

## Article

# Detecting Change in Forest Structure with Simulated GEDI Lidar Waveforms: A Case Study of the Hemlock Woolly Adelgid (HWA; *Adelges tsugae*) Infestation

Peter Brehm Boucher <sup>1,2,\*</sup>, Steven Hancock <sup>3</sup>, David A Orwig <sup>4</sup>, Laura Duncanson <sup>5</sup>, John Armston <sup>5</sup>, Hao Tang <sup>5</sup>, Keith Krause <sup>6</sup>, Bruce Cook <sup>7</sup>, Ian Paynter <sup>8</sup>, Zhan Li <sup>9</sup>, Arthur Elmes <sup>1</sup> and Crystal Schaaf <sup>1</sup>

<sup>1</sup> School for the Environment, University of Massachusetts, Boston, MA 02125, USA

<sup>2</sup> Department of Organismic and Evolutionary Biology (OEB), Harvard University, Cambridge, MA 02138, USA

<sup>3</sup> The University of Edinburgh, Edinburgh EH8 9XP, UK

<sup>4</sup> Harvard Forest, Harvard University, Petersham, MA 01366, USA

<sup>5</sup> Department of Geographical Sciences, University of Maryland, College Park, MD 20740, USA

<sup>6</sup> Battelle, National Ecological Observatory Network (NEON), Boulder, CO 80301, USA

<sup>7</sup> NASA Goddard Space Flight Center, Greenbelt, MA 20771, USA

<sup>8</sup> Universities Space Research Association (USRA), Columbia, MD 21046, USA

<sup>9</sup> German Research Centre for Geosciences (GFZ), 14473 Potsdam, Germany

\* Correspondence: Peter.Boucher001@umb.edu; Tel.: +1-(301)-204-5628

Received: 23 March 2020; Accepted: 16 April 2020; Published: 20 April 2020

**Abstract:** The hemlock woolly adelgid (HWA; *Adelges tsugae*) is an invasive insect infestation that is spreading into the forests of the northeastern United States, driven by the warmer winter temperatures associated with climate change. The initial stages of this disturbance are difficult to detect with passive optical remote sensing, since the insect often causes its host species, eastern hemlock trees (*Tsuga canadensis*), to defoliate in the midstory and understory before showing impacts in the overstory. New active remote sensing technologies—such as the recently launched NASA Global Ecosystem Dynamics Investigation (GEDI) spaceborne lidar—can address this limitation by penetrating canopy gaps and recording lower canopy structural changes. This study explores new opportunities for monitoring the HWA infestation with airborne lidar scanning (ALS) and GEDI spaceborne lidar data. GEDI waveforms were simulated using airborne lidar datasets from an HWA-infested forest plot at the Harvard Forest ForestGEO site in central Massachusetts. Two airborne lidar instruments, the NASA G-LiHT and the NEON AOP, overflew the site in 2012 and 2016. GEDI waveforms were simulated from each airborne lidar dataset, and the change in waveform metrics from 2012 to 2016 was compared to field-derived hemlock mortality at the ForestGEO site. Hemlock plots were shown to be undergoing dynamic changes as a result of the HWA infestation, losing substantial plant area in the middle canopy, while still growing in the upper canopy. Changes in midstory plant area (PAI 11–12 m above ground) and overall canopy permeability (indicated by RH10) accounted for 60% of the variation in hemlock mortality in a logistic regression model. The robustness of these structure-condition relationships held even when simulated waveforms were treated as real GEDI data with added noise and sparse spatial coverage. These results show promise for future disturbance monitoring studies with ALS and GEDI lidar data.

**Keywords:** GEDI; lidar; disturbance; hemlock woolly adelgid; invasive; G-LiHT; NEON

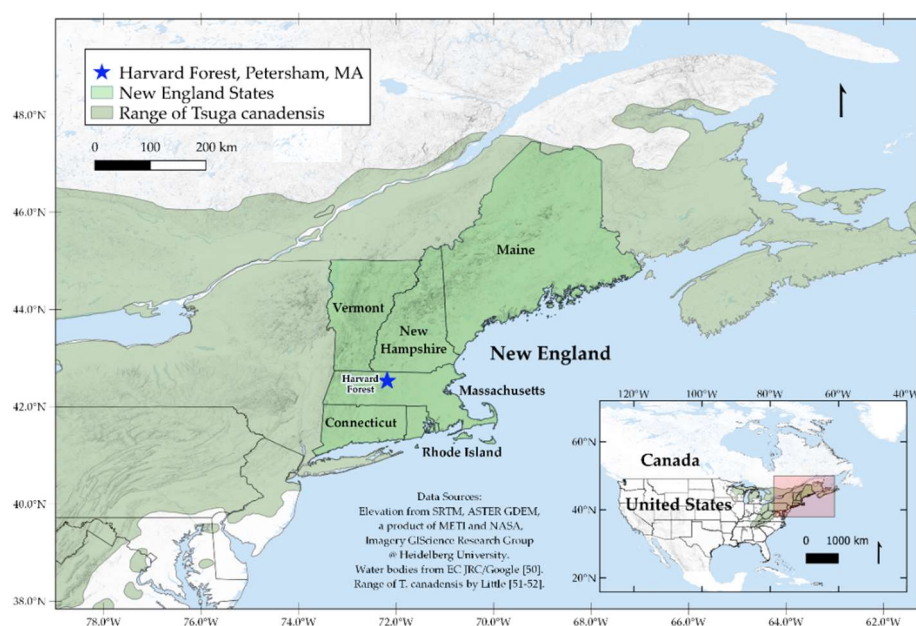
## 1. Introduction

Insect infestations cause unique changes in the structure of forests over time, opening up gaps in the canopy and triggering growth among understory species [1]. Active remote sensing instruments, such as spaceborne and airborne lidar scanning (ALS) instruments, can penetrate dense forest canopies to observe these subtle changes in vertical structure. Thus, lidar has potential to detect the spread and to monitor the severity of insect infestation impacts at the landscape scale.

An invasive insect, the hemlock woolly adelgid (HWA; *Adelges tsugae*), is currently bringing about distinct and lasting impacts on the forest structure of the eastern United States [2,3]. HWA is spreading northward as climate change warms New England winters and reduces the frequency of cold snaps, which are the major limiting factor to adelgid populations in the US [4–11]. As multiple time series of lidar data over HWA infested sites have become publicly available, new opportunities have emerged to track the spread of this insect infestation and to characterize its impacts on forest structure. A new spaceborne lidar mission, the Global Ecosystem Dynamics Investigation (GEDI) [12], is now collecting measurements of vertical canopy structure over the New England region, providing the regional scale coverage that could capture HWA's northern expansion. This study explores the use of waveform lidar for observing structural change and the future potential of GEDI data for monitoring the spread of the HWA into New England.

### 1.1. Structural Signals of the Hemlock Woolly Adelgid Infestation

HWA's impacts are widespread upon eastern and Carolina hemlock trees (*Tsuga canadensis* and *Tsuga caroliniana*) in the eastern United States, extending across the Appalachian Mountains from Georgia to Maine. While HWA does not directly defoliate hemlock trees, colonies of HWA bring about progressive needle loss and branch death by draining hemlocks of their stored sugars [13]. Infested trees die within 5–15 years from the onset of infestation, and are replaced primarily by deciduous tree species and white pine trees (*Pinus strobus*) [2]. Studies project that HWA will spread into the full range of hemlock in New England and southern Canada [14,15] (Figure 1), potentially eliminating this ecological foundation species from the landscape in the coming decades.



**Figure 1.** Map of the study region showing the range of eastern hemlock (*Tsuga canadensis*) in the northeastern United States and southern Canada. Data on water bodies is acquired from Pekel et al. [16], and the range of *T. canadensis* is from Little [17,18].

## 1.2. Lidar Remote Sensing for Monitoring Forest Health

Repeat measurements from lidar instruments have been shown to be effective in detecting disturbances and recording subtle structural changes in forests [19–23]. Lidar sensors are active remote sensing instruments that emit pulses of light to characterize the structure of targets in three-dimensional space [24]. Emitted light reflects off target objects and returns to the instrument, which records the distribution of returning energy over time, a waveform, to measure the range to targets. Most lidar instruments decompose these waveforms into a set of three-dimensional points that correspond to peaks in the distribution of returning energy, known as discrete lidar data or a point cloud. Small-footprint lidar instruments sample a small area with each beam, from the millimeter to decimeter scale. In contrast, a large-footprint lidar instrument might measure a stand of trees with a single beam, capturing vertical structure in a circular footprint of tens of meters in diameter. While waveform and discrete lidar instruments record structure with the same physical system, the specific parameters of the sensor, the sampling scheme, the post-processing techniques, and the data type can make a time series of lidar data from different sensors difficult to compare.

