implicit majority-class bias. Global accuracy decreases slightly when an ODT is applied, but remain around 80% or higher. This global classification accuracy implies a strong relationship between PAD variables and bathymetry, although global classification accuracy is not the sole consideration in this work.

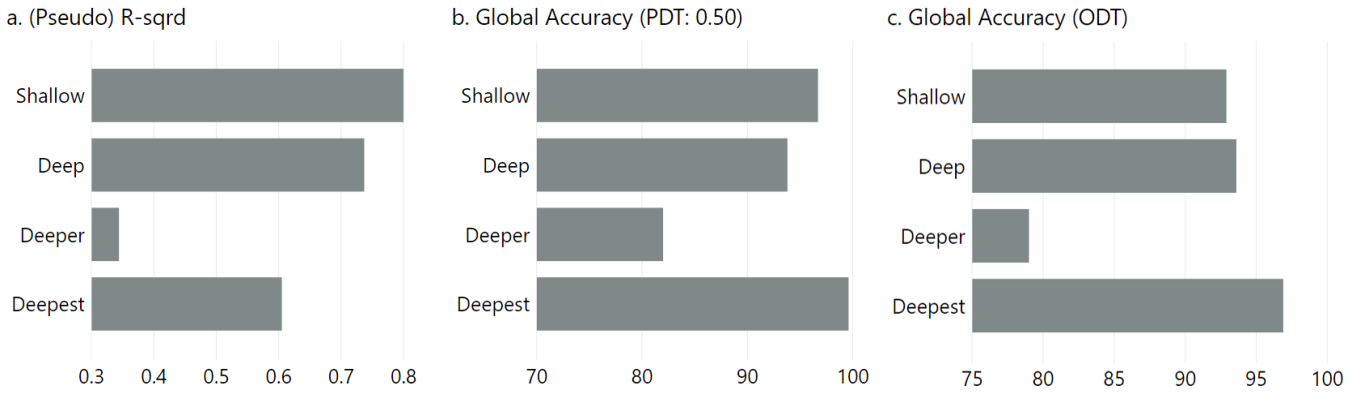


Figure 4. Goodness-of-fit information. a. (Pseudo) R^2 values from model fitting. b. Global accuracy using a PDT (probability decision threshold) of 0.50. c. Global accuracy using the optimal decision threshold (that equalizes the TPR and TNR).

The number of "contributing variables" in each model was at least 13 or 59% of the total available (Table 3). This suggests that the bathymetric signal is not concentrated in a few key PAD variables but is distributed across a large number of variables. This supports the premise that the ML technique XGB is able to describe the relationship between PAD variables and bathymetry.

Table 3. Description of variables included in the XGB models for each tile.

Identifier (Northing)	Variable Set (Total Number of Variables)	Number of Contributing Variables[1]	Five Most Important Contributing Variables[2]	Optimal Threshold
Shallow (2719500n)	Main Effects & selected interactions (22)	13	*abs_scanangl*, *stdYPR*, *last*, *intensity*, *absdevia*	0.87
Deep (2728500n)		17	*abs_scanangl*, *numreturns*, *intensity*, *stdYPR*, *rela_rtrn_no*	0.56
Deeper (2708000n)		16	*abs_scanangl*, *stdYPR*, *intensity*, *absdevia*, stdXYZ	0.40
Deepest (2707500n)		17	*abs_scanangl*, *stdYPR*, *return_no*, *rela_rtrn_no*, *intensity*	0.02

[4] McFadden's pseudo R^2 cannot be tested for statistical significance. This statement is based on the authors' experience with conventional R^2 values with large sample sizes – i.e., greater than 500,000.

[1]See text for explanation of "contributing variables."
[2]Variables are presented in order of decreasing contribution.

The five contributing variables with the highest scores are presented in descending order of importance in Table 3. There are commonalities among models. The return-specific variable *abs_scanangl* was consistently the most important for all tiles. Variables that quantify platform stability appeared important: at least one SBET variable (*StdYPR* and *StdXYZ*) was present in all models and the edge-based variable *absdevia* appeared in two models. *Intensity* was similarly important even for the "Shallow" tile (2719500n) for which it might have been expected that the *intensity* of returns would be less variable than for other tiles. Finally, the return number was generally important with at least one variable that describes return order – *last, return_no,* and/or *rela_rtrn_*num – being among the five most important variables except for the "Deeper" tile (270800n).

