A priori Data Set Screening to Improve Efficiency of LiDAR Processing for Shallow Water Bathymetry

Kim Lowell[1*]

[1]Center for Coastal and Ocean Mapping and Joint Hydrographic Centre, University of New Hampshire,
Durham, NH 03824 UNITED STATES

[*]Corresponding Author: klowell@ccom.unh.edu (ORCID: 0000-0002-8326-4022)

Abstract

High data volumes and the time required to process LiDAR point clouds to identify bathymetric points create a potentially large lag between data acquisition and use for shallow water mapping. In this study, a method was developed for *a priori* identification of areas (500m by 500m "tiles") in a data set that are unlikely to contain bathymetric pulse returns and therefore do not need to be processed. Using an airborne LiDAR data set centred on Key West, Florida (United States) containing 1374 tiles, a logistic regression model was developed to predict if a tile contained extractable bathymetry (according to standard operating procedures of the United States National Oceanic and Atmospheric Agency (NOAA)) using quantifiable characteristics of depth frequency histograms as predictors. Results indicated that tiles that do not contain extractable bathymetric pulse returns could be identified with 90% accuracy. A post-modelling "spatial reassignment" of individual tiles based on characteristics of neighbouring tiles provided only a minor accuracy improvement. The methodology was validated on a Miami Beach LiDAR data set containing 120 tiles. Results were comparable to the Key West results although the logistic regression model had to be re-calibrated for Miami Beach. To operationalize the results and eliminate the need to process all tiles *a priori*, a progressive tile-sampling approach is suggested. Furthermore, operational use of this *a priori* tile screening approach also requires consideration of expected uses of bathymetric maps and risk tolerance relative to the different consequences of false negative (FN) and false positive (FP) errors. For the Key West data set comprised of 1374 tiles of which 36% did not contain extractable bathymetry, screening tiles and then processing non-excluded tiles for bathymetric extraction was estimated to reduce total time by 489 hours (161 human/manual hours and 328 computer hours) compared to not screening and processing all 1374 tiles.

Keywords: Logistic regression, LiDAR point clouds, LiDAR processing, hydrographic mapping

Funding Details

This work was supported by the National Oceanic and Atmospheric Administration (NOAA) under Grant NA20NOS4000196.

Data Availability Statement

The 2016 airborne LiDAR data for the Key West area are available free of charge via NOAA's data access viewer (https://www.coast.noaa.gov/dataviewer/#/). The data set collectively has the name:

- Key West: 2016 NGS Topobathy Lidar: Key West (19,946,934,287 points).

At the time of paper submission, the 2022 airborne LiDAR data for the Miami Beach area were not yet available. These can be obtained by contacting the author (pending approval by NOAA).

1. Introduction

Airborne LiDAR for bathymetry (ALB)[1] data are increasingly being used operationally to map shallow water areas. ("Shallow water" for the purpose of this paper is depths less than 30 m although LiDAR penetration to depths up to 80 m have been reported under ideal conditions (Parker and Sinclair, 2012). This is apparent in the increasing number and geographical and temporal extents of publicly available LiDAR data sets that can be downloaded via, for example, the NOAA (National Oceanic and Atmospheric Administration) Data Access Viewer (NOAA 2024). ALB data are viewed as a useful data source for hydrographic mapping in large measure because their acquisition is not constrained by potentially dangerous shallow water conditions that hinder ships equipped with acoustic/sonar sensors.

Generally ALB data conforming to contractual standards are provided as ".las" files or ".laz" files – a compressed form of ".las" files. Generally, each .las or .laz file covers a 500 m-by-500 m area and is referred to as a "tile." LiDAR tiles are delivered in a recognised format (ASPRS 2019) that supports both point cloud LiDAR data and full waveform LiDAR data. The use of full waveform LiDAR data for a range of ocean mapping purposes has been explored – e.g., Parrish *et al*. (2014),

[1] To improve comprehension, Appendix 1 contains acronyms that may be unknown to some readers or that are specific to this article.

Rogers *et al*. (2015), Eren *et al*. (2019). However, most operational ALB-based hydrographic mapping focuses on point cloud data; this is also the focus of the present article.

ALB point cloud data are essentially 3-dimensional (3D) locations of pulse returns as well as pulse return metadata such as the number of the pulse return (for multiple returns) and the intensity of a return. Though not considered here, the utility of such metadata for hydrographic mapping has been examined (e.g., Lowell and Calder 2021, Lowell *et al*. 2021) as well as for other applications such as building detection (Matikainen 2009).

Once the 3D {x, y, z} coordinates of each pulse return – i.e., known as a "sounding" in hydrographic parlance -- are acquired, it remains to identify and separate the pulse returns that represent ocean depth from those that represent the ocean surface or are noise. Considerable effort continues to be expended and multiple approaches explored to improve the accuracy of this process – e.g., Agrafiotis *et al.* (2019), Yang *et al.* (2020), Ranndal *et al.* (2021), Lowell and Calder (2022).

A topic that has received little attention, however, is "*a priori* screening" to eliminate LiDAR tiles (areas) in which depth or water characteristics make it unlikely that processing ALB data successfully identifies bathymetric pulse returns – i.e., those that represent the ocean floor. Yet *a priori* screening ALB data has the potential to identify tiles that are unlikely to contain extractable bathymetry (usually due to "too deep" depth) thereby reducing backlogs, speeding data availability, and reducing the processing resources required. Some physics-based studies have addressed the estimation of LiDAR extinction depth (Giannakaki *et al.* 2020, Lisenko and Shamanaev 2022) – or the detection of a subsurface ocean layer (Krekov *et al.* 1997; Krekov *et al.* 1998). Such studies do not, however, provide for the identification of areas that exceed extinction depth. Without this knowledge *a priori*, all .las files in a data set must be processed. Hence those .las files covering areas that are beyond extinction depth – i.e., "too deep" – can only be identified *a posteriori* meaning that processing resources will have been unnecessarily expended.

This article is focused on the accurate *a priori* identification of geographic partitions (tiles) in a data set that do not have extractable bathymetric (i.e., "*Bathy")* pulse returns. If identification of such tiles is sufficiently accurate, this *a priori* screening could become an initial time-saving step in a workflow to extract *Bathy* pulse returns from the .las files in an ALB database. As imagined, .las files having a low likelihood of containing extractable *Bathy* soundings would be identified algorithmically and eliminated from time-consuming processing that identifies individual bathymetric pulse returns within each ALB tiles.

Exploring the potential of *a priori* screening to reduce LiDAR bathymetric processing effort appears to be quite novel. Though there has been considerable work on removing individual pulse return outliers from LiDAR point clouds for a number of applications (e.g., Matkan *et* al. 2014, Le *et* al. 2022, Szutor and Zichar 2023), such work assumes that a decision has already been made to process all tiles of the LiDAR data set of interest. The work presented here addresses the decision of whether or not each tile in an ALB survey merits processing at all. If this can be determined accurately prior to processing, considerable time and effort can be saved. Evaluating one approach to this *a priori* screening is the overarching goal of this work.

The approach explored is the quantitative characterization of each tile's frequency distribution of individual pulse return depths and the use of logistic regression modelling to estimate the probability that a given tile has extractable *Bathy* pulse returns. Clearly, such *a priori* identification of tiles that should undergo bathymetric pulse return processing is unlikely to be 100% accurate. Hence this study also quantifies the expected number of false negative (FN) tiles (that represent a loss of recoverable data), and false positive (FP) tiles (that represent an unnecessary use of processing resources). The results present a guide to the trade-offs between potential resource savings and impacts on accuracy.

2. Materials and Methods

To facilitate reader comprehension, Figure 1 provides a schematic workflow of the materials and methods.