The recent launch of GEDI, a spaceborne lidar instrument designed specifically to measure forest structure over large areas, presents new opportunities for monitoring forest disturbances with vertical profiles of forest structure [12]. GEDI is particularly well suited to monitor HWA infestation impacts because of its ability to penetrate gaps in the upper canopy and record lower canopy structural change. Large-footprint waveform lidar data similar to GEDI (~25 m in diameter) have proven useful for observing distinct canopy structures related to disturbances. For example, large-footprint data have been used to infer past hurricane damage [25], to identify developmental stages of forest plots [26–28], to observe phenological cycles [23], and to estimate biomass change [29,30].

While GEDI's spatial resolution will be useful for pursuing ecological research, its spatial coverage is limited by orbital constraints and the physics of an active optical system. GEDI relies on large-footprint samples to maximize the coverage that it can achieve at the given pulse rate. Operating from the International Space Station (ISS), GEDI samples forests with three lasers that are split into four beams, which are optically dithered to create 8 parallel tracks, spaced ~600 m apart on the ground [12]. Each of GEDI's tracks samples the Earth's surface with a waveform every 60 m along its path. By gridding these discrete samples at the 1 km scale, the mission will produce global data products of vegetation structure for regions between ~51.6 degrees N and S latitude [31].

Due to the orbital constraints of the ISS, GEDI will only be able to sense forests in the temperate and tropical regions. The areas will be sampled multiple times during the 2-year mission, but it is rare that GEDI footprints will perfectly overlap, making it infeasible to calculate structural change at an individual footprint scale. While other lidar or field datasets could be compared with GEDI data to calculate change, the geolocation accuracy of GEDI waveform data presents an additional obstacle to comparison. After processing, the average geolocation error is expected to be ~8–11 m [12], which may be too high for accurate comparisons between GEDI footprints and field plots or other coincident datasets.

Fortunately, the obstacles to change detection with GEDI can be overcome with the GEDI Simulator [32], an open-source software developed to calibrate GEDI's data products before launch. The GEDI simulator uses discrete and waveform ALS lidar data to produce large-footprint lidar waveforms. By processing ALS data, the simulator corrects for the specific characteristics of the ALS datasets and the instruments that collected them [32]. In addition, the simulator software provides tools for geo-locating GEDI waveforms to within ~2–3 m on average using a waveform shape-matching technique [33]. As long as a site has spatially overlapping ALS and GEDI data available, studies can use the simulator to (1) geolocate GEDI waveforms, (2) simulate comparable waveforms from ALS data at the same footprint locations, and (3) output measurements of vegetation structure that are comparable between a variety of ALS platforms and GEDI. Thus, the simulator enables the

calculation of structural change for a set of ALS and/or GEDI data at the scale of a ~25 m footprint. If this change detection method proves viable for real GEDI data, the ecological applications of GEDI could vastly increase, enabling future studies of forest disturbance and structural change for any site with ALS data within GEDI's coverage.

### 1.3. Study Design

To investigate the utility of GEDI data for detecting forest disturbances, GEDI waveforms were simulated from ALS data collected in 2012 and 2016 over a mixed temperate forest infested by HWA, the 35-hectare Smithsonian Institute Forest Global Earth Observatory (ForestGEO) plot at the Harvard Forest experimental site in Petersham, MA, USA [34]. By comparing changes in simulated waveforms over an established field site with a well-studied infestation, this study outlines a viable method for change detection in temperate forests with GEDI waveforms. Only a limited set of GEDI data has been collected thus far, and therefore, this study turns to simulated data to evaluate a method for change analysis that can be applied to both multi-temporal ALS data and future real GEDI data as it is accumulated over the next 2 years of the mission. It aims to extract ecologically meaningful observations of the impacts of HWA upon New England forest structure, and to test the viability of GEDI waveforms for detecting disturbance.

## 2. Materials and Methods

### 2.1. Overview

GEDI waveforms were simulated with the GEDI Simulator [32] using two ALS datasets, one collected by the NASA Goddard Lidar Hyperspectral and Thermal instrument (G-LiHT) in 2012 [35], and the other by the NSF funded National Ecological Observatory Network (NEON) Airborne Observation Platform in 2016 [36]. To compare waveforms between years, waveform metrics were derived with the GEDI Simulator and LibCLidar software packages [32,37]. The change in waveform metrics from 2012 to 2016 was calculated for each footprint ( $2016 - 2012 = \text{Change}$ ). The importance of waveform metrics for observing the impacts of the HWA disturbance was assessed by relating metrics to the extensive Harvard Forest ForestGEO plot field data [34]. Metrics were assessed with two statistical methods: 1) one-way analysis of variance (ANOVA) grouped by the dominant tree species of each footprint; and 2) logistic regression of hemlock tree mortality as a function of waveform change metrics.

The analysis has two coincident aims. The first aim is to assess the utility of a large-footprint lidar simulator [32] for comparing ALS acquisitions from different sensors over time. We recognize that the sensor and flight parameters of the ALS datasets can impact derived metrics, and thus can introduce biases into change analysis studies. However, the GEDI simulator has been demonstrated to minimize these biases [32], making ALS datasets from different sources comparable at the scale of a large-footprint lidar waveform. In pursuit of this aim, waveforms are first simulated without noise and at high spatial coverage to identify waveform metrics that correlate with the structural impacts of the HWA disturbance.

The second aim is to test whether the method outlined in this study could be viable for future disturbance studies with real GEDI data. Waveforms are simulated with varying degrees of noise based on pre-launch signal-to-noise estimates. Then, the sparse spatial coverage of real GEDI acquisitions are applied to noisy waveforms. Lastly, statistical relationships between waveform metrics and forest condition are re-assessed to evaluate how the noise parameters and spatial coverage of real GEDI data might affect the observed relationship between waveform metrics and forest condition.

### 2.2. Simulated GEDI Data

Each lidar instrument and acquisition has a specific footprint and pulse shape that affect the derived metrics. To prevent any differences in these properties between GEDI and ALS data, the GEDI simulator aggregates small-footprint lidar data and convolves a large-footprint waveform of

return energy based on the vertical distribution of points within a specified footprint. Simulated GEDI data have been shown to be highly comparable with real large-footprint lidar data, having been validated with the NASA Land Vegetation and Ice Sensor (LVIS) [38] and other large-footprint lidar instruments [32]. The simulated waveforms used in this study are structured as a sum of smooth Gaussian components, and waveform features, such as ground elevation, are extracted following Hofton et al. [39].

In order to mimic the realistic processing of GEDI data, all metrics used in this study are derived directly from the Gaussian model of the GEDI waveform. Because GEDI's ground elevation is derived directly from this smoothed waveform, there can be discrepancies between GEDI's ground elevation and ALS ground elevation that may impact derived metrics. For instance, relative height (RH) metrics are calculated in relation to the lowest elevation of the waveform, defined as the bottom of the ground signal. Thus, ALS and GEDI data need to agree on a common ground height in order to produce RH metrics that originate from the same elevation.

### 2.3. Airborne Lidar Scanner (ALS) Data

The two ALS datasets were collected with different sensors (Table 1): the Riegl VQ-480 on the NASA G-LiHT and the Optech ALTM Gemini on the NSF-funded NEON AOP. Previous research has shown that simulated waveforms from Optech and Riegl sensors with similar flight parameters to the datasets in this study were both able to accurately match large-footprint LVIS waveforms, with LVIS data being similar to that of real GEDI data [32]. Differences in sensor wavelength (1064 nm and 1550 nm) that can cause ALS instruments to vary in their sensitivity to vegetative targets also did not significantly impact simulation results [32]. These findings lay the foundation for comparing lidar datasets with different sensor and flight acquisition parameters at the scale of a GEDI waveform.

To further check for potential sensor biases in simulated waveform data, including those caused by different ALS scan angle and point density distributions, a supplementary analysis was performed using waveforms with the most similar sets of input data between years (Appendix A). Other sensor differences, such as sensor wavelength, were not investigated further in this study, as they were shown to be negligible in a previous study of simulated data [32].

