An Algorithm for Detection of Ground and Canopy Cover in Micropulse Photon-Counting Lidar Altimeter Data in Preparation of the ICESat-2 Mission

Ute C. Herzfeld[(1,2)], *Brian W. McDonald*[(1)], *Bruce F. Wallin*[(2,3)], *Thorsten Markus*[(4)], *Thomas A. Neumann*[(4)] *and Anita Brenner*[(5)]

(1) Department of Electrical, Computer and Energy Engineering, University of Colorado Boulder, Boulder, Colorado, U.S.A.

(2) Cooperative Institute for Research in Enviromental Sciences, University of Colorado Boulder, Boulder, Colorado, U.S.A.

(3) now at : New Mexico Tech, Socorro, New Mexico, U.S.A.

(4) Cryospheric Sciences Branch, 614.1, NASA Goddard Space Flight Center, Greenbelt, Maryland, U.S.A.

(5) Sigma Space Corporation, Lanham, Maryland, , U.S.A.

Corresponding author:
Ute C. Herzfeld, CIRES and ECEE, University of Colorado Boulder, Boulder, Colorado 80309-0449, U.S.A.
e-mail: uch5678@gmail.com, ute.herzfeld@colorado.edu
phone: (303) 735-5164

Abstract. The Ice, Cloud and Land Elevation Satellite-II (ICESat-2) mission has been selected by NASA as a Decadal Survey mission, to be launched in 2016. Mission objectives are to measure land ice elevation, sea ice freeboard/ thickness and changes in these variables and to collect measurements over vegetation that will facilitate determination of canopy height, with an accuracy that will allow prediction of future environmental changes and estimation of sea-level rise. The importance of the ICESat-2 project in estimation of biomass and carbon levels has increased substantially, following the recent cancellation of all other planned NASA missions with vegetation-surveying lidars.

Two innovative components will characterize the ICESat-2 lidar: (1) Collection of elevation data by a multi-beam system and (2) application of micropulse lidar (photon counting) technology. A micropulse photon-counting altimeter yields clouds of discrete points, which result from returns of individual photons, and hence new data analysis techniques are required for elevation determination and association of returned points to reflectors of interest including canopy and ground in forested areas.

The objective of this paper is to derive and validate an algorithm that allows detection of ground under dense canopy and identification of ground and canopy levels in simulated ICESat-2-type data. Data are based on airborne observations with a SigmaSpace micropulse lidar and vary with respect to signal strength, noise levels, photon sampling options and other properties. A mathematical algorithm is developed, using spatial statistical and discrete mathematical concepts, including radial basis functions, density measures, geometrical anisotropy, eigenvectors and geostatistical classification parameters and hyperparameters. Validation shows that the algorithm works very well and that ground and canopy elevation, and hence canopy height, can be expected to be observable with a high accuracy during the ICESat-2 mission. A result relevant for instrument design is that even the two weaker beam classes considered can be expected to yield useful results for vegetation measurements (93.01-99.57% correctly selected points for a beam with expected return of 0.93 mean signals per shot (msp9) and 72.85% - 98.68% for 0.48 msp (msp4)). Resampling options affect

results more than noise levels. The algorithm derived here is generally applicable for analysis of micropulse lidar altimeter data collected over forested areas as well as other surfaces, including land ice, sea ice and land surfaces.

(1) Introduction

Determination of vegetation height of the Earth's forests is an essential requirement in estimation of global and regional biomass and carbon levels. Because of the scale of the problem and the inaccessibility of many of the Earth's forested areas, this is best achieved from satellite. NASA's Ice, Cloud and Land Elevation Satellite (ICESat) mission (2003-2009) has resulted in important new findings in ecology ([17, 28, 29, 30, 27, 32, 34, 37, 40, 33, 39]), in addition to many results in the primary mission objectives in cryospheric sciences (e.g. [47, 45, 46, 48, 41, 42, 22, 24, 5, 7, 14, 15, 31, 43, 25, 26] see also http://icesat.gsfc.nasa.gov/publications), ICESat ceased operation in 2009. The National Research Council's "Decadal Survey" [35] has made ICESat-2 one of its first-tier missions citing the urgent need to observe the rapidly changing cryosphere ([44, 36]), with launch currently planned for 2016 ([1, 2]).

Laser altimetry is suited to observe vegetation height and structure, because returned signals include return from the top of the canopy, from within the canopy and from the ground. Therefore the ICESat-2 mission has an ecosystem science requirement, stated as "ICESat-2 shall produce elevation measurements that enable independent determination of global vegetation height with a ground track spacing of less than 2km everywhere over a 2-year period". Based on results from the ICESat mission, which included canopy height estimates with root-mean-square errors of 2-6m ([29, 34, 37, 40]), it is expected that extending the ICESAt-2 mission into a 91-day continuous measurement will facilitate derivation of a vegetation height product with 3-m accuracy at 1-km spatial resolution, especially if off-nadir pointing can be used to increase the spatial distribution of observations over terrestrial regions. There are, however, different requirements in orbit design and sampling for vegetation science and for the ICESat-2 mission's primary, cryospheric objectives ([2, 13, 16, 23]). Hence a different Decadal Survey Mission, Deformation Ecosystem Structure and Dynamics of Ice (DESDynI) was planned to include a lidar specifically designed to measure vegetation height. The importance of the ICESat-2 project in estimation of biomass and carbon levels has increased substantially, following the recent cancellation of all other planned NASA missions with vegetation-surveying lidars, including the DESDynI mission.

Determination of vegetation height from ICESat-2 measurements will be based on determination of canopy and ground elevations. This is not trivial, because ICESat-2 will operate a so-called next-generation lidar, and identification of ground and canopy in the resultant data requires development of new mathematical methods and algorithms. Two innovative components will characterize the ICESat-2 lidar: (1) Collection of elevation data by a multi-beam system and (2) application of micropulse lidar (photon counting) technology. Other than the classic pulse-limited altimeter, a micropulse photon-counting altimeter yields clouds of discrete points, which result from returns of individual photons, and hence new data analysis techniques are required for elevation determination and association of returned points to reflectors of interest including land and sea ice surfaces, ground, tree canopy, water, clouds and blowing snow.

Identification of tree canopy is especially challenging, because of the fuzzy margin of a tree crown, and detection of ground under possibly dense canopy is difficult, because only a small percentage of the originally transmitted photons penetrate the atmosphere and the tree cover, are reflected from the ground and, after reflection, penetrate tree cover and atmosphere again, before reaching the receiver aboard the satellite. Because reflectance of ice surfaces is much higher than of tree crowns and ground, a lower ratio of signal photons to noise photons must be expected for vegetation level determination, and therefore the vegetation algorithm development poses a mathematically more difficult problem than the ice algorithm design. The objective of this paper is to derive and describe a mathematical algorithm that allows detection of ground and canopy in micropulse photon-counting lidar data, of characteristics similar to those that will be expected from ICESat-2, and to apply these to forest data. So that the most challenging cases can be solved, data stem from the Smithsonian Environmental Research Center (SERC) forest, which has a dense canopy.

(2) ICESat-2 instrument design cases and data description

(a) Micropulse photon-counting lidar data. The sensor used in the ICESat mission, the Geoscience Laser Altimeter System (GLAS) [41] was a pulse-limited laser altimeter. Elevation determination is based on analysis of waveforms fitted to the returned signal, the peak of the waveform is associated with geolocation of the "point" (footprint center) from which the signal is returned, and elevation is derived from 2-way travel time associate with the waveform peak. Micropulse photon-counting technology, as pioneered by [8, 12, 10, 9, 11] is realized in an airborne system built by SIGMASpace corporation (and in other instruments). Data collected with the SIGMASpace system, which operates at the 532nm wavelength that will be used for the Advanced Topographic Laser Altimeter System (ATLAS) that is in development for ICESat-2, form the basis of this analysis.

(b) Design cases for a multi-beam sensor for ICESat-2. Designs of a multi-beam system discussed for ICESat-2 include a combination of stronger and weaker beams. Science requirements in ice observation have led to the observation requirement of a multi beam-system, while energy constraints limit the number of equally strong beams to about 2-4. The fact that a lidar system only penetrates thin clouds, but clouds prevail in the Arctic about half of the time, necessitates at least one strong beam. A larger number of beams is needed to observe spatial variability of the ice surface, which provides characteristics indicative of ice types, ice dynamics, morphogenesis of sea ice and other parameters of interest, and improves accuracy of ice elevation mapping and change detection ([18, 1] and other work cited therein ([2]). The combination of these constraints suggests a design that includes beams of different strengths; the two favorites at times of this research were a 9-beam design (with beam strengths 1-2-1; 2-4-2; 1-2-1; i.e. center beam in each row twice as strong as outer beams, yielding corner beams a quarter of the strength of the center beam) and a 6-beam design (with strengths 2-4; 2-4; 12-4; i.e. a weak beam and a strong beam in each row). In this paper, we analyze under which conditions the design cases for the beams can be expected to yield useful data for observation of ground and canopy levels in forests.

(c) Characteristics of the forest type. The dense forests of the Smithsonian Environmental Research Center (SERC), located at (38.889°N, 79.559°W) in eastern North America, have been selected as the test cases for the algorithm development, because detection of ground under canopy is especially difficult for dense canopies. SIGMA data have been collected there during leaf-on conditions.

The SERC forest contains 3350 trees of 84 recognized species on 16 hectare and is situated adjacent to a sub-estuary of the Chesapeake Bay on the coastal plain near Edgewater, Maryland. The square 16 hectare plot is dominated by mature secondary upland forest but is bisected with a section of floodplain forests, both around 120 years since initiation. The upland forest is an example of the"tulip poplar association with an overstory dominated by tulip poplar (Liriodendron tulipifera), several oaks (Quercus spp.), beech (Fagus grandifolia), and several hickories (Carya spp.); a mid-canopy of red maple (Acer rubrum) and sour gum (Nyssa sylvatica); and an understory composed of American hornbeam (Carpinus caroliniana), spicebush (Lindera benzoin), and paw-paw (Asimina triloba). The flood plain forest is dominated by ashes (Fraxinus spp.), sycamore (Platanus occidentalis), and American elm (Ulmus americana). Installation of the plot began in September 2007. The forest is rather tall (to as high as 40 m) and has a high richness for this part of the temperate zone, with more than 34 species of at least 20.0 cm DBH. As of November 2009, the tagging and censusing of all woody stems 1 cm DBH in about 9.0 hectares of the plot have been completed [38]. At time of the survey with the SIGMA photon-counting sensor in October 2009, the SERC forest had reached a mature state with a closed canopy cover (over 95 % canopy closure) and leaves were still on.

