The Potential Impact of Vertical Sampling Uncertainty on ICESat-2/ATLAS Terrain and Canopy Height Retrievals for Multiple Ecosystems

Amy L. Neuenschwander and Lori A. Magruder *

Applied Research Laboratories, University of Texas at Austin, Austin, TX 78712, USA; amyn@arlut.utexas.edu
* Correspondence: magruder@arlut.utexas.edu; Tel.: +1-512-835-3068

Abstract: With a planned launch no later than September 2018, the Ice, Cloud and land Elevation Satellite-2 (ICESat-2) will provide a global distribution of geodetic elevation measurements for both the terrain surface and relative canopy heights. The Advanced Topographic Laser Altimeter System (ATLAS) instrument on-board ICESat-2 is a LiDAR system sensitive to the photon level. The photon-counting technology has many advantages for space-based altimetry, but also has challenges, particularly with delineating the signal from background noise. As such, a current unknown facing the ecosystem community is the performance of ICESat-2 for terrain and canopy height retrievals. This paper aims to provide the science user community of ICESat-2 land/vegetation data products with a realistic understanding of the performance characteristics and potential uncertainties related to the vertical sampling error, which includes the error in the perceived height value and the measurement precision. Terrain and canopy heights from simulated ICESat-2 data are evaluated against the airborne LiDAR ground truth values to provide a baseline performance uncertainty for multiple ecosystems. Simulation results for wooded savanna and boreal forest result in a mean bias error and error uncertainty (precision) for terrain height retrievals at 0.06 m (0.24 m RMSE) and −0.13 m (0.77 m RMSE). In contrast, results over ecosystems with dense vegetation show terrain errors of 1.93 m (1.66 m RMSE) and 2.52 m (3.18 m RMSE), indicating problems extracting terrain height due to diminished ground returns. Simulated top of canopy heights from ICESat-2 underestimated true top of canopy returns for all types analyzed with errors ranging from 0.28 m (1.39 m RMSE) to 1.25 m (2.63 m RMSE). These results comprise a first step in a comprehensive evaluation of ICESat-2 anticipated performance. Future steps will include solar noise impact analysis and investigation into performance discrepancy between visible and near-infrared wavelengths.

Keywords: LiDAR; ICESat-2; photon-counting; vegetation mapping

1. Introduction

With a launch expected in the latter half of 2018, the ICESat-2 satellite will provide a global distribution of geodetic measurements from a space-based laser altimeter of both the terrain surface and relative canopy heights, which will provide a significant benefit to society through a variety of applications ranging from forest structural mapping to improved global digital terrain models. The Earth's land surface is a complex mosaic of geomorphic units and land cover types resulting in large variations in terrain height, slope, roughness, vegetation height and reflectance, often with the variations occurring over very small spatial scales. Documentation of these landscape properties is a first step in understanding the interplay between the formative processes and response to changing conditions. Characterization of the landscape is also necessary to establish boundary conditions for models that are sensitive to these properties, such as predictive models of atmospheric change that

depend on land-atmosphere interactions. Topography, or land surface height, is the fundamental component for many relative height applications, both to the scientific and commercial sectors. The most accurate global elevation product was produced by the Shuttle Radar Topography Mission (SRTM) launched in 2000; however, elevation data are limited to non-polar regions. Furthermore, the accuracy of SRTM-derived heights range from 5–10 m, depending on the amount of topography and vegetation cover over a particular area [1,2].

In addition to producing a global terrain product, monitoring the amount and distribution of above ground vegetation and carbon pools enables improved characterization of the global carbon budget. The quantification of the global carbon pools has been a focus of research in the recent past with a goal to understand forest change due to growth, degradation and deforestation [3–5]. Uncertainties in the global carbon budget are directly related to the limited knowledge of vegetation structure, specifically canopy height [6]. Airborne LiDAR has been used for small studies to capture canopy height, and in those studies, canopy height variation for multiple forest types is measured with uncertainties ranging from 1–9 m depending on the ecosystem and type of study [6–8]. Although the spatial extent and changes to forests can be mapped with existing satellite remote sensing data, the lack of information on forest vertical structure and biomass limits the knowledge of biomass/biomass change within the global carbon budget. A space-based LiDAR mission can provide a baseline for global carbon stocks [4,9]. In 2011, the ecosystem community outlined the need for a global biomass and 3D vegetation structure product for science and policy decisions [6]. By 2021, there will be several spaceborne systems dedicated to quantifying the carbon stocks, including NASA's Global Ecosystem Dynamics Instrument (GEDI) and NASA's NISAR (the collaborative NASA and Indian Space Research Organization SAR mission). Canopy height measurements from ICESat-2 will complement these other NASA missions through the quantification of global forest carbon stocks.

ICESat-2 will provide the community global elevation measurements to support multiple scientific disciplines. The ICESat-2 scientific goals are similar to those of its predecessor, ICESat, particularly for cryospheric and ecological applications. However, the differences of the ICESat-2 onboard technology from that of ICESat were driven by the goal of providing elevation and slope information with enhanced spatial coverage. The data acquisition operational scenario planned for ICESat-2 over the mid-latitudes will provide a global distribution of geodetic measurements of both the surface and vegetation; creating a terrain and canopy height product intended to enhance existing geophysical data resources and provide opportunities for further understanding of our environment.

The Advanced Topographic Laser Altimeter System (ATLAS) instrument will operate at 532 nm in the green range of the electromagnetic (EM) spectrum and will have a laser repetition rate of 10 kHz and a laser pulse width of <2 ns. Although a space-based laser propagating at a 532-nm wavelength is subject to more atmospheric attenuation than 1064 nm, wavelength-specific differences of ground and canopy characteristics between 532 nm and 1064 nm from airborne data have been reported as relatively minor [10]. The combination of the laser repetition rate and satellite velocity will result in a center to center separation of 70 cm for each laser footprint on the surface. Based on the altitude of the spacecraft (~496 km) and the divergence properties of the laser, the footprint diameter is estimated to be 13 m at the start of the mission and will gradually increase to 17 m as the laser divergence increases toward the end of the mission [11,12] due to inherent loss of laser energy and degradation of the optics. Each Gaussian laser pulse is split by a diffractive optical element within ATLAS's optical path to generate six individual beams, transmitted in a three pair configuration (Figure 1). The beams within each pair have different transmit energies, which constitute the label of the 'strong' beam and the 'weak' beam based on the 4:1 energy ratio. This relative energy difference is anticipated to provide compensation for the detector dynamic range given variations in surface reflectance. Due to the relatively low surface reflectance from vegetation at 532 nm, the signal from the weak beam may be too weak to be detected. The beam pairs are separated by ~3.3 km in the across-track direction, and the strong and weak beams are separated by ~2.5 km in the along-track direction. As ICESat-2 moves along its orbit, the ATLAS beams describe six tracks on the Earth's surface; the array is rotated

slightly with respect to the satellite's flight direction so that tracks for the fore and aft beams in each column produce pairs of tracks, each separated by approximately 90 m. Once on-orbit, ICESat-2 will begin its measurement collection along a planned reference ground track for an orbital repeat cycle of 91 days. However, once the initial 91-day cycle is completed, ICESat-2 will maneuver for off-nadir pointing over the mid-latitudes during subsequent cycles, which will ultimately provide a uniform distribution of measurements over the optimal three-year operational lifetime of the satellite. The planned distribution once complete should have a spatial separation between collection tracks of 1 km at the Equator. As a note to the reader, ICESat-2 will not be in the same orbit as ICESat.
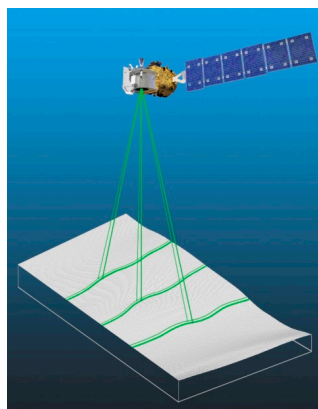

Figure 1. Beam configuration of the Advanced Topographic Laser Altimeter System (ATLAS) instrument on ICESat-2.