**Table 1.** Instrument and flight specifications of ALS data collected over the ForestGEO plot

|                                        | G-LiHT       | NEON               |
|----------------------------------------|--------------|--------------------|
| Instrument                             | Riegl VQ-480 | Optech ALTM Gemini |
| Beam divergence (mrad)                 | 0.3          | 0.8                |
| Altitude (m)                           | ~300         | ~1000              |
| Wavelength (nm)                        | 1550         | 1064               |
| Pulse repetition frequency (PRF; kHz)  | 300          | 100                |
| Max number of returns per pulse        | 8            | 4                  |
| Max scan angle used (degrees)          | 36           | 18                 |
| Point density (points/m <sup>2</sup> ) | 29.6         | 6.8                |
| Acquisition date                       | June 2012    | August 2016        |

### 2.4. Metrics

Vertical profiles of changes in effective plant area index (PAI) and relative height (RH) metrics were derived to compare waveforms between years, following the standard methods of GEDI data processing [12,40]. The effective PAI describes the horizontal projected area of plant elements (foliage, branches, and trunks) per unit of ground area (m<sup>2</sup>m<sup>-2</sup>) within a volume of canopy [41]. The PAI distribution is derived from the vertical distribution of gaps within the canopy (directional gap probability; P<sub>gap</sub>) with a widely used method [40,42–45]. The PAI at any given height is calculated from waveform data with the equations:

$$Pgap(\theta, z) = 1 - \frac{R_v(z)}{R_v(0) + \frac{\rho_v}{\rho_g} R_g} \quad (1) \quad (2) \quad (1)$$

$$PAI(z) = \frac{-1[\ln(Pgap(\theta, z + \Delta z)) - \ln(Pgap(\theta, z))]}{G(\theta) \Omega(\theta)} \quad (2)$$

$R_v(z)$  is the integral of reflected energy from canopy elements from the canopy top down to the given height ( $z$ ).  $R_g$  is the integral of reflected energy from the ground.  $\Delta z$  is the height increment at which PAI is calculated. Constants are set as: a nadir view angle ( $\theta = 0$ ), a spherical leaf angle distribution ( $G = 0.6$ ), a random spatial distribution of canopy elements ( $\Omega = 1$ ), and a canopy-to-ground reflectance ratio ( $\rho_v/\rho_g$ ) of 1.425.

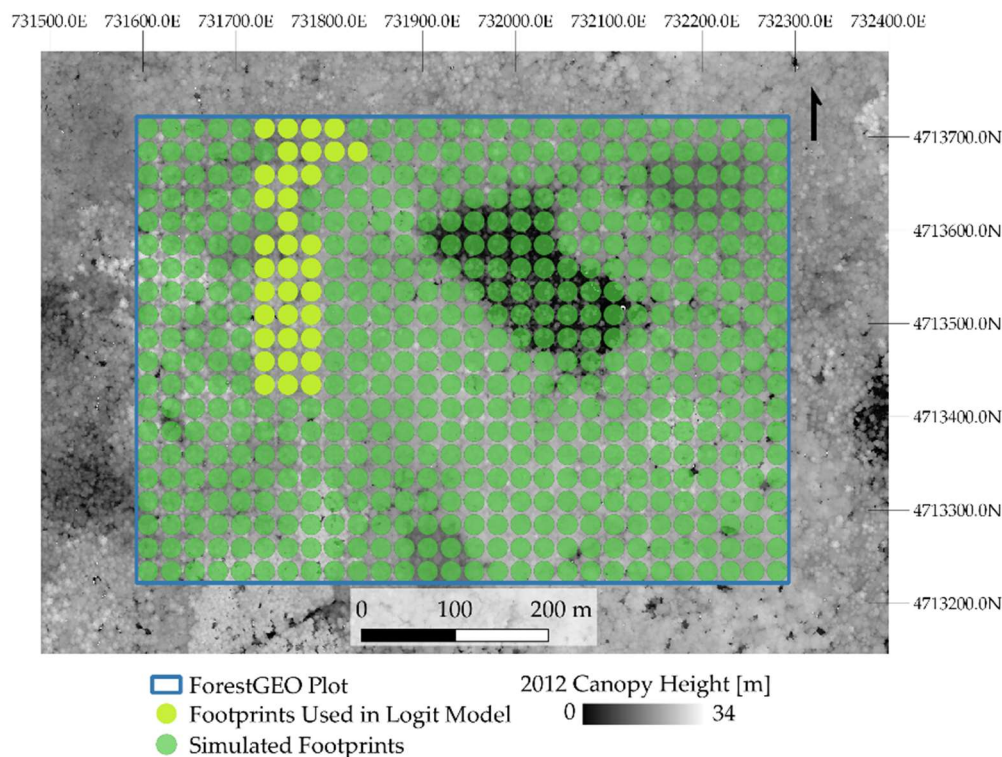
RH metrics describe the percentiles of the vertical distribution of energy in a returned waveform. Heights are derived from the normalized, cumulative distribution of the waveform, summed from the ground to the canopy top [27,46]. For example, RH10 is the height relative to the fitted ground elevation below which 10% of the cumulative waveform energy has been returned. *RH0* represents the bottom of the ground signal, and *RH100* represents the top of the canopy signal. Note that RH metrics can be negative, particularly in sparse conditions where a significant portion of the waveform energy is reflected below the fitted ground elevation.

Changes in RH metrics over time can be indicative of changes in canopy structure, including growth, gap formation, and disturbance [29]. The HWA infestation is hypothesized to shift RH values lower over time in hemlock forests, as foliage loss from HWA increases canopy permeability and allows more laser energy to reach the ground.

## 2.5. Field Data

The Center for Tropical Forest Science (CTFS) Forest Global Earth Observatory (ForestGEO) plot at Harvard Forest is part of an international network of forest sites [47]. It is a georeferenced 35-hectare plot divided into a grid of 20 m square quadrats. The first census of the ForestGEO plot began in 2010 [34], and all quadrats containing hemlock trees in the plot were found to be infested by HWA in 2012. A subset of the plot was reassessed for hemlock tree mortality in 2016 [48]. This 2016 assessment of hemlock mortality only revisited select quadrats (72 quadrats in total) in the mature hemlock area (western portion of the plot) to assess tree condition. Tree condition was recorded as a binary value (alive or dead) based on the presence or complete absence of foliage.

Hemlock mortality from 2010 to 2016 was calculated for a set of simulated GEDI footprints that overlapped with the western area of the ForestGEO plot that had been reassessed in 2016 (highlighted in Figure 2). Mortality was defined as the number of live hemlock trees that died in the period from 2010 to 2016 divided by the total number of live hemlock trees within the footprint in 2010. By 2016, most hemlock trees within the ForestGEO plot were infested with HWA, and thus, even plots with 0% mortality may well be displaying the impacts of HWA.



**Figure 2.** Simulated GEDI footprint locations within the ForestGEO plot, rendered in WGS-84 UTM 18N. They are overlaid on a 1 m canopy height map derived from the G-LiHT 2012 ALS data [35]. The 35 hemlock dominated footprints that were assessed for mortality in 2016 and used in the lasso logistic regression (logit model) are highlighted.

The times of observation of hemlock tree mortality (summer 2010 and summer 2016) are offset from the time of lidar observations (June 2012 and August 2016). This offset was unavoidable, since these data acquisitions were planned with different research goals in mind.

To compare waveform metrics by tree species, the dominant species of each GEDI footprint was derived across the ForestGEO plot with field data. Dominant species were calculated by summing the total basal area (BA) of each species within a lidar footprint and assigning the dominant tree species as that with the majority BA. The top five dominant species used in this analysis were: eastern hemlock (*Tsuga canadensis*; 38% of the 560 total footprints), red oak (*Quercus rubra*; 32%), red maple (*Acer rubrum*; 16%), white pine (*Pinus strobus*; 8%), and red pine (*Pinus resinosa*; 5%). These species comprised the majority of footprints (~99%) in the ForestGEO plot.

## 2.6. Comparing ALS with the GEDI Simulator

First, ALS acquisitions from 2012 and 2016 were compared with the GEDI simulator to identify waveform metrics that correlated with the structural impacts of the HWA disturbance. Waveforms were simulated at the same footprint locations using both the 2012 and 2016 ALS data, and changes in the metrics of simulated waveforms were calculated at each location during the time period. A total of 560 GEDI waveforms were simulated within a rectangular grid using spatially-coincident discrete ALS data from G-LiHT in 2012 and NEON AOP in 2016. Simulated footprints were spaced along a 25 m rectangular grid (Figure 2) in order to maximize coverage within the 35 hectare ForestGEO plot (500 × 700 m).

Waveform footprints were 22 m in diameter and did not overlap in space. This 22 m diameter footprint was defined as the area containing 96% of the Gaussian footprint of the return energy ( $2 \times$  standard deviation). RH metrics at 10% energy intervals and PAI metrics at 1 m height intervals were calculated for all footprints.

