Remote Sensing of Environment xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements

Mary Ellen Miller [a], Michael Lefsky [a,*], Yong Pang [a,b]

[a] Center for Ecological Applications of Lidar, College of Natural Resources, Colorado State University, Fort Collins, CO 80523, USA
[b] Research Institute of Forest Resource and Information Technology, Chinese Academy of Forestry, Beijing 100091, China

ARTICLE INFO

Article history:
Received 10 March 2010
Received in revised form 31 August 2010
Accepted 2 September 2010
Available online xxxx

Keywords:
Lidar
Canopy height
ICESat GLAS
Optimization
Remote sensing
Vegetation structure

ABSTRACT

The Geoscience Laser Altimeter System (GLAS) has collected over 250 million measurements of vegetation height over forests globally. Accurate vegetation heights can be determined using waveform metrics that include vertical extent and extent of the waveform's trailing and leading edges. All three indices are highly dependent upon the signal strength, background noise and signal-to-noise ratio of the waveform, as the background noise contribution to the waveforms has to be removed before their calculation. Over the last six years, GLAS has collected data during thirteen observation periods using illumination from three different lasers. The power levels of these lasers have changed over time, resulting in variable signal power and noise characteristics. Atmospheric conditions vary continuously, also influencing signal power and noise.

To minimize these effects, we optimized a noise coefficient which could be constant or vary according to observation period or noise metric. This parameter is used with the mean and standard deviation of the background noise to determine a noise level threshold that is removed from each waveform. An optimization analysis was used with a global dataset of waveforms that are near-coincident with waveforms from other observation periods; the goal of the optimization was to minimize the difference in vertical extent between spatially overlapping GLAS observations. Optimizations based on absolute difference in height led to situations in which the total extent was minimized as well; further optimizations reduced a normalized difference in height extent. The simplest optimizations were based on a constant value to be applied to all observations; noise coefficients of 2.7, 3.2, 3.4 and 4.0 were determined for datasets consisting of global forests, global vegetation, forest in the legal Amazon basin and boreal forests respectively. Optimizations based on the power level or the signal-to-noise ratio of waveforms best minimized differences in waveform extent, decreasing the percent root mean squared height difference by 25–54% over the constant value approach. Further development of methods to ensure temporal consistency of waveform indices will be necessary to support long-term satellite lidar missions and will result in more accurate and precise estimates of canopy height.

1. Introduction

The ability to estimate forest height, canopy structure and carbon storage from lidar waveforms has been demonstrated for examples of the major forest physiognomic types at local to regional scales (1 to 10^5 km^2) (Lefsky et al., 2002, 2005; Morsdorf et al., 2004; Rosette et al., 2008; Tickle et al., 2006). These studies generally use observations taken by a single sensor, equipped with one laser, during a single period of observation of days to weeks. Current and planned spaceborne lidar missions increase the spatial and temporal extent of measurements by using multiple lasers concurrently (Vegetation Canopy Lidar (VCL), this instrument was planned, but was cancelled Dubayah et al., 1997) or consecutively (Ice, Cloud, and land Elevation Satellite (ICESat) Zwally et al., 2002) or both (Deformation, Ecosystem Structure and Dynamics of Ice (DESDynI), National Research Council, 2007). Lasers of the same type can, over time or as the satellite's environment changes, vary in total power, pulse width and shape, and spatial distribution of power within the beam. Differences in the alignment of the sensor's optics can lead to changes in the size and eccentricity of the area sampled on the ground (the sensor's footprint). Power and noise characteristics of individual waveforms are also due to solar background, aerosol depth, and land surface reflectance at the laser's wavelength. Development of a global dataset of forest structural characteristics from lidar systems will depend on methods for developing indices of waveform extent and distribution that are consistent despite changing instrument and environmental characteristics.

The Geoscience Laser Altimeter System (GLAS), in orbit on the Ice, Cloud and land Elevation Satellite (Schutz et al., 2005) was the first long-term mission using a space based laser altimeter capable of

* Corresponding author. Tel.: +1 970 491 0602; fax: +1 970 491 0279.
E-mail addresses: mmaryellen@gmail.com (M.E. Miller), lefsky@gmail.com (M. Lefsky), caf.pang@gmail.com (Y. Pang).

0034-4257/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2010.09.002

measuring forest canopy height on a global scale (see Sun et al., 2003 for work involving the short term Shuttle Laser Altimeter missions in 1996 and 1997). GLAS began collecting data in February of 2003 using the first of its three lasers (Schutz et al., 2005). Gold wires coated with indium solder were used in the construction of the lasers; these two metals reacted to break individual wire connections, which led to decreased power and laser failure (IGARB, 2003; Jellison, 1979; Leidecker, 2003). Laser 1 operated for 38 days before failing; to extend mission life span the remaining lasers were rescheduled to operate in three 33 day periods per year (Schutz et al., 2005). Laser 2 was operated for three such periods before being shut down after its energy level declined from 80 mJ to 5 mJ. Laser 3 was operated at a lower temperature until its failure in October of 2008; the instrument was retired one year later when Laser 2 which had been operating at lower power levels failed. Because of these and related issues (Abshire et al., 2005) the laser pulses used to measure canopy height varied over time which affected the consistency of resulting measurements. A list of the operation periods considered in this study along with the energy at the start and end of the periods is found in Table 1. Data from Period 1 was not considered here due to problems with the geolocation of the data at the time of these analyses; the geolocation issues have subsequently been addressed.

The GLAS lasers produce a 5 ns pulse at a wavelength of 1064 nm to illuminate an ellipsoidal spot on the earth's surface with a diameter that has ranged between 52 and 90 m for lasers 2 and 3. The laser operates at a rate of 40 Hz and observations are separated along the satellite's track by 172 m. The return pulse is collected through a 1 m diameter telescope and measured by an analog detector. A 1 GHz sampler digitizes the return pulse providing the laser waveform (Schutz et al., 2005).