Figure 5 presents the accuracy and error rates for *Bathy* (considered "positive") and *NotBathy* ("negative") from the XGB models based on a conventional 0.50 PDT. The "Deepest" tile (2707500n) which is the one that has the greatest *Bathy/NotBathy* sample imbalance (Table 1) demonstrates most clearly the difficulty of using a conventional PDT of 0.50 to assess bathymetric signal strength. Despite having a global accuracy near 100% (Figure 4) for a PDT of 0.50, its TPR (percent of correctly classified *Bathy* returns) is extremely low (approximately 10%) while its TNR (percent of correctly classified *NotBathy* returns) is near 100%.

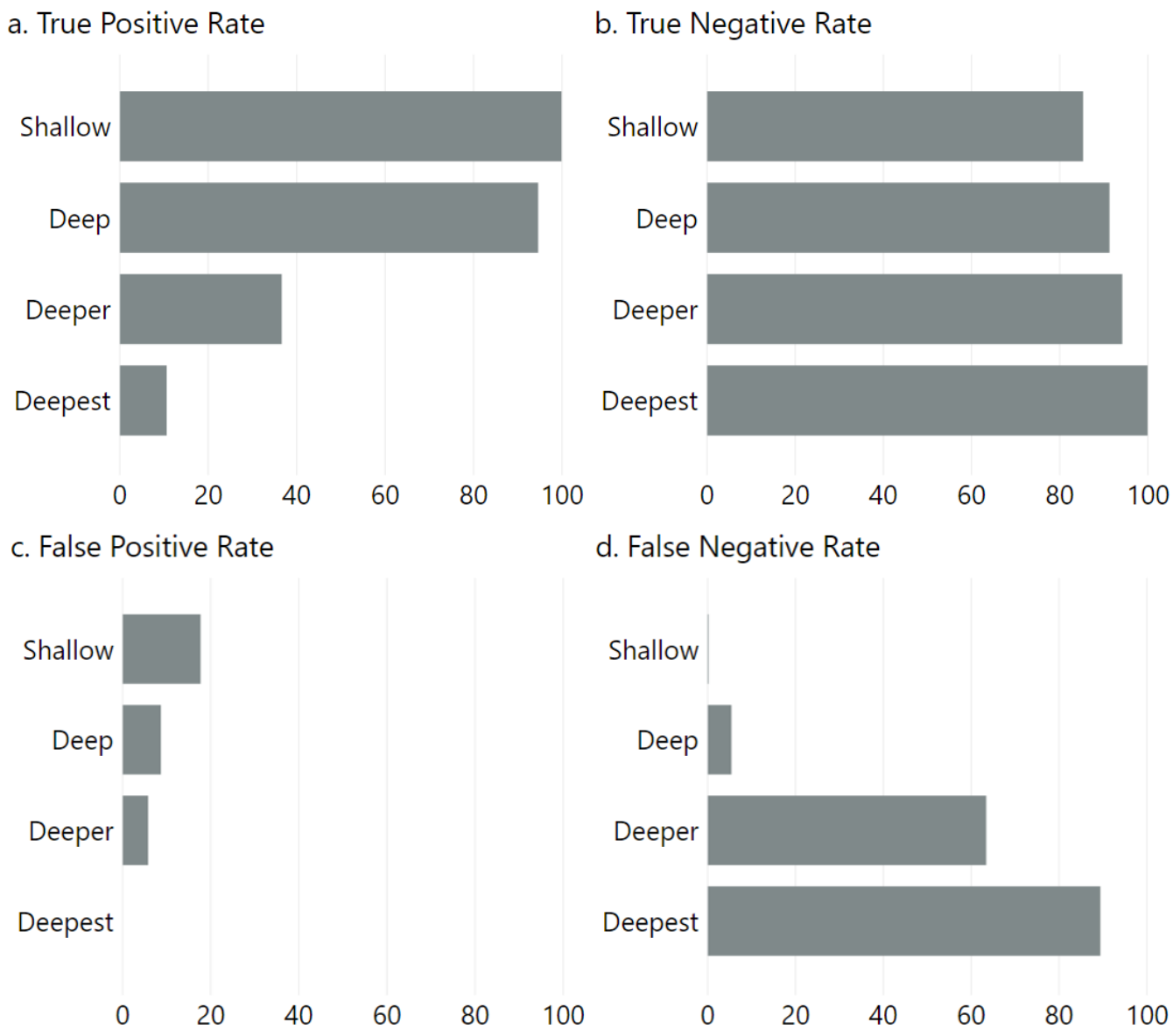


Figure 5. Accuracy and error rates resulting from the use of a probability decision threshold of 0.50.