(Geoffrey Parker, see

`http://www.ctfs.si.edu/site/SERC%3A+Smithsonian+Environmental+Research+Center; 2-10-2012)`

(d) ICESat-2-type simulated data based on airborne SIGMA data. File name conventions. SIGMA data collected over the SERC forests have been simulated into data sets resembling expected ICESat-2 data and vary with respect to noise levels, radius of photon capture, resampling, laser intensity and expected number of mean photons per shot, subarea/ flight track, and several realizations of random processes ([3, 4]). Prior to simulation, the data were prepared by eliminating most returns above canopy and below ground and many

noise photon returns. Since vegetated targets generate diffuse reflections, it is impossible to completely separate signal photons from noise photons. Observations flights were conducted at dusk, which results in reduction of noise compared to noise from day-time ambient light. Elimination of many (and not all) noise photons above canopy and below ground was performed by manually applying a prescribed, spatially variable range that includes trees and ground surface. The resultant data set is termed data base of signal-only photons.

To match the simulated data to expected ICESat-2 data in spatial distribution, straight-line segments were selected along the aircraft ground track and footprint locations defined every 70 cm by interpolating along the aircraft track. For each footprint location, a Poisson-distributed random function was generated, and photons selected within a cylinder of a given radius. The desired number of signal photons returned per footprint was generated using a Poisson-distributed random number with a mean equal to the signal photons per shop (msp) number appropriate for the surface type, in this case, for vegetation and ground under vegetation ([3], see also A. Martino, AtlasPerformance20100421.xls on the ICEsat-2 website). A desired photon location is determined using a Gaussian-weighted random distribution with a 2-sigma diameter of 10 meters to select a radial distance from the footprint center and a uniform random distribution to select an angle with respect to aircraft ground track. The closest point to that location is determined from the data base of signal-only photons. The region of photon selection is limited to an n-meter radius circle around the desired photon location; the circle is termed the cap size. If no photons are found within this distance, none are selected. Selection is limited to within cap size for two reasons: (1) to avoid selecting photons that are too far away to be selected by the ICESat-2 instrument, and (2) to minimize computer time required to select photons. For dense data sets, cap size can be smaller than for sparse data sets. The radius of photon capture is a cut-off value for the entire simulated data set. Photon selection is repeated for each footprint location, and a photon may be used in several or a single footprint location. Thereafter, noise is added. — In the following description file-name extensions are given in brackets in the order in which they occur in the file names.

Radius of photon capture (cap1), defined as the distance surrounding a location of a cylinder from which the photons are reflected, is 1 meter in all data sets.
Resampling (r0 (no resampling), r1 (resampling) indicates the photon reuse flag, $r0$ indicates that no photon is used more than once, $r1$ indicates that a photon can be selected for any footprint even if it was selected for a previous footprint. Effectively, $r0$ results in fewer recorded photons per shot than $r1$
Laser intensity, quantified as "msp" = mean signal photons expected per shot (p4 – 0.48 msp, p9 – 0.96 msp). Laser intensity is used to characterize the different strengths of the beams considered for the ICESat-2 instrument. Intensity is quantified by a floating point number indicating the mean number of signal photons desired per footprint (per shot). The number of photons selected for any footprint is calculated using a Poisson distributed random function with this mean. Note that this study analyzes the two cases of the weaker beams, as these are the cases limiting instrument design; there is also the case of the strong beam (1.93 msp), for which no simulations are included in the 99 ICESat-2 type SERC data sets (version 2010).
Subarea/ flight track (sb0-1, sb0-3, sb0-5). In the airborne experiment conducted by SIGMASpace Corp. over the SERC forests, data along five tracks were collected and three of those were used to create the simulated data sets.
Randomization instance (s1, s2, s3, s4) refers to a new run of the simulation with a different seed. This allows to run simulations for the same ground track that will select different photons.
Noise levels (uz2 (lowest), uz3(middle), uz5(highest)). Random noise is added in the simulations to mimic different atmospheric conditions, typical of night-time conditions (uz2, 0.5MHz), clear day-time conditions, as encountered on a crisp winter day (uz3, 2MHz) and hazy day-time conditions, as encountered on a humid summer day (uz5, 5MHz). The existence of solar background noise and atmospheric scattering provides one of the main challenges in detection of returns from vegetation and ground under canopy.

(3) Mathematical concepts of the algorithm

Problems that must be addressed in the determination of canopy height from photon-counting lidar data include fuzzyness of tree crowns, poor signal-to-noise ratios in many observational cases, roughness of the ground, trends in slope of the ground over larger distances, and specific density of trees per unit area

which varies with forest type. The mathematical approach uses spatial statistical methods and discrete mathematics, building on concepts similar to those developed by the author for other signal processing and spatial classification problems [20, 21, 19] and developing new concepts where needed. A challenge lies in the implementation of an algorithm that facilitates automatization of a "soft" solution that selects those regions as canopy and ground that visually appear as such in the cases of the stronger beams or less noise, and in addition succeeds at ground and canopy detection even in those cases that cannot be interpreted visually any more. This will be achieved by a combination of a density-quantification that uses radial basis functions, and a generalization of the so-called hyper parameter concept of the geostatistical classification method, adapted and applied to histograms of the density-function results. The idea of the hyperparameter concept is to capture those items that stand out visually [21]. The fact that the ground is in principle a simply-connected feature (in the sense of 1-connectedness in mathematical topology), but may appear as disconnected segments of denser areas in the photon data, calls for a topologically motivated algorithm component. In summary, the computational algorithm developed for SIGMA aircraft data analysis includes the following components (1) anisotropy (eigenvectors), spatial density centers, moving window techniques, (2) analysis of cumulative distribution function, filter, hyperparameter method of geostatistical classification adapted to identify ground/ canopy ranges, (3) density threshold functions for canopy/ ground over background scatterers, (4) linear interpolation on-off, (5) several plotting options and optional data output for comparative analyses and validation.

The algorithm uses the following mathematical concepts, that will be explained in the sequel.

(M.1) globalization-localization paradigm

(M.2) radial basis function (rbf)

(M.3) rbf-density

(M.4) geometrical anisotropy

(M.5) geostatistical classification parameters: slope parameter p_1 and significance parameter p_2

(M.6) hyperparameters

(M.7) application of geostatistical classification ideas to the histogram of the density values (rather than the variogram)

(M.1) Globalization-localization paradigm

A new globalization-localization approach is used to overcome a well-known statistical sampling problem, by disconnecting sampling bases in different steps of the algorithm. The idea here is to treat the following problem, typical of statistical analysis: If the data window (in distance along the flight path) is too small, then not enough photons are available to derive sufficient statistical information to identify ground under canopy. If the data window is too large, then ground and canopy may not be separable any more. The globalization - localization idea used here is to disconnect the two problems, by using a large window (here: an entire data set) to derive a suite of statistical parameters, then in another algorithm step employ a local classification or detection algorithm that utilizes the globally derived parameters. Future ICESat-2 data are expected to be much larger data sets than the simulated data analyzed here, and hence the globalization-localization paradigm can be implemented using a large moving window in the along-track direction combined with smaller windows inside that.

(M.2) Radial basis function

A *radial basis function (rbf)* if a real-valued function where the value depends on distance from the origin

$$\Phi(x) = \Phi(\|x\|) \tag{1}$$

for all x in a definition area $\mathcal{D}$, or on distance from a *center* $c \epsilon \mathcal{D}$

$$\Phi(x, c) = \Phi(\|x - c\|) \tag{2}$$

with respect to any norm $\|\cdot\|$.

In the application, we will utilize a Gaussian radial basis function (letting $r = x - c$)

$$\Phi(r) = e^{-(\frac{r}{\sqrt{2}s})^2} \tag{3}$$

where s is derived as given below.

Visualized as a surface in $\mathcal{R}^3$, this rbf has the shape of (half) a Gaussian bell curve rotated around the location of a center $c \epsilon \mathcal{R}^2$. In the photon-data analysis, we have $c \epsilon \mathcal{R}^3$ and the surface is in $\mathcal{R}^4$. More formally, the Gaussian probablility density function is

$$f_{normpdf} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\frac{x-\mu}{\sqrt{2}\sigma})^2} \tag{4}$$

with the standard deviation σ and mean μ; replacing $\sigma = s$ and $\mu = 0$ yields eqn (1):

$$\Phi(r) = \sqrt{2\pi} f_{normpdf} \tag{5}$$

(see [6]).

(M.3) Density centers

Identification of points within tree crowns is motivated by the observation that a tree crown is a diffuse reflector, but points within the tree crown have a high probability of being located within clusters of other parts of the tree-crown, a property that does not hold for reflections of ambient light or noise outside of the tree crowns. To identify points located inside clusters or clouds of points with higher density, the *rbf* concept is applied as follows:

For the photon-data analysis problem, the definition set $\mathcal{D}$ is the set of all photons (in a track or window). For each point $c \epsilon \mathcal{D}$, a density value $f_d(c)$ is calculated by summing up *rbf* values for all neighbors within a 15 m radius:

$$f_d(c) = \sum_{x \epsilon \mathcal{D}_c} \Phi(\|x - c\|_a) \tag{6}$$

with $\mathcal{D}_c = \{x \epsilon \mathcal{D} : \|x - c\|_2 \leq 15m\}$ the set of all points within a given radius (here: 15 meters) from the center point c (note that in this initial distance determination simply the 2-norm (Euclid norm) $\|\cdot\|_2$ is used). In the radial basis function, we use a norm $\|\cdot\|_a$ that takes anisotropy into account. — The concept of density centers is illustrated in Figure 1a,b.