When a return pulse is from a vegetated surface the waveform is a function of the vertical distribution of the illuminated vegetation and ground surfaces within the pulse's footprint. If the terrain is level then the height of the vegetation can be calculated utilizing the time difference between the first returns from the upper canopy and the modal height of the returns from the ground surface (Blair and Hofton, 1999; Harding et al., 2001). As terrain slope and footprint width increase it is more difficult to identify the ground return and methods based on more general indices of waveform distribution become necessary (Lefsky et al., 2005, 2007). Of these indices, waveform extent (the distance between the first and last returns from the waveform) is the most important parameter for determining canopy height. Waveform extent increases with terrain slope and surface roughness necessitating additional information about the terrain and vegetation some of which can be determined from the waveform as well (Lefsky et al., 2007).

To estimate canopy height from the waveform (Harding et al., 2001; Lefsky et al., 1999) a threshold is used to separate energy returned from the forest canopy from background noise that is due to other sources of light (e.g. scattered sunlight) entering the sensor's telescope (Fig. 1). This threshold is a function of the mean and standard deviations of background noise and a "noise coefficient" that sets the threshold at some multiple of the standard deviation above the mean background noise level (variable names follow ICESat standards in Zwally et al., 2003).

$$\text{Noise Threshold} = \text{d\_4nsbgmean} + \text{nc}^*\text{d\_4nsbgsdev} \quad (1)$$

Where:

nc is the noise coefficient
d_4nsbgmean is the mean of the waveform background noise
d_4nsbgsdev is the standard deviation of the waveform background noise

Whereas in some sensors the background noise level must be determined using only the waveform data (Harding et al., 2001; Lefsky et al., 1999), the GLAS sensor uses energy recorded after the waveform signal has ended. The energy recorded at that time contains no signal, and is a combination of only atmospheric returns and sensor noise. After removing the threshold from the waveform values, any return values that are less than zero are set to zero.

Fig. 1 illustrates the influence of the noise coefficient on a hypothetical waveform consisting of background noise and a 4 unit increase in waveform power between bins 250 and 750. From a simple statistical approach, the value of the noise coefficient might be expected to be constant. For instance, for a noise coefficient of three, a Gaussian random background signal should only exceed the threshold once in every 740 bins (probability of 0.134%). However, even this low probability is problematic: each waveform can contain hundreds of background points and the global dataset contains over 250 million waveforms from forests. Compounding this problem, a random point that exceeds the threshold might occur anywhere in the waveform, leading to unacceptable increases in waveform extent (see Fig. 1). Conversely, the noise coefficient should not be set too high, as this increases the probability of missing actual return signals (Harding & Carabajal, 2005). This is especially important when laser power (and total signal) is low, when there are individual tree crowns which emerge above the general canopy level, or when forest cover is high and the ground surface is occluded by vegetation. In each of these cases, the amount of power at the start and finish of the signal is low. Given the wide range of possible scenarios, it is not clear that a "perfect" solution exists; the best solution is one where the noise coefficient balances under- and over-estimation of waveform extent. Given the properties of random noise, there are always circumstances in which noise will be incorrectly defined as signal, or vice versa.

Table 1
Laser observation periods (TBD = to be determined). For latest information: http://nsidc.org/data/icesat/laser_op_periods.html.

Observation period	Start date	Days in period	Laser maximum power (mJ)	Laser minimum power (mJ)	Footprint diameter (m)	Horizontal geolocation accuracy (m)
1	2/20/2003	38	72	51	148.6	4.6 ± 9.3
2A	9/25/2003	55	80	55	86.7	TBD
2B	2/17/2004	34	57	33	89.8	TBD
2C	5/18/2004	35	33	5	88.4	37.7 ± 53.4
3A	10/3/2004	37	67	62	55.8	0.0 ± 2.7
3B	2/17/2005	36	68	54	79.3	17.4 ± 22.8
3C	5/20/2005	35	49	44	55.4	TBD
3D	10/21/2005	35	43	39	52	TBD
3E	2/22/2006	34	TBD	TBD	52.3	TBD
3F	5/24/2006	33	TBD	TBD	51.2	TBD
3G	10/25/2006	34	TBD	TBD	53.4	TBD
3H	3/12/2007	34	TBD	TBD	55.6	TBD
3I	10/2/2007	37	TBD	TBD	57.3	TBD

Please cite this article as: Miller, M.E., et al., Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.09.002

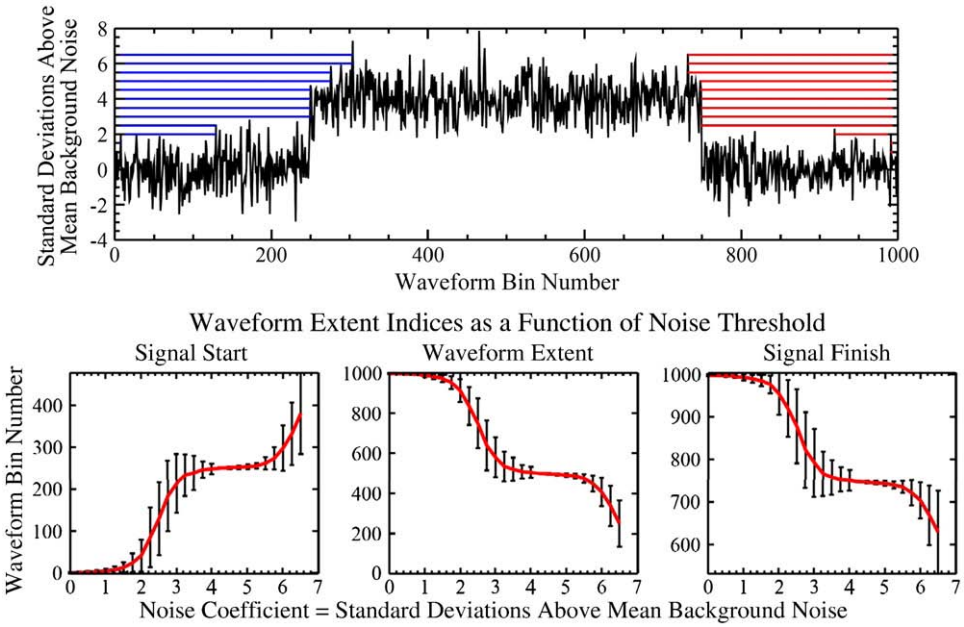

Fig. 1. Synthetic waveform with Gaussian noise and a 4-unit step increase in power between bins 250 and 750. The actual extent of the signal in this synthetic waveform is 500 bins. By removing 3–5 multiples of the standard deviation from the waveform we are able to match the actual extent. If the noise coefficient is too high the waveform extent will be too small and if the coefficient is not high enough the extent will be too large.