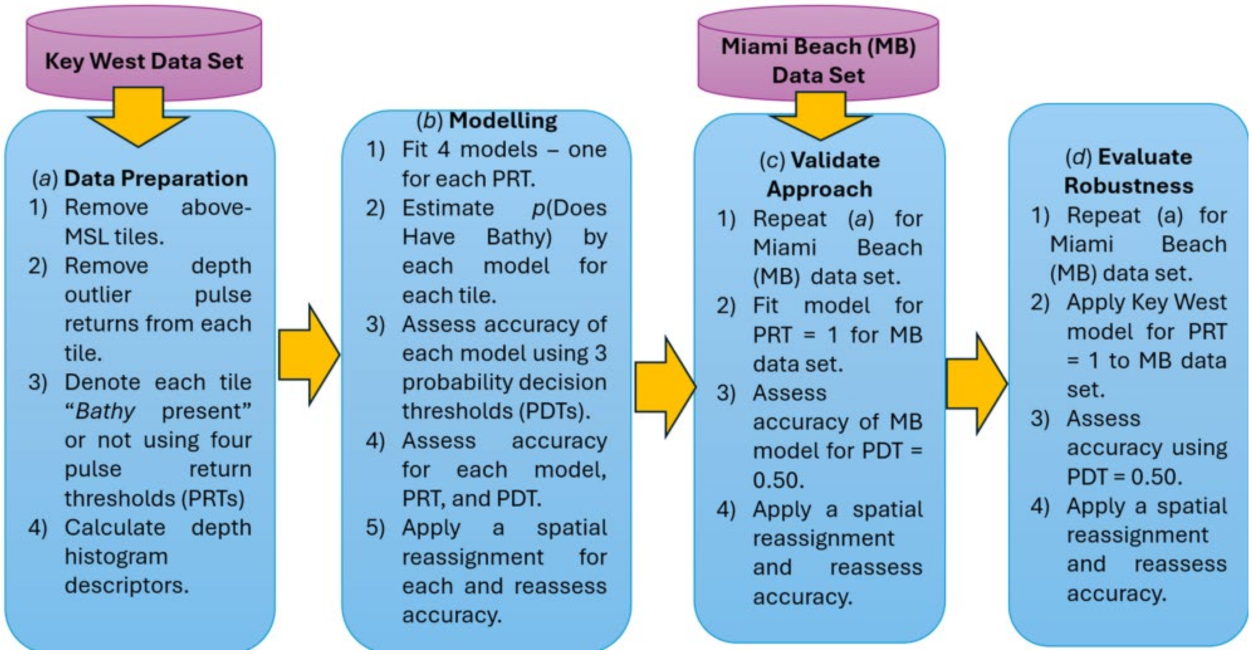

Figure 1. Analytical workflow.

2a. Data

Two ALB data sets captured by NOAA were employed (Figure 2). The calibration/developmental ALB data were captured between April 19 and 25, 2016 in the vicinity of Key West, Florida (24º33’ N latitude and 81º46’ W longitude. The validation/evaluation data set was acquired April 28, 2022 in the vicinity of Miami Beach, Florida (25º57’ N latitude and 80º05’ W longitude). Both datasets were acquired using multiple overlapping flight lines generally having a north-south orientation using a nominal flying height of 400 m that provided a point density of approximately 10 soundings sq m^{-1}. The 2016 Key West data set was acquired using a Riegel™ VQ-880-G sensor; the 2022 Miami Beach data were collected using a Riegl VQ880GII instrument. Both instruments employ a counterclockwise circular scan and a 20^0 scan angle. Data were provided as 500 m-by-500 m tiles aligned north-south and east-west registered to zone 17N of the Universal Transverse Mercator Projection (UTM) referenced to the World Geodetic System 1984 (WGS84) datum. Individual pulse returns in all tiles had been processed using NOAA’s Standard Operating Procedures (SOPs) to classify each pulse return according to standard LiDAR classes (ASPRS 2019). For this study, these classes were collapsed into a binary *Bathy/NotBathy* classification. Depth values had been tide-corrected to Mean Sea Level (MSL). Information on the data sets is provided in Table 1. Notable differences between the two (in addition to location and date) are that the Key West data set has many more tiles, a higher average number of pulse returns and *Bathy* pulse returns per tile, and covers a wider variety of depths according to NOAA SOP's bathymetric extraction. The Key West area also has areas where shipping channels increase water turbidity and a more complex ocean substrate that includes sea grass, sandy bottom, and intersecting channels. The Miami Beach area has consistently clear water, a sandy substrate, and relatively constant slope from land to LiDAR extinction depth.

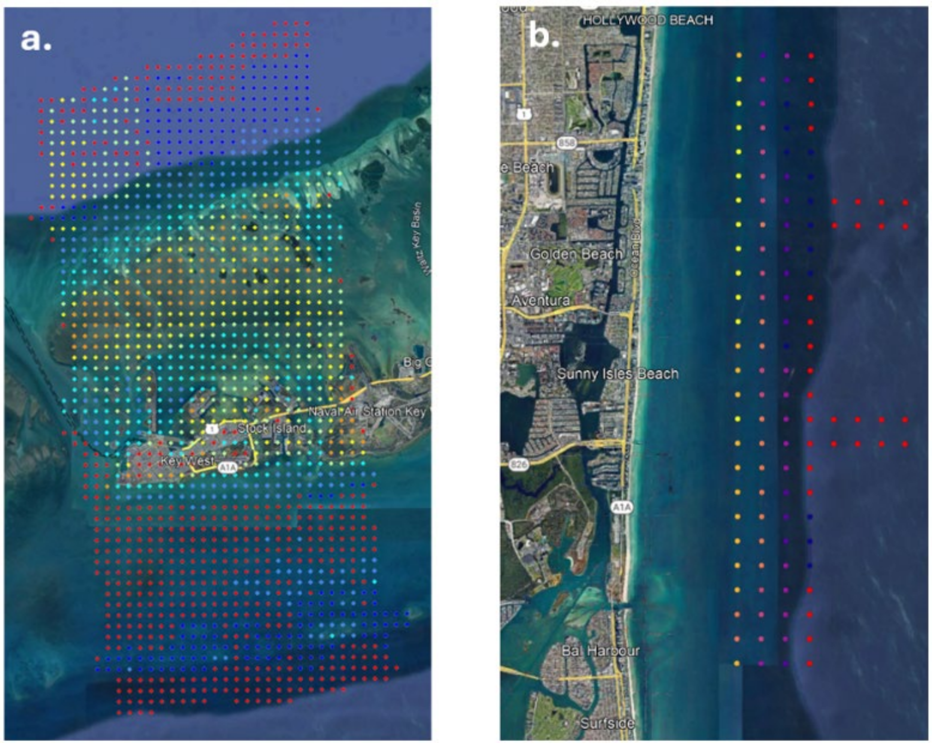

Figure 2. Layout and location of the study areas. Each dot represents the centre of an ALB data tile; distance between dots on both figures is 500 m. Red dots are tiles on which bathymetric pulse returns are not present according to NOAA's SOP classification. (Imagery courtesy of GoogleEarth™.)

Table 1. Descriptive statistics for tiles not having areas above Mean Sea Level (MSL) for the Key West (n = 1374) and Miami Beach (n = 120) data sets.

Key West Data Set (n = 1374)			
	Minimum	**Mean**	**Maximum**
Number of Pulse Returns	85	6731100	29180900
Number of *Bathy* Pulse Returns	0	1670400	9328100
Percent *Bathy* Pulse Returns	0%	21.0%	98.6%

Depth of *Bathy* Pulse Returns (m)	-11.9	-2.4	0.0			
Number of tiles having a certain range of *Bathy* Pulse Returns:						
Range:	**<= 0 (None)**	**> 0**	**>= 1 <= 100**	**>100 & <= 250**	**>250 & <= 500**	**> 500**
Number:	468	906	85	17	14	790
Miami Beach Data Set (n = 120)						
	Minimum	**Mean**	**Maximum**			
Number of Pulse Returns	1454	4165700	8870500			
Number of *Bathy* Pulse Returns	0	784500	2075100			
Percent *Bathy* Pulse Returns	0%	13.9%	33.6%			
Depth of *Bathy* Pulse Returns (m)	-33.6	-18.6	-11.4			
Number of tiles having a certain range of *Bathy* Pulse Returns:						
Range:	**<= 0 (None)**	**> 0**	**>= 1 <= 100**	**>100 & <= 250**	**>250 & <= 500**	**> 500**
Number:	33	87	2	3	2	80

2b. Approach and Procedures

Analysis was confined to those tiles on which NOAA did not detect the presence of any "land soundings" – i.e., those above MSL. Operationally, given that NOAA's pulse return classification would not be available, it is assumed that such tiles could be identified with sufficient accuracy using, for example, satellite imagery and threshold slicing of the normalized difference water index (NDWI) -- e.g., Wen *et* al. 2021, Qi *et al.* 2022.