The ICESat-2 mission will provide several geophysical data products specific to the measured surface type (e.g., ice sheets, ocean and land) and format (e.g., along-track and gridded). The land product, ATL08, will contain the along-track terrain and canopy heights, canopy cover and other descriptive parameters derived from the ATLAS measurements. Height estimates from ATL08 are poised for comparison to other geodetic data and supply the foundation for higher level data products. As such, this effort presents the expected performance of ICESat-2 for vegetation characterization over specific ecosystems, but will also highlight the interpretation approach for these anticipated datasets and how they might be leveraged for additional scientific objectives and realizations.

ATLAS will use a multi-beam, micro-pulse laser and single-photon detection capabilities to measure the range between the instrument and the reflective surface. Of primary interest to the ecosystem community is the accuracy to which the terrain and canopy height can be derived from altimetry in the forest environment. Certainly, there is ambiguity as to where within the vertical column the photon will be reflected, which creates an opportunity for potential biases in the derived heights. Given the differences in the photon-counting and full-waveform altimetry detection approaches, there is not a direct translation to the performance capabilities between the two system types shown in Figure 2. Full-waveform sensors record the entire temporal profile of the reflected energy along the laser line of sight, including the vertical structure of the canopy when present. However, this profile record is based on the signal exceeding a certain energy threshold and is not impacted by significant noise. Features in the waveform correspond to surfaces encountered by the laser energy (e.g., clouds, vegetation or terrain). In contrast, photon-counting systems will record time tags (arrival) for every detection event within the vertical distribution, including both the signal and noise. As such, an accumulation of multiple shots (>100) could produce a histogram of ranges based on the received photons similar to the product provided by the full-waveform systems and statistically representative of a probability distribution function (PDF) for the location of a detection within the vertical extent of the measurement. This statistical assessment of where the photons are actually reflected from within the PDF is referred to as the vertical sampling uncertainty and points to error associated with

the interpretation of the height from the data, as well as the precision of the measurement. Again, the focus of this study is to quantify the vertical sampling uncertainty for various ecosystems in the form of a mean height bias and measurement precision as a first step toward quantifying the potential performance for ICESat-2 once it is on orbit. Effectively, the bias value is an indication of the accuracy of the height retrieval process as compared to ground truth, and the precision is the standard deviation of the measurements over the multiple iterations. Certainly, the results from this study will not provide a comprehensive understanding of the anticipated quality of ICESat-2 data, but will create a performance baseline for future studies that can build upon this first step. Future studies are warranted in order to investigate how the presence of solar noise will potentially impact the vertical sampling errors associated with the height retrieval process and how the use of a visible wavelength might differ from measurements collected using the near-infrared over vegetation.
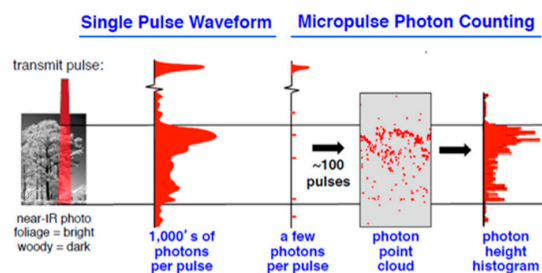

Figure 2. An illustration of the comparison between full-waveform LiDAR and photon-counting LiDAR technologies with an example of how an accumulation of individual photon detections can provide a similar statistical representation of the vertical structure to that produced from the full-waveform data product [13].

The number of detected photon events associated with each outgoing laser pulse is a function of the transmitted laser energy, surface reflectance, solar conditions and scattering and attenuation in the atmosphere. For highly reflective surfaces (e.g., ice sheets) and clear skies, approximately 10 signal photons from a single strong beam are expected to be detected and recorded by the ATLAS instrument for a given transmit pulse. Over vegetated land where the surface reflectance is considerably less than snow or ice surfaces, the signal level is anticipated to be considerably less. This reduction in signal is to be expected based on the design cases analyzed for surface reflectance characteristics at a 532-nm wavelength; snow and ice surfaces have high reflectance (Lambertian), between 0.8 and 0.98 [14], while canopy and terrain surfaces have much lower reflectance (typically around 0.3 for soil and 0.1 for vegetation). As a consequence, there will be 1/3–1/9 as many photons returned from terrestrial surfaces as from ice and snow surfaces. For vegetated areas, the number of reflected signal photon events per transmitted laser pulse is estimated to range between 0 and 3 photons for the strong beam case, depending on land cover and atmospheric conditions.

The ICESat-2 Instrument team has developed a full series of design cases for different generic land cover types in order to support the development of an estimation model for quantitative signal prediction. This model is based on the expected number of photons using a Poisson distribution for each outgoing laser shot using the planned ICESat-2 measurement scenario (footprint size and spacing) and the vegetation and soil reflectance characteristics. Three design cases for vegetation are shown in Table 1. For example, every 70 cm in the along-track direction, the ATLAS instrument will potentially detect zero, one or two returns over boreal forest.

Geophysical algorithms developed by the ICESat-2 Science Definition Team (SDT) will produce estimates of terrain heights and relative canopy heights based on the statistical analysis of detected photons. The objective of this research is to determine the effect of vertical sampling error as a first step towards quantifying the likely performance characteristics and potential uncertainties in the land and canopy height products of ICESat-2/ATLAS for different vegetation types.

Table 1. Design cases and photon numbers.

Vegetation Type	Poisson Mean Photoelectrons per Outgoing Shot	Poisson Distribution
Boreal Forest	1.0	[0 1 2]
Temperate Forest	1.9	[0 1 2 3]
Tropical Forest	0.6	[0 1]

2. Materials and Methods

2.1. Data

In support of the performance study for space-based photon-counting LiDAR specific to terrestrial applications, ICESat-2/ATLAS data were simulated for multiple ecosystem types using existing airborne, small-footprint LiDAR datasets. Five specific areas were selected and include an oak-juniper savanna in central Texas, a pinon-juniper savanna in central New Mexico, boreal forest in central Alaska, deciduous hardwoods in eastern Tennessee and tropical moist forest in western Gabon. Each of these individual areas are discussed below and are also summarized in in Table 2.

Table 2. Summary of test data utilized for the analysis.

Vegetation Type Analyzed	Data Type Available for Simulation	Location	Vegetation Type
Woodland Savanna	Full-waveform airborne LiDAR	Texas, New Mexico	Live oak, Ashe juniper, pinon pine
Boreal Forest	Discrete return airborne LiDAR	Alaska	Black spruce
Temperate Harwood	Full-waveform airborne LiDAR	Tennessee	Deciduous hardwoods (red oak, white oak, scarlet oak, chestnut oak, mockernut hickory, sweet birch and bear oak
Tropical Forest	Discrete return airborne LiDAR	Gabon, Africa (Mondah Forest)	Rain forest

Woodland savanna: Small-footprint, full-waveform airborne LiDAR data were acquired in September 2010 over Freeman Ranch (29°56′N, 98°W) located near San Marcos, Texas. Freeman Ranch (~17 km^2 in size) is a live oak (*Quercus virginiana*)-Ashe juniper (*Juniperus ashei*) savanna woodland and is typical of the vegetation in central Texas. For this study area, live oak is best characterized with trees heights approaching 13 m and having distinct vertical layers within the canopy. For this survey, the average point density of the LiDAR survey was ~10 laser shots per m^2. The accuracy of the small-footprint LiDAR data was determined using kinematic GPS survey data collected within the survey area and was found to have a vertical accuracy of 10 cm or better for open surfaces.

