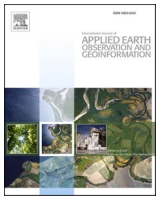
International Journal of Applied Earth Observations and Geoinformation 107 (2022) 102699

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



Operational performance of a combined Density- and Clustering-based approach to extract bathymetry returns from LiDAR point clouds

Kim Lowell *, Brian Calder

Center for Coastal and Ocean Mapping and Joint Hydrographic Center, University of New Hampshire, 24 Colovos Road, Durham, NH 03824 United States

ARTICLE INFO

Keywords:
K-means clustering
Machine learning
Florida Keys
Coral reefs
Shallow water bathymetry

ABSTRACT

Conventional techniques for extracting bathymetric soundings from LiDAR point clouds are at best semi-automated and require considerable manual effort. An algorithm that couples a widely used sonar data processing method with a newly developed machine-learning(ML)-based algorithm was evaluated for accuracy and potential operationalisation. Data representing an operationally realistic range of environmental and data conditions comprised 103 500 m-by-500 m data tiles for method development/calibration and 20 tiles for validation located in the Florida Keys. Tiles are processed individually to classify each LiDAR pulse return ("sounding" in hydrographic terminology) as bathymetry or not. Compared to a reference classification an average agreement of about 90% was produced for the calibration and validation data sets, and accuracy varied depending on ocean bottom and data conditions. The average false negative rate – the most important metric in hydrographic mapping – was about 5%. Processing time for tiles containing the average number of soundings (seven million) on a desktop computer was approximately 100 min. The algorithm does not require *in situ* ground-"truth" data for training or calibration, although its adaptation to other geographic and data conditions might require data-guided adjustment of ML tuning parameters.

1. Introduction

The capture of LiDAR point clouds from airborne platforms and associated data volumes continue to grow. Accordingly, the need to replace processing methods reliant on manual input with fully automated approaches is increasing. A familiar example is the processing of LiDAR data to produce terrestrial digital elevation/terrain models (DEMs/DTMs) (e.g., Xiangyun and Yi 2016, Wang et al. 2017, Zhang et al. 2020). Producing "underwater DEMs" – i.e., mapping bathymetry – is a natural extension of such work particularly for shallower navigable areas. However, mapping bathymetry using LiDAR is less explored than terrestrial DTM production. It also presents different challenges such as the impacts of sea surface roughness on light reflectance and depth-related reduction in light penetration.

LiDAR point clouds comprise three-dimensional {x, y, z} coordinates of individual pulse returns and associated metadata. (See also .las file data standards in ASPRS (2013)). Processing pulse returns for bathymetry means separating those that represent the ocean floor from those representing the ocean surface or noise in the water column. Artificial intelligence (AI) and machine learning (ML) approaches are among recent methods that have been explored to "de-noise" or "filter" LiDAR point clouds for various applications – e.g., analysing statistical characteristics of clusters developed after an initial principal components analysis (PCA) (Duan et al. 2021), coupling auto-labeling with convolutional neural networks (CNNs) (Henzler *et al.* 2020), using a statistical "cell histogram" outlier removal approach (Carrilho et al. 2018), or extending a two-dimensional cell histogram approach to three-dimensional "voxels" (Yong-hua et al. 2017).

Improving filtering of acoustic/sonar point cloud filtering to produce maps/charts that meet international standards (IHO 2020) is an ongoing activity of the hydrographic mapping community. A recent example is a voxel-based approach underpinned by CNNs for which a 97% *Bathy/*

Abbreviations: AI, Artificial Intelligence; BDI, Bathymetric Depth Interval; CHRT, Cube with Hierarchical Resolution Techniques (Calder and Rice 2017); CI, Confidence interval; CNN, Convolutional Neural Network; DEM/DTM, Digital Elevation/Terrain Model; EN, Estimation Node (established by the CHRT algorithm); FPR/FNR, False positive(*Bathy*)/negative(*NotBathy*) rate; LiDAR, Light Detection And Ranging; ML, Machine Learning; MLD, Most Likely Depth of an estimation node (EN); MSL, Mean Sea Level; NOAA, National Oceanic and Atmospheric Administration; PAB, Producer's Accuracy for *Bathy* (This is the same as the TPR.); PCA, Principal Components Analysis; TPR/TNR, True positive(*Bathy*)/negative(*NotBathy*) rate; UAB, User's Accuracy for *Bathy*.

* Corresponding author.
E-mail addresses: klowell@ccom.unh.edu (K. Lowell), brc@ccom.unh.edu (B. Calder).

https://doi.org/10.1016/j.jag.2022.102699
Received 3 December 2021; Received in revised form 14 January 2022; Accepted 24 January 2022
Available online 3 February 2022

NotBathy classification accuracy was reported (Stephens et al. 2019). (See also the review article and associated bibliography of Le Deunf et al. 2020.) Such methods are not readily adapted to LiDAR point clouds, however, due to factors such as a lower signal-to-noise ratio of LiDAR data. Yet the value of operationally viable methods for filtering LiDAR point clouds is immense given the inability of sonar-equipped ships to operate in "shallow-water" – depths less than maximum LiDAR depth penetration (17 m in this study although LiDAR penetration to depths of 80 m has been reported (Parker and Sinclair 2012)).

An early approach for shallow-water bathymetric LiDAR processing combined LiDAR point clouds and multispectral scanner data (Lyzenga 1985). More recently other methods have been explored including interpolation (kriging) for mapping underwater archaeological structures (Doneus et al. 2013, 2015), geomorphometric analysis (Yang et al. 2020), and ML modelling (extreme gradient boosting modelling (Friedman 2001)) of sounding metadata (Lowell et al. 2020).

Such proof-of-concept research has understandably focused on methodological accuracy. Consequently, three important operational issues have received limited attention: methodology robustness, processing time, and the need for *in situ* training/calibration data. In particular, the need for independent *in situ* data increases processing costs, decreases efficiency, and limits the use of LiDAR point cloud data in remote or inaccessible areas. (Exceptions to methods requiring *a priori* calibration data for LiDAR and shallow-water bathymetry are Yang et al. (2020), Lowell and Calder (2021), and Ranndal et al. 2021 for ICESat-2 data.)

This article presents a new method for processing LiDAR point clouds for shallow-water bathymetry that does not require an *in situ* calibration data set and documents its performance relative to operational considerations. The fundamental motivation of this research is to bridge the gap that exists between proof-of-concept and operational LiDAR point cloud processing methodologies.

2. Study area and data

LiDAR data were acquired by the United States National Oceanic and Atmospheric Administration (NOAA) from April 22 to-26 in 2016 in the vicinity of Key West, Florida (United States) (Fig. 1a). Data were acquired using a Riegel™ VQ-880-G instrument from a nominal altitude of 400 m and speed of 200 km hour^{-1}. The sensor emits green and near-infrared pulses (532 nm and 1064 nm, respectively). The pulse frequency of 45,000 s^{-1} results in a point density of about 10 soundings m^{-1} for a single flight path. Acquired originally in generally north–south overlapping swaths, data were subsequently "cut" into 1620 500-m-by-500-m tiles. Each was processed by NOAA to classify soundings as *Bathy/NotBathy* using a multi-stage processing workflow. Primary processing filters point clouds using algorithms based on principles described in, for example, Nagle and Wright (2016). Secondary processing relies on computer-aided manual editing to separate terrestrial areas and ocean surface soundings from soundings representing bathymetry. NOAA determined that approximately 500 tiles were located in areas whose depth exceeded LiDAR penetration; these were removed from further consideration. Note that this means that an initial step in operationalizing the algorithm presented is the elimination of "too deep" tiles; this might be achieved by, for example, analysing sounding depth frequency distributions.