ICESat's primary mission is measuring change in ice elevations. To achieve this objective, its orbit is maintained as close to a reference track as possible. Outside of the polar regions, the ground track is allowed to diverge as much as 1 km from the reference track (Schutz et al., 2005). This limits the number of repeat measurements of forest canopies along track. Overlaps also occur when ascending and descending tracks intersect. We refer to a pair of coincident observations as overlaps or overlapping pairs. A much smaller number of cases with three or more overlapping observations also exist, but are too infrequent to allow for statistical analysis.

Each overlap pair consists of two observations of what we constrain to be the same forest and topographic conditions as modified by the characteristics of the laser and the environment at the time of observation. Using a global database of 545,742 overlap pairs, we can determine the set of noise coefficients that minimizes the difference in waveform extent. In this work, we evaluate several different approaches to minimizing these differences: an optimized constant noise coefficient, separate coefficients for each observation period, and methods that estimate optimal coefficients as a function of waveform power and background noise characteristics.

2. Methods

2.1. Identification of overlap pairs

Overlapping pairs of GLAS observations were found using Delaunay triangulation and subsequent analysis of the associated adjacency list to find each observation's nearest neighbor. The pairs were then filtered to remove those separated by more than 12 m to ensure the observations cover a common area and to remove repeat pairs. Two footprints with a nominal footprint of 70 m and centered 12 m from each other will have 77% of their area overlapping. For a footprint of 55 m (which approximates the size of footprints later in the mission) the overlap percentage would be 71%.

Data collection periods fell into three main groups: winter data collections that generally started in February, late spring data collections that started in May, and Fall data collections that started in October. To identify pairs of observations with contrasting canopy phenology, we used the MODIS Land Cover Dynamics product (MOD12Q2, Zhang et al., 2003) for the year 2005 to classify each observation into a leaf-on or leaf-off state as a function of both location and day of observation. Observations pairs that had one leaf-on and one leaf-off observation account for 0.7% of all pairs. These pairs were kept in the overall analysis but would not have affected the results greatly.

2.2. Optimization approach

A simple looping optimization technique was used to determine the best noise coefficient parameters for obtaining consistent waveform extent estimates. A more sophisticated approach to optimization utilizing Ostrich (Matott, 2005) was also evaluated but the large number of variables to optimize and the high number of observations led to difficulties in finding optimal solutions. In initial attempts, optimizing for differences in waveform extent led to results in which the optimization algorithms favor high noise coefficients which truncate too much signal, driving error in the objective function (in this case, the sum of the squared difference in the two extents) to zero as the waveform extents also approach zero (Fig. 2). To preserve waveform signal, the following objective function was minimized.

$$Objective\ Function = \sum_{i=1}^{n} \left(\frac{Extent_i^1 - Extent_i^2}{Extent_i^1 + Extent_i^2} \right)^2 \quad (2)$$

Where:

Objective Function is the objective function we seek to minimize
$Extent_i^1$ is the extent of the first GLAS observation in overlap pair *i*
$Extent_i^2$ is the extent of the second GLAS observation in overlap pair *i*

Noise coefficients were constrained to be between 2 and 7 in all optimization runs. Values outside of that range are unreasonable in practice as they would result in too large a probability of having false "signal starts" and false "signal ends" (for values below 2) or overly truncating waveforms signal (for values above 7).

Optimizations were evaluated by comparing the mean extents, the root mean square difference (RMSD) of the optimized extents, and their ratio (RMSD%). The reported values are from validation datasets

Please cite this article as: Miller, M.E., et al., Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.09.002

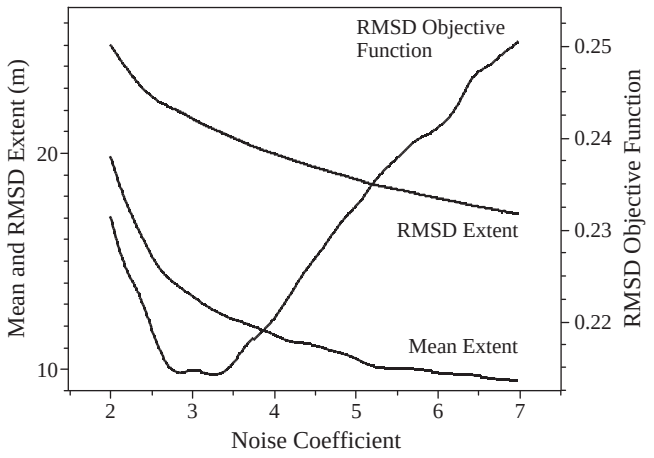

Fig. 2. Effect of noise coefficient on mean extent and RSME extent for Amazon.

of overlap pairs that were not used in the optimizations. Outliers with extent differences greater than two standard deviations from the mean extent difference were removed from consideration, as forest structure can easily change over the course of 6 years and given the global scale of the data it was not possible to investigate all outliers for potential land cover change detected in overlap pairs (Duong et al., 2008). Five optimization approaches were evaluated.

2.2.1. Constant coefficient optimization

In constant coefficient optimization, the objective function was determined for all GLAS pairs in each study area for noise coefficients ranging from 2 to 7 in increments of 0.01. The optimal noise coefficient was the one that minimized the objective function. This optimization provides a baseline level to compare with the other approaches.

2.2.2. Optimization by observation period

We hypothesized that variation in signal power and noise characteristics tends to vary more between periods than within a single period. To minimize these effects, the objective function was determined for all GLAS pairs in each study area using a separate noise coefficient for each observation period. Both intra- and inter-period observations were considered to determine which combination minimizes the objective function.

