Developing Machine Learning Models for Quality Assurance and Continuous Improvement of Bathymetry Extraction from Lidar Point Clouds

Kim Lowell[1], Brian Calder, Anthony Lyons, Centre for Coastal and Ocean Mapping and Joint Hydrographic Centre, University of New Hampshire, Durham, NH 03824 UNITED STATES

Quality assurance and continuous improvement (QA&CI) are essential elements of operational workflows and require information that pinpoints the weakest part(s) of workflow outputs. The focus of this article is the production of such information to improve the accuracy of shallow water bathymetric maps produced from airborne lidar. The collaborating and funding partner for this work is the United States National Oceanographic and Atmospheric Administration (NOAA).

Essentially NOAA's lidar processing workflow classifies each pulse return in a point cloud as bathymetry (*Bathy*) or not bathymetry (*NotBathy*) using a variety of information and data. One type of information that is routinely collected but not currently used is lidar "point attribute data" (PAD) associated with each return – e.g., the intensity of the return, the stability of the airborne platform at the moment of acquisition. To improve the accuracy of NOAA's bathymetric extraction – i.e., its *Bathy*/*NotBathy* classification –the strength of the bathymetric signal in the PAD was evaluated using machine learning (ML) techniques. The data used for evaluation are four 500 m-by-500 m lidar data "tiles" located in the vicinity of Key West, Florida covering an approximate depth range of 0.5m to 20m. It has been concluded that:

- The bathymetric signal is sufficiently strong in the PAD to warrant further exploration for QA&CI.
- The three types of PAD variables – we term these pulse-specific, SBET (Smoothed Best Estimate of Trajectory), and lidar-edge – were related to the bathymetric signal.
- Extreme gradient boosting (XGB) modelled the bathymetric signal better than did neural networks and logistic regression
- XGB was the fastest ML and is therefore preferred for operational implementation of results.

XGB models that measure bathymetric signal strength can also be used to estimate the probability that each pulse return is bathymetry. This "*p(Bathy)*" value can be used in QA&CI in two ways. First, it can be used in NOAA's lidar processing workflow to improve the initial *Bathy/NotBathy* classification. Second, because PAD are not used in NOAA's workflow, *p(Bathy)* can be used to produce a second independent *Bathy*/*NotBathy* classification. This can then be compared to NOAA's classification statistically and spatially to identify areas of disagreement in feature space and geographic space.

Figure 1 shows an example for the area-based rate of false negatives – i.e., pulse returns identified by NOAA as *Bathy* that were *NotBathy* according to the XGB model. In Figure 1a (left), major differences appear as red and dark blue areas with light green areas indicating better agreement between the two classifications. Figure 1b shows the same information statistically; the red line represents the line of perfect agreement and the blue line and surrounding zone indicate the 95% confidence interval around the regression line.

[1] Corresponding Author – klowell@ccom.unh.edu

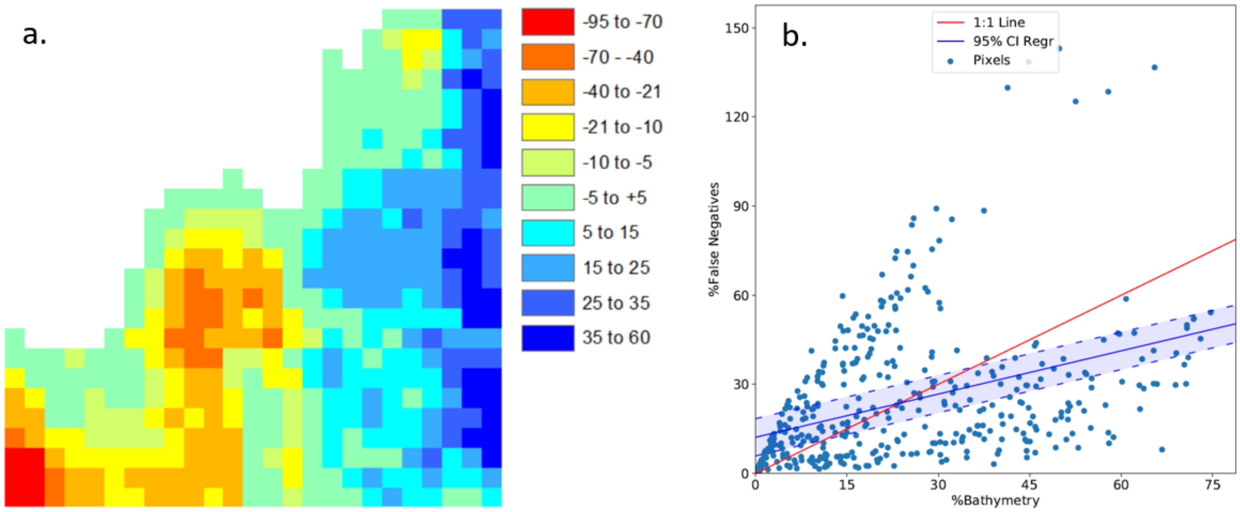

Figure 1. Representation of differences in false negatives – *Bathy* pulse returns incorrectly labelled *NotBathy* – between a NOAA classification and a classification produced by a machine learning model.

Figure 1a provides a clear indication that the southwest and eastern edge of this tile are the areas of maximum disagreement and therefore should be targeted by QA&CI protocols. Similarly, Figure 1b suggests that disagreement is greatest in areas in which NOAA has indicated the highest density of bathymetry – i.e., the blue line and zone diverge the most from the red perfect-agreement line. Moreover, Figure 1b demonstrates that the variance of differences is heteroscedastic which, with further analysis, may provide additional information about why there is disagreement between classifications and how this can be addressed if necessary.

Figure 1 is based on the number of false negatives which is often of greatest interest in bathymetric mapping. However, for QA&CI comparable figures can be produced for false positives (pulses incorrectly identified as bathymetry), true positives (correctly identified *Bathy*), and true negatives (correctly identified *NotBathy*).

In closing, we note that the reason for the disagreements between the two classification –an inaccurate NOAA classification, an inaccurate ML classification, or both – is not apparent from the information provided. However, the information produced provides for targeted improvement of bathymetric maps and a consequent savings of time and effort.