The remaining tiles cover a range of ocean conditions likely to be present in operational LiDAR bathymetric surveys – varying substrate, working harbors, and a range of depths. They also represent a range of data conditions (Fig. 1b, c, and d) – e.g., varying data densities, partial absence of soundings in areas where depth exceeded LiDAR penetration capabilities (about 17 m in this study). Randomly selected for analysis were 103 tiles – 100 in addition to the original four tiles with one tile belonging to both sets – (Fig. 1). These tiles were (subjectively) judged to sufficiently address the range of ocean and data conditions present.

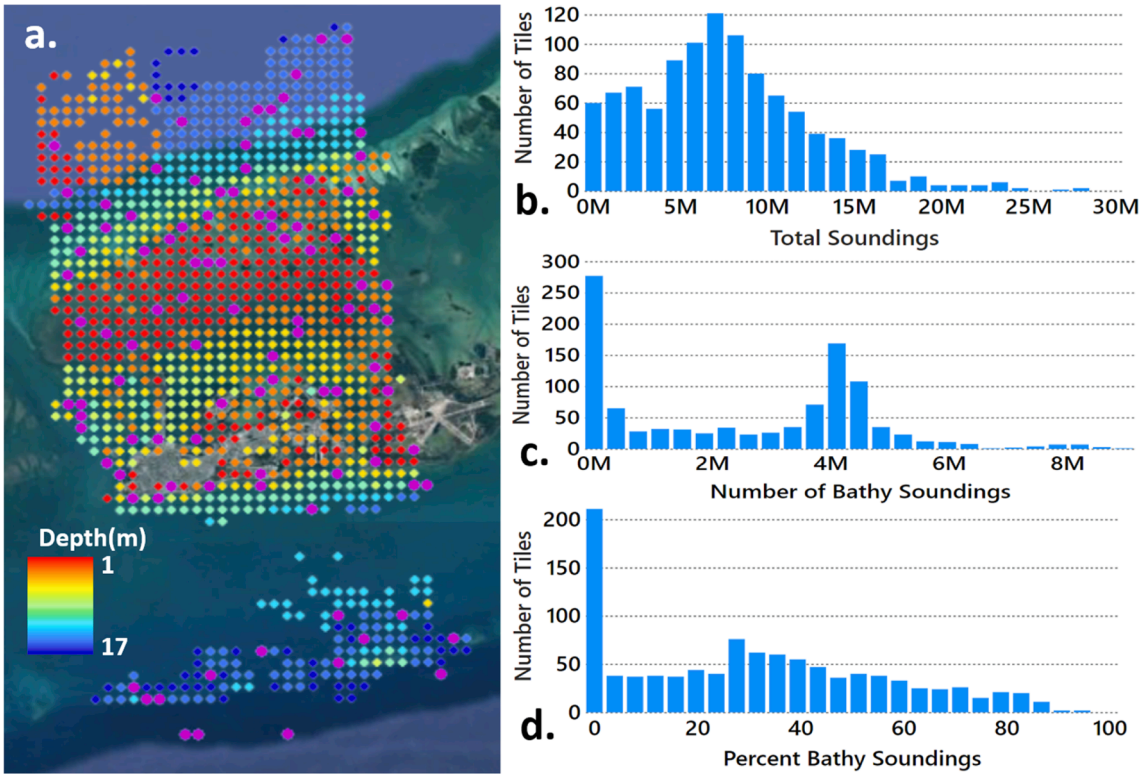


Fig. 1. a. Centres of available LiDAR data tiles overlain on GoogleEarth™ imagery. Purple dots indicate centres of the 103 tiles analysed. The Key West airport is visible in the centre-east (approximately 24.57°N / 81.70°W). Frequency distribution of tiles by b. total soundings, c. number of NOAA-classified *Bathy* soundings, and d. percent of soundings NOAA identified as *Bathy*.

3. Methods

The classifier ("CHRT-ML" for reasons that will be explained) evolved from the unsupervised metadata-based classifier described in Lowell and Calder (2021) that combined ML analysis of LiDAR sounding metadata with a density-based processing method. An average global classification accuracy of 93% (minimum of 84%/maximum of 99.99%) was reported for four data tiles of varying depth compared to the NOAA *Bathy/NotBathy* reference classification. However, expansion of the method to additional tiles indicated two major difficulties:

1. Speed: Processing time for tiles with seven million soundings – the average for tiles having *Bathy* soundings – was approximately 15 h on a desktop computer with an Intel® Xeon® W-2135 CPU @ 3.70 GHz, 3696 Mhz Processer and 32 Gb of RAM.
2. Lack of robustness primarily due to the presence of multiple distinct depth tiers on some tiles due to shallow coral reefs and/or areas cut by deeper channels as illustrated in Fig. 2.

The foundation of the new methodology described in this article and the original Lowell and Calder (2021) approach is the density-based Cube with Hierarchical Resolution Techniques (CHRT) algorithm (Calder and Rice 2017) that is widely employed operationally for sonar data. CHRT establishes a grid of "estimation nodes" (ENs) across an area at a spacing determined by the average sounding density. For each EN, soundings that are within half the distance to a diagonally contiguous EN – i.e., the radius that encloses an entire EN "pixel" (Fig. 3) – are extracted and progressively "ingested." The first sounding defines the initial depth "hypothesis" for the EN. Subsequent soundings ingested create a new hypothesis if their depths are statistically "too far" from the mean depth of existing hypotheses. Otherwise each is assigned to its closest (by depth) existing hypothesis. Disambiguation rules then identify the most likely depth (MLD) of the EN. This approach, whether applied to sonar or LiDAR data, essentially converts a set of geographically proximal soundings to the equivalent of a LiDAR waveform. In the production of terrestrial DEMs, the waveform is analysed to identify the

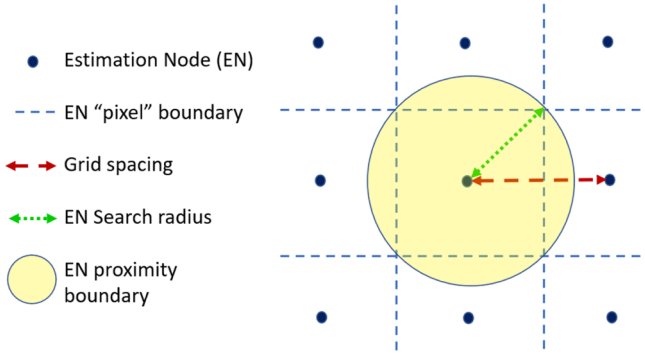


Fig. 3. Schematic for Estimation Nodes (ENs) for CHRT.