2.2.3. Optimization by signal characteristics

For these optimizations, the relationship between an optimal noise coefficient and one of three aspects of waveform noise characteristics (noise, power, and signal-to-noise ratio or SNR) was examined. The first two indices were the background noise level and the mean power of the waveform. Background noise was represented by "d_4nsbgsdev", the standard deviation of the background noise identified by the GLAS processing software (Brenner et al., 2003). Waveform power was defined as the mean value of the waveform after noise is removed using Eq. (1) and the standard GLAS noise coefficient of 4.5. The signal-to-noise ratio (SNR) was calculated as the ratio of mean power to background noise.

Each index (noise, power, and SNR) was divided into 16 levels and the objective function was minimized by varying the noise coefficients for each level. The boundaries of these levels were selected to ensure a roughly equal number of calibration shots in each level. Regression analysis was then used to relate the optimized noise coefficients to the original index. GLAS observations were then reprocessed using noise coefficients predicted using the regression equation and new extents were calculated. Mean extent, RMSD of extent, and the extent RMSD% were calculated for each study area using validation data.

2.2.4. Comparison of optimization results

The statistical significance of differences in RMSD was evaluated by testing the equality of variances of the differences in extent as a function of method. We used the two-sided F-test for the ratio of group variances, and evaluated the ratios for methods taken two at a time.

2.2.5. Comparison of inter- and intra- period results

Observed differences in observation pair extent are only partially due to differences in noise and signal characteristics; there is a substantial fraction of the RMSD that cannot be addressed by signal processing. Differences in extent are very sensitive to the distance between each observation but few observation pairs are absolutely centered on one another. Therefore observations with some distance between them must be accepted to generate a sufficiently large dataset. In this work we analyzed observations whose calculated locations were within 12 m of each other, but we expect that geolocation error generally exceeds this limit (Table 1). Geolocation error can be estimated for individual observation periods but varies continuously within them. Change in land surface conditions was addressed by removing outliers, but this does not address all land surface changes including the yearly growth of trees.

To provide a context for the results of the other analyses, we selected intra-period observation pairs and calculated their RMSD for each of the four study areas (see below). While it would have been preferable to select waveforms from the same observation period that also had similar SNR, there were too few of these observation pairs to do the analysis. We recalculated RMSD values for all analyses after subtracting the intra-period RMSD from each of the RMSD values from the inter-period analyses. To estimate their difference, both RMSD values were first squared and the square root of the difference was then calculated as the "corrected" RMSD.

2.3. Study areas

We defined four study areas to examine the effect of region and physiognomy on the optimizations: natural vegetation (global forests shrublands, savannahs and grasslands), global forests, boreal forests and Amazon forests. Canopy height was assumed to be at equilibrium during the five year data acquisition period. Forest growth and decay should not be too great during a five year period and any radical changes in canopy height would be eliminated as outliers.

2.3.1. Natural vegetation

Natural vegetation was considered to include the classes for evergreen needleleaf and broadleaf forests, deciduous needleleaf and broadleaf forests, mixed forests shrublands, savannas, and grassland as defined by the MODIS land cover product (MOD12Q1, Strahler et al., 1999). Globally there were 545,742 overlap pairs from areas with natural vegetation. The spatial density of ICESat observations increases with latitude; and the number of vegetated shots also depends on the amount of land mass available. To obtain a more balanced subset of natural vegetation, an equal number of overlap pairs were selected from seven zones defined by the absolute value of their latitudes from 0 to 70°. The 20–30° zone contained the fewest overlaps with 10,570 pairs in the calibration set; and all of these pairs were used. An additional 10,570 pairs were randomly selected from each of the other 6 zones for a total of 73,990 overlap pairs, a balanced set of 10,682 pairs (~12.5%) were selected from the reserved validation data set.

2.3.2. Global forests

The global forest overlaps dataset was similar to the Natural Vegetation dataset, but included only the five forested land cover classes. There were 143,671 forested pairs in the overlaps dataset, but again the majority of these pairs were from Boreal regions. Overlap

Please cite this article as: Miller, M.E., et al., Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.09.002

pairs were balanced by latitude to provide a total of 22,856 overlap pairs of which 2791 were kept for validation. The optimized noise coefficients from this dataset should be a good estimate for all forests.

2.3.3. Amazon

As the largest rainforest in the world, the Amazon is an important component of the global carbon cycle (Ometto et al., 2005). Therefore to increase the accuracy and consistency of canopy height estimations within the Amazon basin we optimized noise coefficients for the overlap pairs found within the legal Amazon boundary. There are 3357 overlap pairs to use for calibration and 473 pairs reserved for validation. Overlap pairs were sparse because of the Amazon's equatorial location.

2.3.4. Boreal

We also optimized noise coefficients for the boreal forests. The cold temperatures found in Boreal forests suppress decomposition making this biome an important carbon sink that may be threatened by global warming (ACIA, 2005). We used 263,483 pairs for calibration and reserved 44,795 pairs for validation. The World Wildlife Foundation ecology biome map was used to determine which pairs were in Boreal forests (Olson et al., 2001).

3. Results

3.1. Constant value optimization

A single noise coefficient to be applied to all waveforms was found by determining which noise coefficient minimized the objective function (Eq. 2) for each study area regardless of period or noise characteristics. Similar mean and RMSD extents were observed for natural vegetation, boreal forest, global vegetation and Amazon forests (Table 2). The optimal noise coefficient ranged between 2.7 for global forests and 4.0 for the boreal forest; although Fig. 3 indicates that a broad range of noise coefficients have similar values for the RMSD of the objective function values. Boreal forests in particular were less sensitive to variations in the noise coefficient, perhaps because the relatively shorter stature and more open canopies of these forests leads to higher power within the overall waveforms, especially for the ground return.