Additional airborne LiDAR data were collected over several sites in central New Mexico in 2011. One of the sites is located in a pinon-juniper savanna woodland, 24 km southeast of Willard, NM (34.425489, −105.861545), and this study region exists within a managed rangeland ecosystem, with one-seed juniper (*J. monosperma*) and pinon pine as the two dominant woody species present. The LiDAR point cloud data had an average horizontal point spacing of 20 cm and a point density of approximately 10 points per m^2. The mean horizontal relative accuracies was 25 cm, and vertical accuracy was approximately 10 cm or better for open surfaces.

Boreal forest: The NASA Goddard Lidar, Hyperspectral and Thermal imager (G-LiHT) near Fairbanks, Alaska, in 2014 was also used in this study. The average point density for this dataset is approximately 5 points per m^2. Similar to the approach used with small-footprint, full-waveform data to create a statistical representation of the ICESat-2 signal characteristics, the discrete elevations provided by this type of airborne sensor system can be aggregated to create a pseudo-waveform. However, the concession of using discrete data is effectively a lower digitization rate, meaning the full-waveform data provide a complete vertical history of the propagated laser energy, while the discrete return offers a limited sampling rate along the laser line of sight. Although the discrete return point density is not considered particularly high for this type of simulation, a 13 m-diameter simulated

ICESat-2 footprint would consist of roughly 660 LiDAR points to construct the pseudo-waveform and does provide sufficient information as to the canopy structure and the relative capability of ICESat-2 for this region. Vegetation in the interior boreal forest consist of predominately black spruce.

Temperate hardwoods: Small-footprint, full-waveform airborne LiDAR data were acquired in October 2011 over Bays Mountain, TN (36°31′N, 86°35′W), located near Kingsport, TN. This region in eastern Tennessee is a mixture of deciduous hardwoods (i.e., red oak, white oak, scarlet oak, chestnut oak, mockernut hickory, sweet birch and bear oak). The survey area corresponded to 1400 m × 250 m over the northern region of Bays Mountain. The LiDAR data were acquired with an average point density of 20 points per m^2. The accuracy of the small-footprint LiDAR data was determined using kinematic GPS survey data collected within the survey area and was found to have a vertical accuracy of 10 cm or better for open surfaces.

Tropical forest: Small footprint airborne discrete LiDAR data with a nominal point density of 40 points per m^2 over Gabon, Africa, were used to support analysis over a tropical forest environment. Data with this point density allow for the construction of higher fidelity pseudo-waveforms for canopy structure detail than achieved for the lower density G-LiHT datasets. This airborne LiDAR data were collected in November 2011 over the Mondah forest in Gabon. Information regarding the LiDAR acquisition scenario was not provided with the data; however, the data were determined to have a vertical accuracy of 10 cm or better in open areas. The Mondah forest is on the Libreville peninsula and is located approximately 20 km northwest of Libreville (0°34′N, 9°20′E). The Mondah forest has been traditionally protected; however, recently, peri-urban growth has occurred on its perimeter, and the amount of protected lands has decreased from 10,200 hectares in 1951 to 6750 hectares in 2012 [15]. In the protected area, the habitat is tropical moist forest, and tree heights range from 15–25 m, with emergent species reaching upwards of 45 m.

2.2. Methodology

Small-footprint, full-waveform LiDAR is essentially a surrogate for the statistical distribution of expected photons from ATLAS once on orbit. As such, these temporal profiles of received laser energy can essentially be used to empirically derive simulated ICESat-2 data by using a normalized waveform as a representative probability distribution function of the detected photons along the simulated vertical distribution. In instances where full-waveform data were not available, small-footprint discrete return data having a high point density (i.e., >30 points per m^2) were used as a surrogate to generate a pseudo-waveform. For this simulation, center coordinates of ICESat-2 laser footprints are determined using a 70-cm center-to-center spacing along a transect within an existing airborne LiDAR survey. Using these footprint center coordinates, the algorithm determined all small-footprint waveforms occurring within the radius of the ATLAS measurement. That is, all small-footprint waveforms within a 6.5-m radius of the center are identified and combined into a 13-m diameter waveform using a $1/e^2$ weighting distribution based on the center position. This combined waveform is referred to here as a synthesized waveform, or pseudo-waveform, as mentioned previously. Figure 3 outlines the general methodology used to generate the simulated ICESat-2 data. To generate a probabilistic distribution function (PDF) of laser energy through the canopy, the combined waveform is normalized such that the integrated waveform amplitude sums to a value of one. For each PDF, a height vector is constructed originating at the perceived ground surface to the top of the canopy within each measurement simulation. Based on the design cases developed by the ICESat-2 Project Science Office, a mean number of photons returned for each output laser shot over various ecosystems can be estimated and used to simulate the ICESat-2 data. Using a Poisson random distribution, the number of photons returned for each ICESat-2 footprint, X, was randomly sampled. Next, the height vector is randomly sampled X times using the normalized PDF as the weighting function. For the cases where full-waveform data were not available, small-footprint discrete return data were used. In these cases, the point density was approximately 10 points/m^2 or greater, which should be sufficient to generate a pseudo-waveform to conduct the random sampling. The result of the sampling is a photon (or zero or

two) for each simulated ICESat-2 footprint that is randomly sampled from the ground or within the canopy. The simulations shown here were implemented in MATLAB 2015b using software developed by the author. Statistics for each simulation were also produced in MATLAB.
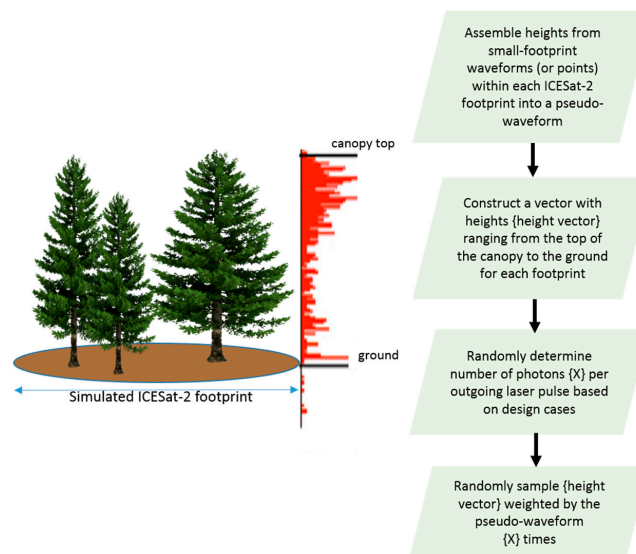

Figure 3. Flow chart illustrating how the small-footprint LiDAR data are combined to create a pseudo-waveform used to simulate ICESat-2 data.

A significant benefit to using airborne LiDAR data in the simulation of ICESat-2/ATLAS data is the role it can provide as ground truth when analyzing the expected space-based results in terms of estimating the terrain and canopy heights for specific regions. The truth height for the terrain is the height associated with the perceived ground (last) peak in the synthesized waveform. The truth height for the top of the canopy is the maximum height associated with the synthesized waveform. In addition to the top of canopy and ground heights, relative canopy height metrics RH50, RH60, RH75, RH90 and RH95 were also computed from the truth data. Relative canopy heights are defined as the percentage of canopy returns above the terrain surface.