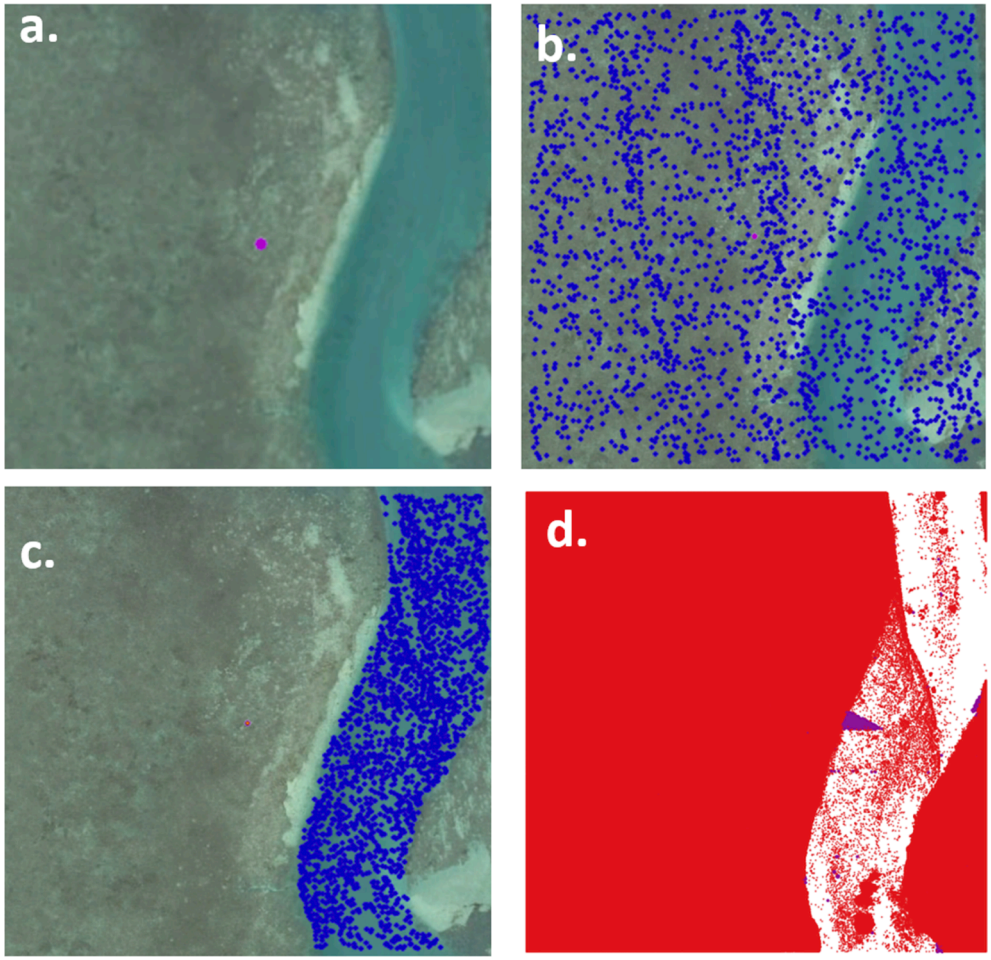


Fig. 2. Illustration of lack of robustness in original density-based/ML hybrid classifier described in Lowell and Calder (2021). a. GoogleEarth™ imagery. b. NOAA reference classification *Bathy* soundings (blue). c. *Bathy* soundings from the original classifier. d. *Bathy* false negatives (red) and false positives (purple). (For Fig. 2b and c, a random sub-sample of soundings is displayed to improve interpretability.)

"hypothesis" representing the ground rather than trees. Equivalently, for sonar or LiDAR bathymetric soundings, the goal is to identify the ocean bottom.

The MLD for an EN does not necessarily represent ocean depth. In the original Lowell and Calder (2021) method, the MLDs of the ENs were filtered and classified as *Bathy/NotBathy* using a workflow comprised of outlier screening, 2-cluster *k*-means clustering (Forgy 1965), and extreme gradient boosting modelling. When applied to the 103 tiles randomly sampled herein, the method's lack of robustness became apparent, and a new ML component was developed (hence the algorithm name "CHRT-ML").

Inherent in CHRT-ML are a number of tuning parameters (Table 1). Note that neither CHRT-ML nor its tuning parameters are "calibrated" using *in situ* data and numerical optimization of a global loss function using methods such as gradient descent (Curry 1944). Instead the tuning parameter values (Table 1) were determined by selecting a set of values, examining the results across all tiles and tiles, adjusting various values, re-examining results, etc. The sensitivity of CHRT-ML to tuning parameters was also determined via this process.

CHRT-ML processing begins with CHRT identifying the MLD for each EN. A two-stage outlier screening/filtering is then applied to the MLDs:

1. Gross: Removal of all ENs whose MLD is 3 m above or 20 m below mean sea level (MSL).
2. Refined: Calculation of Mahalanobis (Mahalanobis 1936) distance based on a variety of EN-related variables and elimination of ENs whose Mahalanobis distance exceeds the limits of the 99.9% confidence interval (CI).

Using the ENs retained, unsupervised 3-cluster *k*-means clustering is performed using two variables – the MLD and the average depth of the non-MLD hypotheses. A set of Bathymetric Depth Interval (BDI) rules is then applied to the characteristics of the clusters to define the BDI for a tile. A gaussian distribution of ENs in each cluster is assumed and the clusters are ordered by relative mean MLD depth – shallow, mid-depth, deep.

The deeper BDI limit is always defined as the lower limit of the 99.9% MLD CI of the deep cluster plus 1.0 m; adding 1.0 m was found to improve the *Bathy/NotBathy* classification of deeper geographically isolated *Bathy* soundings.

The shallower BDI limit is determined by "BDI rules" formulated as a decision-tree (Fig. 4). Initially it was assumed that the shallowest cluster would define the ocean surface, and the mid-depth and deepest clusters would define bathymetry. However, cluster characteristics indicated complex interactions among depth, geomorphometry, and the proportion of MLDs at different depths and other data characteristics. The BDI rules were thus defined subjectively based on these complex interactions and the ability to discern a reasonable geophysical or clustering methodology reason. The resulting BDI rules determine if MLDs representing bathymetry are present in one, two, or all three clusters and define the shallower BDI limit accordingly. The characteristics of each cluster employed are its 1) 99.9% MLD CI, 2) average MLD, and 3) average depth of non-MLD ENs. Initially (top green row of Fig. 4), the overlap among the MLDs of the three clusters is determined. Subsequently, clusters containing ENs whose MLD represents bathymetry are identified by assessing (lower green row of Fig. 5) if a cluster is "narrow" (i.e., well-defined) or "highly variable" as determined by the *Bathy/NotBathy* tolerance hyperparameter (Table 1). Table 2 presents details and rationale for each rule and Fig. 5 presents exemplar frequency distributions.

Finally, soundings for a tile whose depth falls within BDI are classified as *Bathy* and the other soundings classified as *NotBathy*.