3.2. Optimization by period

Noise coefficients which minimized the objective function for each study area were found for each observation period and are shown in Fig. 4. Average mean extent, RMSD, and RMSD% are given in Table 3. The period optimizations showed slight improvements over the constant value optimizations as the extent RMSD% were lower for all study areas. For Laser 3, noise coefficients tend to follow the pattern of laser power, generally decreasing with successive periods. Noise coefficients for Laser 2 also decline as power declines, with the exception of the Amazon study area, although the noise coefficient for period 2B appears to be an outlier. A separate ANOVA indicated that season of data collection had no effect on the optimal noise coefficient.

Table 2
Constant value optimization results.

Study area	Natural vegetation	Global forest	Amazon	Boreal
Mean extent	11	17.5	19.7	10.3
RMSD extent	4.8	8.2	6.9	5
RMSD%	44%	47%	36%	49%
Noise coefficient	3.2	2.7	3.4	4

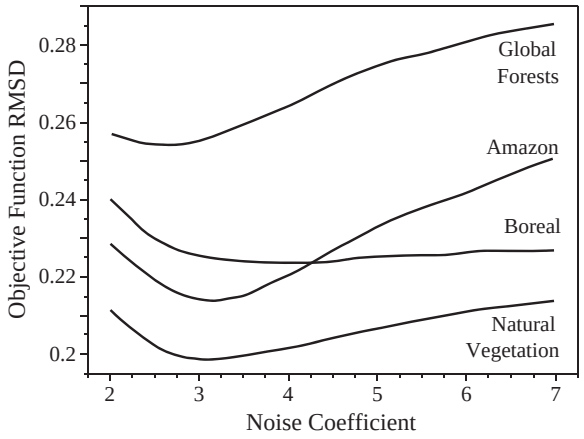

Fig. 3. Constant value optimization, noise coefficient vs. the objective function.

3.3. Optimization by noise characteristics

The optimal noise coefficients for each of the 16 levels of background noise are plotted in Fig. 5. For the global forest, natural vegetation and boreal datasets regression analysis found an exponentially declining relationship between the mean background noise of each level and the optimized noise coefficient, but the relationships for forest, natural vegetation, and the Amazon were weak ($R^2 = 0.23$, 0.27, 0.26) (Table 4.1). Although the results of the observation period analysis suggested a possible relationship between noise and the optimal noise coefficient, this analysis had higher RMSD% values than the period optimization (Table 4.2).

Noise coefficients for the mean power of the waveform had strong linear relationships with power level (Fig. 6). Table 5.1 indicates the higher percent of variance explained in the noise coefficient (R^2 of 0.91 to 0.99), and all the datasets showed evidence of decreased RMSD% (Table 5.2) relative to the period optimization. The power level optimization was the most effective method for reducing RMSD% for the Amazon dataset (Table 5.2).

Linear relationships (Fig. 7, Table 6.1) were found to predict noise coefficients based on a waveform's mean signal-to-noise ratio. The linear relationships were quite strong (R^2 of 0.93 to 0.95) and of the five approaches, optimization based on the SNR values is most effective in reducing the difference between the extents of the overlaps pairs for the natural and boreal datasets (Table 6.2).

Tables 7.1 and 7.2 report the mean extent and RMSD values for all four datasets and five optimization approaches; the last column in Table 7.2 reports the RMSD for the intra-period analysis. Intra-period RMSD is lower than the other analyses for all study areas. Table 7.3 recalculates the RMSD values after removing each dataset's intra-period RMSD from the original RMSD. Tables 7.4 and 7.5 show the

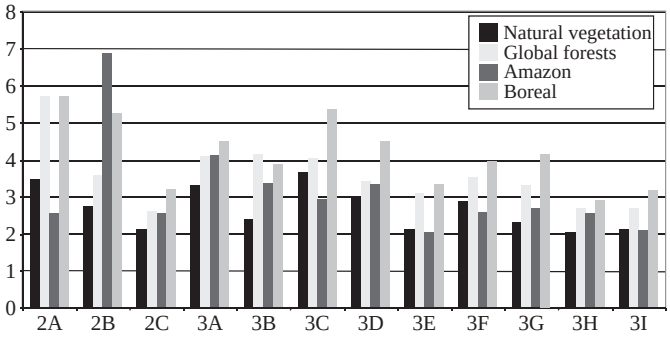

Fig. 4. Optimized noise coefficients by observation period.

Please cite this article as: Miller, M.E., et al., Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.09.002

Table 3
Period optimization results.

Study area	Natural vegetation	Global forest	Amazon	Boreal
Mean extent	10.5	17.7	20.2	10.4
RMSD extent	4.3	7.8	6.7	4.6
RMSD%	41%	44%	33%	44%

original and corrected RMSD% values calculated from 7.1 through 7.3. The original RMSD% values average 40% while the corrected values average 30%. If we consider the results for the constant value analyses as a baseline we can calculate the relative improvement associated with the other methods. Tables 7.6 and 7.7 report, for each set of RMSD%, the percent improvement of each of the four optimization methods and the constant value method (negative values are cases where methods underperformed the constant value approach). Table 7.7 indicates the SNR and power methods are clearly the best approaches, with the SNR slightly outperforming the power method. Relative to the constant value method the SNR and power methods lowered RMSD% by an average of 36% and 34%, respectively. The SNR method reduced RMSD% for the Natural vegetation dataset by 54%.

Tests for the equality of variances indicate that the differences in the RMSD values associated with the power and SNR methods were significantly lower ($P<0.05$) than those obtained using simple, period or noise methods for all study areas. RMSD values for the period method were significantly lower than either the simple method or noise method for the Boreal and Natural Vegetation methods. RMSD values for the SNR method were significantly lower than the power method for the Boreal study area only.

3.4. Applying optimal noise coefficients

Optimal noise coefficients can be obtained by either looking up the appropriate noise coefficients (Table 2, Fig. 4) or by deriving the noise coefficients from the empirical equations provided in Tables 4, 5, and 6. Normalizing GLAS waveforms requires inserting the optimal noise coefficient into Eq. (1). All the necessary information is provided with GLAS waveform data, although the power level and signal-to-noise ratio will have to be derived for each waveform.