Correcting for class imbalance via the application of the ODT (Figure 6) provides a better indication of the strength of the relationship between the PAD and *Bathy* and *NotBathy* individually. Accuracy of *Bathy* identification (TPR) is at least 79% for all tiles compared to, for example, only 10% for the most imbalanced Deepest tile (2707500n) when applying the PDT of 0.50; accuracy of *NotBathy* detection (TNR) is identically high since the ODT is selected to equalize the two. Inaccurate detection of *Bathy* and *NotBathy* are inversely low.

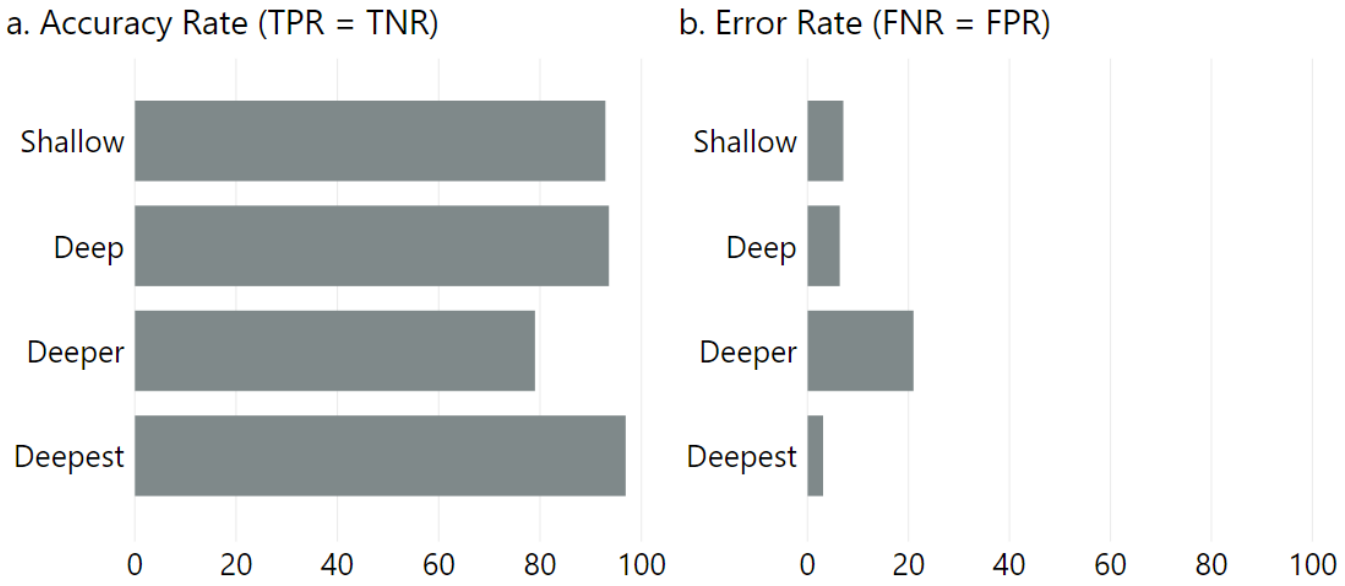


Figure 6. Accuracy and error rates resulting from the use of the optimal decision threshold for each tile.

Results presented thus far suggest that 1)the lidar PAD have a strong relationship to bathymetry and 2)the ML technique XGB is able to model the relationship for individual tiles. Of operational interest for quality assurance and continuous improvement is an examination of model performance in geographic and feature space as described in the Methods section. For brevity, only results for the Deeper/2708000n tile are presented (Figures 7 and 8). The reader is reminded that that negative numbers indicate a relatively high number of errors. The reader is also referred to Figure 1c that indicates the depth and usable data coverage of this tile.