Central to the *a priori* screening approach explored for the remaining/not-excluded "no land" tiles is the expectation that the shape of the depth frequency distributions is indicative of the likelihood that extractable *Bathy* pulse returns are present on a tile. To facilitate obtaining useful measures of depth distribution shape, it is desirable to first eliminate outliers caused by instrument errors, sun glint, etc. In this study, this was done by simply using thresholds that eliminated pulse returns whose depths were clearly above MSL (> 3 m above MSL including consideration of waves), and clearly beyond LiDAR penetration depth (< 70 m below MSL). No effort was made to select optimal outlier screening thresholds, although more objective threshold determination methods such as those based on analytical optimization could be employed. Moreover, more complex outlier screening methods such as Mahalanobis screening (Mahalanobis 1936) or jackknife sampling (Quenouille 1956) could alternatively be employed. It is noted that the use of individual outlier screening methods and thresholds has the potential to limit the transferability of the *a priori* tile screening method explored.

Consider that after removal of depth outliers, a depth frequency distribution for a tile having only *NotBathy* pulse returns that are reflected primarily from the ocean surface would be narrow, highly peaked, nearly symmetrical, unimodal, and have a mean depth near 0.0 (Figures 3a and 3b). Conversely, as the number of *Bathy* pulse returns increases, depth frequency distributions would become wider, less peaked, skewed to the left (i.e., negatively skewed), multimodal (Nason and Sibson 1992), and have a mean depth lower than 0.0 (Figures 3c and 3d). This reflects the observation by others (e.g., Mandlburger and Jutzi 2019; Jung *et* al. 2021) that LiDAR tiles in which pulse returns representing ocean depth are present will have a "vertical gap" in their depth frequency histogram as shown in Figure 3.

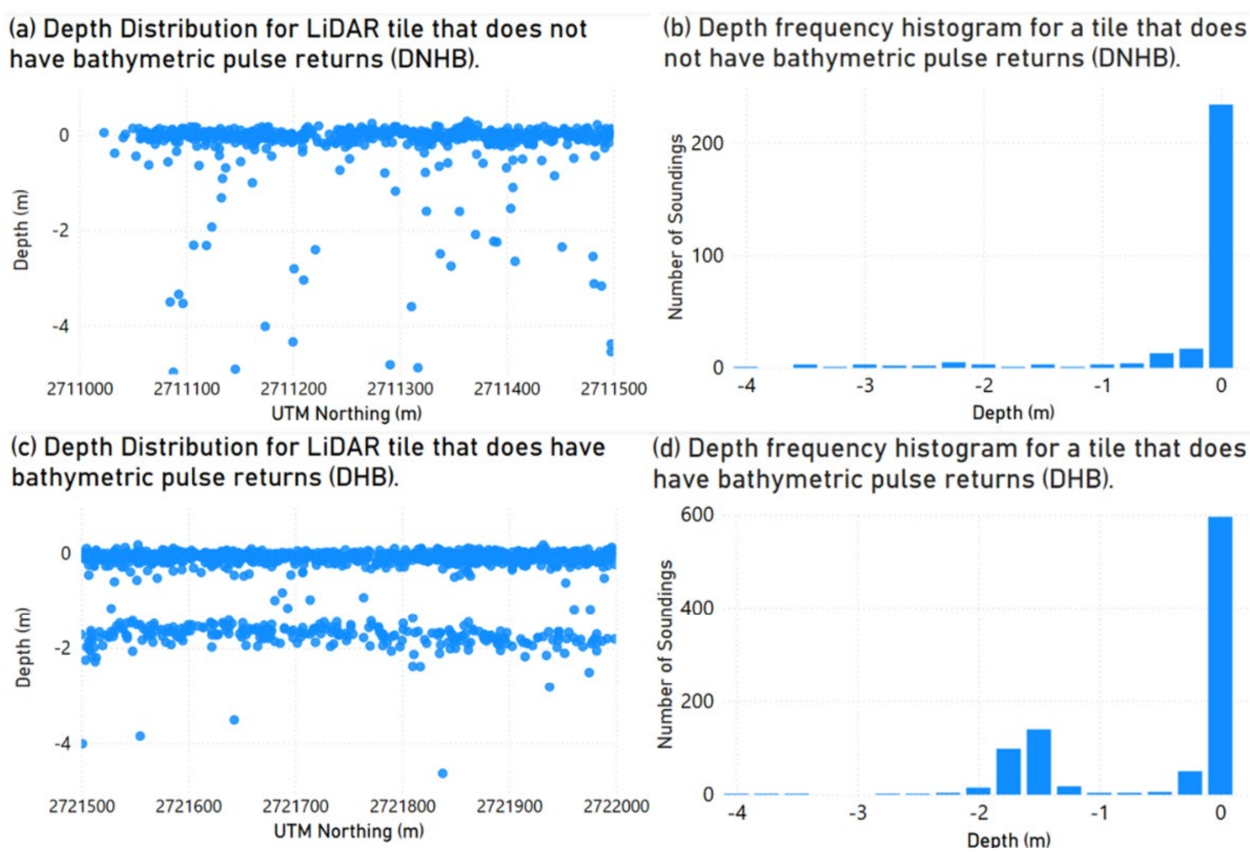

Figure 3. Hypothetical examples of pulse return depths and representative frequency distributions. (*a*) & (*b*) Tile that does not have bathymetric pulse returns present (DNHB); (*c*) & (*d*) Tile that does have bathymetric pulse returns present (DHB).

The following metrics were extracted for each tile's frequency distribution after outliers had been removed:

- Location – minimum, maximum, median, and mean depth.
- Width – standard deviation and coefficient of variation.
- Shape – kurtosis (Balanda and MacGillivray 1988; values greater than 3.0 indicate greater "excessive" peakedness) and skewness (Joanes and Gill 1998; negative values indicate left skew; positive values indicate right skew).
- Modality – the Dip statistic (Hartigan 1985, Hartigan and Hartigan 1985); lower values represent a greater likelihood of unimodality. (The Dip value can be tested for significant unimodality.)

Initially, a value for a single binary dependent variable was assigned to each tile – 0 (zero) if NOAA did not find *Bathy* pulse returns on a tile or 1 (one) if NOAA found at least one *Bathy* pulse return. Operationally, tiles determined to not have at least one bathymetric pulse return are those that *a priori* screening is hoping to identify. (These are termed "Does Not Have *Bathy*" (DNHB) tiles as opposed to tiles that do have at least one bathymetric pulse return ("Does Have Bathy" (DHB)). However, the pulse return threshold (PRT) of "at least 1 Bathy pulse return" does not consider that if there are relatively few *Bathy* pulse returns on a tile, identification of individual bathymetric pulse returns will be less reliable. Hence to examine the "accuracy/effort" trade-off during method development, three more binary "DNHB/DHB" variables were created using additional PRTs that designate tiles having at least 100, 250, and 500 bathymetric pulse returns as 1 (one) (DHB) or 0 (zero) (DNHB) otherwise. The information on the number of tiles having more than each PRT for both data sets is presented in Table 1. Ultimately a potential willingness to process a tile only if it has at least, for example, 250 bathymetric pulse returns implicitly accepts the loss of potentially valuable bathymetric data in exchange for a presumed reduction in the number of tiles processed. PRTs for the relative number of bathymetric pulse returns – i.e., percent of total -- were also examined. However, because models based on such relative PRTs performed poorly, they were not considered further.

