An adaptive ground-filtering technique for noisy high-altitude laser profiling data

Mahmoud Awadallah[1], Ranjith Gopalakrishnan[2], Sherin Ghannam[1], A. Lynn Abbott[1], Randolph H. Wynne[2] and Valerie A. Thomas[2]

[1]Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA {mawadala,sghannam,abbott}@vt.edu

[2]Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA {ranjith7,wynne,Thomas}@vt.edu

Abstract - Reliable global estimates of forest biomass dynamics are critically important for understanding global change. NASA's ICESat (Ice, Cloud and Land Elevation Satellite) mission, which acquired data near-globally from 2003 to 2009 demonstrated the feasibility of using spaceborne lidar for such global estimates. ICESat-2 is the much-anticipated successor to ICESat, featuring the innovative ATLAS photon counting lidar instrument. This new technology brings unique challenges to the detection of ground and top of canopy surfaces, the greatest being the substantial amount of ambient noise introduced into the photon cloud. Data anticipated from ICESat-2 is currently simulated by NASA's multiple altimeter beam experimental lidar (MABEL) system. A key challenge of vegetation characterization (or digital elevation model generation) using lidar is the robust estimation of ground points in the point-cloud. This paper presents an innovative framework for such ground estimation from ICESat-2's noisy high-altitude photon counting profiling lidar data. First, a histogram-based filter is used to reduce noise. Then, a technique of active contour models is utilized to extract terrain points (ground and vegetation). Then, we take advantage of the relative planar symmetry of forest vegetation to simulate a likely lidar point-cloud over a 4-meter-wide strip straddling the original profiling transect. This three-dimensional point-cloud is subjected to a triangular irregular network (TIN) ground-detection algorithm. The advantages of using a TIN-based algorithm are its proven robustness over forested regions and the ability to fine-tune the algorithm. The efficacy of the proposed method is evaluated by estimating the ground along three MABEL profiling transects. The proposed method shows good performance along forested canopies and reasonable performance over non-forested stretches. We show that the results compare favorably with other high-quality reference data and suggest improvements to the proposed method.

Keywords: ICESat-2, Spaceborne lidar, forest canopy height, forest biomass, DEM

1. Introduction

Laser altimetry is well suited to estimate vegetation height and structure. However, at present, there are no operational lidar sensors in space that are designed to measure terrestrial surfaces. This situation will change with NASA's ICESat-2, which is scheduled for launch in 2017. One of the ICESat-2 mission's goals is to enable independent determination of global vegetation height using elevation measurements over a period of two years. ICESat-2 will be equipped with the Advanced Topographic Laser Altimeter System (ATLAS) sensor, which is a multibeam system that will collect elevation data using a photon-counting technology. This approach yields clouds of discrete points, each resulting from the return of an individual green (λ = 532 nm) photon.

During the last two decades, data from airborne laser scanners have been used for biophysical parameter estimation (e.g., Nilsson 1996; Næsset 1997; Means et al. 2000; Popescu et al. 2003; Popescu et al. 2004; Bortolot and Wynne 2005; van Aardt et al. 2006; Thomas et al. 2006). The new photon-counting technology introduced in ATLAS will pose new challenges to the surface finding algorithms. The major challenges are the large amount of noise present over the whole atmospheric column and the weak ground signal under dense canopy covers. This paper describes an enhancement over our previously proposed image-processing method in Awadallah et al. (2013) that has been developed to estimate the top-of-canopy and ground surfaces. The enhancement is carried out by adding a further analysis stage to adjust the previously obtained results. The proposed method in Awadallah et al. (2013) is based on an image processing segmentation technique called active contour models.

2. Data and Study Area

The photon-counting lidar data will be obtained from the Multiple-Altimeter Beam Experiment Lidar (MABEL), a multibeam sensor that was operated at high altitude on NASA's ER-2 platform. This lidar offers insights into the signal qualities that are anticipated for ICESat-2. Only some of the flight lines of MABEL inside Virginia were used in this study.

The reference data source is airborne discrete lidar data collected for the same flight paths using NASA-Goddard's Lidar, Hyperspectral, and Thermal Imager (G-LiHT). The lidar sensor in G-LiHT uses small-footprint scanning analog technology. The study area is shown in Figure 1, with the G-LiHT paths shown in yellow.