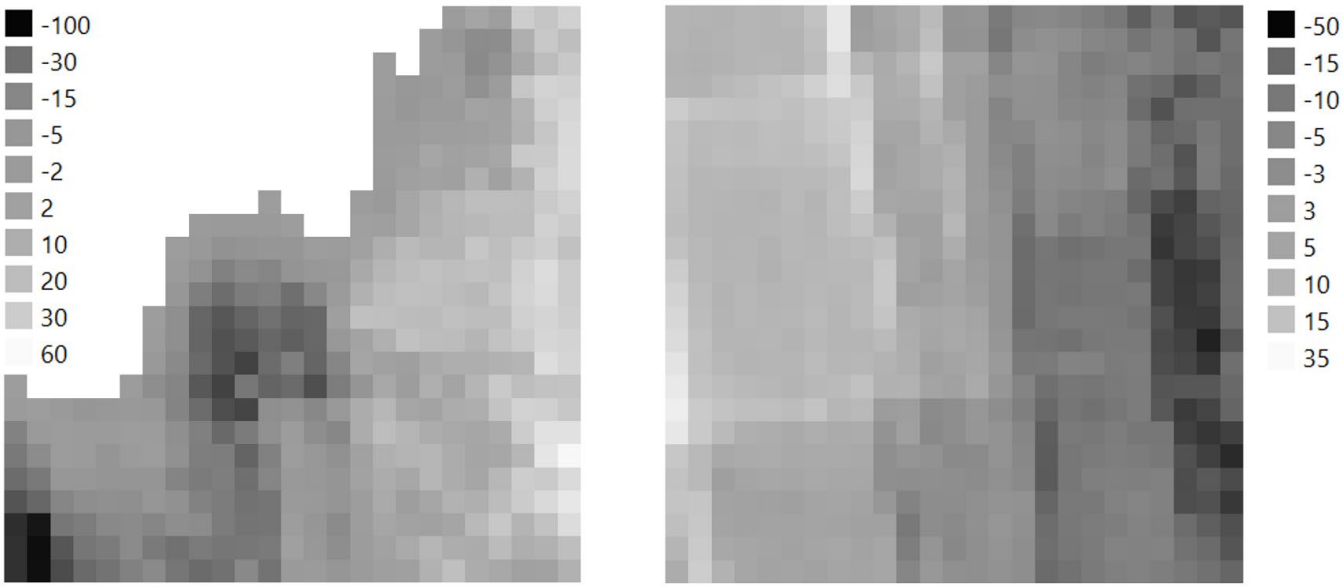


Figure 7. Spatial distribution of classification errors for the "Deeper" tile (2708000n). Legend units are percent multiplied by 100. a (left). % of *Bathy* soundings minus % of false negatives. b (right). % of *NotBathy* soundings returns minus % of false positives. (Negative values indicate "too many" errors.)