Each tile's CHRT-ML classification was cross-tabulated with the NOAA reference classification and various statistics calculated to assess agreement:

- Global Accuracy (GA – % agreement over all classes)
- True Positive Rate/Producer's Accuracy for *Bathy* (TPR/PAB – Proportion of NOAA *Bathy* correctly identified)
- True Negative Rate (TNR – Proportion of NOAA *NotBathy* correctly identified)
- False Positive Rate (FPR – Proportion of NOAA *NotBathy* incorrectly identified as *Bathy*)
- False Negative Rate (FNR – Proportion of NOAA *Bathy* incorrectly identified as *NotBathy*)
- User's. Accuracy for *Bathy* (UAB – Proportion of CHRT-ML-classified *Bathy* that NOAA classified as *Bathy*.

The FNR is most important for real-world hydrographic charting. Whereas the TPR/PAB is a conventional measure of classification accuracy, the UAB measures the confidence that the soundings being used to develop a chart, for example, are truly *Bathy*. (See Congalton and Green (2019).)

Table 1
Tuning parameters for CHRT-ML.

Parameter	Value Employed	Description/Function	CHRT-ML Sensitivity
Minimum depth / Maximum depth in m	−1.5 (above sea level) / 20 m (below sea level)	Defines the depth interval in which individual true *Bathy* soundings can be expected to be found.	Low
Minimum number of MLD soundings	2	Eliminates unstable MLDs from clustering.	High
Minimum cluster overlap in m	0.02	Replaces an absolute definition of cluster "overlap" with a fuzzy tolerance.	Low
Bathy/NotBathy tolerance in m	0.4	Defines the minimum difference that must be observed between the mean depth of the MLD hypothesis and the non-MLD hypotheses to conclude *Bathy* ENs are present in a cluster.	Medium

4. Results

Because it became apparent that tile characteristics strongly impacted *Bathy/NotBathy* classification accuracy, each of the 103 tiles was denoted as being one of seven "types" (Fig. 6) based on the presence of certain features or dominance of certain substrates: 1)"Normal", 2) Reef and Channel, 3)Reef only, 4)Sparse Bathy (Less than 0.5% of total NOAA soundings are *Bathy*), 5)Infrastructure only (human infrastructure such as piers or docks present), 6)Channel only, and 7)Reef, Channel, & Infrastructure.

Individual examination of sparse *Bathy* tiles (orange dots in Fig. 7) indicated all were located in deep areas at the limits of LiDAR penetration. For all but one sparse bathy tile, all accuracy metrics were extreme – very high (near 100%/1.0) or very low (near (0%/0.0). This relates to two factors. First, in deeper areas sparse *Bathy* soundings are usually too spatially dispersed for CHRT to form a *Bathy* depth hypothesis that would be identified as the MLD for an EN. Second, even if some ENs have *Bathy* MLDs, they are not sufficiently numerous to form their own *k*-means cluster and are absorbed into a larger cluster determined to not contain *Bathy* MLDs. Effectively the BDI definition ignores the presence of *Bathy* MLDs and soundings. The result is that all *Bathy* soundings are either erroneously assigned to *NotBathy* (if the erroneous BDI is relatively narrow), or to *Bathy* along with many soundings that are actually *NotBathy* (if the erroneous BDI is relatively wide). The "too narrow" BDI produces high global accuracies, TNRs, FNRs, and UABs,

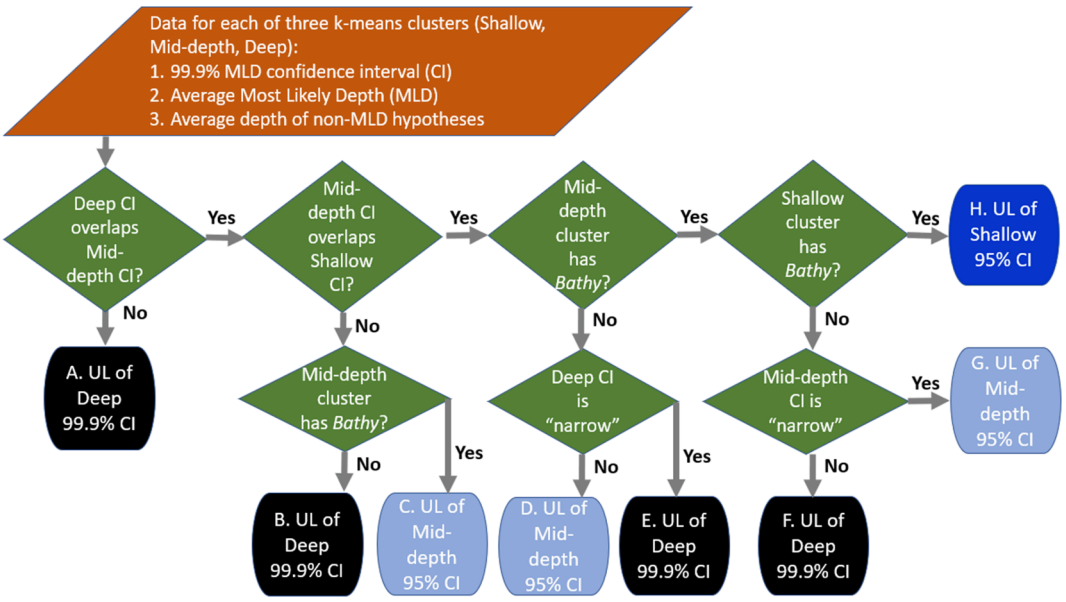


Fig. 4. Flow chart of decision rules to determine the shallow limit of the Bathymetric Depth Interval (BDI). "CI" is confidence interval. Colors in round-edged boxes represent the deep (black), mid-depth (light blue), and shallow (bright blue) clusters as shown in Fig. 5.