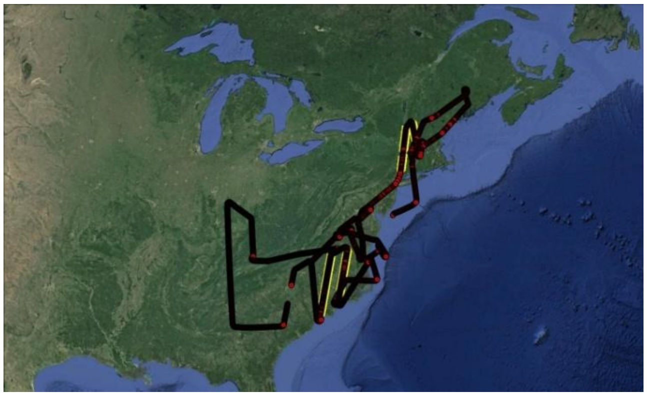


Figure 1: Study area for MABEL (black curves) and G-LiHT (yellow curves).

3. Methodology

The expected outcome of this work is to enhance the ground surface finding results from photon-counting lidar and to compare the results to G-LiHT extracted DTM.

3.1 Detection of Ground and Top of Canopy in MABEL

Over forest canopies, the ICESat-2 laser data will include returns from the top of the canopy and from the ground, so that various canopy height measurements can be derived from the photon distribution between these two levels. However, as can be seen from our preliminary work, ICESat-2 data over a dense forest canopy will pose three major challenges to canopy-ground detection (see Figure 2):

1- Noise will be present throughout the atmospheric column, within the canopy, and below the ground, and will vary according to a) local atmospheric conditions, b) target reflectivity at 532 nm and c) illumination conditions due to day/night changes and, during daylight, due to topography-sun angle interactions.
2- Dense canopies will naturally occlude the ground, making it very difficult to distinguish ground from noise in some instances. This effect will be magnified for closed-canopy forest ecosystems.
3- Gaps in the canopy can be difficult to localize in the presence of signal noise. These gaps could impact the accuracy of height metrics, and therefore affect biomass estimates, under certain conditions.

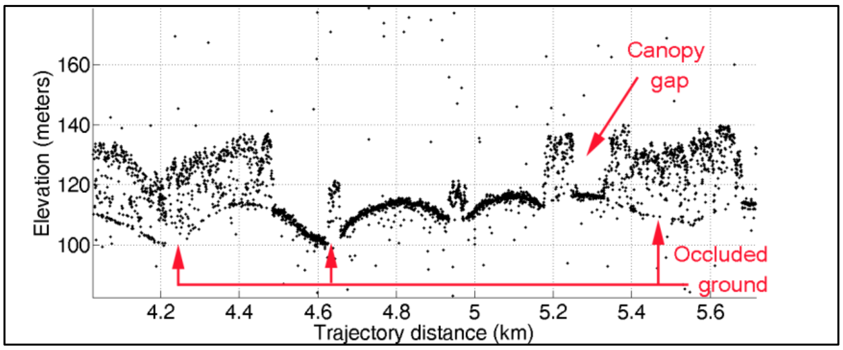


Figure 2: Unique challenges for forestry applications of ICESat-2 data. Noise can be seen throughout the entire vertical column. Canopy gaps and ground occlusion are evident Awadallah et al. (2014).

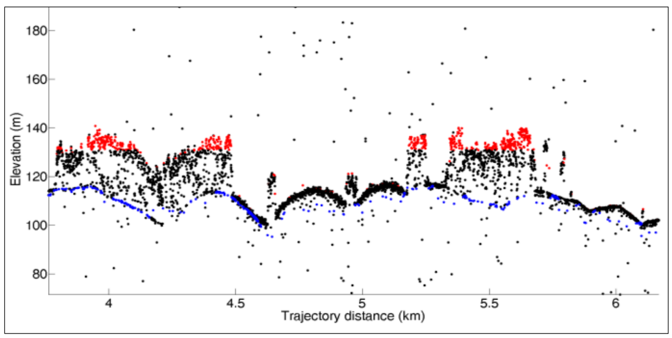
The algorithm begins by using an automated signal-analysis technique that can detect ground and top-of-canopy levels within ICESat-2 data. The main emphasis of the algorithm is on noise immunity and on localization of gaps in the canopy. The techniques addresses the canopy-ground detection problem using a combination of noise filtering, contour detection using deformable-model optimization, and separation of ground and top of canopy. More discussion of the processing techniques is described in Awadallah et al. (2013) and Awadallah et al. (2014).