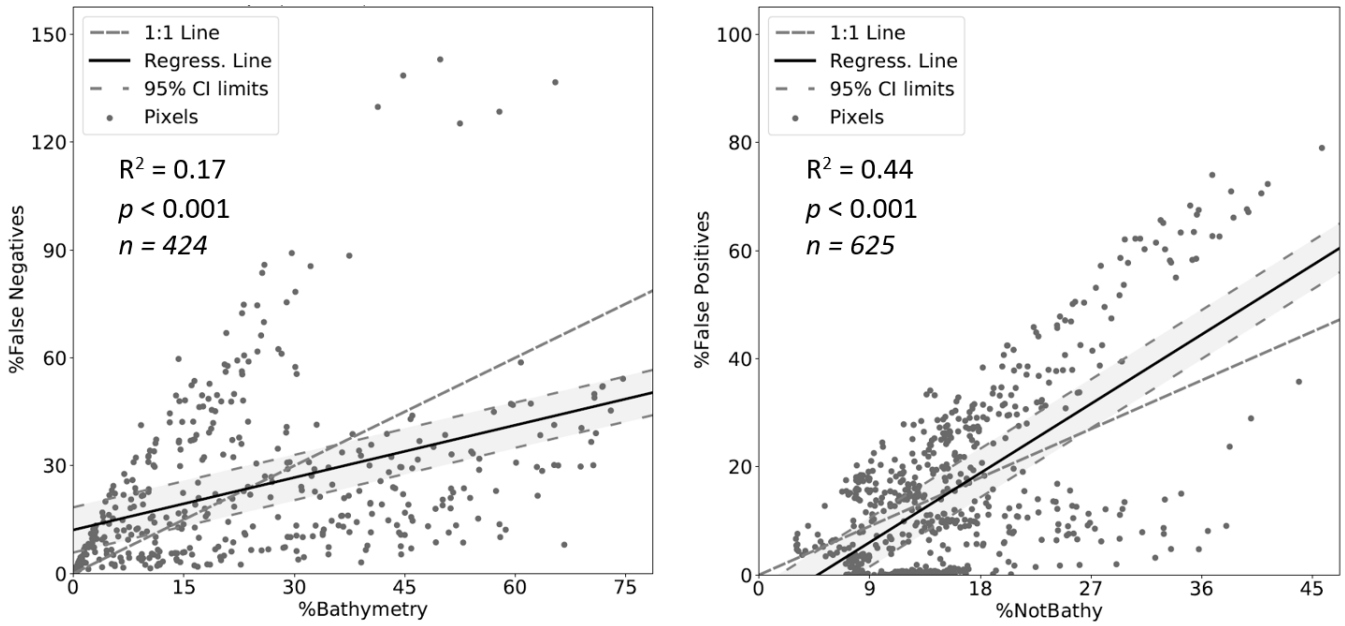


Figure 8. Scatterplot and regression statistics of classification errors for the "Deeper" tile (2708000n). Units of the x and y axes are percent times 100. a. % of *Bathy* soundings vs. % of false negatives. b. % of *NotBathy* soundings returns vs. % of false positives.

Figure 7 provides a means for examining the geographical distribution/concentration of classification errors. For this example, a large number of undetected *Bathy* soundings (FNs; Fig. 7a) are present in the southeast and near the center of this tile in an area close to the limits of lidar penetration (i.e., "extinction") (Fig. 7a). But an "excessive" number of *Bathy* soundings are erroneously being identified (i.e., FPs) along the eastern border (Fig. 7b). Neither the "excessive" FNs nor FPs appear to relate to depth nor proximity to extinction (see Fig. 1c). The utility of the information in Figure 7 is that it can improve continuous improvement efforts – particularly if classification error tendencies are examined over multiple tiles. Figure 8 provides a means to examine ML/AI model performance in feature space. For this tile, R^2s are low though significant, variance around both regression lines is heteroscedastic, FNs are concentrated in pixels where bathymetry soundings are sparse (where the regression line is above the 1:1 line in Fig. 8a), and FPs and *NotBathy* soundings are generally distributed similarly (the 95% confidence band around the regression line overlaps considerably with the 1:1 line in Fig. 8b). Although these are subjective interpretations with which another analyst might not agree, this discussion demonstrates that Figures 7 and 8 are useful descriptors of ML/AI model performance.

This examination of the distribution of errors geographically and in feature space may enable a decrease in manual input by NOAA into the *Bathy*/*NotBathy* classification process and increase accuracy. Areas where errors are "abundant" may be indicative of poor ML/AI model performance, of course. But these may instead (or also) be

indicative of where the NOAA classification accuracy is low. In particular, “high FN” areas may be areas that NOAA could target since these have the greatest potential real-world navigational consequences.

The use of the ODT also could help reduce manual input into the extraction of *Bathy* from lidar point clouds. In the example in Figure 3, although the ODT increases the total errors from 667,000 to 779,000, it reduces the most practically important error – FNs – by 67% from 497,500 to 165,000. Furthermore, the use of the ODT provides a better indication of the strength of the *Bathy* signal in the PAD for severely *Bathy*/*NotBathy* imbalanced point clouds. This could be operationally be important for processing tiles for which a ML/AI model is so weak as to be considered “borderline usable.”.

The ODT is also important in the improvement of the disambiguation rules employed in density-based lidar processing methods such as CHRT. Without consideration of sample imbalance, a pulse return whose *p(Bathy)* value is 0.10, for example, would be classified as*NotBathy*. However, for a tile where the measurable bathymetric signal is weak due to sample imbalance (such as the deepest tile 2707500n where the ODT = 0.02), a *p(Bathy)* of 0.10 would indicate a relatively high likelihood of being bathymetry. Hence the ODT provides a means to scale the *p(Bathy)* appropriately.