4. Discussion

While a capability for absolute referencing of waveform characteristics (e.g. through ongoing collections of airborne lidar data) will be necessary for a global vegetation lidar mission, it seems very likely that methods for using the much larger number of observations available from overlap pairs will be used to ensure the internal consistency of waveform lidar metrics such as extent. The threshold approach to delineating signal and noise in the waveform relies on a noise coefficient whose performance must be considered statistically. Given the properties of random noise, there are always circumstances in which noise will be incorrectly defined as signal, or vice versa. We expected by-period optimization to account for a suite of long-term changes in sensor performance during the ICESat mission. Instead, the best predictor of a suitable noise coefficient was the signal-to-noise ratio (SNR) and the optimal noise coefficient increased with the SNR. At low SNR the optimal noise coefficient was low to avoid missing key parts of the signal, at the expense of possibly overestimating waveform extent. At high SNR, the probability of missing parts of the signal decreased, allowing the noise coefficient to rise to avoid misclassification of noise as signal.

Our results indicate that the SNR and power level methods were by far the best approaches. For the SNR method the RMSD values ranged from 1.41 (natural vegetation) to 4.7 (Amazon forests) when intra-period variance was removed (Table 7.3), an improvement of 25 to 54% (Table 7.7) over results obtained using the constant value method. For the densely vegetated Amazon dataset the power method performed the best with a corrected RMSD value of 4.21, and an improvement of 31% over the simple method. These approaches may be superior because they incorporate both long-period changes in laser characteristics with factors that can vary from shot-to-shot including solar illumination, aerosol depth, and land surface reflectance at the laser's wavelength. The results clearly indicate the noise coefficient should be optimized to the characteristics of each waveform using a variable such as SNR that incorporates all of the observed effects.

Table 4
Noise level optimizations. 1) Best fit exponential equations. The y-axis represents the predicted optimal noise coefficient and the x-axis is background noise (d_4nsbgsdev). 2) Optimization results.

1)

	Exponential Fit	R^2
Natural	$y=3.7216e^{-0.1434x}$	0.27
Forest	$y=3.3809e^{-0.1605x}$	0.23
Amazon	$y=4.8034e^{-0.2281x}$	0.26
Boreal	$y=6.4594e^{-0.3664x}$	0.86

2)

Study area	Natural vegetation	Global forest	Amazon	Boreal
Mean extent	10.9	17.4	19.4	10.0
RMSD extent	4.7	8.3	6.4	5.0
RMSD%	43%	47%	33%	50%

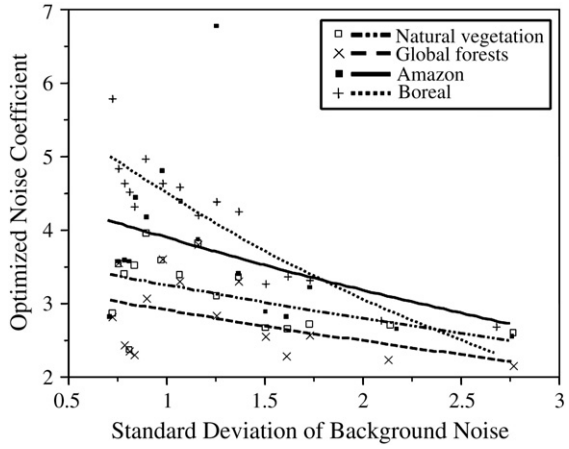

Fig. 5. Optimized noise coefficients by noise level with best fit exponential lines.

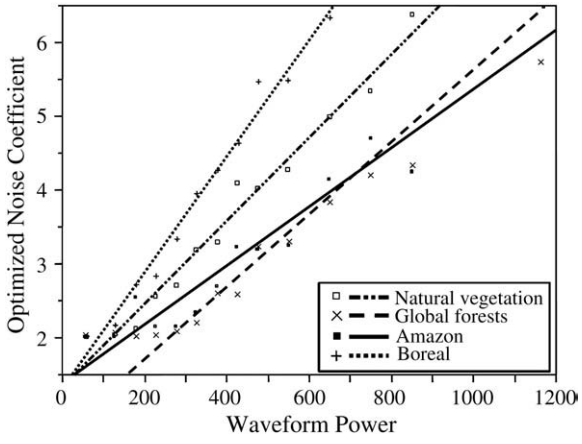

Fig. 6. Optimized noise coefficients by power level with best fit lines.

Please cite this article as: Miller, M.E., et al., Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.09.002

Table 5
Power level optimizations. 1) Best fit linear equations. The y-axis represents the predicted optimal noise coefficient and the x-axis is mean waveform power. 2) Optimization results.

1)

	Linear fit	R^2
Natural	$y = 0.0056x + 1.3398$	0.98
Forest	$y = 0.0036x + 1.3414$	0.96
Amazon	$y = 0.0030x + 1.7673$	0.91
Boreal	$y = 0.0077x + 1.3613$	0.99

2)

Study area	Natural vegetation	Global forest	Amazon	Boreal
Mean extent	10.0	15.9	18.4	9.7
RMSD extent	3.7	6.2	4.6	4.1
RMSD%	37%	39%	25%	42%

Table 6
Signal-to-noise level optimizations. 1) Best fit linear equations. The y-axis represents the predicted optimal noise coefficient and the x-axis is the waveform signal-to-noise ratio. 2) Optimization results.