In this stage the problem is formulated as a problem of two-dimensional (2D) image analysis. Each data file consists of a set of (x, y, z) points from a single flight track. Approximately 10,000 to 25,000 values are present in each file, and the first step in our approach is to map these values onto a 2D grid. When displayed, the resulting image contains many points corresponding to the ground and canopy, along with noise points below the ground and above the canopy. For noise removal, a combination of median filtering, size filtering, and morphological processing have been applied. Much of the noise in the image consists of isolated points, often characterized as impulsive noise, and these techniques are well suited for removing this type of noise.

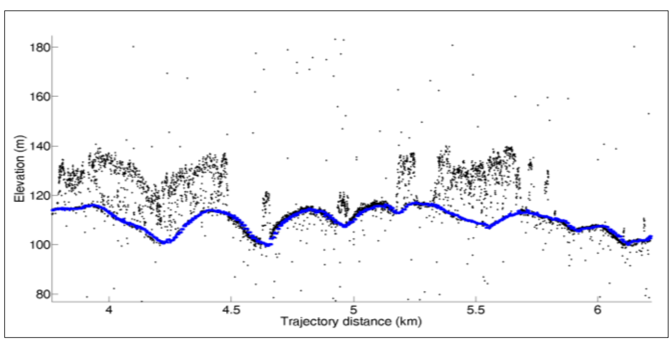
The next major processing step relies on deformable models, which refers to a class of optimization techniques that are widely used in image analysis. The approach used in this paper first introduced by Kass et al. (1988). The fundamental idea is to perform an iterative search for the best fit of a 2D curve to a noisy image. The search is usually a “greedy” algorithm that updates the curve slightly at each iteration step, in such a way that each update is locally optimum according to energy terms that characterize the curve’s shape and its immediate surroundings in the image (Awadallah et al. 2013). The approach does not guarantee that a globally optimum solution will be found, but represents a balance of accuracy and computation speed. These algorithms are sometimes known as “snakes” because of the curve’s appearance, over time, during the optimization procedure. We have extended the method of Chan and Vese (2001), who introduced a type of active contour without the need of edge detectors (Active Contours without Edges). In this approach, the curve is represented implicitly using a “level set” function, rather than the more traditional parametric form. A consequence is that multiple closed curves can be detected in the image. Our approach utilizes regional image statistics rather than more traditional intensity edges as a means of guiding the search. In our experiments, this approach has resulted in better detection of gaps in the canopy.

3.2 Enhanced Ground Surface Extraction from MABEL Profiles

The motivation for this part is as follows. Awadallah et al. (2013) described an algorithm to find the ground and the top of canopy in lidar profiling data with sensor and ambient noise. An energy function is defined, which incorporates the amount of stretching and bending of the curve, while being fitted to the surface in question (could be ground or top of canopy). Then, a solution that minimizes this energy value is searched for. But it turned out that the algorithm was sensitive to whether the segment was a forest patch or one with just short vegetation. The results of this initial implementation of the Awadallah et al. (2013) algorithm can be seen in figure 3.



(a)



(b)

Figure 3: Results from the implementation of the active contour algorithm described in Awadallah et al. (2013). The top part shows the results of the implemented algorithm. The bottom part is reference data, i.e., ground extracted from the G-LiHT flight data. In the short vegetation patch areas, one can see that there is significant difference between the two figures. Hence, it makes sense to identify patches of the transect with short vegetation by other methods, and run a modified version of the active contour algorithm at those patches.

Hence, we were motivated to develop a binary classifier that classifies given transect segments into one of two cases: 1) forest patch, and 2) short vegetation patch.

Firstly, the transect in question was segmented automatically into smaller parts (see figure 4). This was done by a heuristic method of exploring the places which most probably a transition from high vegetation to short vegetation. In this method the difference between the highest point and the lowest point in each 2 meters segment is calculated. The result of this process is a one dimensional vector of heights. An edge detector is used to find the sudden changes in this vector, because these sudden changes are strong indication of transitioning from short vegetation to long vegetation and vice-versa. Then, two methods were used to identify whether the segment was a tree canopy or ground segment.