Logistic regression models were fitted for the Key West data set for each binary DNHB/DHB variable produced by the four PRTs using the frequency distribution descriptors described earlier as dependent variables. Various techniques – e.g., forward and backward stepwise, all possible models -- were employed to identify the optimal logistic model form. Based primarily on the Aikeke Information Criterion (AIC; Cavanaugh and Neath 2019), the most efficient model across all PRTs consistently employed the Dip statistic value, the standard deviation of depth, and skewness. Notably, none of these variables directly address the depth values present in the frequency distribution of each LiDAR tile – i.e., the minimum, maximum, median, or mean depth. This was somewhat surprising given that *a priori* it was speculated that tiles having relatively deep mean depths, for example, would be more likely to have bathymetric pulse returns present. Further examination of the depth frequency distributions suggested that due to depth distributions being dominated by ocean surface pulse returns, the variability in the minimum, maximum, and mean pulse return depths across all tiles was low thereby limiting the predictive utility of these variables.

Each model was used to estimate the probability that each tile was a DHB or DNHB tile. Subsequently, tiles with a model *p(DHB)* value greater than 0.5 were designated DHB and those with a *p(DHB)* value less than 0.5 designated DNHB. While 0.5 is the conventional probability decision threshold employed, alternatives that identify an "optimal" numerical value are possible – e.g., Youden's Index (Youden 1950), simulation-based graphical trade-offs between false positive and false negative errors for imbalanced data sets (Lowell *et al*., 2021). This is addressed in the discussion section.

Ideally for operational purposes, a single logistic *a priori* screening model fitted on a large data set would be "universally" applicable. To explore the robustness of the models developed, two approaches were employed. First, logistic regression models of the same form for the same PRTs were fitted to the Miami Beach data set. The statistical agreement of coefficients for the Key West and Miami Beach models was evaluated. Second, the Key West models for each PRT were applied to the Miami Beach data set and the accuracy compared to the result of applying Miami Beach models to the Miami Beach data set.

To this point, the prediction of each tile as DHB/DNHB (does have/does not have bathymetric pulse returns) was based solely on a logistic model fitted using only measurable depth frequency distribution characteristics. It was surmised that for each tile the *p*(DHB) (the probability of having bathymetric pulse returns) of a tile's neighbours might also be indicative of its "true" DHB/DNHB state. Hence a *post hoc* "spatial reassignment" was applied across all tiles. The logic of this reassignment was that if, for example, a "large proportion" of DHB tile's neighbours were labelled as DNHB, the DHB tile being examined was actually likely to not have extractable bathymetry. Sensitivity testing was undertaken and a threshold of greater than 0.70 was used to define a "large proportion". For each FN and true negative (TN) tile (i.e., tiles designated DNHB based on the modelled *p*(DNHB)) having at least three immediate neighbours (i.e., tiles not on the edge of the study area), the proportion of the tile's immediate neighbours that were designated DHB based on the model prediction was determined. If that proportion was greater than 0.70, the FN/TN tile's designation was changed to DHB regardless of its logistic regression model *p*(DHB) value; this had the effects of increasing the number of tiles requiring bathymetric processing and increasing the FP tiles with a consequent reduction in FN tiles. The same logic and process was applied to FP/TP tiles. Figure 4 provides a visual depiction. Figure 4(*a*) shows the logistic regression probability of a tile having bathymetric pulse returns *p*(DHB) for all tiles from which the green/red tiles in Figures 4(*b*) and 4(*c*) were "erroneously" predicted by the model to be DNHB or DHB based on the *p*(DHB) values of each tile's immediate neighbours. These tiles have a relatively low or high *p*(DHB) in Figure 4(*a*), but their immediate neighbours have the opposite. TN red tiles in Figure 4(*b*) would be (erroneously) reassigned from DNHB to DHB while green tiles in Figure 4(*c*) would be (correctly) reassigned from DNHB to DHB. The change in accuracy caused by spatial reassignment was evaluated in the same manner as the accuracy obtained using the logistic model only.

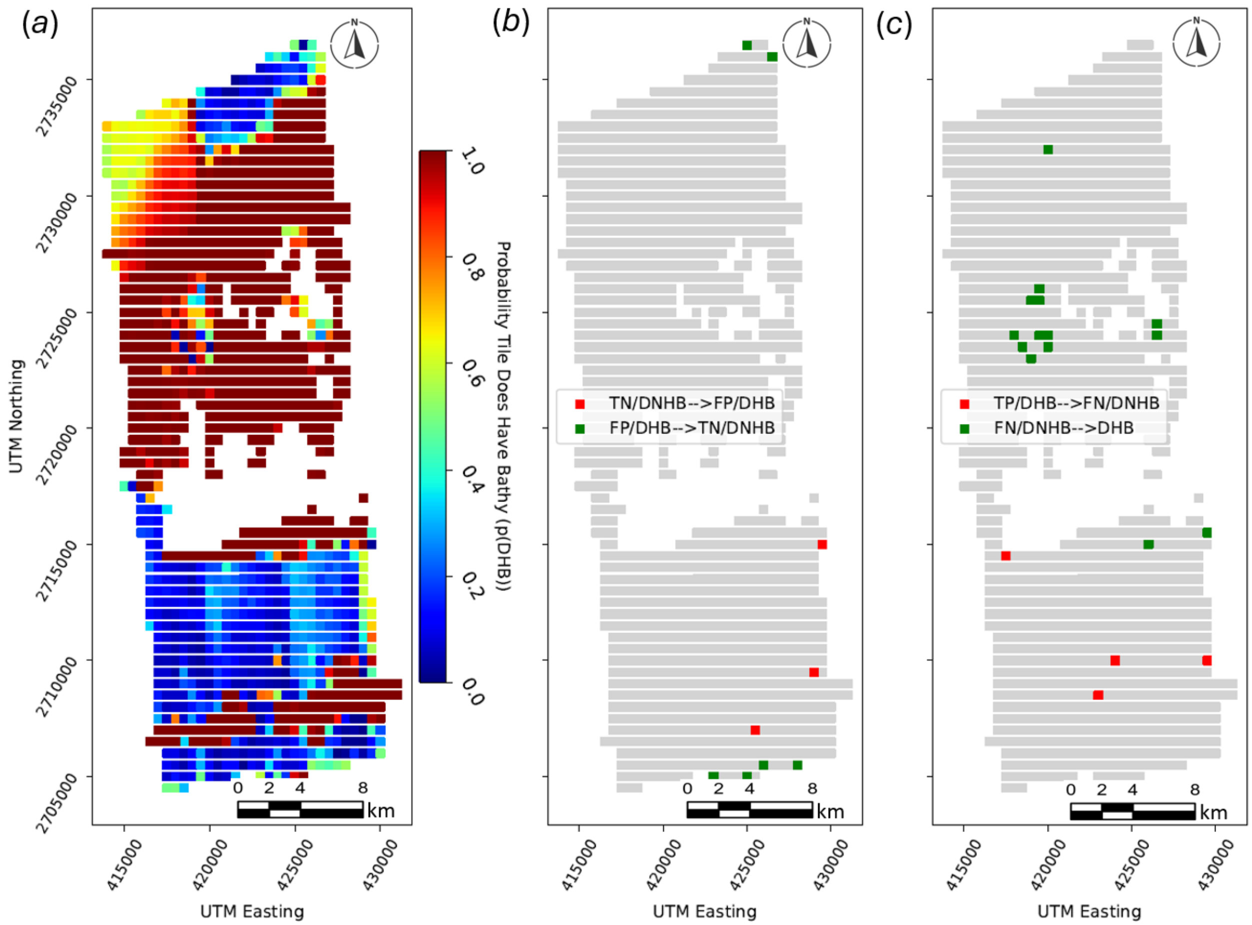

Figure 4. Example of spatial reassignment of individual tiles. (*a*). p(Bathy) for each tile according to model. (*b*). Red tiles were correctly identified as “Does Not Have Bathy (DNHB)” by the model but spatial filtering incorrectly assigns them to “Does Have Bathy (DHB). Green tiles were incorrectly identified as “DHB)” but spatial filtering correctly assigns them to “DNHB”. (*c*). Red tiles were correctly identified as “DHB” but spatial filtering incorrectly assigns them to “DNHB.” Green tiles were incorrectly identified as “DNHB” but spatial filtering correctly identifies them as “DHB”.