FIGURE 1 here – fig:data-density-histo-lines

(M.4) Anisotropy norm

Using an anisotropy norm is motivated by the notion that tree canopy has a tendency to extend more in the horizontal direction than in the vertical direction. When the anisotropy norm is combined with the radial basis function, points found in a horizontal direction from the center point are weighted higher than points found in a vertical direction. The following algorithm implements a matrix multiplication that is an affine transformation of the density function (the radial basis function) into a function of ellipsoidal shape.

This is implemented by the following algorithm: The anisotropy norm is defined as

$$\|v\|_a = \|A(v)\|_2 \tag{7}$$

for any vector $v\epsilon\mathcal{R}^3$, with a transformation matrix

$$A = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{8}$$

This is applied to the density centers c and all their neighboring points in eqn. (6) as

$$\|x - c\|_a = \|A(x - c)\|_2 \tag{9}$$

Points of the same *rbf* value $\Phi(\|x - c\|_a)$ are now located on an ellipsoid with axes (3,3,1) around the center point c and (half) Gaussian bell curves along each radial line. The density value $f_d(c)$ then reflects the tendency of tree crowns to connect horizontally into forest canopies. (The same anisotropy norm is used for ground, as ground continues more in horizontal direction. For terrain with a high topographic relief, the anisotropy matrix A can be set to a different value, or to identity.)

(M.5-M.7) Geostatistical classification ideas and their application to histogram analysis

Several algorithm concepts are inspired by concepts of the geostatistical classification method ([21, 20]) and modified to solve the lidar-data analysis problem. Analysis of the variogram or its generalization, the vario function, lies at the basis of the geostatistical classification, but some of the principles transfer to any function that is affected by noise and here are applied to the histogram of the data and the histogram of density. More generally, we may consider any positive real-valued discrete function, $f(x_i)$, defined for values x_i, i=1,n.

The geostatistical classification proceds by analysis of sequences of minima and maxima in the vario function, derivation of parameters from those sequences, construction of a feature vector from the parameters, and classification or class association based on the feature vector. A related problem in signal processing is the analysis of a time series or recording of a time-variable signal, which is often based on the analysis of the minima and maxima of the signal.

(M.5) Geostatistical classification parameters

Let $f(x_i)$ be a positive real-valued discrete function defined for values x_i, i=1,n. This function may be a histogram, a variogram or a vario function. We introduce classification parameters used in the photon-classification problem. The *mindist* parameter is defined as the lag of the first minimum after the first maximum in the function. *mindist* gives the spacing of parallel features recorded in the function. We further define the significance parameters $p1$ and $p2$:

$$p1 = \frac{f(x_{max_1}) - f(x_{min_1})}{x_{max_1} - x_{min_1}} \tag{10}$$

$$p2 = \frac{f(x_{max_1}) - f(x_{min_1})}{f(x_{max_1})} \tag{11}$$

$p1$ is the slope parameter and $p2$ the relative significance of the first minimum min_1 after the first maximum max_1. In this notation,

$$mindist = x_{min_1} \tag{12}$$

Parameters of types $p1$ and $p2$ can be calculated for any max-min sequence, defining

$$pt1(max_i, min_j) = \frac{f(x_{max_i}) - f(x_{min_j})}{x_{max_i} - x_{min_j}} \tag{13}$$

and an analogon to eqn. (11) for $p2$-type parameters, for $i \leq j$ and the convention that minimum min_i always follows maximum max_i. Note that slope parameters involve distance and $p2$-type parameters do not.

(M.6) Hyperparameters

A problem typical of the analysis of complex and noisy processes or data sets lies in the fact that the maxima and minima that tell the "story" of the problem can be identified visually because they stand out, but are numerically obscured by noise or by other processes that may interfere with the main process of interest. In the lidar data analysis, we use a robust search algorithm to automatically identify "bigger" max-min sequences and associated generalized parameters, as described in [21]. We determine *bigmax*, the largest maximum in a group of g maxima, and then *bigmin*, the smallest minimum in a group of g minima following *bigmax*. For a fixed groupsize g, a sequence of *bigmax*s and *bigmins* can be determined, and the selected ones are those which survive several increases of the groupsize. The optimal groupsize for a given problem can be determined automatically, here we have applied a criterion to find stable groupsizes such that the *bigmax-bigmin* pair stays the same for 3 consecutive groupsizes. The so-determined parameters $bigmax_i, bigmin_i$, $i = 1, n$ are termed hypermaxima and hyperminima. For these selected hypermaxima and hyperminima, hyperparameters are defined as generalizations of eqns. (10, 11,13):

$$pt1(bigmax_i, bigmin_j) = \frac{f(x_{bigmax_i}) - f(x_{bigmin_j})}{x_{bigmax_i} - x_{bigmin_j}} \tag{14}$$

and

$$pt2(bigmax_i, bigmin_j) = \frac{f(x_{bigmax_i}) - f(x_{bigmin_j})}{f(x_{bigmax_i})} \tag{15}$$

(M.7) Application to histograms of forest lidar data and density

The hyperparameter concept is applied to identify the two main maxima in the histogram, which represent ground and canopy. It is a necessary piece in the analysis, because even after filtering, many histograms of forest lidar data have several maxima that may be identified as ground or canopy (see Figure 1b). – The geostatistical classification concepts are applied to the histograms of elevation values and to the histograms of density values (see sections M.3 and M.4).

(4) Algorithm steps

The algorithm proceeds by the following steps:

(1) *Import data:* Data are recordings returns of individual photons, with $P = (x_1, x_2, z)$ the location of the reflector in three dimensions, $z = z(X) = z(x_1, x_2)$ is the elevation value of a photon in location P and $X = (x_1, x_2)$ is the projection of the photon's location onto the ground. Data are loaded into the program.

(2) *Identification of ground and canopy elevation ranges by histogram analysis of photon elevation data:*

(2a) A histogram of the elevation values of received photons is created, grouped by elevation bins. Here, we used 100 elevation bins for a total elevation range of 100 m (bin size 1 m).

(2b) The histogram is filtered using a Butterworth filter with $\alpha = (0.0625, 0.25, 0.375, 0.25, 0.0625)$.

(2c) In the next step, two hypermaxima are identified ($bigmax_1$ and $bigmax_2$). These are the two maxima that stand out visually, and will represent ground and canopy elevation centers (see mathematical concept *hyperparameters*). For the ground and canopy-range-detection problem, the hyperparameter location algorithm is adapted from that described in Herzfeld et al. 2006b for hyperparameters of vario functions.

For the ground and canopy-range-detection problem, the following algorithm is used to determine the hypermaxima locations by an iterative process: In the first iteration, group size is $g_1 = 1$, and

all local maxima in the histogram are identified and written into an index list. To go from step $(n-1)$ to step n of the iteration, the following is used: Given a list of maxima in the index set I_{n-1}, the group size is increased $g_n = g_{n-1} + 1$ and the largest maximum within each group of g_n maxima in the original list is determined and written into Index set I'_n. A maximum is retained in list I_n, if it was already in the previous list:

$$I_n = I_{n-1} \cap I'_n \tag{16}$$

Iteration is continued until at most 2 maxima are left (n_b is the index of the break point of the iteration):

$$|I_{n_b}| < 2 \tag{17}$$

Noting that I_{n_b-1} may contain more than two maxima, the two most significant maxima in I_{n_b-1} are selected, using a $param_2$-type criterion (see mathematical concept significance parameter p_2), with the constraint that the final two maxima must be at least 8 histogram bins apart. (This corresponds to 8 m in the SERC study and is easily changed.)

The hypermaxima are identified in the histograms in Figure 2; panel b demonstrates that it is necessary to determine the hypermaximum in a series of maxima that remain after Butterworth filtering. After application of this step, two "elevation centers" are identified, $bigmax_g$ and $bigmax_c$ with $bigmax_g < bigmax_c$ and corresponding x-locations $bigmax_g(x_g)$ and $bigmax_c(x_c)$.

FIGURE 2 here (histo-cases-fig)

(2d) The process for determining a canopy elevation range and a ground elevation range, described in this step, is illustrated in Figure 2 (histo-fig, panels c and d). Colored lines are used for illustration. First, the minimum $z_min(x_0)$ between the ground and canopy centers, $bigmax_g$ and $bigmax_c$, is determined. Then the minimum is mirrored around the ground and canopy center locations, as $x_{green_g} = x_g - (x_0 - x_g)$ and $x_{green_c} = x_c + (x_c - x_0)$. The green lines are placeholders for finding the range values. Three local minima closest to the green lines are identified in I_0, the one with the lowest minimum is termed $z_{red}(x_{red})$ (this is a hyperminimum), the one with the steepest slope to the associated hypermaximum is termed $z_{yellow}(x_{yellow})$ (this utilizes a p_1-type criterion). Finally, the range limits are determined using the slope values from the "red" and "yellow" points to the hypermaxima:

$$z_{final}(x_{final}) = z_{red}(x_{red}) \tag{18}$$

if

$$0.8pt1(bigmax_c, z_{yellow}(x_{yellow})) < pt1(bigmax, z_{red}(x_{red})) < 1.2pt1(bigmax_c, z_{yellow}(x_{yellow})) \tag{19}$$

and

$$z_{final}(x_{final}) = z_{yellow}(x_{yellow}) \tag{20}$$

otherwise. The elevation range for ground is determined analogously.

(3) *Segmentation of the data set into ground and canopy range sets.* The ground and canopy elevation ranges determined in step (2) are applied to segment the photon data set into a canopy range set and a ground range set, and a rest class (elevations higher than canopy range or lower than ground range). It is worth emphasizing that the ground and canopy range sets are not a classification of photons into ground and canopy returns, but a *segmentation* of the global data set into sets in which ground and canopy can be found.

The next analysis steps are then carried out separately for the ground range set and the canopy range set.