1)

	Linear fit	R^2
Natural	$y = 1.4129x + 0.662$	0.93
Forest	$y = 1.1070x + 0.5011$	0.94
Amazon	$y = 1.0057x + 1.0589$	0.93
Boreal	$y = 1.5145x + 1.0295$	0.98

2)

Study area	Natural vegetation	Global forest	Amazon	Boreal
Mean extent	9.9	15.6	18.5	9.6
RMSD extent	3.6	6.2	5	3.9
RMSD%	36%	40%	27%	41%

Our prior experience using waveform lidar over smaller spatial extents indicated a single noise coefficient (4.5, Harding et al., 2001; Lefsky et al., 2007) could be used for all waveform processing. A single noise coefficient might still be appropriate under those circumstances, but this work suggests that value was too conservative for forested systems and that if a single noise coefficient is used a threshold of 2.7 would be best, values of 3.4 or 4.0 should be considered for the Amazon or boreal forests respectively, although individual study areas may differ. The sensor and its platform also need to be considered in selecting optimal noise coefficients. The higher noise coefficients found for boreal forests may reflect an underlying problem with the assumptions of this work. From the perspective of accuracy we might want lower thresholds which would have a higher probability of detecting the tops of conical crowns. However the optimization process favors consistency in estimates of extent which may favor higher thresholds that remove the difficult-to-observe crown tops from all waveforms. One way to address this problem would be to weigh results from observations with higher signal-to-noise ratios which would tend to make the final set of noise coefficients give results that are more like the highest quality waveforms. In addition, combining overlap comparisons with comparisons of selected waveforms and airborne lidar measurement could ensure both internal consistency and accuracy. Future spaceborne lidar missions may have signal-to-noise ratios high enough to reduce the probability of missing the crown top to a level where this effect is no longer important.

Future spaceborne lidar missions dedicated to mapping vegetation will probably have smaller laser footprints, designed to more accurately measure vegetation structure. This might make obtaining overlap pairs more difficult; however future missions will probably collect more data as they should not be plagued with the same power issues that limited ICESat-1 data collections.

In this work, the use of multiple indices (e.g. observation period and SNR ratio) to estimate noise coefficients was impractical due to the high number of observations required, however again future missions are likely to collect more data and may not have such a wide range of SNR values observed here. Physical modeling of the problem may identify an index more suitable than SNR or power level for estimating optimal noise coefficients.

Initial attempts at optimization used the difference in waveform extent as the variable to be minimized, which led to very high noise coefficients, and associated low extents, as the difference in extents declines with absolute waveform extent. In this work, we considered a normalized difference between the two extents in a pair. While this approach seems to have led to reasonable results in this work, there is no reason to believe that the objective function (Eq. 1) is itself optimal. For instance, we might be more concerned with the difference between stands with larger extents and could rewrite the equation to include that factor. What is clear is that the objective function we used did work as it reduced the RSMD. One complication in interpreting the noise coefficients reported in this work is that a broad range of noise coefficients result in objective functions that are at or near the minimum RMSD value (Fig. 3).

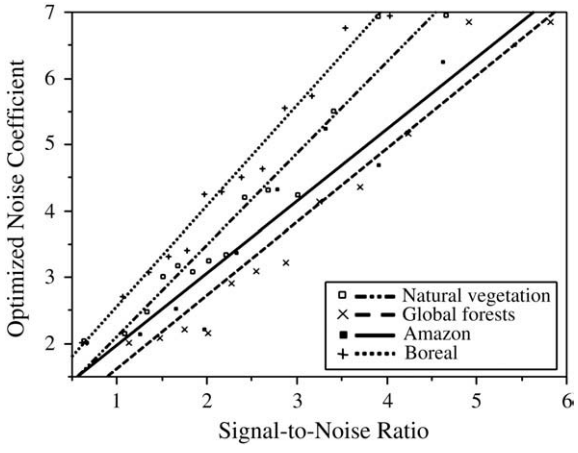

Fig. 7. Optimized noise coefficients by mean SNR level with best fit lines.

5. Conclusion

Radiometric correction is an element of all satellite remote sensing missions and spaceborne lidar missions will require it as well. Current waveform processing approaches require that a threshold return energy level be defined to separate signal from noise. At signal-to-noise ratio (SNR) values similar to those of current sensors, small surface areas of foliage or ground can return an amount of energy that might also represent atmospheric returns. Given this circumstance, no threshold value can be 100% accurate and the best we can achieve is a balance between misidentifications of signal for noise or its opposite.

Increasing the noise coefficient as a function of SNR was overall the best approach of those we examined, however power level was a close second and was the best approach for the Amazon dataset. These results indicated that when power or the SNR is low, the threshold should be a low multiple of the background noise level, independent of the value of the noise level. As signal-to-noise ratio increases, the probability of missing signal decreases and a higher multiple of background noise can be used.

These approaches improve the consistency of near-coincident waveform extent measurements but not necessarily their accuracy. Future work should combine analysis of near-coincident satellite observations with near-coincident satellite and field or airborne lidar

Please cite this article as: Miller, M.E., et al., Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.09.002

Table 7
(1) Mean waveform extent for each method and each study area. (2) RMSD values for each optimization method and the intra-period comparison. (3) RMSD values corrected for intra-period RMSE. (4) RMSD as a percent of mean extent. (5) Corrected RMSD as a function of waveform extent. (6) Percent reduction of RMSD% relative to the constant value method using data from Eq. (4). (7) Percent reduction of RMSD% relative to the constant value method using data from Eq. (5). See text for details of statistical significance between optimization methods.

1. Mean extent

	Simple	Period	Noise	Power	SNR
Natural	11.0	10.5	10.9	10.0	9.9
Forest	17.5	17.7	17.4	15.9	15.6
Amazon	19.7	20.2	19.4	18.4	18.5
Boreal	10.3	10.4	10.0	9.7	9.6

2. RMSD

	Simple	Period	Noise	Power	SNR	Intra-period RMSD
Natural	4.8	4.3	4.7	3.7	3.6	3.3
Forest	8.2	7.8	8.3	6.2	6.2	4.6
Amazon	6.9	6.7	6.4	4.6	5.0	1.7
Boreal	5.0	4.6	5.0	4.1	3.9	3.0

3. Corrected RMSD

	Simple	Period	Noise	Power	SNR
Natural	3.43	2.74	3.39	1.62	1.41
Forest	6.84	6.32	6.89	4.21	4.12
Amazon	6.69	6.49	6.18	4.31	4.70
Boreal	3.96	3.54	3.96	2.77	2.53

4. RMSD %

	Simple	Period	Noise	Power	SNR
Natural	43%	41%	43%	37%	36%
Forest	47%	44%	47%	39%	40%
Amazon	35%	33%	33%	25%	27%
Boreal	48%	45%	50%	42%	41%