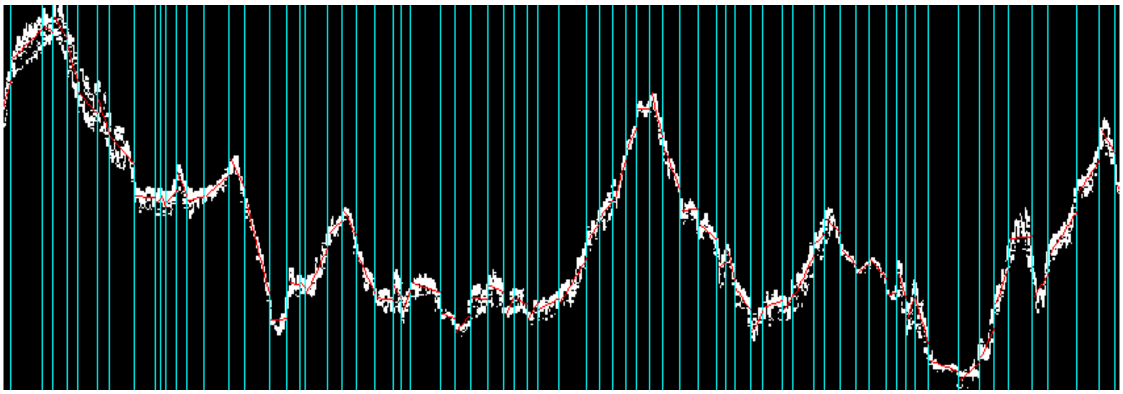


Figure 4: The segmentation of the lidar transects. Each vertical green line represents a separate lidar transect.

3.2.1 Percentile method:

The method can be described as follows. For each segment:

1. Identify the mid-point of the distribution of points. For example, in the "C" section of figure 5, this is marked by the thin light blue line. This is done by finding the median height of the points in that segment.
2. Identify the height bin (1-dimensionally symmetrical around the mid-point) that contains 95% of the points in that segment. This is denoted by the thick orange lines of segment "G" and the thick green lines of segment "C", in figure 5. Find the height of the bins for each segment (h_{bin}).
3. If $h_{bin} > h_{threshold}$ then classify as canopy, else classify as ground.

3.2.2 Regression method:

This method can be described as follows. For each segment:

1. Identify the simple linear regression line that fits the set of points best.
2. Calculate the "goodness" of fit, with an RMSE statistic.
3. If $RMSE_{segment} > RMSE_{threshold}$ then classify as canopy, else classify as ground.

3.2.2 Regression method:

In order to assess the accuracy of the two proposed methods, three random transects were selected and each transect is divided into a set of segments. Each segment was classified manually (also using the G-Liht data) to either short vegetation or forest patch.

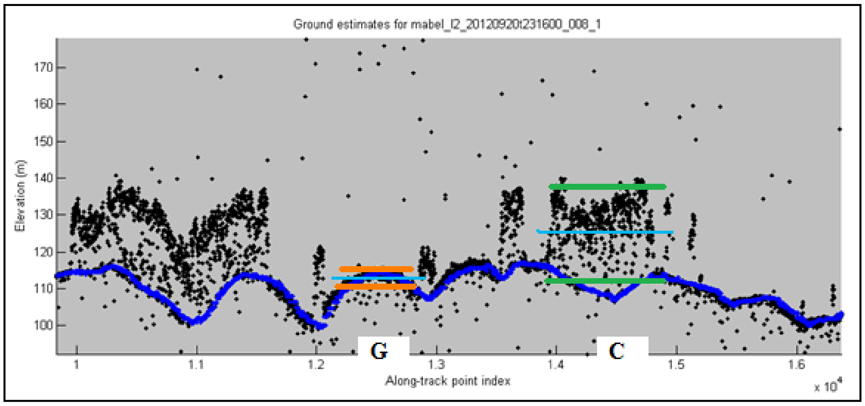


Figure 5: A schematic representation of the percentile-based method. Note that there are two segments of interest: short vegetation patch (labelled G, with the orange horizontal lines) and forest patch (labelled C, associated with the green horizontal lines). For each segment, height bins that enclosed 95% of the points were made. The height of these bins gives an indication of whether the patch is a forested patch or not.