Globalization -localization. Note that the segmentation algorithm can be applied in a window. For the SERC data, the algorithm steps 1-3 have been applied globally. The following steps (4)-(9) are applied

in a localization. This allows to use properties of a larger window, or the whole data set, for a first identification of elevation points in a likely range, based on the histogram analysis. Then in the second part of the algorithm, different mathematical concepts are applied to identify points that are ground and canopy reflectors.

(4) *Apply density function for canopy center identification.* Density values $f_d(c)$ are calculated as described in section (M.3), using the radial basis function (equation (6)) for all points in a 15 m radius. For the function evaluation, the distance values transformed according to the anisotropy norm described in section (M.4) are employed. The sum of all rbf values of all neighbors of a point is called (rbf-)density of that point.

(5) *Histogram of photon density in the canopy and ground region.* A histogram $H(d)$ of the density values, d (Step (4)) for photon events in the canopy region is calculated in 100 evenly spaced bins and filtered using the Butterworth filter with the same values $\alpha = (0.0625, 0.25, 0.375, 0.25, 0.0625)$ as in step (2b). The maximal histogram value is identified as $H_{max}(d_m)$ where d_m is the density value for which the maximum occurs.

Then a canopy threshold d_c is set:

Let $H_c = 0.8H_{max}$ and determine d_c as the density value with $H(d_c) = H_c$ and $d_c \geq d_m$.

For ground threshold d_g, a factor of 0.5 is used: Let $H_g = 0.5H_{max}$, where d_g is the density value with $H(d_g) = H_g$ and $d_g \geq d_m$. Note that a lower percentage of the histogram's maximum results in a higher threshold. Figure 3 illustrates this step for ground detection.

FIGURE 3 here

---

(6) *Apply noise filter.* The density value d_m, for which the largest density count occurs (as defined in step (5)) is used as a noise threshold and points with density less than d_m are rejected.

(7) *Re-compute density function.* To eliminate possible high-density noise clusters, the density function (eqn. (6)) is applied a second time, as described above (including anisotropy norm). A high-density point with only noise-type density neighbors will be reassigned a much lower density value in this second run of density, compared to the first run. — We write d_x for the density value of a point $(z(x), x)$ after the second run of the radial basis function f_d.

(8) *"Build line" : Canopy class association (more clearly: define the set of discrete points that are in the canopy class).* A point $(z(x), x)$ is identified as a member of the canopy set, if all of the following hold:

(i) $d_x > d_c$ (and $d_x > d_g$ for ground)

(ii) $(z(x), x)$ is the point with maximal density in a 10 m along-track interval

(iii) a rigidity criterion is satisfied.

The rigidity criterion fixes a maximal elevation difference, that is likely to occur among photons reflected from the same tree or neighboring tress, as $|z(x_i) - z(x_{i-1})| \leq rig$ for $z(x)$ elevation in location x. — The rigidity condition may be adjusted to match forest types, for instance, mapping needle trees in sparse stands may require a higher rigidity number than leaf trees in dense forests. The rigidity condition can be relaxed entirely.

(9) *Ground detection.* To detect ground under canopy and associate discrete photon points to the ground class, steps (4)-(8) are repeated, using the ground parameters. — Canopy and ground lines are illustrated in Figure 1c.

(4) Results

In this section, we analyze under which conditions the design cases for the beams (see section (2)) can be expected to yield useful data for observation of ground and canopy levels in forests. We present results of several case studies, selected from a total of 99 test cases of simulated data. In the first case study, typical cases of the medium-strong beam, labeled "p9" are investigated (Figure 4).

FIGURE 4 here

This figure demonstrates that the algorithm works for the medium-strong beam, the two options of resampling (without (r0) and with (r1) resampling), and increasing noise levels. The plots show the simulated data in the top panel and the interpretation of ground and canopy by the detection algorithm. Points that are original signal points in SERC forest observations are colored red, while noise points resultant from the simulation are shown in black. The information signal-versus-noise was not used in the algorithm, but aid in visually assessing validity of the algorithm. Information on ground versus canopy, or reflections of other items (birds, rocks or other features, atmospheric reflections) is not provided. in this section, visual validation is used; statistical validation is given in section (5).

In all cases, the level of the canopy is well-detected by the algorithm. The canopy assumes similar shapes in all cases, despite of increasing noise and two different sampling strategies. The resampling flag "r1" indicates that resampling is allowed, which increases the signal to noise ration. Cases labeled "r0" (no resampling) constitute a weaker signal, given the same noise level (left column of figure panels, a, c, e). At the start of the window, no canopy data are identified, however, this matches the visual impression. In the case studies, an entire flight segment of 2500 m is analyzed. For actual satellite or aerial observation data sets, a moving window algorithm will be implemented, which will eliminate edge effects that occur in the shorter segments analyzed here.

Detection of ground level under canopy also works well, the number of points identified as canopy, however, shows some variability. The software includes a simple, piecewise-linear interpolation option, that allows to continue ground level across large gaps (over 400 m in 4a, over 600 m in 4b). Even in the worst case of combining no resampling of beams with highest noise levels, the detection of ground and canopy works. Since the ground and canopy detection works for both resampling options, science or engineering criteria can be employed for deciding between the two resampling options.

To investigate whether it may be possible to utilize the weaker beams in the ICESat-2 sensor panel and still expect to detect ground under canopy in observations of forests, a case study for the weakest beams ("msp4") is conducted, using the same algorithm as for stronger beams. Figure 5 illustrates how the algorithm performs in the worst cases of the weakest beam (mp4 and r0), increasing noise. Notably the algorithm functions for all three noise levels, and automated detection exceeds the possibilities of visual detection of ground and canopy. In comparison to results in figure 4, ground can still be detected in sufficiently many locations to derive ground level, but there is a tendency for noise clusters to be misidentified as ground. Canopy continues to be correctly identified. Quantifications of these statements will be given in section (5) on validation. The results are encouraging to include the weakest beams in the instrument panel for ICESat-2.

FIGURE 5 here

Introduction of a flexible racking rigidity parameter serves two purposes: (a) options in detection of canopy and ground for weak noise in case of weakly non-stationary ground and canopy levels, and (b) a possibility to match forest-type specific characteristics. To give an example of the latter, a forest with wide-standing conifers or pines may result in lidar data that show individual trees, hence a large slope outlining tree shape may be appropriate, and consequently a large rigidity parameter will be helpful. A forest with a dense leaf-

tree canopy of homogeneous age typically has a narrow range of crown-top elevations, which is better detected by a lower rigidity parameter. Figure 6a,b illustrates the effect of using two different rigidity parameter values for analysis of the same data set. Figure 7 shows that the rigidity parameter can be employed to improve ground and canopy detection for the weakest beam (mp4) combined with the no-resampling option (r0), that effectively yields fewer signal photons, and the highest noise level (uz5) (cf. Figure 5).

FIGURE 6 here

FIGURE 7 here

---

(5) Validation

To facilitate algorithm validation, the original signal points are flagged in the simulated ICESAT-2-type data sets (column with a 0-1 flag). This information was not used in the detection and classification algorithm and can therefore serve for validation of the algorithm. Results of the validation are given in Table 1, for the following statistical parameters, calculated separately for ground and canopy: (1) percentage of points selected that are signal points; and (2, 3) distance in meters from a point that has been identified as a signal point to the nearest point that is a signal point, given as mean and median of nearest-neighbour distances in 3-dimensional space. Note that the distances in (2, 3) are not elevation errors. — Results listed in Table 1 are summarized from results obtained for all 99 data sets, so that performance for weak beams (msp4), medium-strength beams (msp9), resampling options and the three noise levels can be analyzed.

TABLE 1 here

For ground, the percentage of correctly selected points is 94.7% to 99.47% for all groups of msp9-strength beams and 85% to 98.81% for all groups of msp4-strength beams. The average value over all data sets in a group is 95.89% for $(p9, r0)$ for any noise level, 99.17% for $(p9, r1)$ and 95.53% for all $p9$ cases. The average value over all data sets in a group is 88.44% for $(p4, r0)$ for any noise level, 98.34% for $(p4, r1)$ and 92.68% for all $p4$ cases. The median distance from a point in the selected set to the nearest neighbor in the signal points set is always zero, the mean distance is 0.20 m to 0.55 m; the resampling option has a stronger effect than the noise level. The validation demonstrates that the algorithm works very well for detection of ground under canopy. The elevation error is a lot smaller than the distance numbers, but has not been calculated directly, because the piece-wise linear interpolation is only included for visualization of the ground and canopy lines, and the objective of the paper is to design a ground detection algorithm, not an interpolation algorithm.

For ground, the percentage of correctly selected points is 93.01% to 99.57% for all groups of msp9-strength beams and 72.85% to 98.68% for all groups of msp4-strength beams. The average value over all data sets in a group is 94.23% for $(p9, r0)$ for any noise level, 99.19% for $(p9, r1)$ and 96.71% for all $p9$ cases. The average value over all data sets in a group is 80.26% for $(p4, r0)$ for any noise level, 98.07% for $(p4, r1)$ and 87.89% for all $p4$ cases. The median distance from a point in the selected set to the nearest neighbor in the signal points set is zero for all p9 cases and for all $(p4, uz2)$ cases, it is 0.25 m for the average of all p4 cases. The mean distance is 0.44 m for the average of all msp9 cases, lowest for $(p9, r1, uz2)$ at 0.18 m and highest for $(p9, r0, uz5)$ at 0.83 m. The mean distance is 1.02 m for the average of all msp4 cases, lowest for $(p4, r1, uz2)$ at 0.25 m and highest for $(p4, r0), uz5$ at 2.26 m. In all cases, the resampling option has a stronger effect on the accuracy than the noise level. This is a good result, because the resampling option can be set in the instrument-level detection algorithm, whereas noise from ambient light and atmospheric conditions is an environmental constraint that is corrected for in the data analysis.

The results of the detection algorithm are also very good for the medium strength beams, with similarly good values as the results for the ground detection. For the weakest beams (msp4), the canopy detection is not quite as accurate, which may be explained by the fact that canopy is fuzzy and has a much larger diameter than the ground (the theoretical diameter of the ground line is zero, but the practical is not) and that the sparse canopy returns have to be extracted from many noise points. Even in this hardest case, the average distance is 1.02 m.