5. Corrected RMSD %

	Simple	Period	Noise	Power	SNR
Natural	31%	26%	31%	16%	14%
Forest	39%	36%	39%	27%	26%
Amazon	34%	32%	32%	23%	25%
Boreal	38%	34%	40%	29%	26%

6. RMSD %: percent reduction from Simple method

	Simple	Period	Noise	Power	SNR
Natural		6%	0%	15%	16%
Forest		6%	−1%	17%	16%
Amazon		5%	6%	28%	23%
Boreal		7%	−3%	12%	15%

7. Corrected RMSD% percent reduction from Simple method

	Simple	Period	Noise	Power	SNR
Natural		16%	0%	48%	54%
Forest		9%	−1%	32%	32%
Amazon		5%	6%	31%	25%
Boreal		11%	−3%	25%	31%

data to improve both precision and accuracy. Rigorous evaluations of possible objective functions should also be incorporated into the next set of analyses.

References

Abshire, J. B., Sun, X., Riris, H., Sirota, M., McGarry, J., Palm, S., et al. (2005). Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit measurement performance. *Geophysical Research Letters*, *32*, L21S02. doi:10.1029/2005GL024028.

ACIA (2005). *Arctic climate impact assessment*. Cambridge, UK: Cambridge Univ. Press.

Blair, J., & Hofton, M. (1999). Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophys. Res. Lett.*, *26*(16), 2509−2512.

Brenner, A. C., Zwally, H. J., Bentley, C. R., Csathó, B. M., Harding, D. J., Hofton, M. A., et al. (2003). Derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights. *GLAS ATBD Version 4.1* 92 pp..

Dubayah, R., Blair, J. B., Bufton, J. L., Clark, D. B., JaJa, J. R., Knox, G., et al. (1997). The vegetation canopy lidar mission. *Land satellite information in the next decade II: Sources and applications* (pp. 100−112). Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

Duong, H., Lindenbergh, R., Pfeifer, N., & Vosselman, G. (2008). Single and two epoch analysis of ICESat full waveform data over forested areas. *International Journal of Remote Sensing*, *29*(5), 1453−1473.

Harding, D. J., Lefsky, M. A., Parker, G. G., & Blair, J. B. (2001). Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment*, *76*, 283−297.

Harding, D. J., & Carabajal, C. C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, *32*, L21S10. doi:10.1029/2005GL023471.

IGARB report, NASA (2003). Executive summary. available from. http://icesat.gsfc.nasa.gov/docs/IGARB.pdf.

Jellison, J. E. (1979). Gold–indium intermetallic compounds: Properties and growth rates. In H. Leidecker (Ed.), *NASA GSFC Code 313, Materials Control and Applications Branch (MC&AB), 11/8/79* 7/03.

Leidecker, H. W. (2003). Indium solder encapsulating gold bond wire leads to fragile gold–indium compounds and an unreliable condition that results in wire interconnection rupture. *GSFC NASA Advisory, Advisory # NA-GSFC-2004-01*.

Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). LiDAR remote sensing for ecosystem studies. *Bioscience*, *52*(1), 19−30.

Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Espirito-Santo, F. D., et al. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, *32*, L22S02.

Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., & Shugart, H. H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of Eastern Maryland, USA. *Remote Sensing of Environment*, *67*, 83−98.

Lefsky, M. A., Keller, M., Pang, Y., de Camargo, P., & Hunter, M. O. (2007). Revised method for forest canopy height estimation from the Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, *1*, 013537.

Matott, L.S. (2005). OSTRICH: An optimization software tool; documentation and user's guide. Version 1.6. Technical report, Department of Civil, Structural, and Environmental Engineering, University at Buffalo.

Morsdorf, F., Meier, E., Kotz, B., Itten, K. I., Dobbertin, M., & Allgower, B. (2004). LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, *92*(3), 353−362.

National Research Council (2007). Earth science and applications from space: National imperatives for the next decade and beyond. *Committee on Earth Science and Applications from Space: A Community Assessment and Strategy for the Future, Space Studies Board, Division on Engineering and Physical Sciences, National Academy of Sciences* (pp. 437). Washington, D.C.: National Academies Press.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, *51*, 933−938.

Ometto, J., Nobre, A. D., Rocha, H. R., Artaxo, P., & Martinelli, L. A. (2005). Amazonia and the modern carbon cycle: Lessons learned. *Oecologia*, *143*(4), 483−500.

Rosette, J., North, P., & Suárez, J. (2008). Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *International Journal of Remote Sensing*, *29*(5), 1475−1493.

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., & DiMarzio, J. P. (2005). Overview of the ICESat Mission. *Geophysical Research Letters*, *32*, L21S01. doi:10.1029/2005GL024009.

Strahler, A., Muchoney, D., Borak, J., Gao, F., Friedl, M., Gopal, S., et al. (1999). *MODIS land cover product, algorithm theoretical basis document (ATBD), Version 5.0*. Boston, MA: Center for Remote Sensing, Department of Geography, Boston University.

Sun, G., Ranson, K. J., Kharuk, V. I., & Kovacs, K. (2003). Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. *Remote Sensing of Environment*, *88*(4), 401−411.

Tickle, P. K., Lee, A., Lucas, R. M., Austin, J., & Witte, C. (2006). Quantifying Australian forest floristics and structure using small footprint LiDAR and large scale aerial photography. *Forest Ecology and Management*, *223*(1–3), 379−394.

Zhang, X. Y., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., et al. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, *84*(3), 471−475.

Zwally, H. J., Schutz, B., Abdalati, W., Abshire, J., Bently, C., Brenner, A., et al. (2002). ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, *34*(3–4), 405−445.

Zwally, H. J., Schutz, B., Bentley, C., Bufton, J., Herring, T., Minster, J., et al. (2003). *GLAS/ICESat L2 Antarctic and Greenland Ice Sheet Altimetry Data V001*. Boulder, CO: National Snow and Ice Data Center. Digital media.

Please cite this article as: Miller, M.E., et al., Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.09.002