The accuracy of the classification for a given method is defined as the percentage of segments it marked accurately (with respect to the reference classification). The above two methods of classification of segments into forest patches or short vegetation patches was run on three transects. The accuracy of classification can be seen in figure 6.

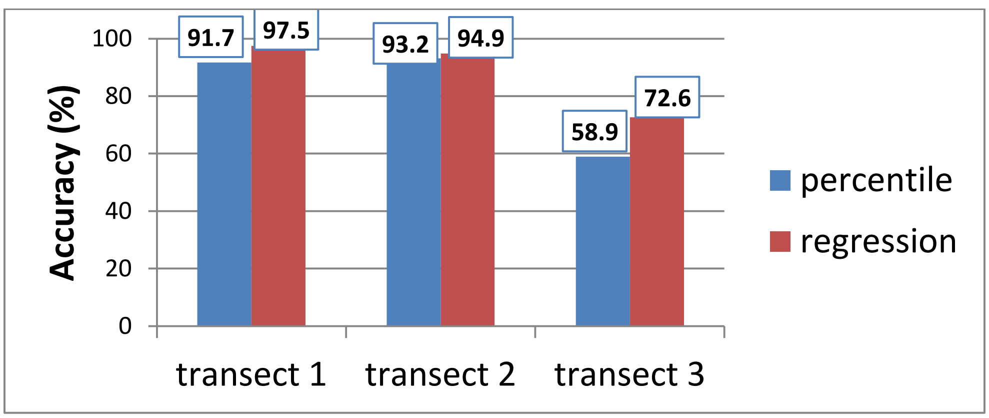


Figure 6: The accuracy of classification for the two methods considered here.

We note that the overall accuracy is pretty high (~90%) for transects 1 and 2, for both methods. Also, the regression method gives better overall results. Also, the results for transect3 are less promising, and more analysis is needed to understand the reason for this lower accuracy.

3.3 Simulation of a 3D Point Cloud Corresponding to the Transect

In this part, we first simulate a three-dimensional point-cloud around the original transect. Here, we try to reproduce the point cloud that would have resulted if a scanning lidar instrument was used, rather than the profiling one used. This is done by taking advantage of the relative planar symmetry of forest vegetation. We replicate each point in the original transect with slightly jittered horizontal and vertical co-ordinates (that is, *x*, *y*, & *z* values). The jittering is done following a normal distribution of mean zero and standard deviation two meters. Each point is replicated 50 times, to give a new transect of width ~4 meters (see figure 7). After this, we ran a standard implementation of an adaptive TIN-based ground filtering algorithm (Axelsson, 2000).

This was attractive because of its overall robustness and its efficacy over forested regions (Sithole & Vosselman 2004). We used an initial seeding step size of 5 meters.