Results

Table 2 indicates that for Key West the use of a depth frequency distribution-based model is a viable approach to identifying tiles for which bathymetric processing is necessary or unnecessary by virtue of being able to accurately separate DHB from DNHB tiles. Model goodness-of-fit (R^2) values are all statistically significant and accuracy values are 0.90 or above and comparable across all PRTs. (Readers are reminded that the number of tiles that achieved each PRT is presented in Table 1.) The relatively high, and roughly equal, F1 scores for the DHB/DNHB tiles are indicative of models that perform well and equally across both classes. The importance in the models of the variables is reasonably consistent. The positive signs for coefficient values are as expected and indicate that as frequency distributions become wider, more negatively skewed (i.e., skewed to the left), and more multi-modal, tiles are more likely to be identified as DHB. Note that it is the global accuracy and F1 scores that are of most interest as these are indicative of the number of FN tiles (that represent "needlessly" lost data) and FP tiles (i.e., "needlessly" processed tiles) that can be expected operationally – i.e., when the true FN and FP rates are not known.

Table 2. Details of logistic models fitted for Key West.

Data / Model	**Soundings Threshold**	**Pseudo R^2 [1]**	**Global Accuracy**	**F1 Score**		**Variable Importance / Coeff. Sign**		
				Does Not Have *Bathy*	**Does Have *Bathy***	**Most**	**Mid**	**Least**
Key West	**1**	0.65[1]	0.91	0.87	0.93	Depth std. dev. / +	Skewness / +	Dip / +
	100	0.65	0.90	0.88	0.91	Depth std. dev. / +	Dip / +	Skewness / +
	250	0.67	0.90	0.89	0.91	Depth std. dev. / +	Dip / +	Skewness / +
	500	0.69	0.91	0.89	0.92	Depth std. dev. / +	Dip / +	Skewness / +

Figure 5(*a*) (the light blue line) shows that as the PRT increases, the number of tiles requiring bathymetric processing decreases. For example, suppose that one is willing to accept that tiles having fewer than 100 bathymetric pulse returns cannot be processed accurately given that the average number of pulse returns on Key West tiles is 6.7 million (Table 1). For the data set employed adopting a PRT of 100 would decrease the number of tiles one must process from about 880 to 770 – a reduction of about 12%. This is accompanied by an increase in FNs (Figure 5(*b*)) from 75 to 100 (a 33% increase) and a reduction in FPs (Figure 5(*c*)) from 49 to 44 (10%). This requires acceptance that any extractable bathymetric pulse returns present in the 85 tiles that have fewer than 100 bathymetric pulse returns are lost – i.e., one is accepting the existence of 85 FNs or approximately 6% of total tiles. If one accepts this, a PRT of 100 or 250 seems to be the optimal PRT value as this is where the PRT-related decrease in tiles to process (Figure 5(*a*)) appears to asymptote. A higher PRT of 500 does not reduce the number of tiles to process substantially, nor does it change the number of FNs and FPs markedly.

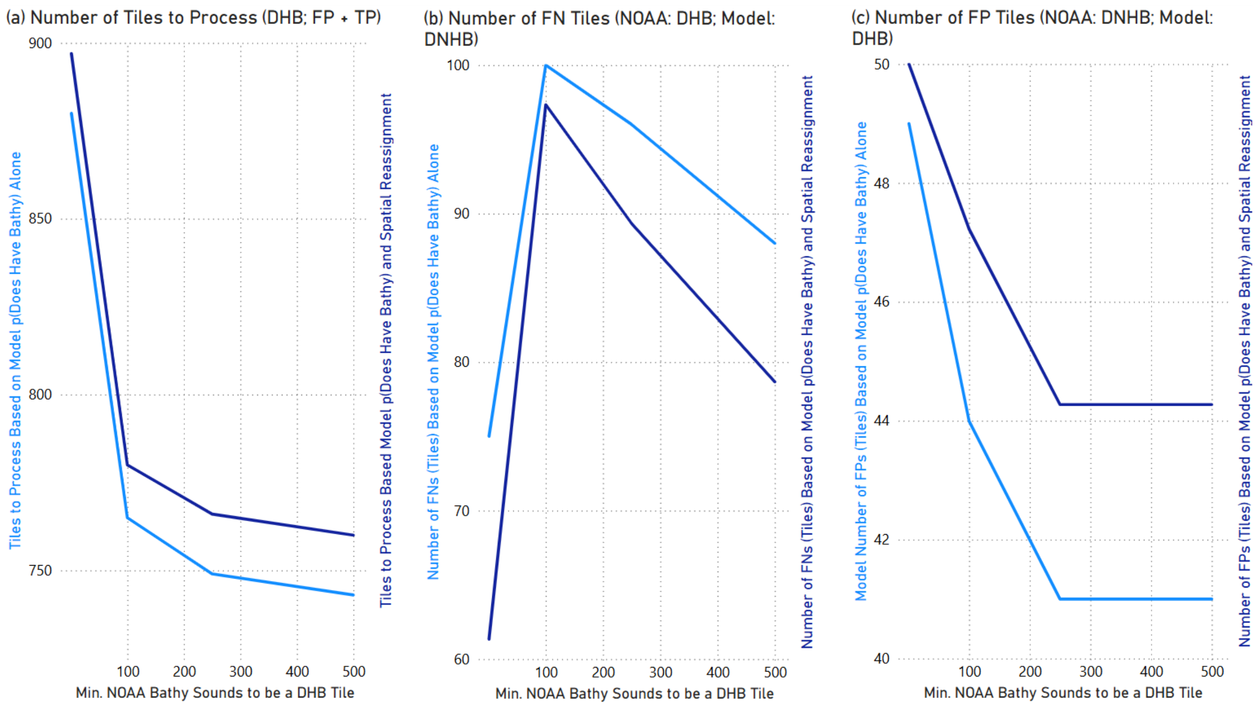

Figure 5. Trade-offs for Key West data and models between (*a*) the number of tiles processed and the (*b*) FN and (*c*) FP error rates.

Combining the results of logistic regression with spatial reassignment had a moderate impact on results. The number of tiles to process increased by about 20 tiles across all PRTs (about 2%), FNs decreased by about 10% and FPs increase by about 8%. Despite these modest gains, spatial reassignment is completely algorithmic and requires no human time and very little machine time (seconds per tile) making it an operationally viable step in an *a priori* tile screening workflow.

Concerning the broader applicability of the model form, Table 3 summarizes the logistic models fitted to the Miami Beach data set. Comparison with comparable information provided for the Key West models presented in Table 2 indicates comparable performance across all PRTs. These results suggest that the logistic model form developed using Key West depth frequency distributions is applicable to other areas for *a priori* screening to identify tiles that are likely to contain bathymetric pulse returns and therefore should progress to bathymetric processing of individual pulse returns.

Table 3. Details of logistic models fitted for Miami Beach.

Soundings Threshold	Pseudo R^2	Global Accuracy	F1 Score		Variable Importance / Coeff. Sign		
			Does Not Have *Bathy*	Does Have *Bathy*	Most	Mid	Least
1	0.78	0.96	0.93	0.97	Skewness / -	Depth std. dev. / +	Dip / +
100	0.87	0.97	0.94	0.98	Depth std. dev. / +	Skewness / -	Dip / +
250	0.93	0.98	0.96	0.98	Skewness / -	Depth std. dev. / +	Dip / +
500	1.00	1.00	1.0	1.0	Skewness / -	Depth std. dev. / +	Dip / +

That the model form may be broadly applicable for *a priori* screening does not mean the same for the calibrated model. Table 4 indicates that coefficient values for each of the variables are generally not significantly different for the models for the two data sets. However, skewness shows a clear difference – both in value and sign. This results from a lack of variability in skewness for the Miami Beach LiDAR tiles whose distributions are only either narrow and highly peaked (DNHB) or clearly bimodal -- i.e., tiles that clearly are DHB tiles.

Table 4. Coefficient values for Key West (KW) and Miami Beach (MB)models. Bold *p* values indicate significantly different coefficients at a = 0.05.