The truth heights and height metrics extracted from the synthesized waveforms are compared to those derived from the simulated ICESat-2 data to examine the performance characteristics of the photon-counting system for terrain and canopy height recovery. The initial comparison analyzes the validity to which the signal photons are identified as terrain or canopy. Classifying the signal as ground or canopy is a challenging process given the vertical distribution along a surface inherent to photon-counting systems. The vertical distribution, or point spread function (PSF), of the ATLAS instrument is estimated to be 25 cm based on the pulse width of the laser and expected uncertainties in the timing. As such, ground photons and top of canopy photons were identified as follows using the PSF parameter. The two equations below provide the justification for identifying signal photons as ground (G) or as top of canopy (TOC) based on a criteria associated with the elevations of the aggregated data (step size). The step size was empirically determined as the quantity that would provide a statistically-relevant assessment of the terrain without diminishing the topographic detail.

$$Photon_G = \{photons < min_{20} + 3PSF\}$$

$$Photon_{TOC} = \{photon < (abs(max_{20} - 1))\}$$

Thus, using an optimized step size of 20 photons along the theoretical ICESat-2 ground track, ground photons were identified as those within three times the PSF relative to the minimum height in the sample bin. The top of canopy photons were identified as those within 1 m of the sample

bin (20 photons) simulated ICESat-2 maximum height. The identified ground photon heights were differenced from the truth ground height for each simulated ICESat-2 footprint to examine the potential capability of ICESat-2 in terms of estimating terrain elevations. Similarly, the top of canopy photons were differenced from the true top of canopy height within each simulated footprint. In addition to true top of canopy height, height metrics, such as RH50, RH60, RH75, RH90 and RH95, were computed. Terrain height residuals for the individual iterations (1 or 100) were evaluated for each test area as a function of percent canopy cover. Canopy cover was computed from the input, small-footprint data and is defined as the ratio of the number of canopy returns to the total number of returns.

3. Results

The results from the residual analysis described in the previous section are shown in Table 3 for each ecosystem type considered in this effort. Table 3 specifically highlights the theoretical mean RMS and mean bias values for both the terrain and canopy height estimates for ICESat-2 simulations compared to the airborne LiDAR ground truth dataset. In all cases, as would be expected, the RMS values for terrain and canopy heights increase as the simulated ICESat-2 footprint diameter increases. Disparate footprint diameters were investigated to understand how anticipated changes in the divergence of the laser over the mission lifetime might affect the accuracy of the land products as footprint size increases. Results here indicate that ecosystems dominated by conifers should yield ICESat-2 results that compare well with truth for the entire course of the ICESat-2 mission. For these ecosystems in the simulations, a sufficient number of photons is reflected from the ground surface enabling a consistent retrieval of terrain height and subsequent relative canopy height. In ecosystems with and an increased percentage of canopy cover, i.e., broadleaf forest and tropical forest, the mean bias and RMS value of the terrain estimates are higher at 1.93 m/1.66 m and 2.52 m/3.18 m, respectively. For these ecosystem simulations, it is clear that very few photons are received from the ground surface as the majority of the signal will be reflected from the canopy. This obscuration associated with the higher percentage canopy cover make it difficult to resolve relative terrain and canopy heights as accurately as those environments with less vegetation cover since the lowest elevation detected is often from within the canopy rather than the ground surface. This situation will result in an incorrect terrain height and subsequent incorrect canopy height on the ICESat-2 data product if there is not a mitigation strategy for these known errors.

3.1. Savanna Woodland

Although all of the quantitative results are shown in Table 3, in the interest of brevity, the details and discussion of the analysis will be provided for test sites with unique results since some of the areas have obvious similarities. For example, the Freeman Ranch, Texas, results are quite comparable to the pinon-juniper woodland in New Mexico, as these ecosystems exhibit similar structural characteristics, as well as simulation results. The first test area that will be described in detail is Freeman Ranch, Texas. The results for this region (Table 3) indicate that the terrain heights from the simulated data agree well with the ground truth. When evaluating the residuals for 100 iterations assuming a 13-m diameter footprint, the ground height was biased by 0.06 m and had an RMS uncertainty of 0.24 m. Considering the 0.25 m PSF of the ATLAS instrument, there is confidence that the ICESat-2/ATLAS interpretation of terrain heights in woodland/savanna ecosystems will be accurate to that level. The top of canopy residuals reveal a bias of 0.28 m with an RMS error of 1.39 m for this ecosystem. In this case, the top of the canopy is biased low by approximately 0.28 m indicating that canopy heights in this ecosystem will be under-represented by a small amount. Figure 4 provides the simulated ICESat-2 data for the Freeman Ranch dataset with the truth terrain and top of canopy points included for comparison. The pink dots in the lower profile represent the uppermost canopy points of the simulated ICESat-2 data, whereas the dark green represent the actual top of canopy from the airborne LiDAR "truth" data. Blue dots represent the interpreted ICESat-2 ground points, and these agree well with the "truth" ground photons, shown in red.

Table 3. Statistics computed on 100 realizations of residuals between sampled ICESat-2/ATLAS and truth. A footprint diameter of 13 m is expected for the beginning of the ICESat-2 mission, whereas 17 m is anticipated at the end of the mission from laser degradation.

| Location	Forest Type	13-m Footprint Diameter				17-m Footprint Diameter			
		Ground Bias (m)	Ground RMS (m)	Canopy Top Bias (m)	Canopy Top RMS (m)	Ground Bias (m)	Ground RMS (m)	Canopy Top Bias (m)	Canopy Top RMS (m)
Texas	Oak Juniper Woodland	0.06	0.24	0.28	1.39	0.03	0.29	0.49	1.55
New Mexico	Pinon-Juniper woodland	0.09	0.28	1.03	1.58	0.07	0.32	1.09	1.57
Alaska	Boreal forest	−0.13	0.77	1.25	1.69	−0.32	0.96	1.73	1.82
Tennessee	Broadleaf Forest	1.93	1.66	1.25	2.63	2.23	1.89	1.53	2.85
Gabon	Tropical Forest	2.52	3.18	0.61	2.61	2.68	3.29	0.96	2.97
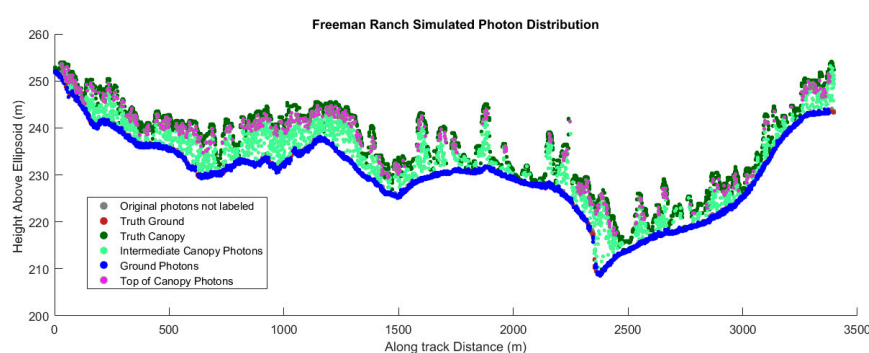
Figure 4. Simulated ICESat-2 profile at Freeman Ranch with the truth data included for comparison.