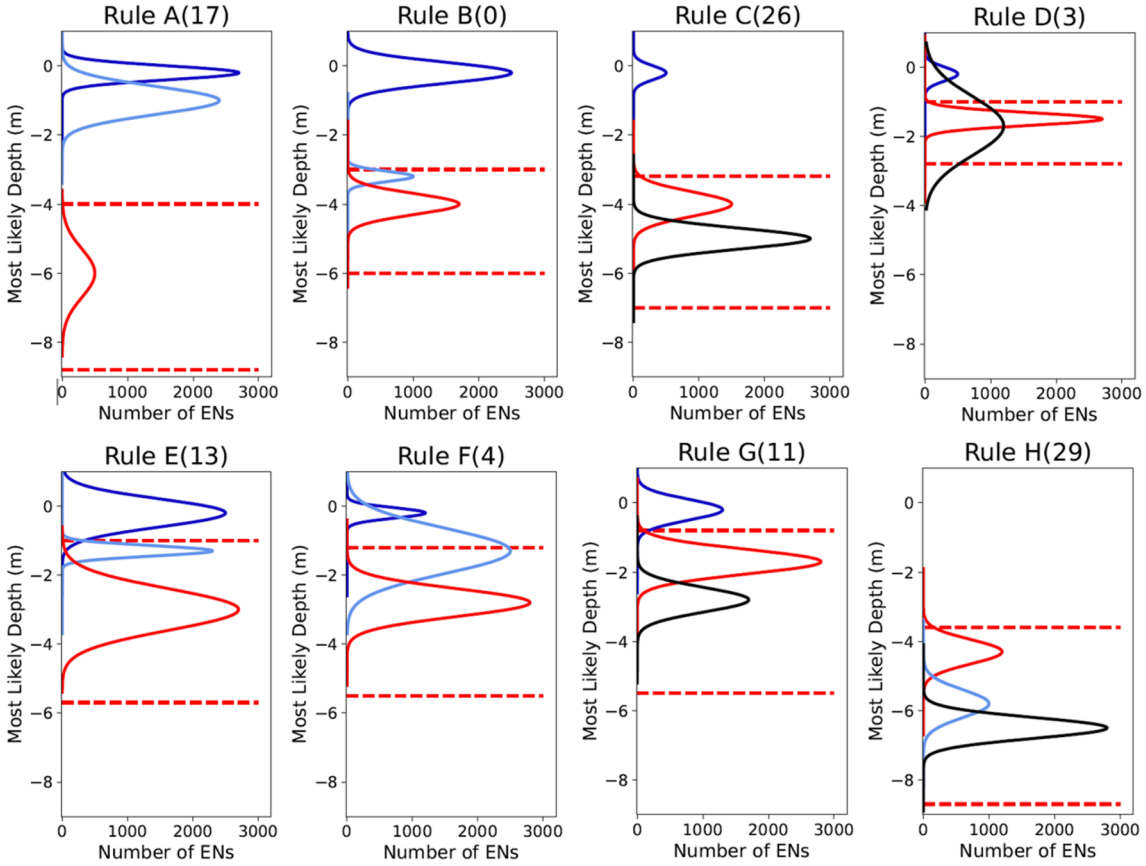


Fig. 5. Exemplar frequency distributions for different BDI rules (Fig. 4; Table 1). Red lines represent the derived BDI (dashed lines) and the cluster on which the shallow limit is based (solid lines). Black, light blue, and dark blue lines represent deep, mid-depth, and shallow (bright blue) clusters, respectively. Rule heading values are the number of tiles to which a rule was applied.

but low FPRs, and TPRs/PABs; the "too broad" BDI produces the opposite. Sparse *Bathy* also produces erroneous estimates of mean depth. This reinforces the *a priori* need to identify tiles at the limits of LiDAR penetration.

Four of the five infrastructure tiles (red dots in Fig. 7) had relatively high accuracy. Fig. 8 shows the anomalous infrastructure tile. The reference NOAA classification apparently identified *Bathy* soundings based on the infrastructure present rather than depth alone. Hence the

Table 2
Explanation of rules that determine the shallow limit of the bathymetry depth interval (BDI). (See also Figs. 4 and 5).

Rule	Primary Condition	Sub-condition	BDI Shallow Limit	Rationale	Number of Tiles[1]
A	No overlap between deep and mid-depth clusters	NA	Shallow limit of 99.9% MLD CI of deep cluster	Shallow and mid-depth clusters represent ocean surface	17
B	Only deep and mid-depth clusters overlap	Mid-depth cluster does not contain *Bathy* ENs	Shallow limit of 99.9% MLD CI of deep cluster.	Mid-depth cluster is noise.	0
C		Mid-depth cluster does contain *Bathy* ENs.	Shallow limit of 95% MLD CI of mid-depth cluster.	Bathymetry present on the tile is relatively variable.	26
D	All three clusters overlap	Mid-depth cluster does not appear to contain *Bathy* ENs.	Shallow limit of 95% MLD CI of mid-depth cluster.	High variability or few ENs in the deep cluster; mid-depth cluster is more reliable than deep cluster.	3
E		Deep cluster is narrow/ highly certain	Shallow limit of 99.9% MLD CI of the deep cluster	Deep cluster is "narrow"; mid-depth cluster variable suggests water column noise	13
F		Mid-depth cluster contains *Bathy* ENs, but the shallow cluster does not	99.9% MLD CI of the deep cluster	Mid-depth cluster is broad and noisy	4
G		Mid-depth cluster contains *Bathy* ENs; shallow cluster does not	Shallow limit of 95% MLD CI of the mid-depth cluster	Mid-depth cluster is narrow/ reliable suggesting mid-depth cluster truly does contain *Bathy* ENs	11
H		All three clusters contain *Bathy* ENs	Shallow limit of 95% CI of the shallow cluster	Multiple statistically identifiable depths on this tile	29

[1] The number of tiles whose BDI shallow limit was determined by this rule.

disagreement in the NOAA and CHRT-ML classifications is not due to CHRT-ML failure.

The relatively high accuracy for 12 of the 13 Reef and Channel tiles (dark blue dots in Fig. 7) suggests that the CHRT-ML classifier successfully addresses the problem of two bathymetric depth tiers (plus ocean surface) illustrated in Fig. 2. The anomalous tile (FNR of 0.8; Fig. 7c) is traversed by a relatively deep channel (Fig. 9). This suggests that CHRT-ML performs well unless channel slopes are "too steep" and the range of depth is "too large." Interestingly, such tiles may also be problematic for the NOAA classifier; relatively few *Bathy* soundings were extracted in the deeper southwest portion of the channel (Fig. 9b).

CHRT-ML performed reasonably well on most of the 11 Reef tiles (mauve dots in Fig. 7) despite depths less than 2 m. Fig. 10 shows the Reef tile on which CHRT-ML performed worst – most notably in the shallowest southeast area (Fig. 10c). A comparable shallow-to-deeper-water transition zone was also present in three poorly performing Normal tiles (green dots in Fig. 7). Most misclassification – *Bathy* false negatives especially – on these tiles occurred near such depth transition zones.

CHRT-ML performs reasonably well for the remaining Normal tiles (green dots in Fig. 7), lone Channel tile (light blue dot in Fig. 7), and lone Reef, Channel, and Infrastructure tile (black dot in Fig. 7). For these tiles the average global accuracy is about 90%, average TNRs and TPRs about 0.90, and average FNRs and FPRs near 0.05. Also encouraging for chartmakers is that for the non-anomalous tiles the average UAB is about 0.90. Moreover, although low accuracy values for the anomalous tiles may have been caused by poor CHRT-ML performance, they may alternatively relate to uneven quality of the NOAA reference classification. Fig. 8 provides an (admittedly extreme) example. Exploring the spatial and statistical structures of misclassified soundings as suggested by Lowell et al. (2020) might provide insight into the cause(s) of misclassification.

CHRT-ML produced comparable results for the validation/"test" data set (Fig. 11). CHRT-ML performed well for the Normal tiles sampled, all of the Reef tiles, and all but one of the Reef and Channel tiles (Fig. 11c FNR of 0.8). Visual examination of this tile suggested the presence of a deeper channel than the one in Fig. 9 that produced comparable accuracy values. On both tiles the deepest *k*-means cluster did not overlap with the two shallower clusters; such clusters are usually formed when only the deep cluster contains *Bathy* ENs. For the single sparse *Bathy* tile sampled, CHRT-ML performed better than expected. The erroneous classification of many NOAA *Bathy* soundings produced an FNR of about 0.40 that was better than for many of the sparse *Bathy* tiles in the development data set. Overall, therefore, it was concluded that CHRT-ML performed comparably for the development and validation data sets.