Results of this study suggest a sufficiently strong PAD-based bathymetric signal for improvement of operational procedures. However, incorporating results into a data processing workflow seemingly raises a paradox that precludes operationalization: fitting ML models to obtain a precise estimate of *p(Bathy)* for subsequent use in disambiguation rules requires a pre-existing classification. There are two potential ways to overcome this apparent circularity. First, it might be possible to fit a single “universally” applicable ML model from a sample of areas across a region and then apply it to individual areas. However, the prevalence of SBET and lidar-edge variables that likely are specific to ocean surface conditions at the time of data capture suggests this may be difficult. Moreover, local depth-related or turbidity interactions might also exist. Nonetheless, it may be possible to fit a “universal model” that is not statistically optimal, but that is operationally useful. This would be useful subject of future research

Second, ML models could become part of an iterative workflow. In its current form, CHRT, for example, is capable of distinguishing between “highly certain” and “less certain” bathymetry soundings. Such a distinction could be formalized as an initial *Bathy/NotBathy* classification. This would be used to train an initial ML model whose estimates of *p(Bathy)* would then be provided to CHRT for a second classification. This would provide for fitting a second set of ML models whose *p(Bathy)* values would again be fed back to CHRT, and the process would repeat until some convergence threshold is achieved. This would also be a useful subject for future research.

The work presented might be extended to other applications. For example, the extraction of terrestrial elevation from lidar data to produce "bare-earth" digital elevation models (DEMs) has received considerable attention (e.g., Mongus and Zalik 2012; Liu 2008) including recent use of ML/AI techniques (Hu and Yuan 2016). Using point-based/discrete lidar for this requires addressing the same conceptual problem as density-based ocean mapping methods: identifying lidar returns that are reflected from the earth surface (ocean bottom) rather than vegetative cover (the water column). Though modelling the vegetative cover using full waveform data is one way to address this, the increase in data volume makes this operationally daunting for a large area. The density-based methods for mapping ocean depth essentially create and analyze area-based waveforms derived from discrete lidar returns for grid points thereby overcoming the problem of waveform data volumes. It would seem that density-based bathymetry mapping from lidar and the enhancements documented in this study could be adapted to the problem of mapping terrestrial elevation.

Finally, an additional topic for potential future research is identified. It is known that the geomorphometry of an area – e.g., its rugosity, slope, slope direction -- impacts lidar pulse returns; it would be useful to examine how geomorphometry impacts the ocean depth signal strength in the pulse return metadata.

Summary and Conclusions

The principal goals of this research were to evaluate if metadata for airborne lidar that are currently not exploited contain information about ocean depth/bathymetry for shallow water (less than 20 m), and to explore how ML/AI can be used to model this relationship. The overarching operational goal was to enhance density-based methods of extracting bathymetry from discrete-return – i.e., not waveform -- lidar point clouds thereby potentially improving map accuracy, reducing person-hours, and decreasing subjectivity. Density-based methods establish a dense grid of estimation nodes (EN) across an area and use the lidar pulse returns of each EN's neighbours to estimate ocean depth for each EN. This entails being able to separate a given EN's neighbouring pulse returns that reflect from the ocean bottom from those that represent the ocean surface or water column noise. Because this is akin to the terrestrial application of separating pulse returns reflected from the Earth's surface from those reflected by vegetative cover, there may be potential for using this approach to produce bare-Earth digital terrain models without the need for voluminous lidar waveform data.

It was concluded that:

- The bathymetric signal strength in the point attribute metadata examined is sufficiently strong to enhance shallow-water bathymetry extraction from lidar point clouds.
- The scan angle, intensity of return, stability of the aircraft, and consistency of heading are the most important variables for determining if a return is bathymetry.

- The application of an "optimal" probability decision threshold that equalizes the TPR and TNR provides a better means to evaluate bathymetric signal strength than a conventional threshold of 0.50. due to potential *Bathy/NotBathy* imbalance in the lidar data. Such a threshold also provides adjusted estimates of the probability of each sounding being *Bathy* – *p(Bathy)* – that may be more useful than unadjusted values in operational density-based lidar processing workflows.
- The distribution of classification errors in geographic and feature space can be explored to potentially improve estimates of *p(Bathy)* and the classification itself.

Data and codes availability statement: The codes that support the findings of this study are available at the link https://figshare.com/s/8c563c3032927af0b25d. 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 to demonstrate how codes function. 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/htdata/lidar2_z/geoid12b/data/6246 as compressed .laz files. These can be decompressed using the LASzip tool which can be downloaded from laszip.org.

Declaration of interest: The authors have no potential competing nor financial interest in the work presented.

Acknowledgements

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

References

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

Beyan, C., Browman, H., 2020. Setting the stage for machine intelligence era in marine science. *ICES Journal of Marine Science*, 77(4):1267-1273, doi:10.1093/icesjms/fsaa084.