Thresh-old	Intercept			Skewness			Dip			Depth Std. Dev.		
	KW	MB	*p*	KW	MB	*p*	KW	MB	*p*	KW	MB	*p*
1	-4.1	-9.9	0.33	0.041	-0.122	**0.001**	152.	137.	0.95	7.7	5.7	0.42
100	-4.5	-13.0	0.28	0.021	-0.206	**0.006**	137.	60.	0.79	5.8	10.3	0.28
250	-4.9	-22.8	0.19	0.018	-0.285	**0.039**	128.	248.	0.77	6.5	13.1	0.33
500	-5.3	-373.1	0.99	0.018	-6.29	0.99	136.	1593.	0.99	6.8	198.3	0.99

Given the significant differences for the skewness coefficients, and the lack of variability in frequency distributions for Miami Beach, it is not surprising that the Key West models perform poorly when applied to the Miami Beach data (Fig. 6). Not only is global accuracy relatively low compared to the application of the Key West and Miami Beach models to the data from which each model was developed (see Tables 2 and 3), but F1 values for the DNHB tiles are much lower than the F1 values for the DHB tiles. This suggests an overabundance of tiles designated as DHB at the "expense" of relatively few DNHB tiles being correctly designated. Examination of individual confusion matrices confirmed that this was due to the Key West model classifying about 95% of Miami Beach tiles as DHB whereas NOAA classified only 68% of Miami Beach tiles as DHB. Figure 6 also indicates that spatial reassignment has minimal impact on these results.

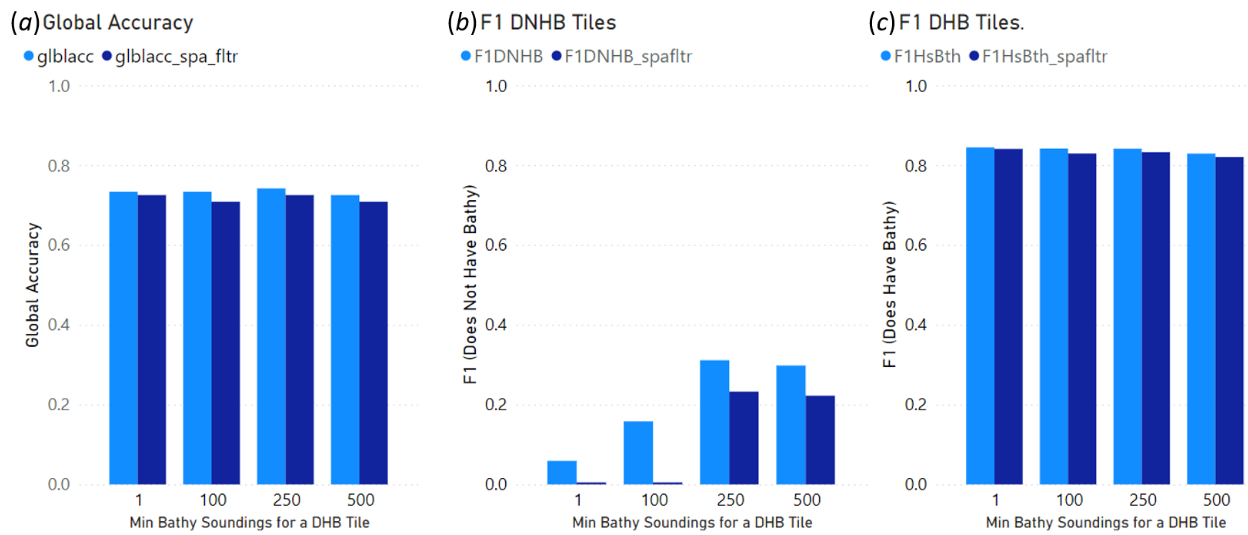

Figure 6. Accuracy results from the application of the Key West logistic model to the Miami Beach data set. (Legend: Light blue bars are based on the *p*(DHB) from the logistic regression models alone. Dark blue bars are based on the model *p*(DHB) and spatial reassignment.)

It is therefore concluded that though the model form developed may be broadly applicable, this approach to *a priori* screening will require local calibration of such models.

3. Discussion

In this study NOAA's processing of LiDAR tiles was used as the high-quality reference data set. Consequently, many results and their interpretation implicitly assume that NOAA's results have a high degree of accuracy – most critically for tiles on which bathymetric pulse returns were "rare".

This was evaluated by independently processing a limited sample of tiles having fewer than 100, 250, and 500 bathymetric pulse returns according to NOAA SOPs. The independent processing was done using a two-stage machine learning-based algorithm known as "CHRT-ML."; CHRT (CUBE[2] with Hierarchical Resolution Techniques) is an algorithm developed by Calder and Rice (2017) for processing sonar data. It was adapted to extract bathymetric soundings from LiDAR point clouds by Lowell and Calder (2022) through the use of machine learning (ML) clustering techniques. CHRT-ML employs a completely different approach to identifying bathymetric pulse returns in LiDAR point clouds than NOAA's SOPs (Nagle and Wright 2016). As expected, the two methods had the greatest disagreement for those tiles on which bathymetric pulse returns were rarest. However, the bathymetric pulse returns identified by NOAA SOPs and CHRT-ML on the sample tiles were sufficiently similar to provide broad confidence in NOAA's processing.

A major goal of this work was to evaluate if quantifiable characteristics of depth frequency histograms could accurately identify airborne LiDAR tiles that do not contain extractable bathymetric pulse returns. Results suggest that this is true. To provide a time-based estimate of the potential savings, experience with bathymetric processing using CHRT-ML was employed – recognising, of course, that each organisation's set of SOPs will require different amounts of time. CHRT-ML required approximately 1 hour per tile with 20 minutes being human time and 40 minutes being computer processing time. Results for the Key West data set indicated that the proposed approach would reduce the number of tiles that would be processed from 1374 tiles (that do not contain land -- i.e., have above MSL areas) to 880 tiles. This suggests a reduction of time of 494 hours of which 330 hours (13.75 24-hour days assuming batch processing) are machine/computer time and 164 hours (20.6 8-hour workdays) are human time. It is acknowledged that these estimates do not include the time cost associated with conducting the *a priori* screening. However, this time cost would be minimal since the process – gathering information on pulse return frequency distributions, fitting a predictive model, and identifying tiles that are unlikely to have bathymetric pulse returns -- is completely automated. Manual verification of results would be advised including considering factors such as location of navigable channels; this would increase the time cost of the *a priori* screening.

[2] Combined Uncertainty and Bathymetry Estimator (Calder and Meyer 2003)

A second major goal of this work was to evaluate the trade-offs between processing effort and the number of resultant FNs and FPs in order to improve operational bathymetric processing of LiDAR tiles. This was explored explicitly through the use of four PRTs. If one prioritizes minimizing the processing resources required, one can employ a higher PRT which effectively eliminates from processing tiles having low numbers of bathymetric pulse returns from processing. The consequence of this is an increase in FN and FP errors. This could potentially be offset, however, by changing the probability decision threshold (PDT) applied to *p*(DHB) values produced by the logistic model. Conventionally in the use of logistic regression, a PDT of 0.5 is employed to assign observations to one of two classes. However, the number of FNs, for example, might be decreased by reducing the PDT – although this would likely cause an increase in the number of FPs. And it would undoubtedly have the impact of necessitating processing of a greater number of tiles. A quantitatively optimal PDT could be identified using a receiver operating characteristic (ROC) curve (Nahm 2022) that shows the global accuracy performance of a binary classification model over all PDTs. However, in the case of identification of bathymetric pulse returns, the consequences of FNs (loss of data) and FPs (unnecessary processing) are organisationally quite different making global accuracy an inappropriate metric. Hence there is little choice in this case but for an organisation to set its PDT based on its qualitative tolerance to risk. In operational contexts, this would change with the area of interest, the targeted end-user, model performance, and other factors.

The previous point emphasizes the need for careful evaluation of the consequences of decreasing the number of tiles processed thereby changing the number of FNs and FPs that occur. In ocean mapping for navigation, for example, FN errors are considered more serious than FPs. FPs represent a “waste” of resources due to “needless” processing of tiles that are unlikely to contain bathymetric pulse returns. Conversely, FNs are a “failure” to capture data that could extend the geographical extent and accuracy of hydrographic maps. Alternatively, in clear water, FNs are likely to occur in the deepest parts of a survey area where sonar-equipped ships may be able to traverse. Hence to optimise operational processing of LiDAR data, one must consider the value of the information produced, the ability to obtain data using an alternative method, the types of errors likely to result, and the consequences of each type of error as well as the processing resources expended.