The data set provided does not identify a ground data set and a canopy data set, hence the classification part of the algorithm cannot be validated numerically. Visual inspection of the results indicates that the canopy-class signal points fall in the upper layer and the ground-class identified points fall in the lower layer, and the continuity of the layers indicates that the classification works correctly. As component of the experimental part of the pre-launch phase ICESat-2 project, validation data sets and instrument test data sets will be collected. To complement future flights with the airborne Multiple Altimeter Beam Experimental Lidar (MABEL), the first photon-counting multi-beam sensor, validation flights with vegetation lidars of a known performance are planned to be carried out.

(6) Summary and Conclusions

In this paper, a set of algorithms has been developed and validated that allows detection of ground under dense canopy and identification of ground and canopy levels in simulated ICESat-2-type data. These data constitute a new type of lidar altimeter data that will be collected during the ICESat-2 mission with a next-generation multi-beam micropulse lidar altimeter. Data analyzed in this paper are based on airborne observations with a SigmaSpace micropulse lidar and simulations vary with respect to signal strength, noise levels, photon sampling options and other properties. To consider the mathematically most difficult cases, (a) data stem from dense forests observed during leaf-on conditions and (b) the cases of the two weaker beam types are analyzed; these are: (1) a beam with expected return of 0.93 mean signals per shot (msp9) and (2) a beam with 0.48 msp (msp4). The third case is a beam with 1.93 msp; this will be used in the ICESat-2 instrument design in any case. The stronger beam (msp9) corresponds to the weaker beam in a design of a 6-beam proposed sensor for ICESat-2, whereas an alternative proposed 9-beam design for an ICESat-2 sensor includes 4 corner beams of strength msp4, 4 middle beams of strength msp9 and a center beam with a signal rate of 1.93msp.

A mathematical algorithm is developed using an approach that combines spatial statistical and discrete mathematical concepts, including radial basis functions, density measures, geometrical anisotropy, eigenvectors and geostatistical classification parameters and hyperparameters. Piecewise linear interpolation is provided as an option to bridge between identified ground points and analogously, canopy centers. The software allows flexibility with respect to output types, which include graphics options and data output for validation and canopy height/ ground elevation determination.

Validation using 99 simulated data sets shows that the algorithm works very well and that ground and canopy elevation, and hence canopy height, can be expected to be observable with a high accuracy during the ICESat-2 mission. A result relevant for instrument design is that even the two weaker beam classes considered can be expected to yield useful results for vegetation measurements (93.01-99.57% correctly selected points for a beam with expected return of 0.93 mean signals per shot (msp9)) and 72.85% - 98.68% for 0.48 msp (msp4). The median distance from a point in the selected set to the nearest neighbor in the signal points set is zero for all msp9 cases and for low-noise msp4 cases, 0.25 m for the average of all msp4 cases. The mean distance is 0.44 m for the average of all msp9 cases and 1.02 m for the average of all msp4 cases. Notably, this is a three-dimensional distance error and not an elevation error; the expected elevation error average is lower. In all cases, the option of resampling versus using each detected photon exactly once has a stronger effect on the accuracy than the noise level. Following our analysis, ground and canopy detection and hence determination of canopy height is possible in all noise conditions. The resampling option can be set in the instrument-level detection algorithm.

Because detection of ground and canopy in forested areas presents a technically and mathematically harder problem than detection of the surface in data collected over land ice and sea ice and most other land surfaces, the algorithm presented here can be expected to be applicable also for land ice, sea ice and land surface

detection and elevation determination. As tree canopy may be considered a diffuse reflector, the algorithm may be generalized for other complex and diffuse reflectors, such as rough ice surfaces and atmospheric reflectors including as clouds and blowing snow. In summary, the algorithm derived here can be used as a basis for an algorithm for the analysis of data from the ICESat-2 mission, data from the mission's airborne precursor instrument, the Multiple Altimeter Beam Experimental Lidar (MABEL), and for analysis of micropulse lidar altimeter data in general.

Acknowledgements. Support provided by the ICESat-2 Project, NASA Goddard Space Flight Center, Awards NNX10AR46G and NNX11AH43G S01 to U.C. Herzfeld is gratefully acknowledged. Work presented here was conducted in collaboration and exchange with the ICESat-2 Project, the ICESat-2 Science Definition Team and SigmaSpace Corporation. Thanks for helpful discussions to Waleed Abdalati, Bea Csatho and the ICESat-2 Science Definition Team, especially the vegetation subgroup of the team, Ross Nelson, Amy Neuenschwander, Birgit Peterson, Michael Lefsky, and to Kristine Barbieri, David Harding and Anthony Martino (NASA GSFC).

References

[1] W. Abdalati, , R. Bindschadler, C. Carabajal, B. Csatho, M. DiJoseph, H.A. Fricker, D. Harding, D. Hancock, U. Herzfeld, W. Krabill, R. Kwok, M. Lefsky, T. Markus, A. Marshak, S. Martin, A. Neuenschwander, S. Palm, J. Ranson, R. Schutz, M. Simard, B. Smith, J. Spinhirne, T. Urban, C. Webb, and J. Zwally. Report of the ad-hoc science definition team for the ice cloud and land elevation satellite-ii (icesat-ii). 2008.

[2] W. Abdalati, H.J. Zwally, R. Bindschadler, B. Csatho, S. Farrell, H. Fricker, D. Harding, R. Kwok, M. Lefsky, T. Markus, A. Marshak, T. Neumann, S. Palm, R. Schutz, B. Smith, J. Spinhirn, and C. Webb. The icesat-2 laser altimetry mission. *Proceedings of the IEEE*, 98(5):735–751, 2010.

[3] K. Barbieri, A. Brenner, T. Markus, T. Neumann, J. Saba, and D. Yi. Description of subsampling algorithm for sigma space data. Technical report, NASA Goddard Space Flight Center, November 2009.

[4] K. Barbieri, A. Brenner, T. Markus, T. Neumann, J. Saba, D. Yi, and K. Brunt. Description of icesat-ii simulated data created from sigma space mpl laser data. Technical report, NASA Goddard Space Flight Center, August 2010.

[5] A.C. Brenner, J.P. DiMarzio, and H.J. Zwally. Precision and accuracy of satellite radar and laser altimeter data over the continental ice sheets. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(2):321–331, 2007.

[6] M. D. Buhmann. Radial basis functions. *Acta Numerica*, pages 1–38, 2000.

[7] C.C. Carabajal and D.J. Harding. Icesat validation of srtm c-band digital elevation models. *Geophysical Research Letters*, 32(22):L22S01, 2005.

[8] J.J. Degnan. Unified approach to photon-counting microlaser rangers, transponders, and altimeters. *Surveys in geophysics*, 22(5):431–447, 2001.

[9] J.J. Degnan. A conceptual design for a spaceborne 3d imaging lidar. *e & i Elektrotechnik und Informationstechnik*, 119(4):99–106, 2002.

[10] J.J. Degnan. Photon-counting multikilohertz microlaser altimeters for airborne and spaceborne topographic measurements. *Journal of Geodynamics*, 34(3-4):503–549, 2002.

[11] J.J. Degnan, R. Machan, E. Leventhal, D. Lawrence, G. Jodor, and C. Field. Inflight performance of a second-generation photon-counting 3d imaging lidar. In *Proceedings of SPIE*, volume 6950, page 695007, 2008.

[12] J.J. Degnan, J. McGarry, T. Zagwodzki, P. Dabney, J. Geiger, R. Chabot, C. Steggerda, J.Marzouk, and A. Chu. Design and performance of an airborne multikilohertz photon-counting microlaser altimeter. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY REMOTE SENSING AND SPATIAL INFORMATION SCIENCES*, 34(3/W4):9–16, 2001.

[13] R. Dubayah and J. Drake. Lidar remote sensing for forestry. *J. Forestry*, 98(6):44–46, 2000.

[14] R. Forsberg and H.Skourup. Arctic ocean gravity, geoid and sea-ice freeboard heights from icesat and grace. *Geophys. Res. Lett*, 32(21), 2005.

[15] H.A. Fricker and L. Padman. Ice shelf grounding zone structure from icesat laser altimetry. *Geophys. Res. Lett*, 33:15, 2006.

[16] F.G Hall, K. Bergen, J. B. Blair, R. Dubayah, R. Houghton, J. Kellndorfer, M. Lefsky, J. Ranson, H.H Shugart S. Saatchi, , and D. Wickland. Characterizing 3d vegetation structure from space: Mission requirements. *Remote Sens. Environ*, 115:2753–2775, 2011.

[17] D.J. Harding and C.C. Carabajal. Icesat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical research letters*, 32(21):L21S10, 2005.

[18] U. Herzfeld, B. Wallin, and M. Stachura. Applications of geostatistics in optimal design of satellite altimetry orbits and measurement configurations. *J. Astronautical Sciences*, 2011.

[19] U. C. Herzfeld. Vario functions of higher order–definition and application to characterization of snow surface roughness. *Computers & Geosciences*, 28(5):641–660, 2002.

[20] U. C. Herzfeld. Master of the obscureautomated geostatistical classification in presence of complex geophysical processes. *Mathematical Geosciences*, 40(5):587–618, 2008.

[21] U. C. Herzfeld, J. A. Maslanik, and M. Sturm. Geostatistical characterization of snow-depth structures on sea ice near point barrow, alaska-a contribution to the amsr-ice03 field validation campaign. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11):3038, 2006.

[22] U. C. Herzfeld, P. J. McBride, H.J. Zwally, and J. Dimarzio. Elevation changes in pine island glacier, walgreen coast, antarctica, based on glas (2003) and ers-1 (1995) altimeter data analyses and glaciological implications. *International Journal of Remote Sensing*, 29(19):5533–5553, 2008.