### 2.7. Selecting and Evaluating Waveform Disturbance Metrics

A lasso regression of hemlock tree mortality was employed to identify the important variables for predicting HWA disturbance impacts. Lasso regressions were fit with the subset of 35 hemlock dominant footprints (Figure 2) that were assessed for hemlock mortality in 2016 and that also had comparable ground heights (less than 1 m difference between G-LiHT and NEON waveforms). Lasso regression employs a penalty parameter, lambda, to reduce the magnitude of insignificant coefficients in a model to 0 [49,50]. RH and PAI change metrics at different canopy heights (42 independent variables in total) were then selected with a 10-fold cross-validation method [49]. Cross-validation results identified an optimal model by using the largest lambda value within 1 standard deviation of the minimum statistical deviance of all models.

Field-measured hemlock mortality was regressed on the waveform change metrics with a logistic regression. The dependent variable, hemlock mortality, was modelled with a binomial distribution. The number of live hemlocks that died from 2010 to 2016 represented the number of observations, and the number of live hemlocks in each footprint in 2010 represented the number of trials. Independent variables were z-score normalized before input into the model in order to better compare coefficient magnitudes.

In addition, the disturbance metrics identified by the lasso regression were assessed for their correspondence with infested hemlock footprints. Each metric was run through an ANOVA grouped by dominant species. Waveform change metrics within hemlock-dominated footprints were expected to be significantly different from those of other species, since HWA only affects hemlock trees and was ubiquitous in the field site at the time of study.

### 2.8. Simulating the Impact of GEDI's Noise

The effect of noise on the relationship between waveform metrics and forest condition was investigated, because without noise, simulated waveforms represent unrealistically clean GEDI waveforms. Gaussian noise was incrementally added to the 560 simulated waveforms in the ForestGEO plot (Figure 2). Noise was incrementally added to waveforms by decreasing the beam sensitivity, a metric that describes the canopy cover at which a noisy waveform would have a 90% probability of detecting the ground and a 5% probability of causing a false positive (GEDI ground height > true ground) [32]. Noise was simulated from 90% to 99% beam sensitivity to cover the expected range of real GEDI data, estimated to be 92–99.6% beam sensitivity.

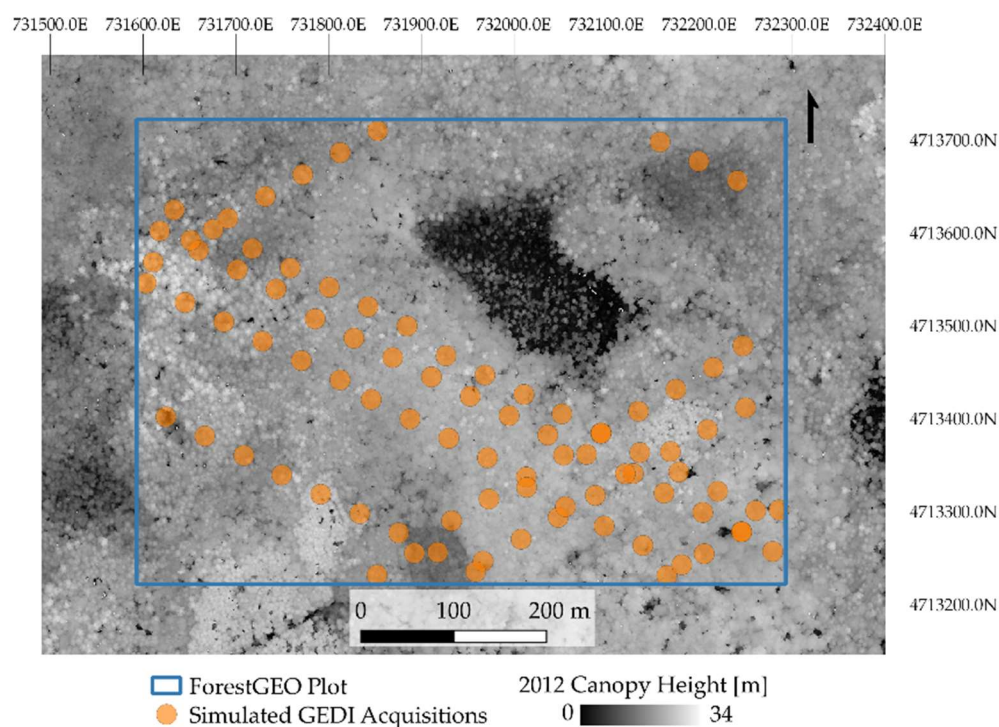
Noise was added to both the G-LiHT 2012 and NEON 2016 waveforms. Then, change metrics were calculated by subtracting noisy waveforms from their noiseless counterparts (e.g., noisy NEON 2016 – noiseless G-LiHT 2012, or noiseless NEON 2016 – noisy G-LiHT 2012). An ANOVA was used to test the effect of species on the PAI change of each dataset with degraded beam sensitivity. The F-statistic and *p*-value of the ANOVA were used to measure how the progressive addition of noise altered the relationship between PAI change and forest condition.

### 2.9. Simulating GEDI's Noise and Spatial Coverage

Lastly, waveforms were simulated over the site with both the sparse spatial coverage and average noise parameters of GEDI's beams. Simulated GEDI tracks were acquired for the Harvard Forest region [12], and a set of 80 footprints (Figure 3) that intersected the study site were selected for further analysis. Footprint locations were spaced to mimic GEDI's spatial coverage in the ForestGEO plot. Waveforms were simulated from the G-LiHT 2012 and NEON 2016 ALS data at these realistically spaced footprint locations. In addition, realistic noise was added to the waveforms using the estimated parameters of the link margin for GEDI's power beam at night (link margin = 4.956 at 95% canopy cover) and coverage beam during the day (link margin = -2.039 at 95% canopy cover).

The real noise parameters of GEDI will vary depending on the power of the laser beam and the time of acquisition. Of GEDI's three laser beams, two are used at higher power (power beams), while one is split into two beams with lower power (coverage beams) [12]. Waveform noise will be lowest when the power beams are operating during the nighttime, while noise will be the highest when the

weaker coverage lasers are operating during the day. Thus, noise was added to the simulated waveforms to realistically represent the best (night-power) and worst (day-coverage) signal-to-noise ratio scenarios of future GEDI acquisitions.



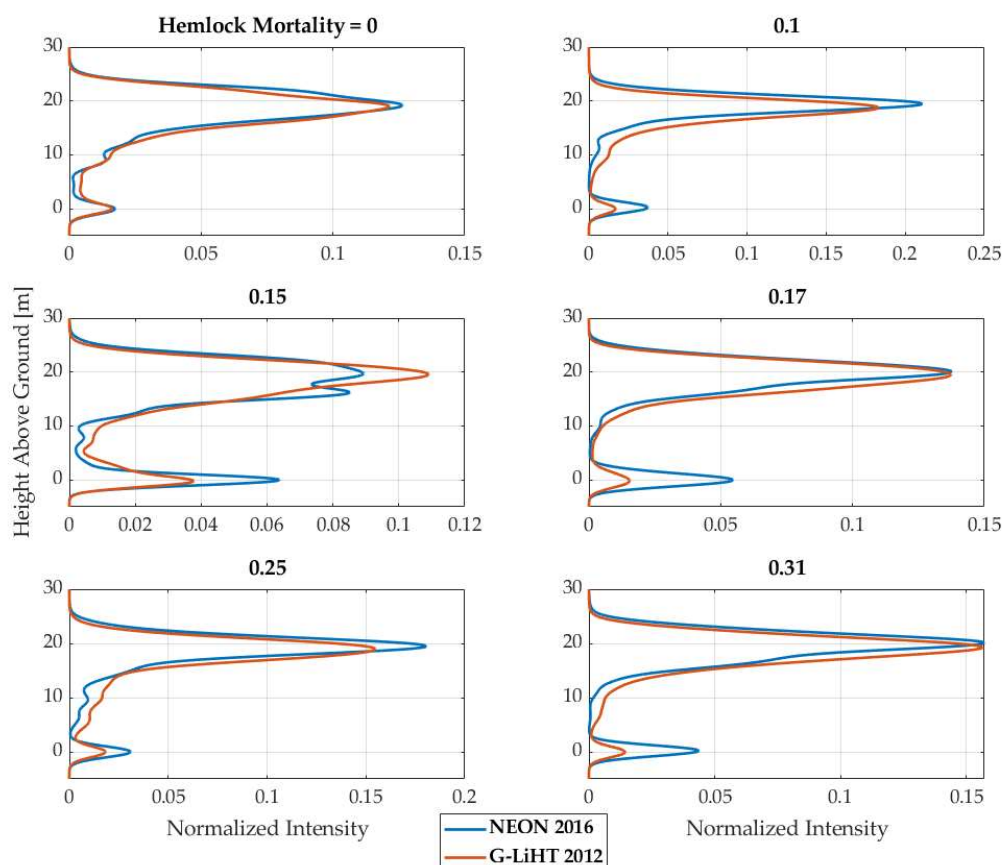
**Figure 3.** GEDI's Simulated Spatial Coverage within the ForestGEO plot. These 80 realistically spaced footprints were produced from the pre-launch simulated orbits of GEDI over New England.