On the left side of Figure 5 is a plot of canopy cover against terrain height residuals. The Freeman Ranch analysis resulted in ground height residuals less than 1 m for canopy cover amounts up to 75%. The right side of Figure 5 is a histogram of the terrain residuals. The histogram for this simulation has an approximate Gaussian distribution centered at 6 cm above the truth.
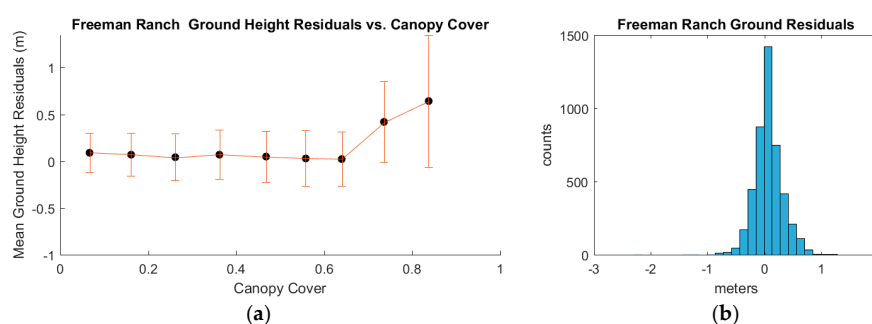

Figure 5. (**a**) Mean residuals (based on 100 iterations) of terrain height as a function of canopy cover for an oak-juniper savanna; the error bars represent the standard deviation; (**b**) histogram of terrain residuals for an oak-juniper savanna.

3.2. Boreal Forest

The second focus area is a Boreal forest located in central Alaska. This particular area analysis yielded interesting results based on the presence of increased vegetation height and topographic relief. Despite the characteristics of the vegetation in this region, there is still a sufficient number of identified

ground photons to produce reasonable estimates of terrain heights. For this simulation, the ground heights exhibited a mean bias of −0.13 m and an RMS uncertainty of 0.77 m. Considering that the entire simulated flight line had an elevation relief of over 230 m, it is anticipated that the terrain height estimates would contain more variability/uncertainty due to the increased topography within each simulated ICESat-2 footprint. That is, if there is significant topography within the footprint, the terrain estimate will have a higher uncertainty since the specific photon location (XY) within the 17-m diameter ICESat-2 footprint is ambiguous without a record of detected intensity [16]. Top canopy heights for boreal forest in this simulation were underestimated with a mean bias of 1.25 m and an RMS uncertainty of 1.69 m. Figure 6 provides a subset of the simulated ICESat-2 along-track data and the corresponding photon classification. In this figure, it is apparent that the ICESat-2 simulated height for top of canopy (pink dots) falls well below the truth top of canopy height.
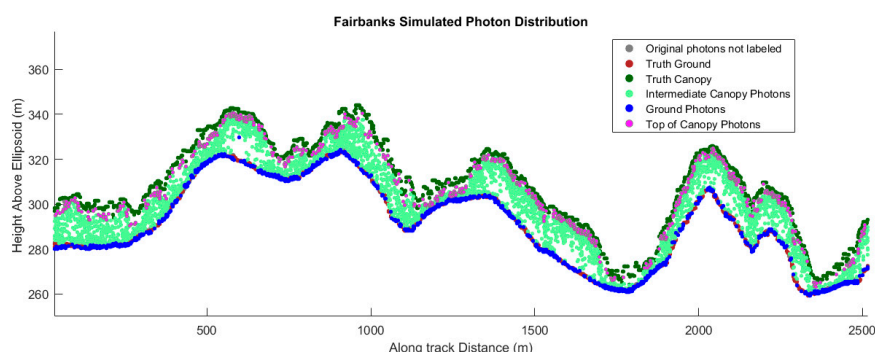

Figure 6. Profile of simulated ICESat-2 data. Pink dots represent the upper canopy points in the simulated ICESat-2 data, and blue dots represent the presumed ground points in the simulated ICESat-2 data.

An interesting trend for the terrain residuals can be observed in Figure 7 (left). This trends shows that the mean value of terrain residuals dips below zero, which indicates that the simulated minimum ground elevations are lower than the truth values. The likely cause for this trend is the significant amount of topography within this particular study area. The truth elevation is the last return of the pseudo-waveform rather than the minimum of an individual laser return within a footprint. In the case of topography, the topographic signal will broaden the last peak of the pseudo-waveform, which in turn will cause the truth elevation to be higher than an individual simulated photon within the footprint. Despite an increase in canopy cover, as compared to the oak-juniper savanna, the histogram of the terrain residuals also has a near Gaussian distribution, which indicates that the terrain surface in the simulations has an adequate number of photons reflected from the surface.
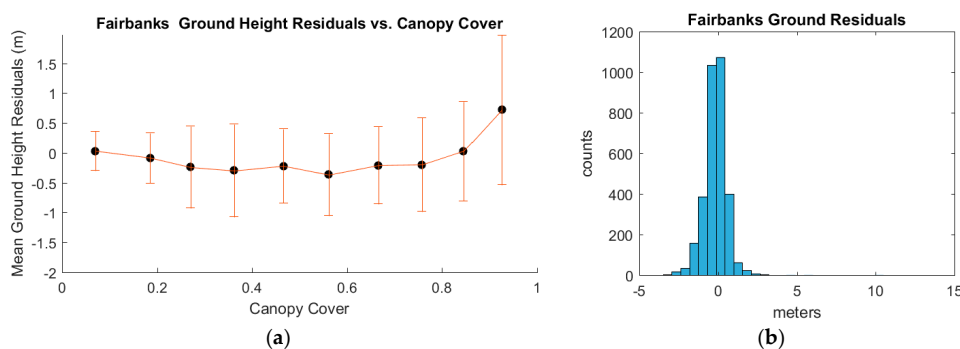

Figure 7. (**a**) Mean residuals (based on 100 iterations) of terrain height as a function of canopy cover for a Boreal forest located in central Alaska; the error bars represent the standard deviation; (**b**) histogram of terrain residuals for a Boreal forest.

3.3. Deciduous Forest

The third focus area to discuss is a deciduous forest in eastern Tennessee. The simulated ICESat-2 data and the corresponding photon distribution are presented in Figure 8. In this particular landscape, the canopy cover ranges from 40%–95%, and as such, we anticipate fewer photons penetrating through the canopy and reflecting from the ground. The simulation results validate this expectation and show a mean bias of 1.93 m and an RMS value of 1.66 m for the terrain heights. The increase in both the bias and RMS of terrain points indicates that simulated ICESat-2 photons are not reaching the ground surface; rather, they are being returned from within the canopy. Similarly, the top of canopy recovery had a bias of 1.25 m with a 2.63-m RMS uncertainty. The bias of 1.25 m indicates that, on average, the top portion of the canopy is not being fully captured. The RMS value of 2.63 m for top of canopy heights, however, exceeds the RMS value achieved from ICESat. Examination of Figure 8 indicates that there are a few portions along the flight line where the lowest simulated photon elevation occurs well within the canopy.
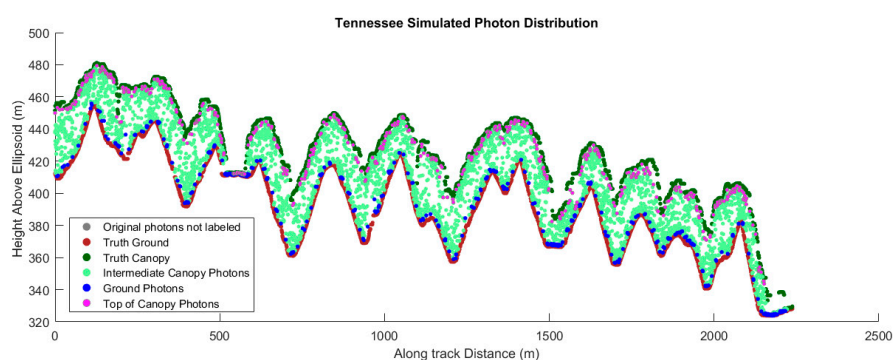

Figure 8. Profile of simulated ICESat-2 data. Pink dots represent the upper canopy points in the simulated ICESat-2 data, and blue dots represent the presumed ground points in the simulated ICESat-2 data.