Results indicate clearly that a logistic regression model fitted for “Area A” (i.e., Key West) is unlikely to be applicable to “Area B” (Miami Beach). Potentially, this could be addressed three ways in operational contexts.

First, the characteristics of the two areas and depth frequency histograms could be examined for similarity. In this study, the Miami Beach data set clearly had less variance and different histogram forms than the Key West study site. This alone indicated potential difficulties in applying the Key West model to the Miami Beach data set.

Second, a machine learning method whose underlying approach is different from logistic regression might be employed. For this study, though logistic regression was employed because of superior results, classification and regression tree (CART) modelling was also evaluated (Breiman *et al.* 1984). Whereas logistic regression focusses on general trends across all variables, CART models “micro trends” by progressively splitting the data into branches and sub-branches based on optimal variables and split points at each step. Though prone to overfitting, it captures non-linear relationships in a way that logistic regression cannot. CART was examined using multiple PRTs, variables, optimization criteria, and maximum number of branches. No CART models performed better than the logistic regression models. Nonetheless, there are other machine learning techniques such as neural networks and support vector machines that are not decision-tree based that were not examined in this study. These could be of interest in part because logistic regression – that models broad trends – did not produce models that were geographically robust. Machine learning techniques other than CART are less likely to produce geographically robust models. However, such techniques might produce models that are better for a single area such as the Key West or Miami Beach areas employed in this study. Note that the risk of model overfitting would increase given the underlying approach to model development of many machine learning techniques.

Third, “progressive sampling” could be employed. Suppose one has 1000 LiDAR tiles to process. One could sample a “representative” number – e.g., 50 tiles -- according to an appropriate scheme – e.g., randomly, weighted by distance from land or likely depth. Each tile in the sample would be processed for bathymetric pulse returns according to normal SOPs. A logistic model would be fitted using the sample, and global accuracy and other accuracy metrics calculated for the 50 tiles. An additional sample of 50 tiles would be taken (without-replacement), processed for bathymetric pulse returns and added to the original sample. A new logistic regression model would be fitted using the 100 tiles, and the selected accuracy metric(s) calculated for the 100 tiles. This process would continue for 150, 200, etc. tiles until the accuracy metric(s) converged – i.e., stabilised. That is, it is expected that the accuracy metric(s) would change "considerably" with each additional sample of 50 tiles at lower sample sizes. However, sampling an increasing number of tiles would increase the representativeness of the tiles sampled causing the accuracy metric(s) to change little with each additional sample of 50 tiles. Once this occurred, the final model fitted would be applied to all 1000 tiles. If 250 tiles (five samples of 50 each) were required to achieve convergence, undoubtedly a certain number would be tiles having no bathymetric pulse returns and would have been "needlessly" processed. However, this approach would identify tiles in the unsampled 750 tiles that do not require processing for bathymetric pulse returns thereby reducing the number of tiles requiring processing.

Finally, a critical part of operational implementation of any *a priori* screening scheme should be the design and implementation of a quality assurance/continuous improvement (QA/CI) program. Results of this study demonstrated that though one can eliminate tiles from "needless" processing, there is a consequent increase in errors that would need to be continually monitored. Each of the three potential operational winnowing strategies would base a tile's "Does/Does Not Have Bathy" decision on a model's estimate of the probability that an unprocessed tile does have bathymetric pulse returns. This could be checked by actually processing a certain number or percentage of randomly chosen "QA/CI tiles" and evaluating if the error rate was consistent with the probabilities; Lowell and Mitchell (1987) described one analytical approach for doing this. Notably, however, aside from the model-based probability, as discussed, the amount of screening to be applied must be decided based on organizational consequences of errors and risk tolerance. An informed decision can only be made if one constantly monitors the screening process including evaluating model goodness-of-fit, magnitude and type of errors, spatial characterization of uncertainty, and operational factors such as time and cost.

4. Conclusions

With an overarching operational goal of reducing processing time and resources, this study demonstrated the viability of using a frequency distribution-based approach to identifying LiDAR tiles whose processing is unnecessary. It also quantified the trade-off between reducing resources required to process LiDAR point clouds to extract bathymetric pulse returns and the consequent increase in errors. For example, by developing and applying a logistic regression model based on quantifiable characteristics of depth frequency histograms, the number of tiles processed for a LiDAR data set for Key West (Florida, United States) could be reduced from 1374 to 900 (34%) with an increase in false negative (FN) tiles from 0 (zero) to 75 tiles (5% of total) and false positive (FP) tiles from 0 (zero) to 48 (4% of total). Reassigning FN and TN tiles from "Does Not Have Bathy" to "Does Have Bathy" based on the *p*(DHB) values of each tile's neighbours reduced the number of FN tiles from 75 to 61 (4% of total) while increasing the number of FP tiles from 48 to 49 (4% of total). The use of "pulse return thresholds" to eliminate from processing the tiles on which bathymetric pulse returns were "rare" was found to reduce the number of tiles requiring processing and the number of FN and FP tiles only modestly. Finally, it was concluded that application of the method developed to other data sets will require the fitting of geographic- and water-condition-specific logistic regression models.

This work is novel in that it is not focused on eliminating or de-noising points in individual LiDAR tiles being processed. Instead, it demonstrates that it is possible to effectively de-noise a set of tiles/areas and eliminate those whose bathymetric processing is unlikely to extract pulse returns/soundings that represent the depth of the ocean floor.

Acknowledgement

The author gratefully acknowledges financial support from the National Oceanic and Atmospheric Administration (NOAA) Grant NA20NOS4000196 and logistical support from NOAA's Remote Sensing Division.

Disclosure of Interest

The author reports no conflict of interest.

References

Agrafiotis, P., Skarlatos, D., Georgopoulos, A., Karantzalos, K. 2019. "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: 2225, 31 pp. DOI: 10.3390/rs11192225.

ASPRS 2019. *LAS Specification Version 1.4-R15*, Bethesda, Maryland, United States: American Society for Photogrammetry and Remote Sensing.

Balanda, K., MacGillivray, H. 1988. "Kurtosis: a Critical Review." *The American Statistician* 42(2): 111.

Breiman, L., Friedman, J., Olshen, R., Stone, C. 1984. *Classification and Regression Trees.* New York, NY, United States: Chapman and Hall/CRC.

Calder, B., Meyer, L., 2003. "Automatic processing of high-rate, high-density multibeam echo sounder data." *Geochemistry, Geophysics, Geosystems* 4: 6, 22 pp. DOI: https://doi.org/10.1029/2002GC000486.

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

Cavanaugh, J., Neath, A. 2019. "The Akaike Information Criterion: Background, Derivation, Properties, Application, Interpretation, and Refinements." *WIREs Computational Statistics* 11(3): e1460, 11 pp., DOI: https://doi.org/10.1002/wics.1460.

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: doi.org/10.1016/j.rse.2017.12.035.

Giannakaki, E., Kokkalis, P., Marinou, E., Bartsotas, N., Amiridis, V., Ansmann, A., Komppula, M 2020. "The Potential of Elastic and Polarization Lidars to Retrieve Extinction Profiles" *Atmospheric Measurement Techniques* 13(2):893–905, DOI: https://doi.org/10.5194/amt-13-893-2020.

Hartigan, P. 1985. "Computation of the Dip Statistic to Test for Unimodality." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 34(3): 320-325.

Hartigan, J., Hartigan, P. 1985. "The Dip Test of Unimodality." *The Annals of Statistics* 13(1): 70-84.

Joanes, D., Gill, C. 1998. "Comparing Measures of Sample Skewness and Kurtosis." *The Statistician* 47(1): 183–189.

Jung, J., Lee, J., Parrish, C. 2021. "Inverse Histogram-based Clustering Approach to Seafloor Segmentation from Bathymetric Lidar Data." *Remote Sensing* 13(18): 3665 15 pp. DOI: https://doi.org/10.3390/rs13183665.