### 3. Results

#### 3.1. Simulation Results

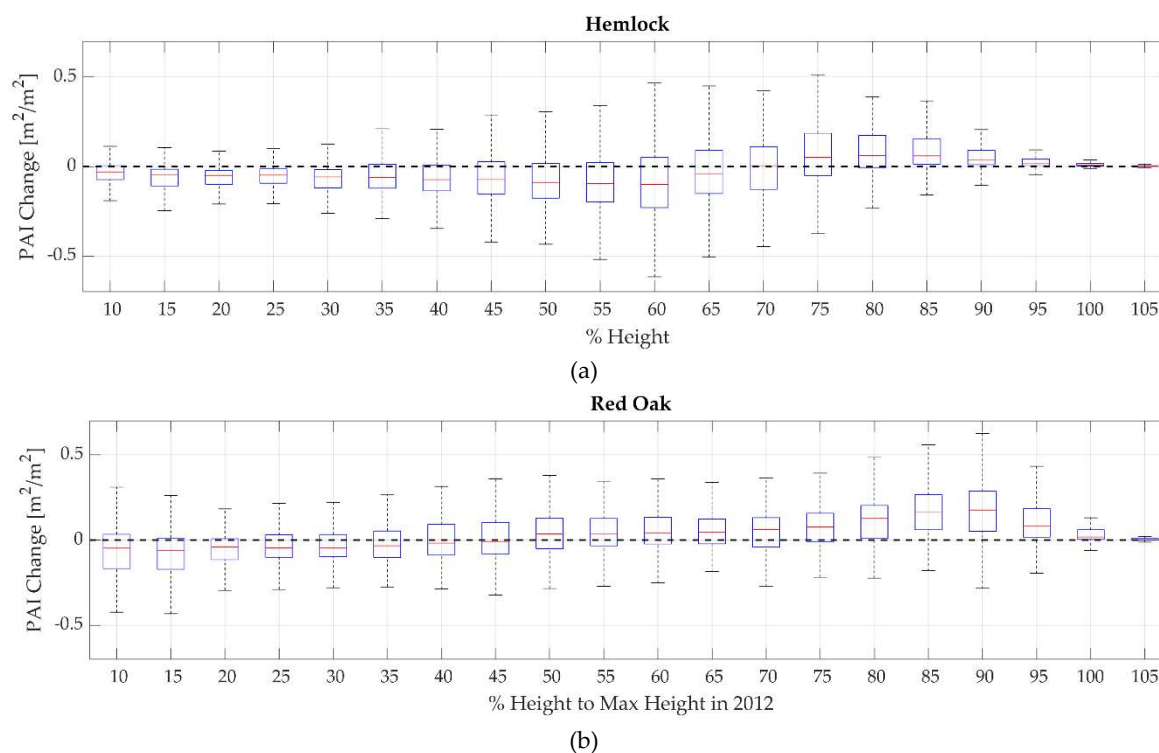
Waveforms were simulated for 2012 and 2016 in 560 footprint locations along a rectangular grid within the ForestGEO plot. Simulated waveforms that were assessed for hemlock mortality in 2016 were plotted by aboveground height against normalized intensity (Figure 4). The highest peak is characteristic of the forest canopy shape, while the lowest peak identifies the ground.

Footprints containing higher hemlock mortality showed increased amplitude of ground return peaks in 2016 (blue) as compared to the peaks of the ground returns in the 2012 waveforms (orange; Figure 4). This change in ground amplitude indicates that the canopy was more permeable in 2016 as compared to 2012, since more energy was able to penetrate the canopy and reach the ground, producing a larger ground return. The middle portion of the waveforms, from about ~5–20 m above ground, has also declined in amplitude between years, indicating a loss of canopy structure within those layers. In addition, the upper canopy tops of waveforms appear to have increased slightly from 2012 to 2016, indicating some growth in the overstory during this time period.



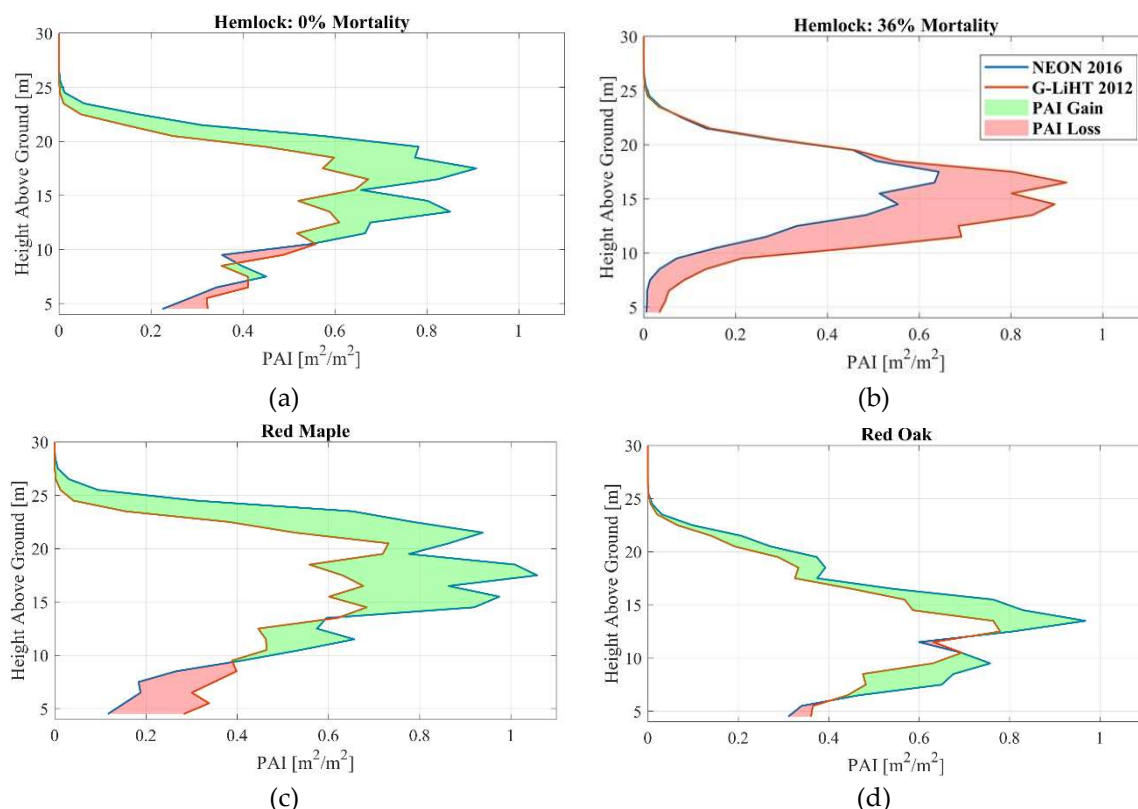
**Figure 4.** Simulated GEDI waveforms from G-LiHT 2012 (orange) and NEON 2016 (blue) over the mature hemlock section of the ForestGEO plot plotted by height above ground. Plots are titled and sorted from left-right and top-bottom by the proportional hemlock mortality within the footprint of each waveform (ex: 0.1 = 10% of live hemlocks in 2012 died by 2016).

A plot of the vertical profile of PAI change for both the eastern hemlock and the red oak footprints showed that the trend in PAI loss in the mid-canopy is ubiquitous within the infested hemlocks of the ForestGEO plot (Figure 5). PAI change at 1 m height intervals was binned to percent height relative to the maximum canopy height (RH100) recorded in 2012. While both tree species displayed a slight loss of PAI in the lower canopy (10–40% of max height) and a slight gain of PAI in the upper canopy (above 70% height), hemlock dominated footprints showed a distinct loss of PAI in the mid-canopy (40–70% height).



**Figure 5.** Profiles of hemlock (a)  $N = 212$  footprints and red oak (b)  $N = 176$  footprints PAI change plotted by % height to maximum height in 2012. Boxes show the quartiles and ranges of the distributions. Vertical profiles of PAI Change for footprints dominated by hemlock and red oak follow similar trends in the lower and upper canopy, but hemlock shows a distinct loss of PAI in the mid-canopy (40–70% height).

Plotting the vertical PAI profiles of healthy and infested hemlock footprints in comparison to those of red oak and red maple shows distinct mid-canopy foliage loss in the hemlock dominated footprints (Figure 6). An example of a healthier hemlock footprint (with 0% mortality) displays a similar increase in PAI compared to that of other species, which show gains of PAI in the upper canopy from 2012 to 2016. In contrast, an infested footprint with 36% hemlock mortality displayed distinct decreases in PAI in its middle and lower canopy.