[23] P. Hyde, R. Dubayah, B. Peterson, J. B. Blair, M. Hofton, C. Hunsaker, R. Knox, , and W. Walker. Mapping forest structure for wildlife habitat analysis using waveform lidar: validation of montane ecosystems. *Remote Sens. Environ.*, 96:427–437, 2005.

[24] N.T. Kurtz, T. Markus, D.J. Cavalieri, W. Krabill, J. Sonntag, and J. Miller. Comparison of icesat data with airborne laser altimeter measurements over arctic sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):1913–1924, 2008.

[25] R. Kwok, G.F. Cunningham, H.J. Zwally, and D. Yi. Icesat over arctic sea ice: Interpretation of altimetric and reflectivity profiles. *Journal of geophysical research*, 111(C6):C06006, 2006.

[26] R. Kwok, G.F. Cunningham, H.J. Zwally, and D. Yi. Ice, cloud, and land elevation satellite (icesat) over arctic sea ice: Retrieval of freeboard. *J. Geophys. Res*, 112:C12013, 2007.

[27] M.A. Lefsky. A global forest canopy height map from the moderate resolution imaging spectroradiometer 10 and the geoscience laser altimeter system 11. *Geoscience Research Letters*, 37, 2010.

[28] M.A. Lefsky, D.J. Harding, M. Keller, W.B. Cohen C. C. Carabajal F. Del Bom Espirito Santo M. O. Hunter, and R. de Oliveira Jr. Estimates of forest canopy height and aboveground biomass using icesat. *Geophysical Research Letters*, 32(22):L22S02, 2005.

[29] M.A. Lefsky, M. Keller, Y. Pang, P. de Camargo, and M.O. Hunter. Revised method for forest canopy height estimation from geoscience laser altimeter system waveforms. *Journal of Applied Remote Sensing*, 1(1), 2007.

[30] M.A. Lefsky and M. McHale. Volume estimates of trees with complex architecture from terrestrial laser scanning. *Journal of Applied Remote Sensing*, 2:19, 2008.

[31] S.B. Luthcke, D.D. Rowlands, T.A. Williams, and M. Sirota. Reduction of icesat systematic geolocation errors and the impact on ice sheet elevation change detection. *Geophys. Res. Lett*, 32(21), 2005.

[32] M.E. Miller, M.A. Lefsky, and Y. Pang. Optimization of geoscience laser altimeter system waveform metrics to support vegetation measurements. *Remote Sensing of Environment*, 115:298–305, 2011.

[33] R. Nelson, K. J. Ranson, G. Sun, D. S. Kimes, V. Kharuk, and P. Montesano. Estimating siberian timber volume using modis and icesat/glas. *Remote Sens. Environ.*, 113(3):691–701, 2009.

[34] A. L. Neuenschwander. Evaluation of waveform deconvolution and decomposition retrieval algorithms for icesat/glas data. *Can. J. Remote Sens.*, 34(suppl. 2):S240–S246, 2008.

[35] National Research Council (US). Committee on Earth Science, Applications from Space, A Community Assessment, and Strategy for the Future. *Earth science and applications from space: national imperatives for the next decade and beyond*. National Academy Press, 2007.

[36] R.K. Pachauri and A. Reisinger. Climate change 2007: Synthesis report. contribution of working groups i, ii and iii to the fourth assessment report of the intergovernmental panel on climate change. *Geneva: IPCC*, 2007.

[37] Y. Pang, M. Lefsky, H. E. Anderson, M. E. Miller, and K. Sherrill. Validation of the icesat vegetation product using crown-area-weighted mean height derived using crowd delineation with discrete return lidar data. *Can. J. Remote Sens.*, 34(suppl. 2):S471–S484, 2008.

[38] G. Parker. Serc:smithsonian environmental research center. *http://www.ctfs.si.edu/site/SERC%3A+Smithsonian+Enviro* February 2012.

[39] S. Popescu, K. Zhao, A. Neuenschwander, , and C. Lin. Satellite lidar vs. small-footprint airborne lidar: comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sens. Environ.*, 115:2786–2797, 2011.

[40] J. A. B. Rosette, P. R. J. North, and J. C. Suareze. Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *Int. J. Remote Sens.*, 29(5):1475–1493, 2008.

[41] B.E. Schutz, H.J. Zwally, C.A. Shuman, D. Hancock, and J.P. DiMarzio. Overview of the icesat mission. *Geophys. Res. Lett*, 32(21), 2005.

[42] C.A. Shuman, H.J. Zwally, B.E. Schutz, A.C. Brenner, J.P. DiMarzio, V.P. Suchdeo, and H.A. Fricker. Icesat antarctic elevation data: Preliminary precision and accuracy assessment. *Geophysical research letters*, 33(7):L07501, 2006.

[43] B.E. Smith, C.R. Bentley, and C.F. Raymond. Recent elevation changes on the ice streams and ridges of the ross embayment from icesat crossovers. *Geophys. Res. Lett*, 32:21, 2005.

[44] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Avery, M. Tignor, and H.L. Miller. *Climate change 2007: the physical science basis: contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2007.

[45] H.J. Zwally, T. Abdalati, K. Herring, J. Larson, J. Saba, and K. Steffen. Surface melt-induced acceleration of greenland ice-sheet flow. *Science*, 297(5579):218, 2002.

[46] H.J. Zwally, M. Giovinetto, J.Li, H. Cornejo, M.Beckley, A. Brenner, J. Saba, and D. Yi. Mass changes of the greenland and antarctic ice sheets and shelves and contributions to sea-level rise: 19922002. *Journal of Glaciology*, 51(175):509–527, 2005.

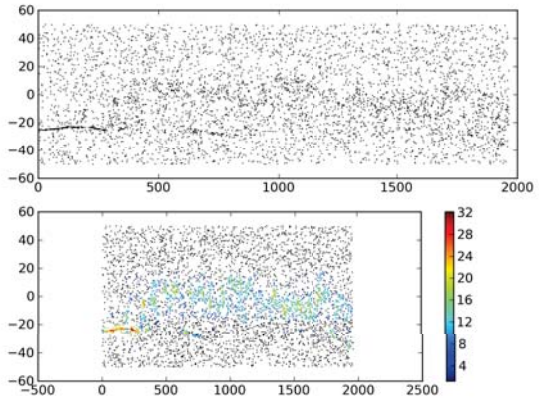
[47] H.J. Zwally, B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, J. Bufton, J. Dezio, D. Hancock, D. Harding, T. Herring, B. Minster, K. Quinn, S. Palm, J. Spinhirn, and R. Thomas. Icesat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34(3-4):405–445, 2002.

[48] H.J. Zwally, D. Yi, R. Kwok, and Y. Zhao. Icesat measurements of sea ice freeboard and estimates of sea ice thickness in the weddell sea. *J. Geophys. Res*, 113, 2008.

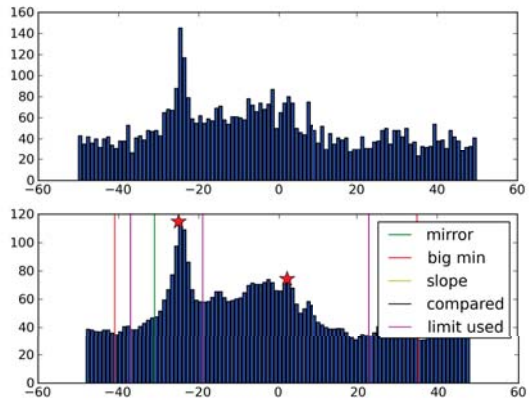
An Algorithm for Detection of Ground and Canopy Cover in Micro-Pulse Photon-Counting Lidar Altimeter Data in Preparation of the ICESat-2 Mission

Ute C. Herzfeld[(1,2)], *Brian W. McDonald*[(1)], *Bruce F. Wallin*[(2,3)], *Thorsten Markus*[(4)], *Thomas A. Neumann*[(4)] *and Anita Brenner*[(5)]