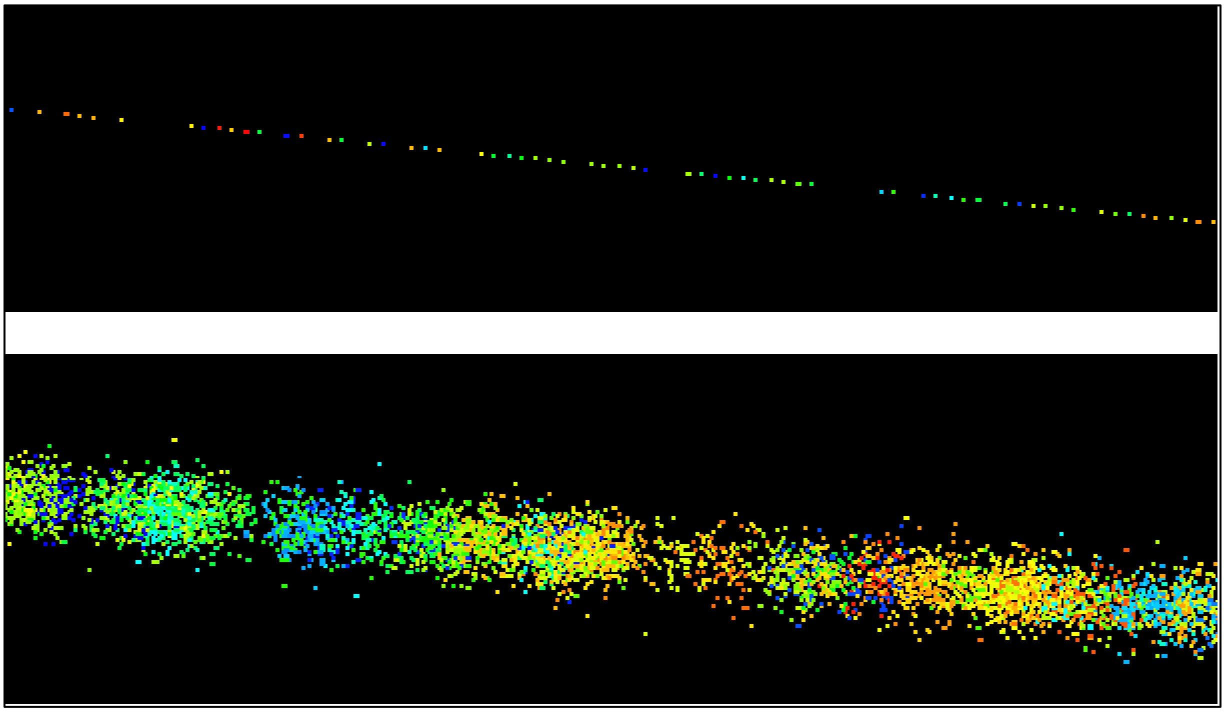


Figure 7: The upper portion shows the top-view of a typical MABEL transect, with profiling lidar data. The lower part shows the top-view a part of the same transect, but with similated scanning lidar data. The bottom transect is ~4 meters wide, and can be processed by ground extraction algorithms developed for three-dimensional data.

3.4 Co-registering Datasets

The first step in processing lidar data from multiple sources is to adjust the geolocation and co-registration among different datasets. The geolocation of MABEL data was inaccurate due some operational problems (see Figure 8). These geolocation errors have been resolved by manually aligning flight lines using Merrick MARS software and a MATLAB script.

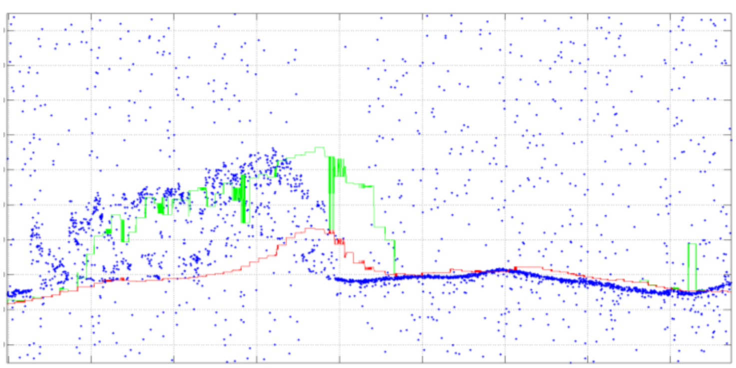


Figure 8: Geolocation problems of MABEL data (blue point cloud) compared to G-LiHT (green and red curves).

4. Experiments and Results

In order to assess the performance of the proposed algorithm, it was applied to a set of MABEL transects. The ground surface estimation was evaluated and compared visually to the reference data obtained from G-LiHT. Figure 9 shows example result of the enhanced DTM and DSM extraction algorithm from three different MABEL transects that differ in topology and vegetation structure.