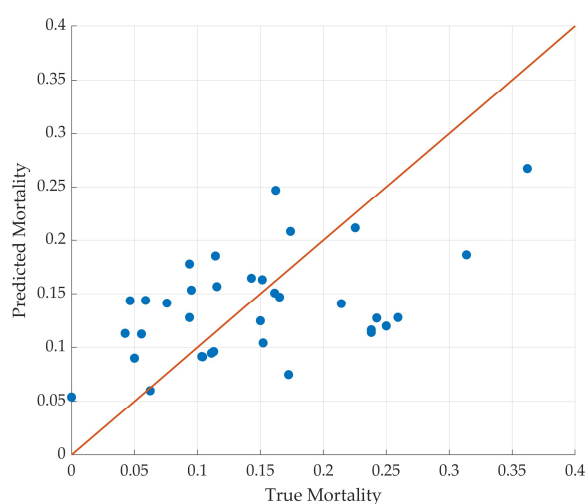


**Figure 6.** Examples of hemlock (a, b), red maple (c), and red oak (d) PAI change by height. GEDI waveforms were simulated using G-LiHT ALS data in 2012 (orange) and NEON ALS data in 2016 (blue), and vertical profiles of PAI were derived for each year in the same footprint locations. An infested hemlock stand with severe mortality (36%; b) has primarily lost PAI in the midstory and understory, and its overstory shows little change. In contrast, an infested hemlock stand with 0% mortality during the time period (a) displays a pattern of growth similar to that of stands of other tree species, including red maple (c) and red oak (d).

### 3.2. Variable Selection

A series of lasso logistic regressions were used to identify waveform metrics that could be important predictors of mortality. Out of 42 independent variables, the negative changes in PAI 11–12 m, PAI 17–18 m, and RH10 were selected as optimal variables. Decreases in RH10 are indicative of increased permeability throughout the canopy, while decreases in PAI highlight foliage and branch loss at specific heights in the canopy.

All three variables were further evaluated in a variety of combinations, and a final model was produced with PAI 11–12 m and RH10. This model had the lowest RMSE, the highest  $R^2$ , and the lowest correlation between independent variables. In a final model, the change in PAI 11–12 m and RH10 both had coefficients with significant fits ( $p < 0.01$ ) of relatively equal magnitude (Figure 7; Table 2). The independent variables were not highly correlated ( $R^2 = 0.07$ ). The variation in PAI 11–12 m and RH10 explained 60% of the variation in hemlock mortality, with a predictive accuracy of 8% mortality (RMSE = 0.08).



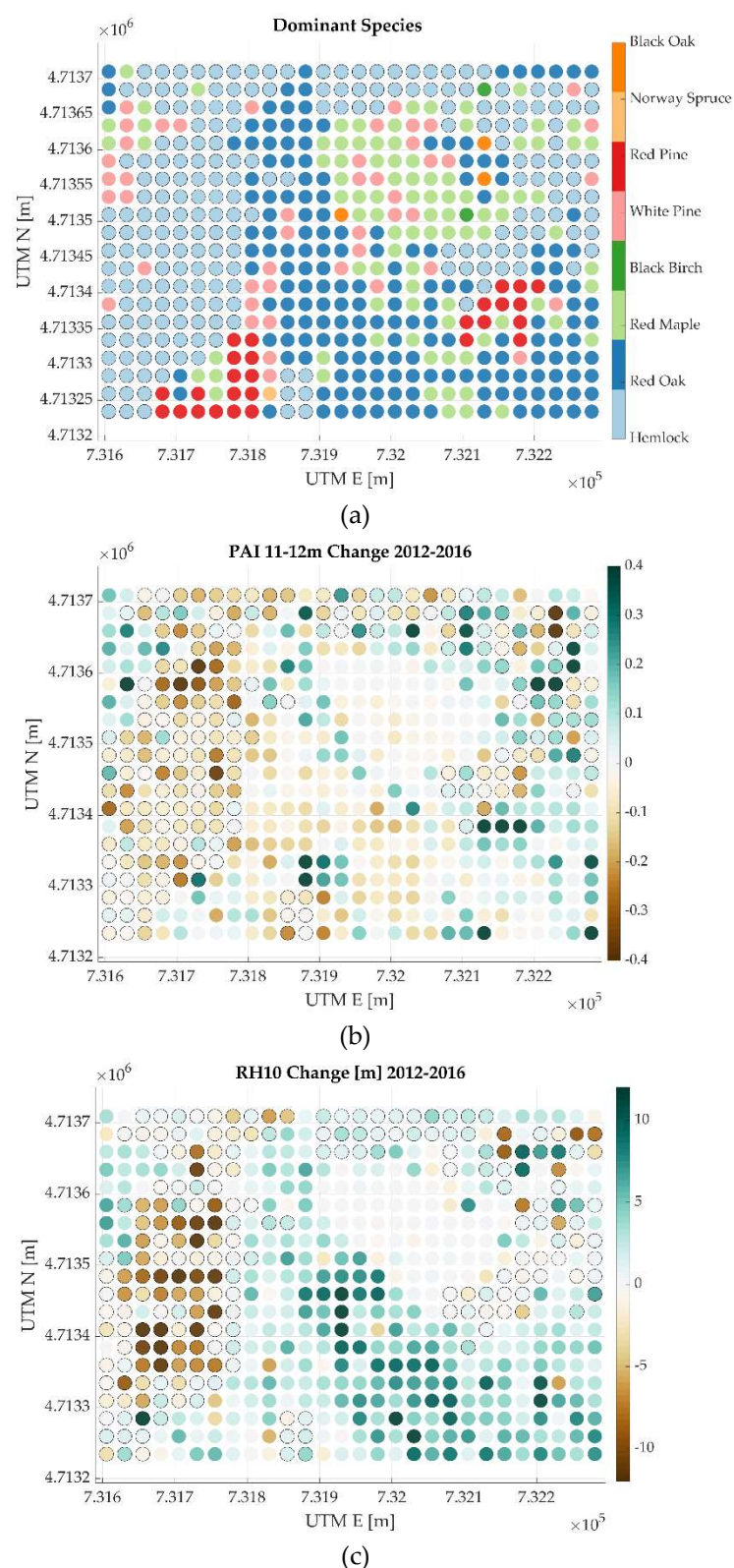
**Figure 7.** Predictions of a logistic regression of hemlock mortality using RH10 and PAI 11–12 m.

**Table 2.** Coefficients and statistical properties of a logistic regression predicting hemlock mortality

| Variable               | Value            | <i>p</i> -value |
|------------------------|------------------|-----------------|
| B <sub>RH10</sub>      | $-0.23 \pm 0.08$ | 0.002           |
| B <sub>PAI11-12m</sub> | $-0.29 \pm 0.08$ | <0.001          |
| Intercept              | $-1.89 \pm 0.08$ | <0.001          |
| R <sup>2</sup>         | 0.6              | N/A             |
| RMSE                   | 0.08             | N/A             |

### 3.3. Waveform Change Metrics

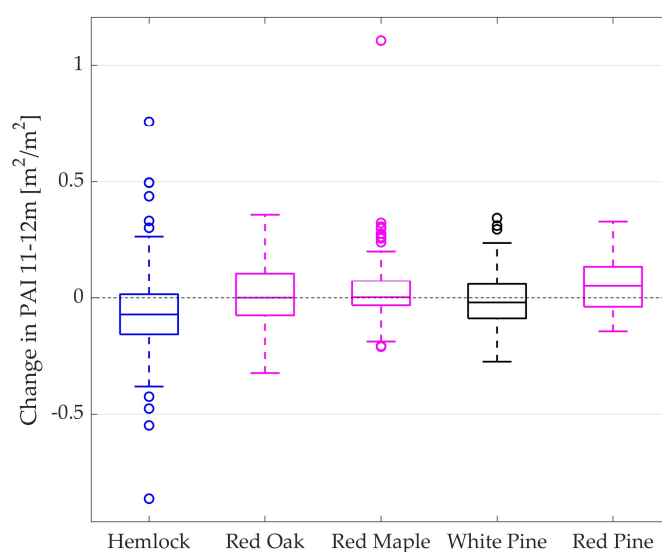
Mid-canopy PAI 11–12 m change showed a loss (brown) in areas dominated by hemlocks, whereas it generally showed an increase (green) or no change (white) in footprints dominated by other tree species (Figure 8a and 8b). In addition, areas dominated by hemlock trees showed decreases in RH10 between years, while areas dominated by healthy tree species, particularly red oak, showed positive increases in RH10 between years (Figure 8c). Decreases in RH10 in hemlock areas are indicative of increased canopy permeability and foliage loss. In contrast, the increases in RH10 in areas of healthy deciduous trees signify growth and increased canopy cover.