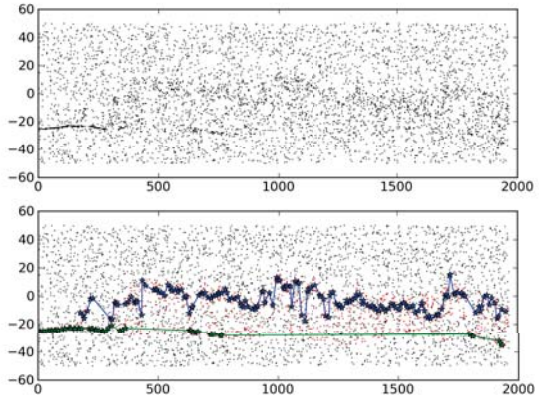
FIGURES



(a) lidar data and density

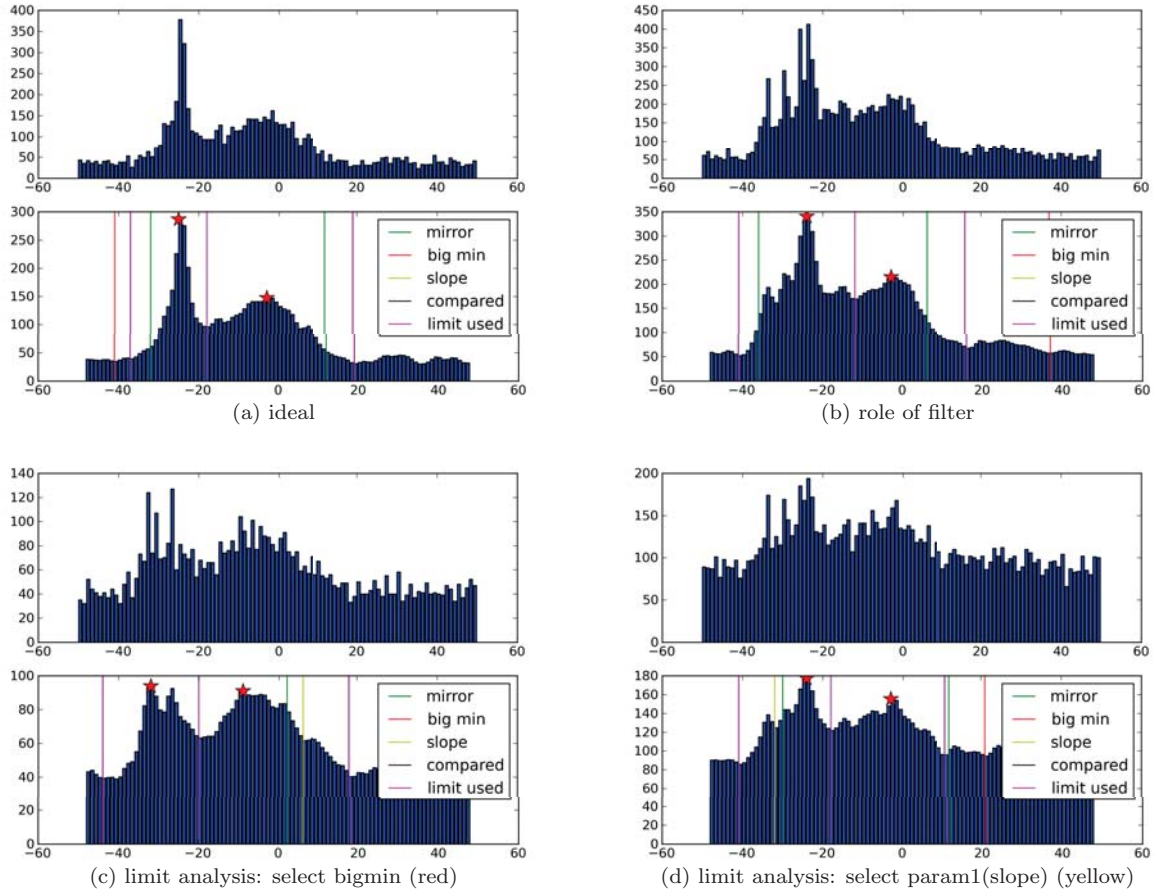


(b) histogram analysis



(c) canopy and ground elevations

Figure 1. Analysis steps. (a) Simulated lidar data (top) and density (bottom), derived from summation of radial basis function values for anisotropic neighborhood. (b) Histogram of lidar data elevation values (top) and filtered histogram (bottom). (c) Simulated lidar data (top) and result of analysis: stars mark density centers within the two classes of ground and canopy, derived using histograms in (b) and lines are piecewise-linear interpolations of density centers within each class.
(a) SERC_V2cap1r0p4_sb0-1-s4.dat_wnoise_uz2._signal_ascii.density..cluster.v11.png
(b) SERC_V2cap1r0p4_sb0-1-s4.dat_wnoise_uz2._signal_ascii.hist..cluster.v11.png
(c) SERC_V2cap1r0p4_sb0-1-s4.dat_wnoise_uz2._signal_ascii.cluster.v11.png



(a) ideal

(b) role of filter

(c) limit analysis: select bigmin (red)

(d) limit analysis: select param1(slope) (yellow)

Figure 2. Histogram analysis.
(a) Ideal situation with strong and single maximum for ground (highest) and canopy (second-highest).
(b) Case where Butterworth filter smoothes out outlying maxima.
(c) Case where *bigmin* criterion is used for canopy-range determination.
(d) Case where *param*1 (slope) criterion is used for canopy-range determination.
Note: Color bars are plotted in the order mirror around selected (starred) maximum (green), *bigmin* (red), *param*1 (slope) (yellow), compared (black), limit used (magenta); earlier lines may be hidden.

(a) SERC_V2cap1r1p4_sb0-1-s3.dat_wnoise_uz2._signal_ascii.hist..cluster.v11.png
(b) SERC_V2cap1r1p4_sb0-5-s2.dat_wnoise_uz2._signal_ascii.hist..cluster.v11.png
(c) SERC_V2cap1r0p4_sb0-3-s1.dat_wnoise_uz2._signal_ascii.hist..cluster.v11.png
(d) SERC_V2cap1r0p4_sb0-5-s1.dat_wnoise_uz3._signal_ascii.hist..cluster.v11.png

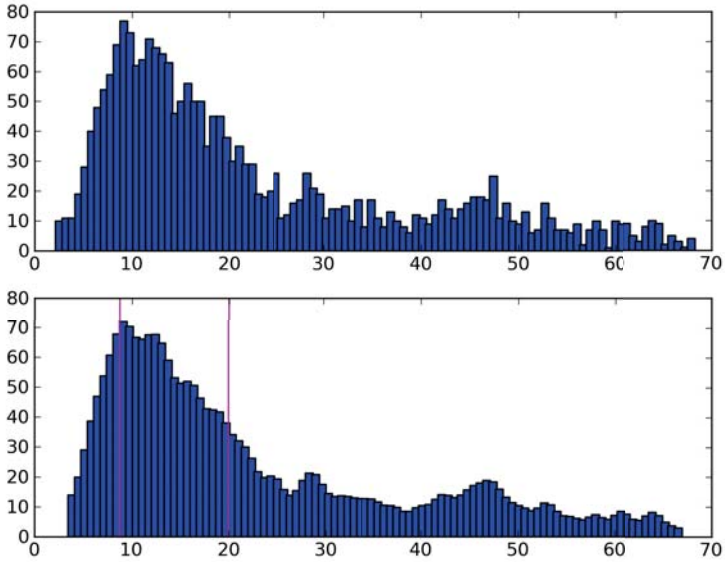
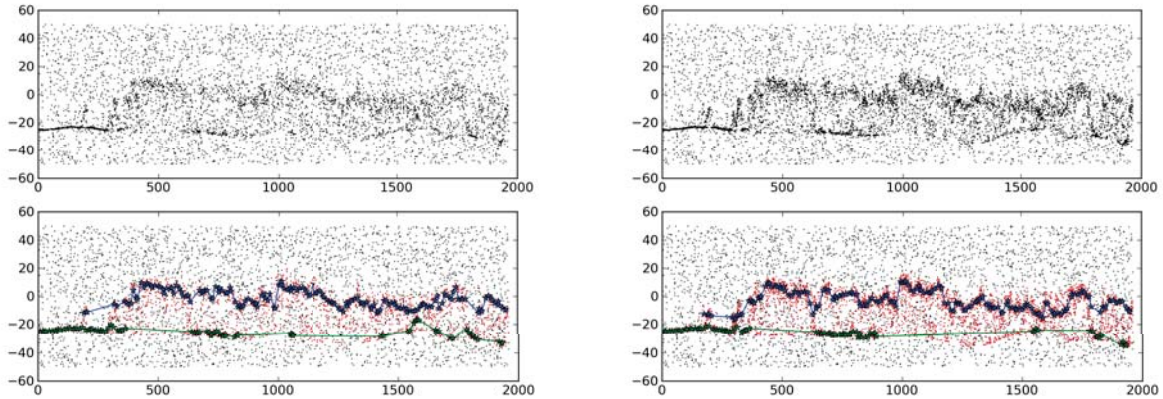
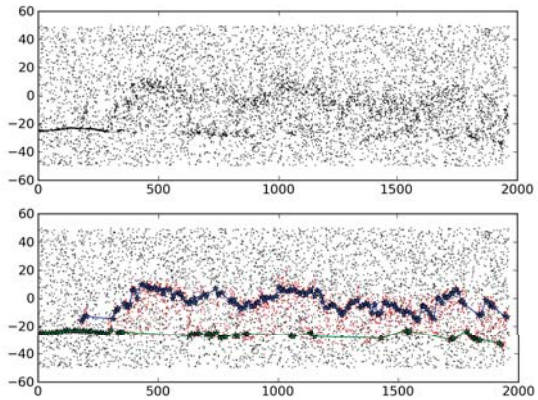


Figure 3. Threshold analysis, demonstrated for ground detection. The threshold used is the bin associated with 0.5 of the histogram value of the bigmax of density of the ground range data set. In this example, 0.5 times 72 = 36 for the histogram values, the ground threshold then becomes 20. The noise threshold is the bin associated with the bigmax, it is 9 in this case.