Boots, B., Getis, A., 1988. *Point Pattern Analysis,* SAGE Publications Scientific Geography Series (Series Editor: G. Thrall), 8, 93 pp.

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

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

Chust, G., Galpasoro, B., Borja, A., Franco, J., Uriarte, A., 2008. Coastal and estuarine habitat mapping, using LIDAR height and intensity and multi-spectral imagery. *Estuarine, Coastal, and Shelf Science*, 78, 633-643.

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

Collin, A., Archambault, P., Long, B., 2011. Predicting species diversity of benthic communities within turbid nearshore using full-waveform bathymetric LiDAR and machine learners. *PLoS ONE,* 6(6), e21265, doi:10.1371/journal.pone.0021265.

Collin, A., Ramambason, C., Pastol, Y., Casella, E., Rovere, A., Thiault, L., Espiau, B., Gilles Siu, G., Lerouvreur, F., Nakamura, N., Hench, J., Schmitt, R., Holbrook, S., Troyer, M., Davies, N., 2018. Very high resolution mapping of coral reef state using airborne bathymetric LiDAR surface intensity and drone imagery. *International Journal of Remote Sensing*, 39(17), 5676-5688.

Dixon, P., 2002. Ripley's K Function. *In Encyclopedia of Environmetrics*, John Wiley and Sons , 3, 1796-1803.

Eren, F., Pe'eri, S., Rzhanov, Y., Ward, L., 2019. Bottom characterization by using airborne lidar bathymetry (ALB) waveform features obtained from bottom residual analysis. *Remote Sensing of Environment,* 206, 260-274, doi.org/10.1016/j.rse.2017.12.035.

Fischler, M., Bolles, J., 1981. Random sample consensus – a paradigm for model fitting with applications to image analysis and automated cartography. *Communication of the ACM*, 24(6), 381-395.

Guo, B., Huang, X., Zhang, F., Sohn, G., 2015. Classification of airborne scanning data using JointBoost. *ISPRS Journal of Photogrammetry and Remote Sensing*, 100, 71-83, doi:dx.doi.org/10.1016/j.isprsjprs.2014.04.015.

Hu, X., Yuan, Y., 2016. Deep-learning-based classification for DTM extraction from ALS point cloud. *Remote Sensing*, 8(9), 730 (16 pp.), doi:10.3390/rs8090730.

Kang, Z., Yang, J., Zhong, R., 2016. A Bayesian-network-based classification method integrating airborne LiDAR data with optical images. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 9(12), 1–11.

Kim, H., Sohn, G., 2011. Random forests based multiple classifier system for power-line scene classification. *In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVIII-5/W12*, ISPRS Workshop, Calgary, Canada (August 29-31), 6 pp.

Liu, X., 2008. Airborne lidar for DEM generation: some critical issues. *Progress in Physical Geography,* 32(1), 31-49, doi:10.1177/0309133308089496.

Lowell, K., Calder, B., 2019. Machine learning strategies for enhancing bathymetry extraction from imbalanced lidar point clouds. *In: Proceedings Oceans 2019,* Seattle, Washington, United States, 8 pp.

Macon, C., Wozencraft, J., Park, J., Tuell, G., 2008. Seafloor and land cover classification through airborne lidar and hyperspectral data fusion. *In: Proc: IEEE International Geosciences and Remote Sensing Symposium*, Boston, Massachusetts, United States, 77-80.

Malde, K., Handegard, N., Eikvil, L., Salbery, A.-B., 2020. Machine intelligenceand the data-driven future of marine science. *ICES Journal of Marine Science* 77(4):1274-1285, doi:10.1093/icesjms/fsz057.

Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), S71–S84.

Matikainen, L., Hyyppä, J., Kaartinen, H., 2009. Comparison between first pulse and last pulse laser scanner data in the automatic detection of buildings. *Photogrammetric Engineering and Remote Sensing*, 75(2), 133–146.

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

Mohamed, H., Negm, A., Zahran, M., Saavedra, O., 2016. Bathymetry determination from high resolution imagery using ensemble learning algorithms in shallow lakes: case study El_Burullus Lake. *International Journal of Environmental Science and Development,* 7(4), 7 pp., doi:10:7763/IJESD.2016.V7.787.