For this study area, both the terrain height mean error and uncertainty increase with increasing canopy cover. As one might expect, the increasing canopy cover obscures the laser energy penetration, as well as any reflected energy from the terrain surface. In this simulation, a canopy cover of 50% results in potential terrain errors greater than 2 m, and a canopy cover percentage of 90% could entice errors on the order of 5 m. The histogram of the terrain residuals are no longer Gaussian in shape and are skewed in the positive direction. These height residual plots for the deciduous hardwood region are presented in Figure 9.
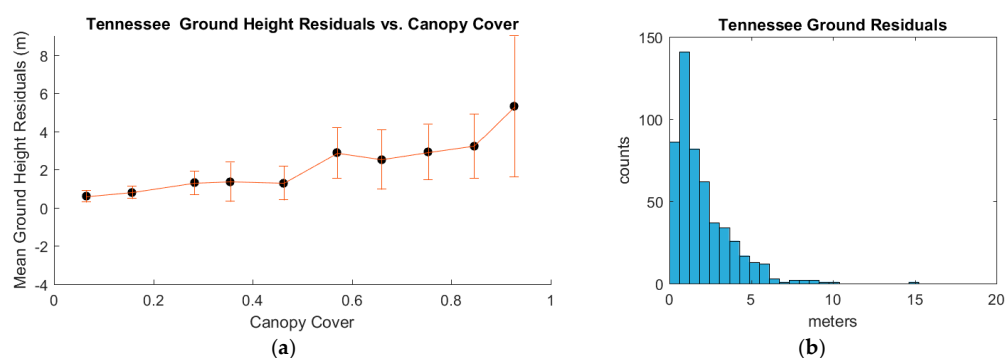

Figure 9. (**a**) Mean residuals (based on 100 iterations) of terrain height as a function of canopy cover for a deciduous hardwood forest located in eastern Tennessee; the error bars represent the standard deviation; (**b**) histogram of terrain residuals for a deciduous hardwood forest.

3.4. Tropical Forest

In Gabon, the representative site for tropical forests, the simulation results determined a decrease in elevation retrieval accuracies due to the diminished number of ground returns presumably from the lack of laser energy penetration through the dense canopy. The simulation results show a bias in the ground estimate of 2.52 m with an RMS error of 3.18 m for the 100 realizations. This bias indicates that the simulated photons perceived as ground are actually photons reflected from within the canopy; this is illustrated in Figure 10. Thus, accurate topographic heights will be difficult to extract from the ICESat-2 data without a mitigation strategy to address the decrease in true ground returns. In contrast to the terrain results, the top of canopy recovery had a bias of 0.61 m with a 2.61-m RMS uncertainty. The bias of 0.61 m indicates that, on average, the top portion of the ICESat-2 simulated canopy is quite close to the true maximum canopy height.
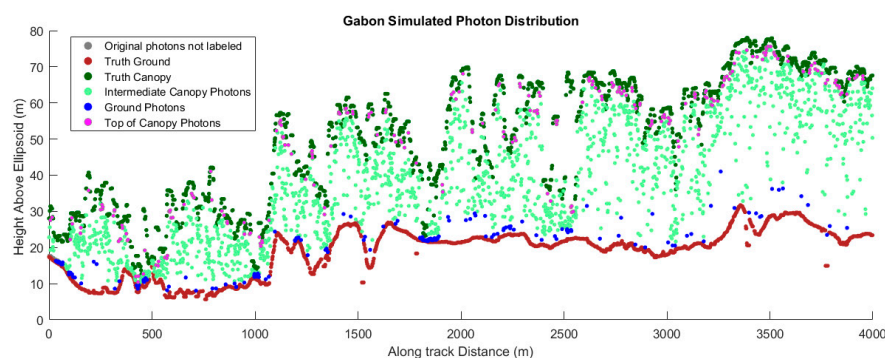

Figure 10. Profile of simulated ICESat-2 data. Pink dots represent the upper canopy points in the simulated ICESat-2 data, and blue dots represent the presumed ground points in the simulated ICESat-2 data.

Looking at the results presented in Figure 11, there is distinct similarity to those determined for the deciduous forest. This similarity is most likely due to the presence of significant canopy cover at both sites. As such, for this this study area, the mean error and uncertainty of the terrain heights both increase with increasing canopy cover. As one might expect, the increasing canopy cover prevents laser energy from returning from the terrain surface. In this simulation, a canopy cover of 70% results in potential terrain errors greater than 2 m, and for 80% canopy cover, the errors are on the order of 4 m. The histogram of the terrain residuals is no longer Gaussian in shape; rather these are bimodal, which indicates that many simulated ICESat-2 ground returns fell 4–8 m above the true ground surface and are reflected from understory vegetation.
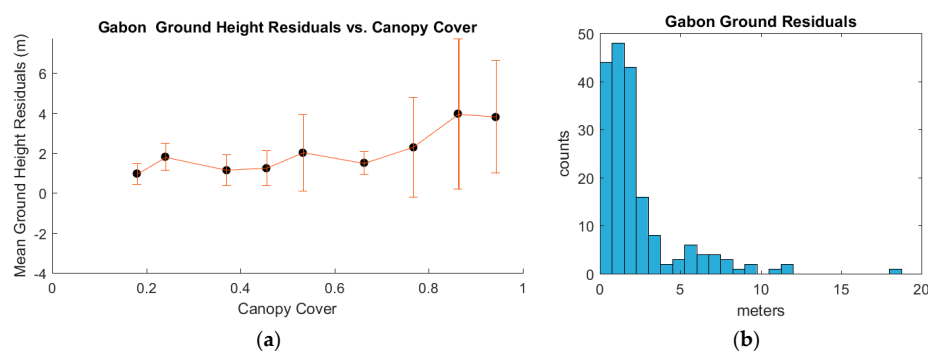

Figure 11. (**a**) Mean residuals (based on 100 iterations) of terrain height as a function of canopy cover for a tropical forest located in Gabon; the error bars represent the standard deviation; (**b**) histogram of terrain residuals for tropical forest.

4. Discussion

Evident from the results presented here, the deciduous hardwoods and tropical forest ecosystems will potentially see the largest errors associated with terrain height retrievals. These errors are most likely associated with the substantial extent of vegetated cover, as the signal received by the ATLAS instrument is dominated with reflections from within the canopy rather than the ground surface, creating a difficult scenario for accurate terrain height estimation and subsequent canopy heights within the ATL08 product. Any errors in the terrain height will propagate to errors in the canopy height and canopy metrics planned for the data product. One mitigation strategy involves introducing alternate sources of topographic information into the ATL08 processing scheme. Terrain models from other space-based systems, such as TandemX or GEDI, could be utilized; however, there are no plans at present for these data to be integrated into the ATL08 processing algorithm. These alternate terrain models come with their own errors and uncertainties; possibly confusing the errors associated from ICESat-2.