5. Discussion

This work was undertaken to evaluate the potential operationalisation of a previously developed proof-of-concept method of extracting bathymetry from LiDAR point clouds. In addition to accuracy, processing time is another operational consideration. Unsurprisingly processing time for each tile increases with the number of soundings (Fig. 12). Extrapolating the results in Fig. 12 suggests that, using the computing resources employed herein, it would take about 65 days to process all 1100 tiles. This processing time could be shortened by using enhanced computing resources as well as by improving existing software code. Though no time comparisons are provided relative to NOAA's existing processing methods, it is notable that CHRT-ML time is computational as opposed to NOAA's that requires considerable human input.

Another advantage of CHRT-ML is the elimination or at least reduction of the need for calibration data. The 3-cluster approach embedded in CHRT-ML is underpinned by an understanding of the interaction of LiDAR reflectance, shallow water geomorphometry, data characteristics, and ML algorithm behaviour. Hence adapting CHRT-ML to other areas should at most require a relatively small amount of

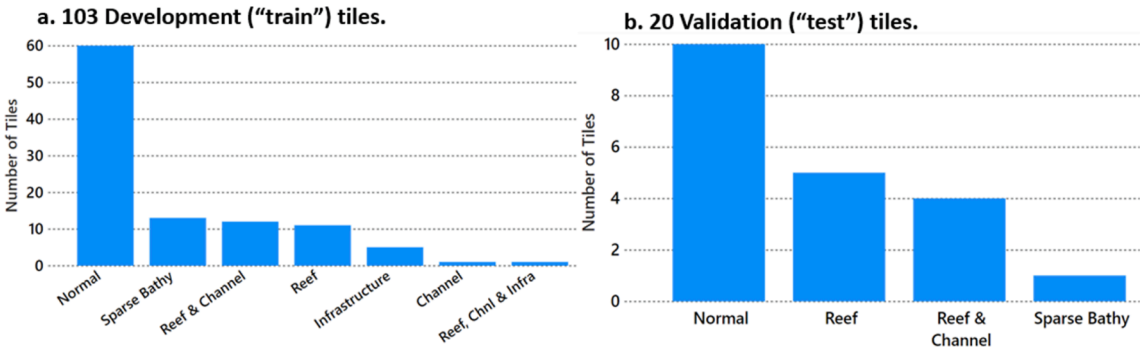


Fig. 6. Number of tiles by type.

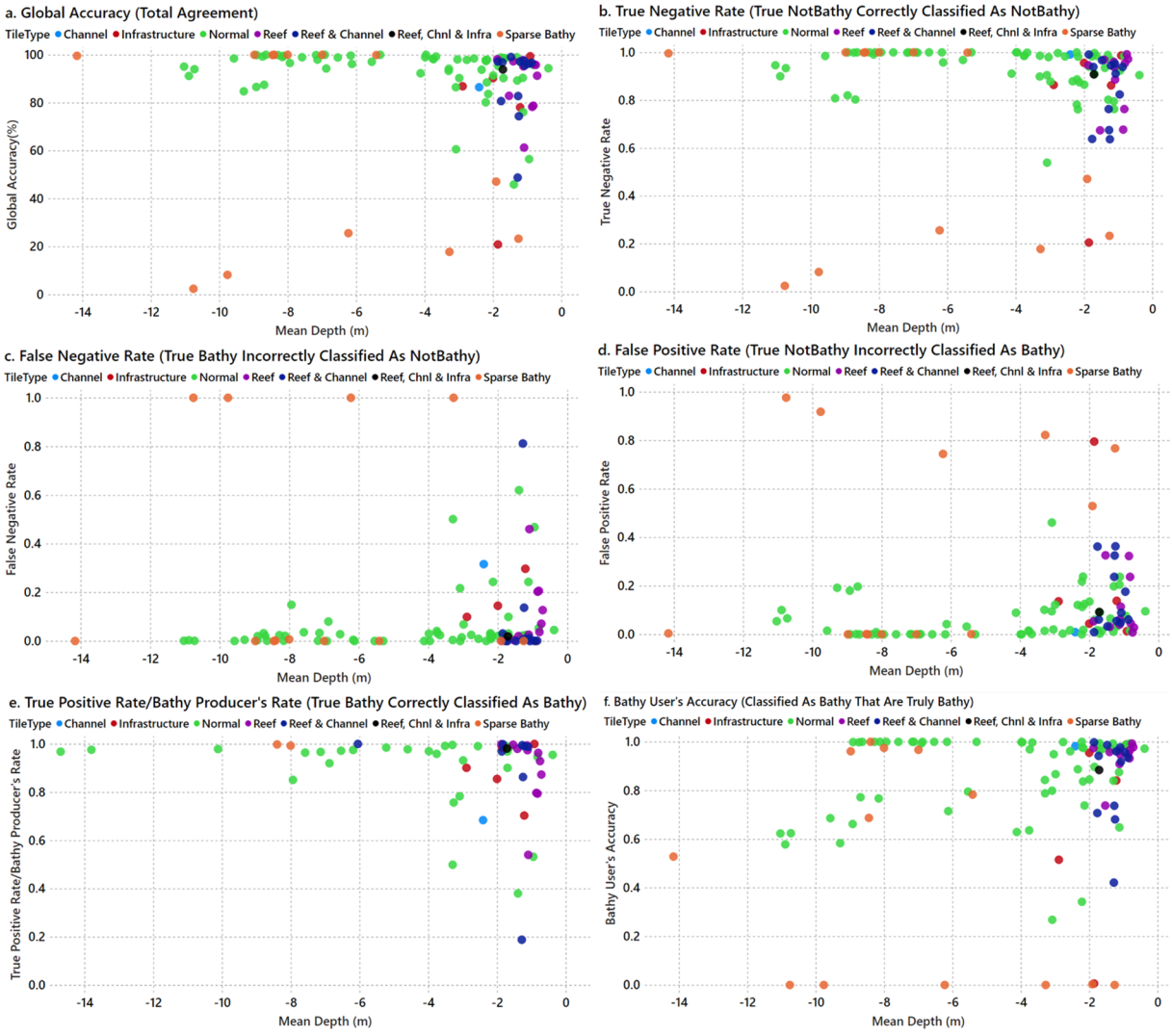


Fig. 7. Accuracy metrics for the development ("train") dataset.

reference data for verification rather than exhaustive numerical optimisation. Should CHRT-ML be found to perform poorly in other areas, adjustment of CHRT-ML's tuning parameters (Table 1) may suffice for successful adaptation.

An avenue for future research might be the use of adaptive tuning parameters. These might be based on readily identifiable *a priori* factors such as "local area type" or based on sounding metadata such as described in Lowell et al. (2020). If successful, the potential disadvantage of increased processing time might be compensated for by accuracy improvement.