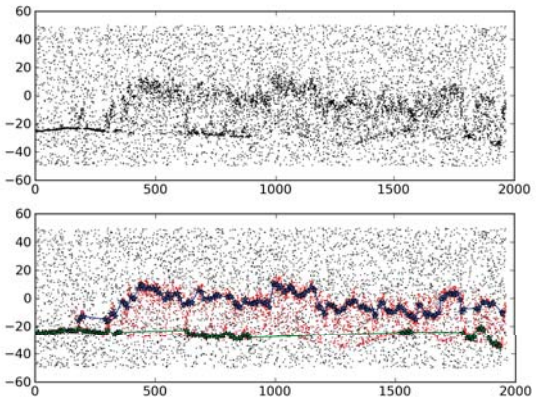


(a) mp9, r0, uz2

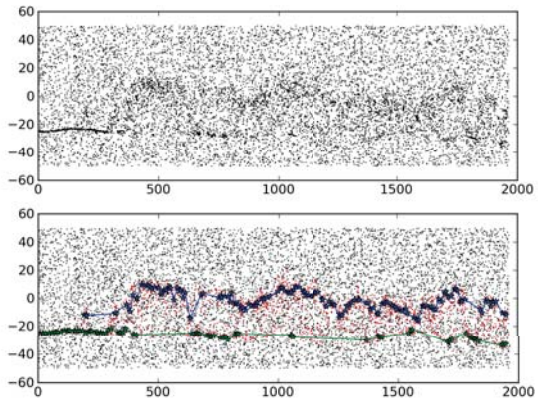
(b) mp9, r1, uz2



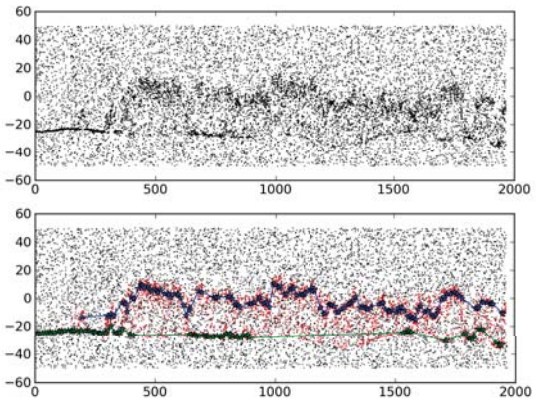
(c) mp9, r0, uz3



(d) mp9, r1, uz3

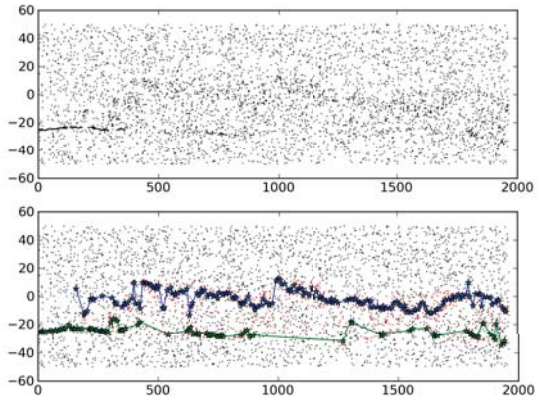


(e) mp9, r0, uz5

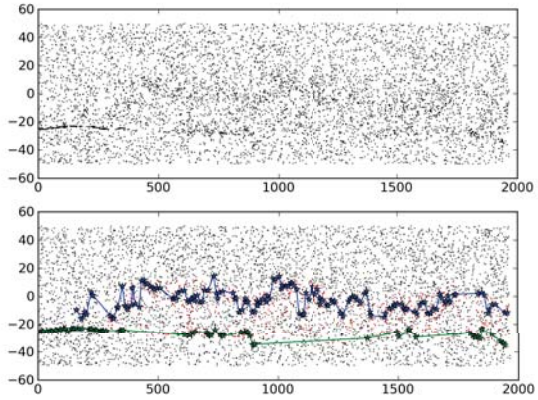


(f) mp9, r1, uz5

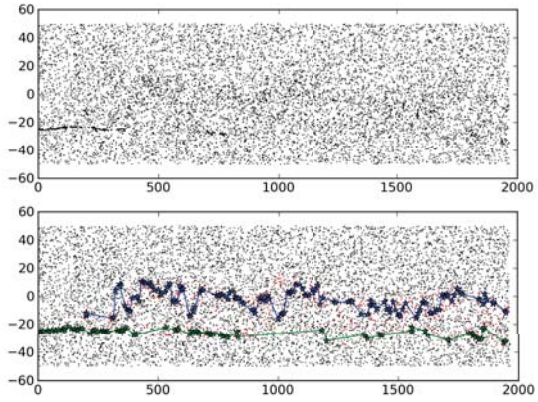
Figure 4. Data and ground/canopy detection for the strong beam (mp9), without and with resampling (r0, r1) [columns] and increasing noise levels (uz2, uz3, uz5) [rows]. All examples are SERC forests, section 1, simulation 2.
(a) SERC_V2cap1r0p9_sb0-1-s2.dat_wnoise_uz2._signal_ascii.cluster.v11.png
(b) SERC_V2cap1r1p9_sb0-1-s2.dat_wnoise_uz2._signal_ascii.cluster.v11.png
(c) SERC_V2cap1r0p9_sb0-1-s2.dat_wnoise_uz3._signal_ascii.cluster.v11.png
(d) SERC_V2cap1r1p9_sb0-1-s2.dat_wnoise_uz3._signal_ascii.cluster.v11.png
(e) SERC_V2cap1r0p9_sb0-1-s2.dat_wnoise_uz5._signal_ascii.cluster.v11.png
(f) SERC_V2cap1r1p9_sb0-1-s2.dat_wnoise_uz5._signal_ascii.cluster.v11.png



(a) mp4, r0, uz2



(b) mp4, r0, uz3



(c) mp4, r0, uz5

Figure 5. Data and ground/canopy detection for the weak beam (mp4), without resampling (r0) and increasing noise levels (uz2, uz3, uz5). All examples are SERC forests, section 1, simulation 1.

(a) SERC-V2cap1r0p4-sb0-1-s1-dat-wnoise-uz2-signal-ascii-cluster-v9.png
(b) SERC-V2cap1r0p4-sb0-1-s1-dat-wnoise-uz3-signal-ascii-cluster-v9.png
(c) SERC-V2cap1r0p4-sb0-1-s1-dat-wnoise-uz5-signal-ascii-cluster-v9.png

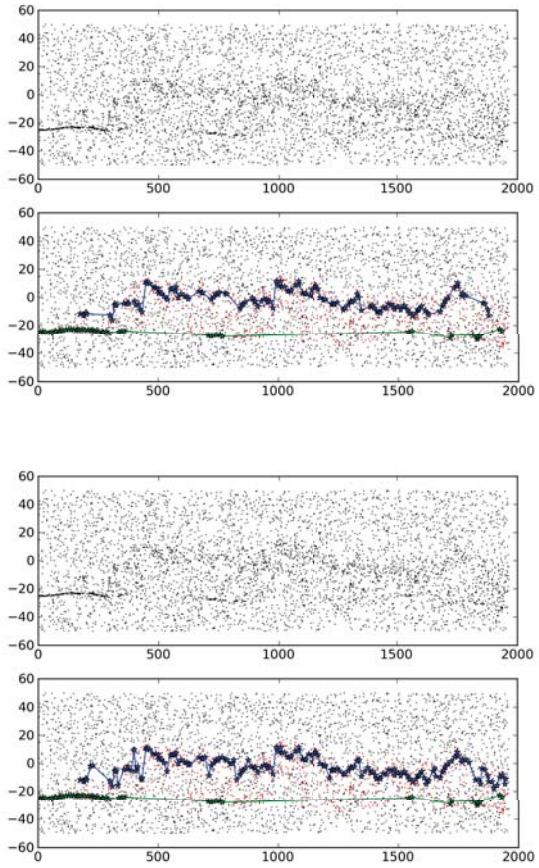


Figure 6. Experiments using tracking rigidity. Same algorithm, except with higher rigidity parameter (top panel, 6a) and lower rigidity parameter (lower panel, 6b), applied to the same data set (=msp p4, resampling r0, uz2, v11).
Top: Higher rigidity parameter (SERC-V2cap1r0p4-sb0-1-s3-dat-wnoise-uz2-signal-ascii-cluster-v11.png);
Bottom: Lower rigidity parameter (SERC-V2cap1r0p4-sb0-1-s3-dat-wnoise-uz2-signal-ascii-cluster-v11a.png)

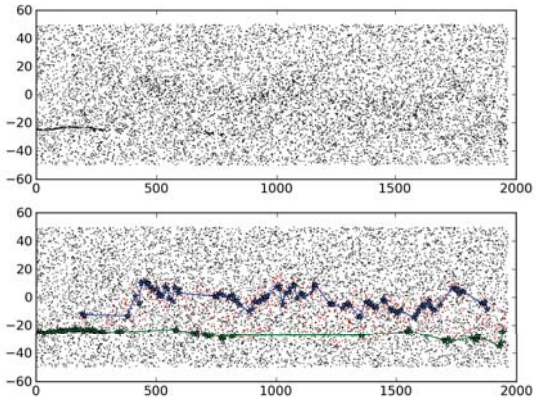


Figure 7. Application of tracking rigidity to improve detection for weak beams and most noise. Data and ground/canopy detection, msp p4, resampling r0, uz5, v11 (SERC-V2cap1r0p4-sb0-1-s3-dat-wnoise-uz5-signal-ascii-cluster-v11.png)

Ground												
Case	**uz2**			**uz3**			**uz5**			**uzAll**		
	mean	median	%	mean	median	%	mean	median	%	mean	median	%
p9, r0	0.45	0.00	97.20	0.49	0.00	95.78	0.55	0.00	94.70	0.49	0.00	95.89
p9, r1	0.20	0.00	99.47	0.24	0.00	99.22	0.33	0.00	98.81	0.26	0.00	99.17
p9, rAll	0.33	0.00	98.33	0.36	0.00	97.50	0.44	0.00	96.75	0.38	0.00	97.53
p4, r0	0.89	0.00	90.25	0.93	0.00	89.79	0.82	0.00	85.28	0.88	0.00	88.44
p4, r1	0.31	0.00	98.51	0.38	0.00	98.48	0.38	0.00	98.04	0.36	0.00	98.34
p4, rAll	0.64	0.00	93.79	0.70	0.00	93.52	0.63	0.00	90.75	0.66	0.00	92.68

Canopy												
Case	**uz2**			**uz3**			**uz5**			**uzAll**		
	mean	median	%	mean	median	%	mean	median	%	mean	median	%
p9, r0	0.38	0.00	96.00	0.57	0.00	93.70	0.83	0.00	93.01	0.59	0.00	94.23
p9, r1	0.18	0.00	99.57	0.30	0.00	99.04	0.37	0.00	98.95	0.28	0.00	99.19
p9, rAll	0.28	0.00	97.78	0.43	0.00	96.37	0.60	0.00	95.98	0.44	0.00	96.71
p4, r0	0.94	0.00	85.92	1.29	0.13	82.01	2.26	1.20	72.85	1.50	0.44	80.26
p4, r1	0.25	0.00	98.68	0.35	0.00	98.06	0.53	0.00	97.46	0.38	0.00	98.07
p4, rAll	0.65	0.00	91.39	0.89	0.07	88.89	1.52	0.69	83.40	1.02	0.25	87.89

Table 1: Summary of Validation of Ground Set (Top) and Canopy Set (Bottom). Correctly identified photons. mean – mean distance in meters from a point identified as ground/canopy to the nearest signal point in the validation set; median – median distance in meters from a point identified as ground/canopy to the nearest signal point in the validation set; % – percent of points identified as ground/canopy points that are also in the validation set of signal points; values are for groups of data sets, p9 – all sets associated with medium-strong beams, p4 – all sets associated with weak beams, r0, r1 – all sets with resampling r0, r1 resp., rAll - all resampling options, uz2,uz3,uz5 – low, medium and high noise levels reps, uzAll – all noise levels.