In addition to true top of canopy height, height metrics, such as RH50, RH60, RH75, RH90 and RH95, were computed. Although these simulation results indicate that the ICESat-2 land product will underestimate the true top canopy height, there is a strong correlation between ICESat-2 heights and relative canopy height metrics (e.g., RH95). In fact, for all of the simulations shown here, the ICESat-2 canopy top was strongly correlated with RH95. Figure 12 shows the relationship of the ICESat-2 canopy height and the RH95 height from the truth data. It is hypothesized that the correlations between simulated ICESat-2 heights and canopy height metrics are a function of both canopy cover and vegetation physiology (e.g., leaf size/shape), which contributes to the horizontal and vertical structure of the vegetation. As canopy cover decreases, the probability of a photon being reflected from the top of canopy decreases, although the details of the relationship between the extent of the canopy cover and canopy height as they relate to ICESat-2 warrant further investigation. Recently, studies using both small-footprint discrete return LiDAR, as well as small-footprint full waveform LiDAR data have shown that RH90 and RH95 heights are highly sensitive to biomass, and RH90 and RH95 were identified as strong predictors in regression models [17,18]. Similarly, in the literature, other canopy metrics, such as canopy density, canopy area and canopy cover [19], have also been shown to be good secondary predictors (i.e., used in conjunction with a canopy height metric), as they are related to the horizontal and vertical structure of vegetation on the landscape [20,21]. LiDAR data, both spaceborne and airborne, have been utilized in conjunction with optical imagery to quantify the spatial extent of biomass and estimate forest attributes, including forest cover [22–25]. Therefore, although ICESat-2 heights may not reflect the true top of canopy, they remain poised to be an important measurement for estimating biomass globally, particularly in sparse or degraded ecosystems.

The results shown in this paper are empirically derived from existing airborne LiDAR data and indicate the potential performance based on the vertical sampling uncertainty due to the photon-counting detection rate of ATLAS. Another simulation, however, of ICESat-2 has been implemented in a modeling environment, which provides a complementary understanding to the potential performance from ICESat-2 once it is on orbit. Montesano et al. [26] tested the FLIGHT [27] radiative transfer model to simulate ICESat-2 performance for the tundra-taiga ecosystem, where vegetation is short and sparse in coverage. The results from Montesano indicate that for ecosystems with a sparse canopy cover, retrieval of canopy heights is problematic, as there is not a sufficient number of photons reflected from the canopy. The simulation results presented in this study did not examine ecosystems with sparse canopy cover; however, we would anticipate similar results to the Montesano study. Certainly, Boreal forest and savanna woodlands can have a wide range of canopy cover; however, the data utilized in these simulations happened to occur over areas with a moderate amount of canopy cover. Just as terrain heights will be challenging to retrieve in dense vegetation cover, canopy height retrieval will be challenging to identify and quantify in sparse vegetation cover. The results discussed in this paper are the theoretical uncertainties of vertical sampling error due to using a photon counting detector with the instrument specifications (e.g., laser power, laser pulse

repetition rate) for ICESat-2. Once ICESat-2 is on orbit, other factors, such as solar background noise and algorithm performance, will contribute and may increase the uncertainty of terrain and canopy heights.
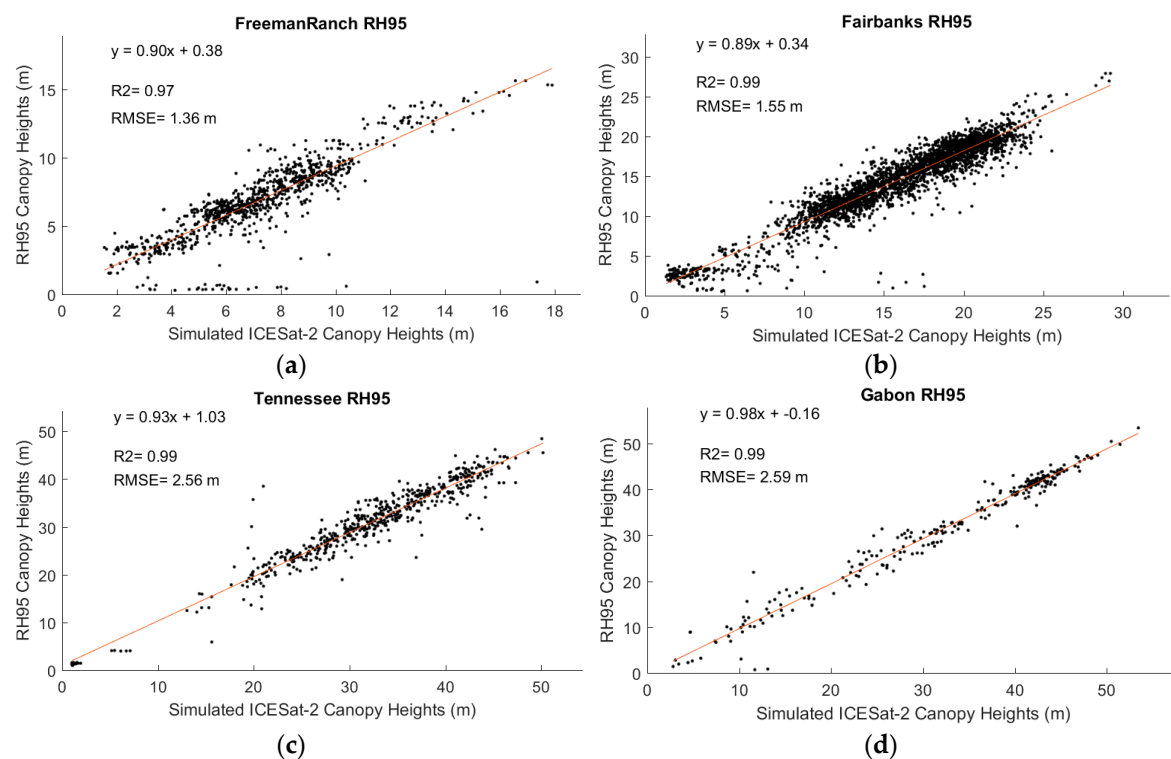

Figure 12. Correlation plot between simulated ICESat-2 Canopy heights and the RH95 canopy height. (**a**) Input data are airborne LiDAR flown over Freeman Ranch, Texas, 2010; (**b**) input data are G-LiHT collected near Fairbanks, Alaska, 2014; (**c**) input data are airborne LiDAR collected over deciduous hardwoods in eastern Tennessee in 2011; (**d**) input data are airborne LiDAR flown over Mondah Forest, Gabon, 2011.

5. Conclusions

This paper is intended to provide the science user community of ICESat-2 land/vegetation data products with a realistic understanding of the performance characteristics and potential uncertainties for the land and canopy height products over four different ecosystems; wooded savanna, boreal forest, deciduous forest and tropical forest, based solely on the sampling strategy from the ATLAS photon-counting detector. For this analysis, data from small-footprint waveform LiDAR are used as a surrogate by empirically deriving simulated ICESat-2 data using the known temporal profile of the laser energy within the waveform as a proxy distribution of received photons. Simulated ICESat-2 data top of canopy heights are evaluated against the true canopy heights derived from source small-footprint LiDAR data to provide a baseline uncertainty for each specific ecosystem. Simulation results for wooded savanna and boreal forest indicate that there should be a sufficient number of photons reflected from the ground to accurately retrieve terrain heights. The computed error and error uncertainty for terrain height retrievals for these two ecosystems are 0.06 m (0.24 m RMSE) and −0.13 m (0.77 m RMSE), respectively. In contrast, results over ecosystems with dense vegetation (i.e., deciduous forest and tropical forest) show terrain errors of 1.93 m (1.66 m RMSE) and 2.52 m (3.18 m RMSE), indicating problems extracting terrain height from the ICESat-2 data without a mitigation strategy to address the decrease in true ground returns. Simulated top of canopy heights from ICESat-2 underestimated true top of canopy returns with errors ranging from 0.28 m (1.39 m RMSE) to 1.25 m (2.63 m RMSE). The simulated ICESat-2 heights had the strongest correlation with the

RH95 height metric for all ecosystems tested. Again, the simulated errors shown here only describe the uncertainty associated with the photon detection sampling strategy from ATLAS. Once ICESat-2 is on orbit, solar background noise will potentially add additional uncertainty in the recovered surface photons. As such, future studies related to further understanding of ICESat-2 performance over land/vegetation surfaces will include the investigation of solar background noise impact on canopy and terrain height retrievals and an evaluation of potential differences associated with the green (532 nm) ICESat-2 wavelength versus a near-infrared (1064 nm) altimeter.