Adaptive tuning parameters might also facilitate continual processing. CHRT-ML was developed for 500 m-by-500 m data tiles. Being able to process airborne LiDAR by flightpath rather than tile would eliminate the need for pre-processing segmentation into tiles although issues of flightpath overlap would have to be addressed. Continual processing might be achieved by progressively segmenting and processing portions of a flightpath including adjusting tuning parameters. The use of overlapping segments might provide comparative classifications that would

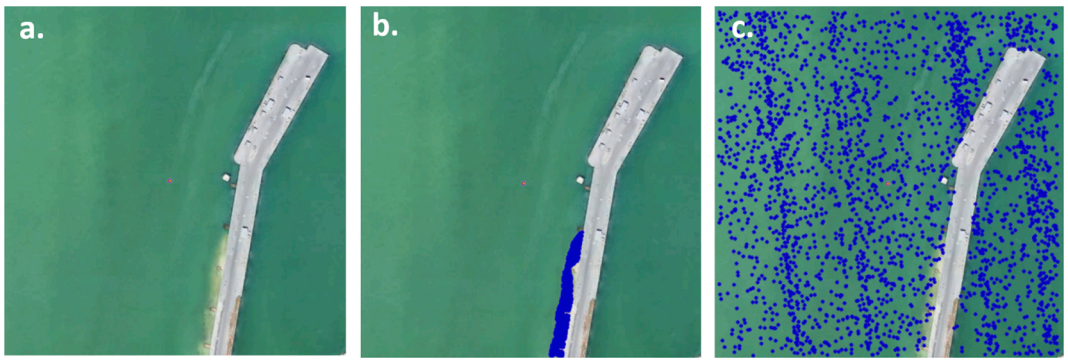


Fig. 8. Anomalous "Infrastructure" having a dock/pier. a. GoogleEarth™ imagery. Soundings (blue) classified as *Bathy* by (b.) NOAA and (c.) CHRT-ML. (A random sub-sample of soundings is displayed to improve interpretability.)

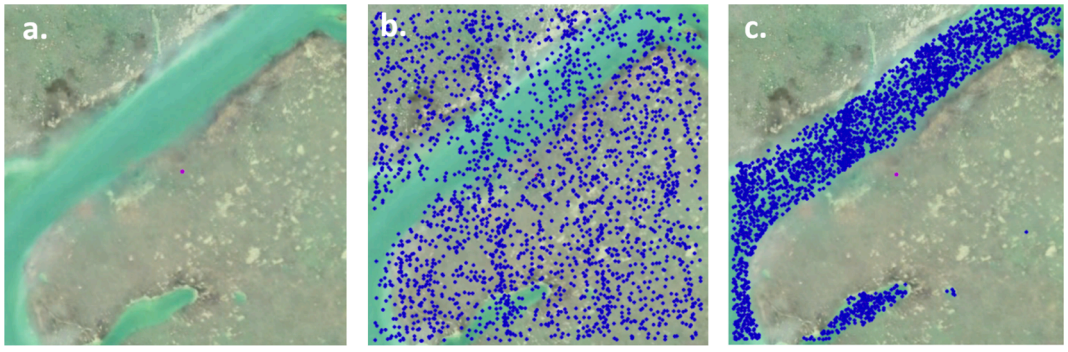


Fig. 9. Anomalous "Reef and Channel" tile a. GoogleEarth™ imagery. Soundings (blue) classified as *Bathy* by (b.) NOAA and (c.) CHRT-ML. (A random sub-sample of soundings is displayed to improve interpretability.)

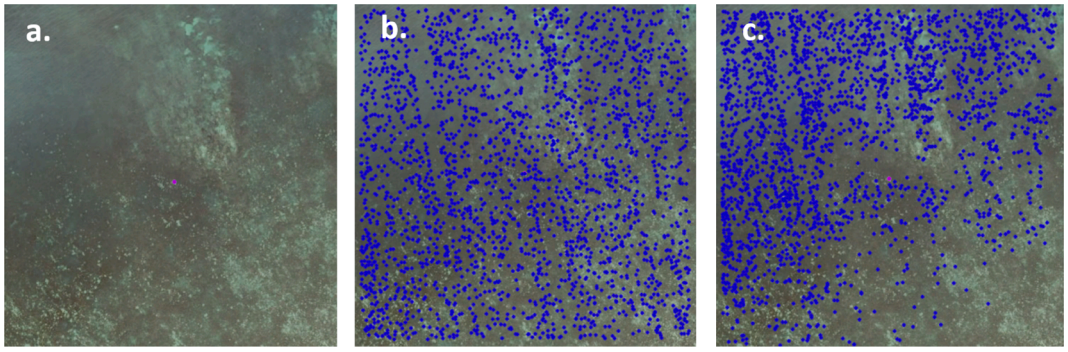


Fig. 10. Distribution of *Bathy* soundings for a poorly performing "Reef" tile. a. GoogleEarth™ imagery. b. Soundings (blue) classified as *Bathy* by NOAA and (c.) by the CHRT-ML approach. (A random sub-sample of soundings is displayed to improve interpretability.)

improve accuracy and facilitate continuous improvement.

A final potential avenue of future research is assessment of the accuracy of depth maps/charts produced by soundings classified as *Bathy*. It is more useful for navigation to achieve various international charting standards (IHO 2020) than it is to produce 100% accurate *Bathy/Not-Bathy* classifications. Analysing the ability of CHRT-ML to produce accurate nautical charts would also support continuous improvement.

6. Conclusions

Three major contributions of this work are identified. First, not only has a LiDAR classifier for shallow-water bathymetry been developed, but operational considerations of robustness, accuracy, and processing time have been evaluated. Second, the classifier does not require ground-truth data for calibration as do most remote sensing-based methods (e.g., Lyzenga et al. 2006; Pacheco et al. 2015). Thus CHRT-ML supports a seamless workflow from data ingestion to *Bathy/*

Fig. 11. Accuracy metrics for the validation ("test") dataset.

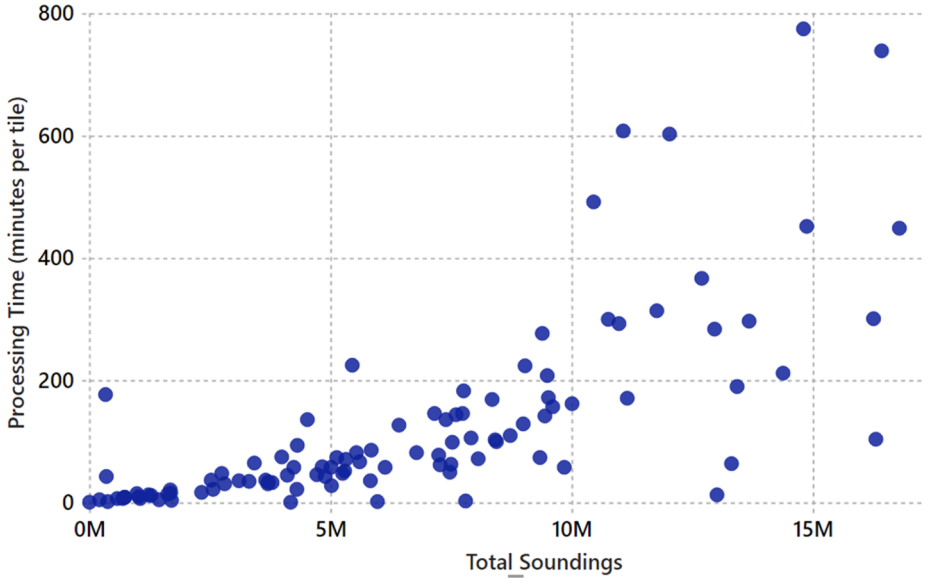


Fig. 12. Processing time for each tile relative to the total number of soundings.