Krekov, M., Krekova, M., Shamanaev, S. 1997. "Laser Sensing of a Subsurface Oceanic Layer. I. Effect of the Atmosphere and Wind-driven Sea Waves." *Applied Optics* 36: 1589–1595. DOI: https://doi.org/10.1364/AO.37.001589.

Krekov, G., Krekova, M., Shamanaev, S. 1998. "Laser Sensing of a Subsurface Oceanic Layer. II. Polarization Characteristics of Signals." *Applied Optics* 37(9): 1596-1601. DOI: https://doi.org/10.1364/AO.37.001596.

Le, M-H., Cheng, C-H., Liu, D-G., Nguyen, T-T., 2022. "An adaptive group of density outlier removal filter: snow particle removal from LiDAR data." *Electronics* (*Topic:Artificial Intelligence in Sensors)* 11:2993, 18 pp. DOI: https://doi.org/10.3390/electronics11192993

Lisenko, A., Shamanaev, V. 2022. "Experimental Determination of the Sea Water Extinction Coefficients in Inhomogeneous Subsurface Sater Layers from the Depth Profiles of the Airborne Lidar Return Signal Power." *Proceedings: SPIE 12341, 28th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics 123414H*, edited by G. Matvienko, O. Romanovskii (7 December 2022; Tomsk Russia). DOI: https://doi.org/10.1117/12.2644800.

Lowell, K., Calder, B. 2021. "Extracting Shallow-water Bathymetry from Lidar Point Clouds Using Pulse Attribute Data: Merging Density-based and Machine Learning Approaches." *Marine Geodesy* 44(4):259-286. DOI: https://doi.org/10.1080/01490419.2021.1925790.

Lowell, K., Calder, B. 2022. "Operational Performance of a Combined Density- and Clustering-based Approach to Extract Bathymetry Returns from LiDAR Point Clouds." *International Journal of Applied Earth Observation and Geoinformation (Special Issue: Recent Advances in Geocomputation and GeoAI for Mapping)* 107: 102699. DOI: https://doi.org/10.1016/j.jag.2022.102699.

Lowell, K., Calder, C., Lyons, A. 2021. "Measuring shallow-water bathymetric signal strength in lidar point attribute data using machine learning." *International Journal of Geographical Information Science* 35(8):1592-1610. DOI:10.1080/13658816.2020.1867147.

Lowell, K., Mitchell, R., 1987. "Stand growth projection: simultaneous estimation of growth and mortality using a single probabilistic function." *Canadian Journal of Forest Science* 17(11): 1466-1470. DOI: https://doi.org/10.1139/x87-2.

Mahalanobis, P., 1936. "On the generalized distance in statistics." *Proceedings of the National Institute of Science of India* 2: 49-55.

Mandlburger, G.; Jutzi, B., 2019. "On the Feasibility of Water Surface Mapping with Single Photon Lidar." *ISPRS International Journal of GeoInformation* 8(4): 188. 23 pp. DOI https://doi.org/10.3390/ijgi8040188.

Mason, G., Sibson, R., 1992. “Measuring Multimodality.” *Statistics and Computing* 2: 153-160.

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.

Matken, A., Hajeb, M., Mirbagheri, B., Sadeghian, S., Amadi, M., 2014. “Spatial analysis for outlier removal from LiDAR data.” *International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences* XL-2/W3: 187–190 DOI: https://doi.org/10.5194/isprsarchives-XL-2-W3-187-2014

Nagle, D., Wright, C., 2016. *Algorithms used in the Airborne Lidar Processing System (ALPS)*. Reston, Virginia, United States: Geological Survey (USGS) Open File Report 2016-1046. U.S. Department of the Interior.

Nahm, F., 2022. “The receiver operating characteristic curve: overview and practical use for clinicians.” *Korean Journal of Anaesthesiology* 75(1): 25-36. DOI: https://doi.org/10.4097/kja.21209

NOAA 2024. Data Access Viewer. https://www.coast.noaa.gov/dataviewer/#/ Last visited March 6 2024.

Parker, H., Sinclair, M. 2012. “The Successful Application of Airborne LiDAR Bathymetry Surveys Using Latest Technology.” *Proceedings: 2012 Oceans – Yeosu* (21-24 May, Yeosu, South Korea), 1-4. DOI: 10.1109/OCEANS-Yeosu.2012.6263588.

Parrish, C., Rogers, J., Calder, B. 2014. “Assessment of Waveform Features for Lidar Uncertainty Modelling in a Coastal Salt Marsh Environment.” *IEEE Geoscience and Remote Sensing Letters* 11(2): 569-573. DOI: 10.1109/LGRS.2013.2280182.

Qi, Y., Dou, H., Wang, Z. 2022. “An Adaptive Threshold Selected Method from Remote Sensing Based on Water Index.” *Journal of Physics: Conference Series* 2228 6 pp. DOI: 10.1088/1742-6596/2228/1/012001.

Qeunouille, M., 1956. “Notes on bias in estimation” *Biometrika* 43(3/4): 353-360. DOI: https://doi.org/10.2307/2332914.

Ranndal, H., Christiansen, P., Kliving, P., Andersen, O., Nielsen, K. 2021. “Evaluation of a Statistical Approach for Extracting Shallow Water Bathymetry Signals from ICESat-2 ATL03 Photon Data.” *Remote Sensing* 13: 3548, 19 pp., DOI: https://doi.org/10.3390/rs13173548.

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.

Szutor, P., Zichar, M., 2023. “Fast radius outlier filter variant for large point clouds.” *Data (Section Information Systems and Data Management)* 8(10): 149, 13 pp. DOI: https://doi.org/10.3390/data8100149.

Wen, Z., Zhang, C., Shao, G., Wu, S., Atkinson, P. 2021. “Ensembles of Multiple Spectral Water Indices for Improving Surface Water Classification.” *International Journal of Applied Earth Observation and Geoinformation* 96: 102278, 13 pp. DOI: https://doi.org/10.1016/j.jag.2020.102278.

Yang, A., Wu, Z., Yang, F., Dianpeng, S., Yue, M., Zhao, D., Qi, C. 2020. “Filtering of Airborne LiDAR Bathymetry Based on Bidirectional Cloth Simulation.” *ISPRS Journal of Photogrammetry and Remote Sensing* 163, 49–61. DOI: https://doi.org/10.1016/j.isprsjprs.2020.03.004.

Youden, W., 1950. “Index for rating diagnostic tests.” *Cancer* 3, 32-35. DOI: https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.

Appendix 1. Acronyms.

Acronym	Definition
ALB	Airborne LiDAR for Bathymetry
Bathy/NotBathy	A designation for individual pulse returns indicating that an individual pulse return does or does not represent bathymetry according to NOAA standard operating procedures (SOPs).
CHRT	An algorithm developed for sonar processing.
CHRT-ML	A machine learning algorithm that applies clustering to information produced by CHRT to identify bathymetric pulse returns in LiDAR point clouds.
DHB/DNHB	Designation for a tile that does have (DH) or does not have (DNHB) bathymetric pulse returns. For model fitting, this designation is based on bathymetric processing using NOAA standard operating procedures. For determining a tile's model-based designation, this is based on p(DHB) -- the probability that a tile does have bathymetric pulse returns.
FN	False Negative. A tile designation that indicates that NOAA SOPs found bathymetric pulse returns to be present, but that a logistic regression model identified as having no bathymetric pulse returns present.
FP	False Positive. A tile designation that indicates that NOAA SOPs determined that bathymetric pulse returns were not present, but that a logistic regression model identified as having bathymetric pulse returns present.
NOAA	United States National Oceanic and Atmospheric Administration
p(DHB)/p(DNHB)	Probability that a tile does have (DH) or does not have (DNH) bathymetric pulse returns according to a logistic regression model fitted to a particular data set.
PDT	Probability Decision Threshold. The minimum model probability (p(DHB))
PRT	Pulse Return Threshold. The minimum number of bathymetric pulse returns required to designate a tile as "Does Have Bathy" (DHB).
SOPs	Standard Operating Procedures
TN	True Negative. A tile designation indicating that NOAA SOPs and a logistic regression model agreed that no bathymetric pulse returns were present.
TP	True Positive. A tile designation indicating that NOAA SOPs and a logistic regression model agreed that bathymetric pulse returns were present.