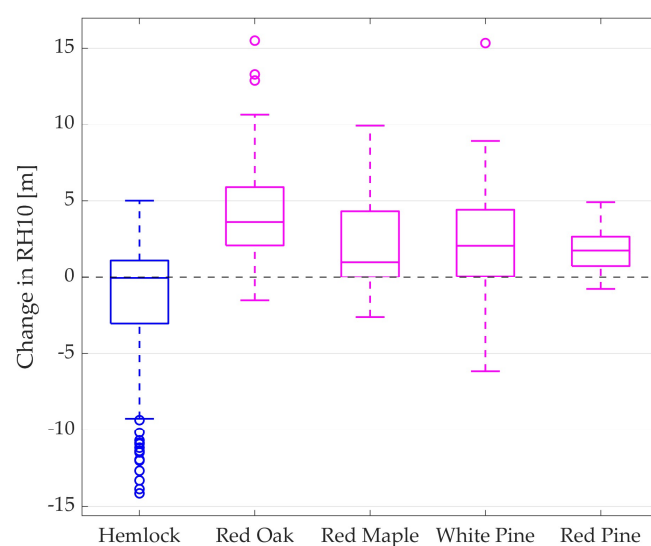
**Figure 8.** Dominant species (a), change in PAI 11–12 m (b), and change in RH10 (c) in ForestGEO plot footprints. Hemlock footprints are outlined in black. A negative RH10 change (shift down in 2016: brown) occurs within infested hemlock footprints (western portion of the plot), while a positive RH10 change (shift up in 2016, green) appears to be associated with hardwood areas (center and SE of the plot). Losses of PAI 11–12 m (brown) also highlight hemlock areas, while deciduous tree species primarily display no change (white) or a gain in PAI (green). In contrast, a swamp containing red maple trees and shrubs (northeast of center) shows little change in either metric during the time period.

To test the effect of species on waveform variables, ANOVA tests were run on 470 footprints with comparable ground heights between years (within 1 m). There was a significant effect of species on PAI 11–12 m change ( $F(4,466) = 10.8$ ,  $p < 0.001$ ; Figure 9a and Table 3). A Tukey–Kramer post hoc test revealed a significant difference ( $p < 0.001$ ) between hemlock and red oak, red maple, and red pine, but not with white pine ( $p = 0.12$ ). There was also a significant effect of species on RH10 change ( $F(4,466) = 54.7$ ,  $p < 0.001$ ). A post hoc test revealed a significant difference ( $p < 0.001$ ) between the RH10 change of hemlock and that of all other tree species (Figure 9b and Table 4).

The ANOVA results identify that unique structural changes are occurring in infested hemlock stands, demonstrated by a loss of PAI at 11–12 m and a decrease in RH10. These distinct changes even separate infested hemlock stands from stands of evergreen trees of similar structure, such as white pine and red pine.



(a)



(b)

**Figure 9.** Change in PAI 11–12 m (a) and RH10 (b) by Dominant Species. Boxplots are colored by their significant difference from hemlock (magenta;  $p < 0.05$ ), as determined by post-hoc tests.

**Table 3.** ANOVA and post-hoc test results of change in PAI 11–12 m by species

| Species    | Median | Mean           | N   | <i>p</i> -value from Hemlock |
|------------|--------|----------------|-----|------------------------------|
| Hemlock    | −0.072 | −0.064 ± 0.012 | 186 | N/A                          |
| Red Oak    | 0.001  | 0.020 ± 0.013  | 138 | < 0.001                      |
| Red Maple  | 0.003  | 0.041 ± 0.017  | 82  | < 0.001                      |
| White Pine | −0.020 | 0.001 ± 0.025  | 40  | 0.15                         |
| Red Pine   | 0.051  | 0.057 ± 0.031  | 25  | < 0.001                      |

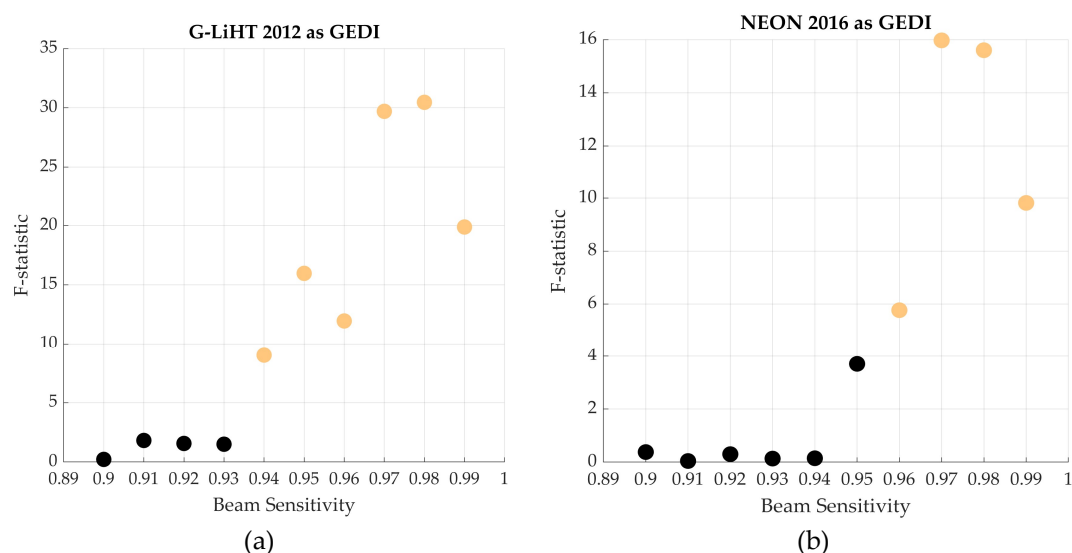
**Table 4.** ANOVA and post-hoc test results of change in RH10 by species

| Species    | Median | Mean           | N   | <i>p</i> -value from Hemlock |
|------------|--------|----------------|-----|------------------------------|
| Hemlock    | −0.085 | −1.563 ± 0.263 | 186 | N/A                          |
| Red Oak    | 3.610  | 4.200 ± 0.305  | 138 | < 0.001                      |
| Red Maple  | 0.995  | 2.166 ± 0.396  | 82  | < 0.001                      |
| White Pine | 2.070  | 2.507 ± 0.567  | 40  | < 0.001                      |
| Red Pine   | 1.760  | 1.695 ± 0.717  | 25  | < 0.001                      |

### 3.4. Simulating GEDI's Noise

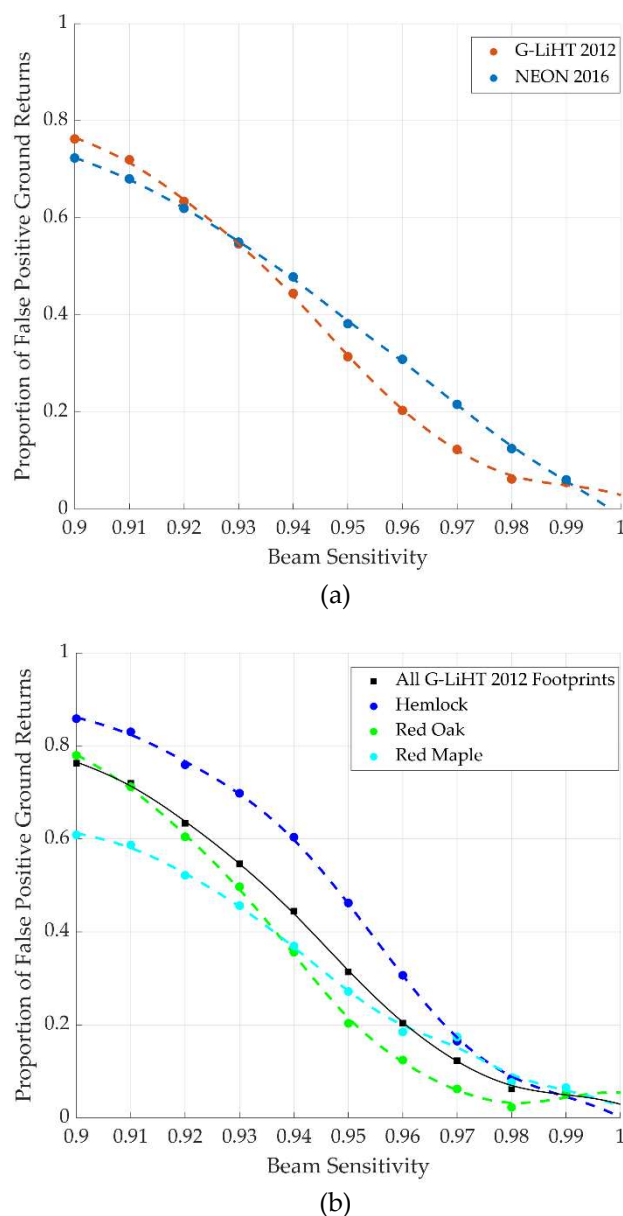
An ANOVA test was performed upon a series of noisy data files to test the effect of species (hemlock and red oak) upon the PAI 11–12 m change with added noise. Noise was added to both the G-LiHT and NEON waveforms. Change was calculated by subtracting metrics from noisy files with their corresponding noiseless counterpart (i.e., Change = noisy NEON 2016 – noiseless G-LiHT 2012). This analysis only used waveforms that had comparable ground heights between 2012 and 2016 (within 1 m), and that also agreed with the ground height derived from the ALS data (within 1 m). The mean canopy cover was  $0.90 \pm 0.18$ .