NotBathy classification reliant only on the data being analysed. Third, the analyses presented provide a means of identifying potentially problematic tiles for a more detailed *a posteriori* examination. Operationally, an *a priori* determination of tile type could be made or tile and sounding metadata analysed. For example, tiles with a decrease in soundings whose depth approaches the limit of LiDAR penetration, or that are "Reef and Channel" tiles, or that produce a channel-based pattern (e.g., Fig. 2c or Fig. 9c) might warrant an *a posteriori* evaluation of classification accuracy.

Funding Sources

This work was supported by the United States National Oceanic and Atmospheric Administration (NOAA) Grants NA15NOS400020 and NA20NOS4000196.

8. Data Availability

All .las data tiles are available at the NOAA site https://coast.noaa.gov/htdata/lidar1_z/.

CRediT authorship contribution statement

Kim Lowell: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Brian Calder:** Conceptualization, Data curation, Funding acquisition, Project administration, Investigation, Resources, Validation, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgements

The authors appreciate the insightful comments of three anonymous reviewers.

References

ASPRS, 2013. LAS Specification Version 1.3-R13, 15 July 2013. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland.
Calder, B.R., Rice, G., 2017. Computationally efficient variable resolution depth estimation. Comput. Geosci. 106, 49–59.
Carrilho, A., Galo, M., dos Santos, R., 2018. Statistical outlier detection method for airborne LiDAR data. ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial. Inf. Sci. Vol. XLII-1, 87–92. https://doi.org/10.5194/isprs-archives-XLII-1-87-2018.
Congalton, R., Green, K., 2019. Assessing the Accuracy of Remotely Sensed Data – Principles and Practices, 3rd Edition. CRC Press, Boca Raton, Florida.
Curry, H., 1944. The method of steepest descent for non-linear minimization problems. Q. Appl. Math. 2 (3), 258–261.
Doneus, M., Doneus, N., Briese, C., Pregesbauer, M., Mandlburger, G., Verhoven, G., 2013. Airborne laser bathymetry-detecting and recording submerged archaeological sites from the air. J. Archaeol. Sci. 40 (4), 2136–2151. https://doi.org/10.1016/j.jas.2012.12.021.
Doneus, M., Miholjek, I., Mandlburger, G., Doneus, N., Verhoven, G., Briese, C., Pregesbauer, M., 2015. Airborne laser bathymetry for documentation of submerged archaeological sites in shallow water. ISPRS Arch. 40 (5), 99–107. http://hdl.handle.net/1854/LU-5933247.
Duan, Y., Yang, C., Chen, H., Yan, W., Li, H., 2021. Low-complexity point cloud denoising for LiDAR by PCA-based dimension. Opt. Commun. 482 https://doi.org/10.1016/j.optcom.2020.126567.
Forgy, E., 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. Biometrics. 21 (3), 768–769.
Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.
International Hydrographic Standards, 2020. Standards for Hydrographic Surveys, S-44 Edition 6.0.0. International Hydrographic Organization, Monaco.
Le Deunf, J., Debese, N., Schmitt, T., Billot, R., 2020. A review of data cleaning approaches in a hydrographic framework with a focus on bathymetric multibeam echosounder datasets. Geosciences 10, 254. https://doi.org/10.3390/geosciences10070254.
Lowell, K., Calder, B., Lyons, A., 2020. Measuring shallow-water bathymetric signal strength in lidar point attribute data using machine learning. Int. J. Geogr. Information Sci. 35 (8), 1592–1610. https://doi.org/10.1080/13658816.2020.1867147.
Lowell, K., Calder, B., 2021. Extracting shallow-water bathymetry from LiDAR point clouds using pulse attribute data: merging density-based and machine learning approaches. Mar. Geod. 44 (4), 259–286. https://doi.org/10.1080/01490419.2021.1925790.
Lyzenga, D., 1985. Shallow-water bathymetry using combined lidar and passive multispectral data. Int. J. Remote Sens. 6 (1), 115–125.
Lyzenga, D., Malinas, N., Tanis, F., 2006. Multispectral bathymetry using a simple physically based algorithm. IEEE Trans. Geosci. Remote Sens. 44 (8), 2251–2259. https://doi.org/10.1109/TGRS.2006.872909.
Mahalanobis, P., 1936. On the generalized distance in statistics. Proc. Natl. Inst. Sci. India 2 (1), 49–55.
Nagle, D., Wright, C., 2016. Algorithms used in the Airborne Lidar Processing System (ALPS). Open File Report 2016-1046. United States Dept. of the Interior/ United States Geological Survey, Washington, D.C.
Pacheco, A., Horta, J., Loureiro, C., Ferreira, O., 2015. Retrieval of nearshore bathymetry from Landsat 8 images: a tool for coastal monitoring in shallow waters. Remote Sens. Environ. 159, 102–116, 10.106/j.rse.2014.12.004.
Parker, H., Sinclair, M., 2012. The successful application of airborne LiDAR bathymetry surveys using latest technology. Proceedings: 2012 Oceans – Yeosu pp. 1-4. 21-24 May, Yeosu, South Korea. 10.1109/OCEANS-Yeosu.2012.6263588.
Randdal, H., Christiansen, P., Kliving, P., Andersen, O., Nielsen, K., 2021. Evaluation of a staitsitcal approach for extracting shallow water bathymetry signals from ICESat-2 ATL03 photon data. Remote Sens. 13, 3548. https://doi.org/10.3390/rs13173548.
Stephens, D., Smith, A., Redfern, T., Talbot, A., Lessnoff, A., 2019. Using three dimensional convolutional neural networks for denoising echosounder point cloud data. Appl. Comput. Geosci. 5, 1000016. https://doi.org/10.1016/j.acags.2019.1000016.
Wang, X., Glennie, C., Pan, Z., 2017. An adaptive ellipsoid searching filter for airborne single-photon LiDAR. IEEE Geosci. Remote Sens. Lett. 14 (8), 1258–1262. https://doi.org/10.1109/LGRS.2017.2704917.
Xiangyun, H., Yi, Y., 2016. Deep-learning-based classification for DTM extraction from ALS point cloud. Remote Sens. 8 (9), 730. https://doi.org/10.3390/rs8090730.
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 J. Photogramm. Remote Sens. 163, 49–61. https://doi.org/10.1016/j.isprsjprs.2020.03.004.
Yong-hua, S., Xu-qing, Z., Xue-feng, N., Guo-dong, Y., Ji-kai, Z., 2017. Denoising algorithm of airborne LiDAR point cloud based on 3D grid. Int. J. Sig. Process. Image Process. Pattern Recogn. 10 (3), 85–92. https://doi.org/10.14257/ijsip.2017.10.3.09.
Zhang, J., Hu, X., Dai, H., Qu, S., 2020. DEM extraction from ALS point clouds in forest areas via graph convolution network. Remote Sens. 12 (1), 178. https://doi.org/10.3390/rs12010178.