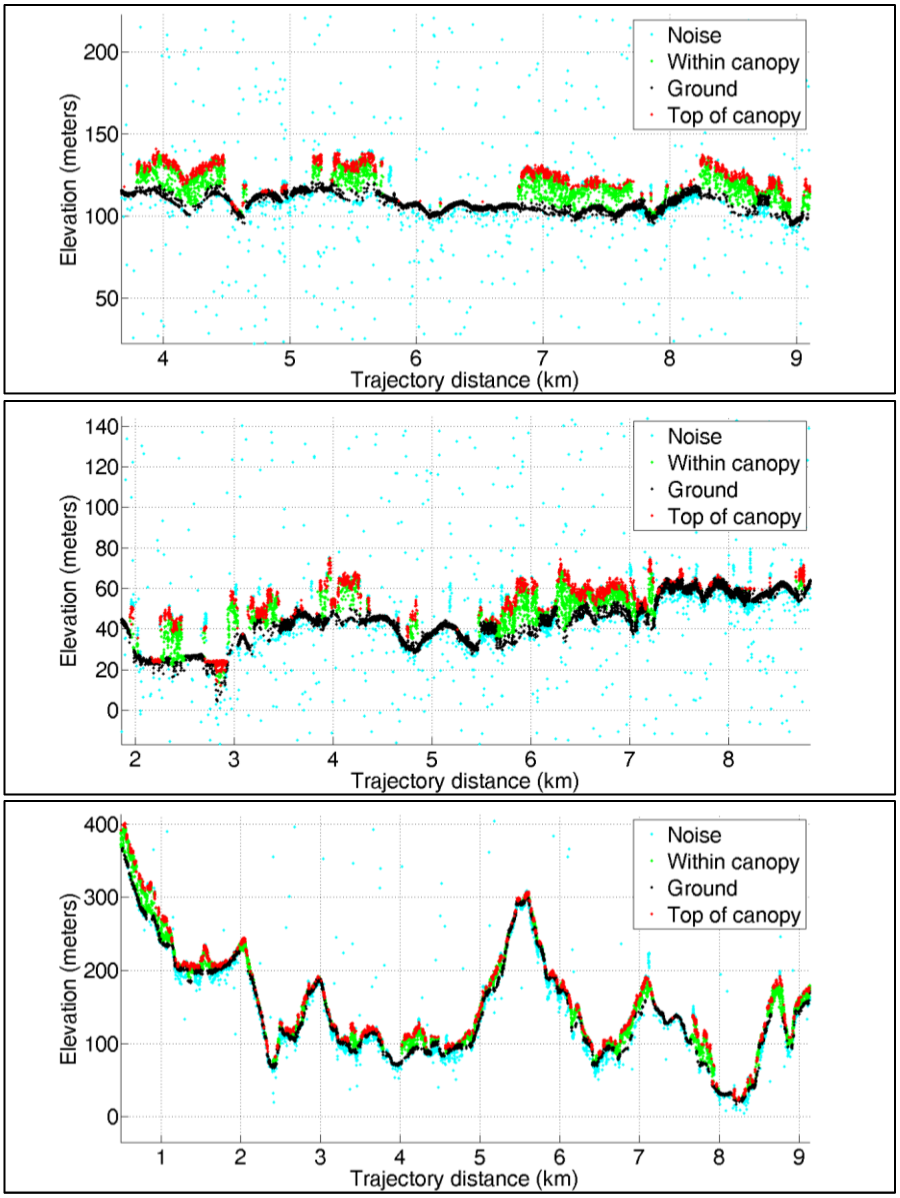


Figure 9: Sample results of DTM and DSM extraction algorithm from 3 different MABEL transects.