Author Contributions: A.L.N. is the developer of the ICESat-2 simulation algorithms and the associated software for the ICESat-2 geophysical land product. She performed the data processing and designed the research approach. A.L.N. and L.A.M. analyzed the results and determined relevant metrics to investigate with respect to the potential performance of ICESat-2. Both authors wrote the article and responded to reviewers' comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carabajal, C.; Harding, D. ICESat validation of SRTM C-band digital elevation models. *Geophys. Res. Lett.* **2005**, *32*, L22S01. [CrossRef]
2. Magruder, L.A.; Leigh, H.; Neuenschwander, A.L. Evaluation of terrain and canopy height products in central African tropical forests. *Int. J. Remote Sens.* **2016**, in press. [CrossRef]
3. Clark, D. Tropical forest and global warming: Slowing it down or speeding it up? *Front. Ecol. Environ.* **2004**, *2*, 73–80. [CrossRef]
4. Houghton, R.A.; Goetz, S.J. New satellites help quantify carbon sources and sinks. *Eos Trans. AGU* **2008**, *89*. [CrossRef]
5. Dubayah, R.; Sheldon, S.; Clark, D.; Hofton, M.; Blair, J.; Hurtt, G.; Chazdon, R. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *JGR Biogeosci.* **2010**, *115*. [CrossRef]
6. Hall, F.; Bergen, K.; Blair, J.; Dubayah, R.; Houghton, R.; Hurtt, G.; Kellndorfer, J.; Lefsky, M.; Ranson, K.; Saatchi, S.; et al. Characterizing 3D vegetation structure from space: Mission requirements. *Remote Sens. Environ.* **2011**, *115*, 2753–2775. [CrossRef]
7. Hunter, M.O.; Keller, M.; Victoria, D.; Morton, D. Tree height and tropical forest biomass estimation. *Biogeosciences* **2013**, *10*, 8385–8399. [CrossRef]
8. Montesano, P.; Sun, G.; Dubayah, R.; Ranson, K. The uncertainty of plot-scale forest height estimates from complementary spaceborne observations in the Taoga-Tundra ecotone. *Remote Sens.* **2014**, *6*, 10070–10088. [CrossRef]
9. Goetz, S.; Dubayah, R. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Manag.* **2011**, *2*, 231–244. [CrossRef]
10. Swatantran, A.; Tang, H.; Barrett, T.; DeCola, P.; Dubayah, R. Rapid, high-resolution forest structure and terrain mapping over large areas using single photon lidar. *Sci. Rep.* **2016**, *6*, 28277. [CrossRef] [PubMed]
11. Abdalati, W.; Zwally, H.J.; Bindschadler, R.; Csatho, B.; Farrell, S.; Fricker, H.; Harding, D.; Kwok, R.; Markus, T.; Marshak, A.; et al. The ICESat-2 laser altimetry mission. *Proc. IEEE* **2010**, *98*, 735–751. [CrossRef]
12. Markus, T.; Neumann, T.; Martino, A.; Abdalati, W.; Brunt, K.; Csatho, B.; Farrell, S.; Fricker, H.; Gardner, A.; Harding, D.; et al. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept and implementation. *Remote Sens. Environ.* **2016**. under review.
13. Harding, D. Pulsed laser altimeter ranging techniques and implications for terrain mapping. In *Topographic Laser Ranging and Scanning: Principles and Processing*; Shan, J., Toth, C., Eds.; CRC Press: New York, NY, USA, 2009; pp. 173–195.
14. Martino, T.; (NASA Goddard Space Flight Center, Greenbelt, MD, USA). Personal communication, 2015.
15. Walters, G.; Ngagnia Ndjabounda, E.; Ikabanga, D.; Biteau, J.P.; Hymas, O.; White, L.J.T.; Ndong Obiang, A.M.; Ndong Ondo, P.; Jeffery, K.J.; Lachenaud, O.; et al. Peri-urban conservation in the Mondah forest of Libreville, Gabon: Red list assessments of endemic plant species, and avoiding protected area downsizing. *Oryx* **2015**. [CrossRef]

16. Magruder, L.; Brunt, K. Performance analysis of airborne photon-counting lidar data in preparation of ICESat-2. *Geosci. Remote Sens. Lett.* **2016**. under review.

17. Cao, L.; Coops, N.; Hermosilla, T.; Innes, J.; Dai, J.; She, G. Using small-footprint discrete and full-waveform airborne lidar metrics to estimate total biomass and biomass components in subtropical forests. *Remote Sens.* **2014**, *6*, 7110–7135. [CrossRef]

18. Margolis, H.; Nelson, R.; Montesano, P.; Beadoin, A.; Sun, G.; Anderson, H.; Wulder, M. Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. *Can. J. For. Res.* **2015**, *45*, 838–855. [CrossRef]

19. Krofcheck, D.; Litvak, M.; Lippitt, C.; Neuenschwander, A. Woody biomass estimation in a southwestern US juniper savanna using lidar derived clumped tree segmentation and existing allometries. *Remote Sens.* **2016**, *8*, 453. [CrossRef]

20. Naesset, E.; Gobakken, T. Estimation of above and below ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens. Environ.* **2008**, *112*, 3079–3090. [CrossRef]

21. Hall, S.; Burke, I.; Box, D.; Kaufmann, M.; Stoker, J. Estimating stand structure using discrete return lidar: An example from low density, fire prone Ponderosa pine forests. *For. Ecol. Manag.* **2005**, *208*, 189–209. [CrossRef]

22. Hudak, A.; Lefsky, M.; Cohen, W.; Berterretche, M. Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sens. Environ.* **2002**, *82*, 397–416. [CrossRef]

23. Chen, X.; Vierling, L.; Rowell, E.; DeFelice, T. Using lidar and effective LAI data to evaluate IKONOS and Landsat 7 ETM+ vegetation cover estimates in a ponderosa pine forest. *Remote Sens. Environ.* **2004**, *91*, 14–26. [CrossRef]

24. Wulder, M.; White, J.; Alvaraez, F.; Han, T.; Rogan, J.; Hawkes, B. Characterizing boreal forest wildfire with multi-temporal Landsat and lidar data. *Remote Sens. Environ.* **2009**, *113*, 1540–1555. [CrossRef]

25. Glenn, N.; Neuenschwander, A.; Vierling, L.; Spaete, L.; Li, A.; Shinneman, D.; Pilliod, D.; Arkle, R.; McIlroy, S. Landsat8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sens. Environ.* **2015**. [CrossRef]

26. Montesano, P.; Rosette, J.; Sun, G.; North, P.; Nelson, R.; Dubaya, R.; Ranson, K.; Kharuk, V. The uncertainty of biomass estimates from modeled ICESat-2 returns across a boreal forest gradient. *Remote Sens. Environ.* **2015**, *158*, 95–109. [CrossRef]

27. North, P. Three-dimensional forest light interaction model using a Monte Carlo method. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 946–956. [CrossRef]