Mongus, D., Zalik, B., 2012. Parameter-free ground filtering of LiDAR data for automatic DTM generation. *ISPRS Journal of Photogrammetry and Remote Sensing,* 67, 1-12, doi:doi.org/10.1016/j.isprsjprs.2011.10.002.

Nagle, D., Wright, C., 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.

Nahhas, F., Shafri, H., Sameen, M., Pradhan, B., Mansor, S., 2018. Deep learning approach for building detection using LiDAR-orthophoto fusion." *Journal of Sensors,* Article ID*:* 7212307, 12 pp., doi:doi.org/10.1155/2018/7212307.

Narayanan, R., Kim, H., Sohn, G., 2009. Classification of SHOALS 3000 bathymetric LiDAR signals using decision tree and ensemble techniques. *In: Proc. TIC-STH 2009: IEEE Toronto International Conference -- Science and Technology for Humanity*, Toronto, Ontario, Canada, pp. 462-467.

Nayegandhi, A., Brock, J., C. Wright, C., 2009. Small-footprint, waveform-resolving lidar estimation of submerged and sub-canopy topography in coastal environments. *International Journal of Remote Sensing,* 4, 861-878.

Odland, J., *Spatial Autocorrelation*, SAGE Publications Scientific Geography Series (Series Editor: G. Thrall), 9, 87 pp.

Parrish, C., Rogers, J., Calder, B., 2014. Assessment of waveform features for lidar uncertainty modeling in a coastal salt marsh environment. *IEEE Geoscience and Remote Sensing Letters*, 11(2), 569-573, doi:10.1109/LGRS.2013.2280182.

Pittman, S., Costa, B., Battista, T., 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.

RIEGL, 2017. *LAS extrabytes implementation in RIEGL software* (whitepaper). RIEGL Laser Measurement Systems.(http://www.riegl.com/uploads/tx_pxpriegldownloads/Whitepaper_LASextrabytes_implementation_in-RIEGLSoftware_2017-12-04.pdf), 17 pp.

Rogers, J., Parrish, C., Ward, L., Burdick, D., 2015. Evaluation of field-measured vertical obscuration and full waveform lidar to assess salt marsh vegetation biophysical parameters. *Remote Sensing of the Environment,* 156, 264-275, doi:10.1016/j.rse.2014.09.035.

Salameh, E., Frappart, F., Almar, R., Baptista, P., Heygster, G., Lubac, B., Raucoles, D., Almeida, L. P., Bergsma, E., Capo, S., De Michele, M., Idier, D., Li,. Z., Marieu, V., poupardin, A., Silva, P., Turki, I., Laignel, B., 2019. *Remote Sensing*, 11(19), 2212, 32 pp., doi: 10.3390/rs11192212.

Tuell, G., Park, J. Aitken, J., Ramnath, V., Feygels, V., Guenther, G., Kopilevich, Y., 2005. SHOALS-enabled 3D benthic mapping." *In: Proc. SPIE 5806: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, Orlando, Florida, United States, pp. 816-826, doi: 10.1117/12.607010.

Tulldahl, H., 2014. Processing of airborne lidar bathymetry data for detailed sea floor mapping. *In Proc. SPIE 9250: Electro-optical Remote Sensing, Photonic Technologies and Applications VIII; and (SPIE 92500E) Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing II*, Amsterdam, Netherlands, 16 pp., doi:10.1117/12.2068940.

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

Vojinovic, Z., Abebe, Y., Ranasinghe, R., Vacher, A., Martens, P., Mandl, D., Frye, S., van Ettinger, E., de Zeeuw, R., 2013. A machine learning approach for estimation of shallow water depths from optical satellite images and sonar measurements. *Journal of Hydroinformatics,* 15(4), 17 pp., doi:10.2166/hydro.2013.234.

Wozencraft, J., Millar, D., 2005. Airborne lidar and integrated technologies for coastal mapping and nautical charting. *Marine Technology Society Journal,* 39(3), 27-35.

Zhang, Z., Zhang, L., Tong, L., Guo, B., 2016. Discriminative-dictionary-learning-based multilevel point-cluster features for ALS point-cloud classification. *IEEE Transactions on Geoscience & Remote Sensing*, 54(12), 7309–7322, doi:10.1109/TGRS.2016.2599163.

Zhao, R., Pang, M., Wang, J., 2018. Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network. *International Journal of Geographical Information Science*, 32(5), 960-979,doi:doi.org/10.1080/13658816.2018.1431840.