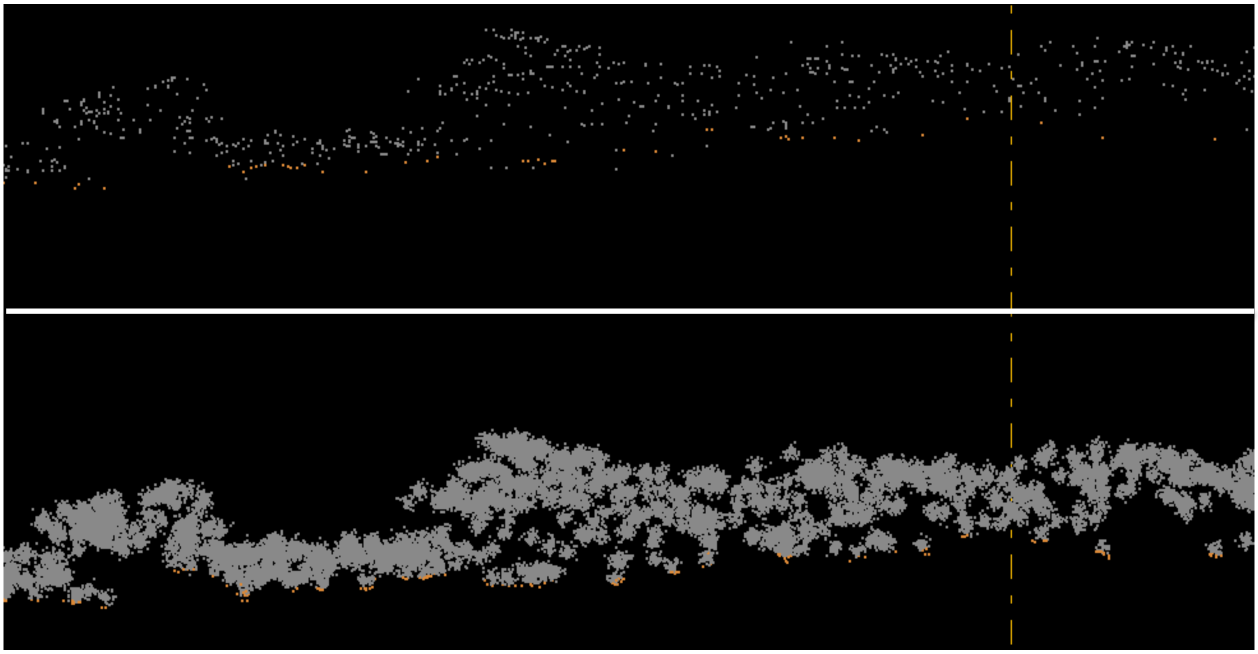


Figure 10: Improvement of ground classification when using simulated 3D transect data. The top part shows ground classification (ground points in orange, non-ground in grey) when using the active contour algorithm. The bottom part shows the ground classification when using an adaptive TIN-based ground filtering algorithm.

The improvements using the simulated 3D point cloud can be seen in figure 10. Firstly, we see that more ground points have been detected. Secondly, one can notice that the ground points in the simulated cloud follow the lower envelope of the point cloud much more closely. Both approaches avoided losing ground surface even under dense canopy in which the ground signal is very weak.

5. Conclusions and Future Work

In this paper, we introduced a novel study of lidar data analysis in which two different lidar datasets were used to assess the accuracy of the areas of intersection between them. We discussed two methods by which forest patches may be distinguished from short vegetation ones in data similar to that expected from ICESat-2. It is noticed clearly that the regression method is better of the two, with accuracies above 90% in most cases. Hence, such classifiers could be used to improve the accuracy of algorithms such as Awadallah et al. (2013). Finally, we introduced a simple method to build a simulated 3D point-cloud, and use an adaptive TIN-based ground detection algorithm.

In the future a more robust and quantitative data comparison should be used to estimate the ground surface extraction accuracy from MABEL as compared to reference data obtained from G-LiHT. Also, better methods to parametrize the adaptive TIN-based ground detection algorithm based on initial estimates of ground vegetation characteristics can be explored.

References

Awadallah, M.; Ghannam, S.; Abbott, A. L.; Ghanem, A., 2013. Active contour models for extracting ground and forest canopy curves from discrete laser altimeter data, *Proceedings: 13th International Conference on LiDAR Applications for Assessing Forest Ecosystems (SilviLaser 2013), Beijing, China*.

Awadallah, M.; Ghannam, S.; Abbott, A. L.; Ghanem, A., 2014. A two-stage algorithm for extraction of ground and top of canopy in photon-counting profiling-LiDAR data in preparation for the ICESat-2 mission, *Proceedings: IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1353–1356.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, 33(B4/1; PART 4), 111–118.

Bortolot, Z. J.; and Wynne, R. H., 2005. Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59, 342–360.

Kass, M.; Witkin, A.; and Terzopoulos, D., 1988. Snakes: active contour models. *International Journal of Computer Vision*, 1 (4), 321–331.

Means, J. E.; Acker, S. A.; Fitt, B. J.; Renslow, M.; Emerson, L.; and Hendrix, C. J., 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing*, 66(11), 1367–1371.

Næsset, E., 1997. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 61, 246–253.

Nilsson, M., 1996. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, 56, 1–7.

Popescu, S. C.; Wynne, R. H.; and Nelson, R. F., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29 (5), 564–577.

Popescu, S. C.; Wynne, R. H.; and Scrivani, J. A., 2004. Fusion of small footprint LiDAR and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. *Forest Science*, 50 (4), 551–565.

Sithole, G.; and Vosselman, G., 2004. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens*. 59, 85–101.

Thomas, V.; Treitz, P.; McCaughey, J. H.; and Morrison, I., 2006. Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density. *Canadian Journal of Forest Research*, 36 (1), 34–47.

Van Aardt, J. A. N.; Wynne, R. H.; and Oderwald, R. G., 2006. Forest volume and biomass estimation using small footprint lidar-distributional parameters on a per-segment basis. *Forest Science*, 52, 636-649.