Different quality thresholds were found for each dataset using a series of ANOVAs and the change in PAI 11–12 m. When noise was added to G-LiHT 2012 waveforms, there was a significant effect of species on PAI change ( $p < 0.05$ ) for beam sensitivities 0.99–0.94 (Figure 10a). When noise was added to NEON 2016 waveforms, PAI change was significantly different ( $p < 0.05$ ) by species for beam sensitivities 0.99–0.96 (Figure 10b). Species was not significant as a grouping factor for waveforms with beam sensitivities less than or equal to 0.93 (G-LiHT) and 0.95 (NEON). Thus, each dataset varied in its response to added noise.



**Figure 10.** ANOVA with varying beam sensitivities. A series of one-way ANOVA tests tested the effect of species (hemlock and red oak) upon PAI 11–12 m change with added noise. F-test results are plotted against depreciated beam sensitivity values ( $x$ -axis), with significant test results ( $\alpha = 0.05$ ) highlighted in orange. Higher beam sensitivities represent waveforms with less noise added, while lower beam sensitivities have more noise. At the 5% significance level, the effect of species as a grouping factor stops becoming significant below a beam sensitivity of 0.94 for G-LiHT data with noise (a) and 0.96 for NEON data with noise (b).

Adding noise to waveforms reduced the number of waveforms with comparable ground heights between the years. This can be illustrated by plotting the proportion of false positive ground returns in noisy GEDI data by dominant species (Figure 11). A false positive was defined as the number of instances in which GEDI finds ground to be greater than 1 m above the ground determined by ALS data. The proportion of false positive ground returns represents the proportion of waveforms that could not correctly identify the ground due to added noise.



**Figure 11.** The proportion of false positive ground returns with added noise by dataset (a) and dominant species (b). Waveforms from different datasets and from different species have varied reactions to noise, with hemlocks accumulating ground errors at a higher rate than other species.

The different lidar datasets (Figure 11a) accumulate errors at different rates with added noise. When the beam sensitivity was 0.90, more than 70% of waveforms had issues detecting the true ground height, while with a beam sensitivity of 0.99, only about 5% of waveforms could not identify the ground accurately (Figure 11a). NEON waveforms tended to accumulate ground errors at a higher rate than did G-LiHT waveforms with beam sensitivity values from 0.99 to 0.93, even though both datasets were sampling with the same spatial distribution of footprints. This finding indicates that waveforms from the same areas can respond differently to noise, even when there are only subtle variations in shape between them.

Different species (Figure 11b) also accumulate ground errors at different rates, likely a result of the differences in canopy structural properties (canopy cover, complexity, etc.). With added noise, waveforms over hemlock stands tend to accumulate a higher proportion of false-positive ground returns than do other species, indicating a higher rate of failure to identify the true ground through a complex canopy.

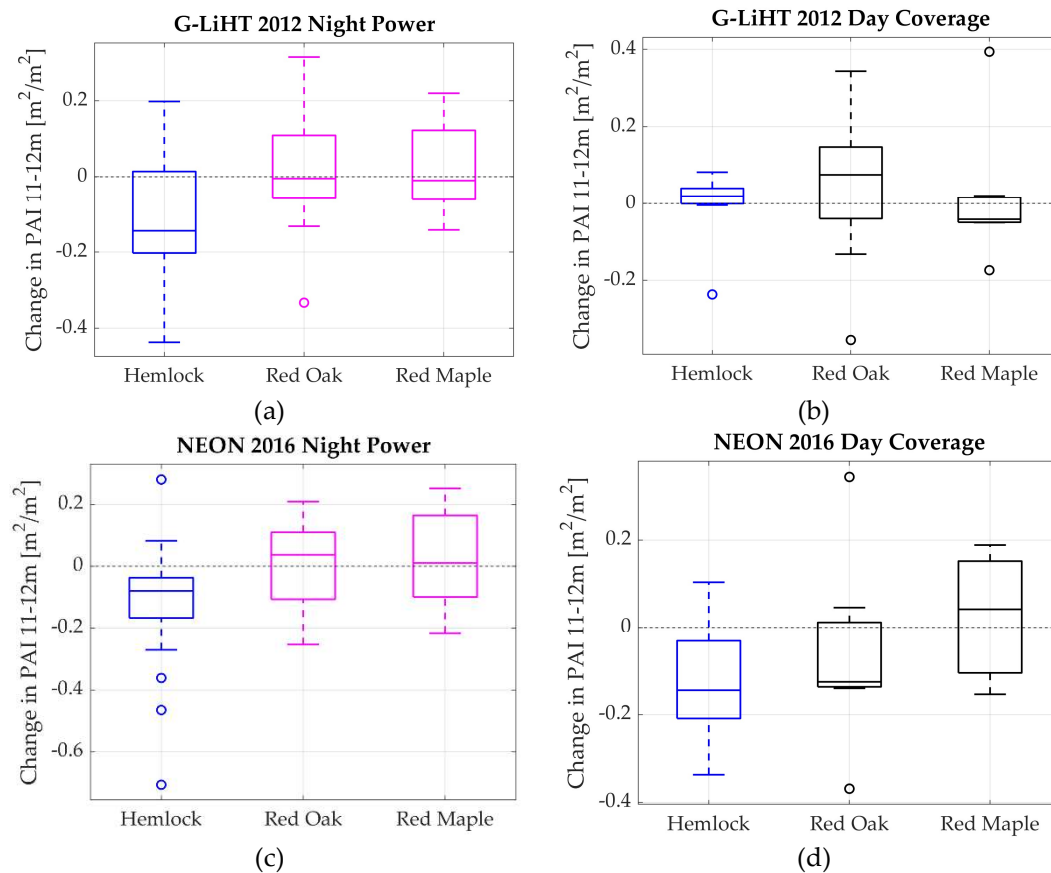
Misidentifying the ground height can affect waveform metrics by causing simulated waveforms from different datasets to be vertically offset from each other. Spatially coincident waveforms with offset ground heights either need to be filtered out before analysis (as is done in this study) or reprocessed with common ground heights before comparison.

### 3.5. Simulating GEDI's Noise and Spatial Coverage

Next, the realistic spatial coverage of GEDI (Figure 3) and the noise parameters of the best (night-power) and the worst (day-coverage) signal-to-noise ratio scenarios of GEDI were applied. When G-LiHT 2012 waveforms were degraded with GEDI's noise parameters and spatial coverage, the PAI change of waveforms using the power beam at night were significantly different by species (night-power:  $F(2,60) = 6.79$ ,  $p = 0.002$ ), while waveforms using the coverage beam during the day were not significantly different (day-coverage:  $F(2,32) = 0.43$ ,  $p = 0.65$ ; Figure 12a, 12b, and Table 5). In the night-power scenario, there was a significant difference between the PAI change of hemlock and red oak ( $p = 0.002$ ) and between that of hemlock and red maple ( $p = 0.03$ ). In the noisier day-coverage scenario, there was no significant difference between hemlock and red oak ( $p = 0.66$ ), nor with red maple ( $p = 0.96$ ).

When NEON 2016 waveforms were degraded with GEDI's noise parameters, waveforms using the power beam at night could be separated by species at the 5% significance level (night-power:  $F(2,58) = 5.39$ ,  $p = 0.007$ ), while waveforms using the coverage beam during the day could not (day-coverage:  $F(2,24) = 1.47$ ,  $p = 0.25$ ; Figure 12c, 12d, and Table 5). In the night-power scenario, there was a significant difference between hemlock and red oak ( $p = 0.02$ ), as well as hemlock and red maple ( $p = 0.02$ ). However, again in the day-coverage scenario, there were no significant differences between hemlock and red oak ( $p = 0.62$ ), nor with red maple ( $p = 0.27$ ).

With night-power GEDI waveforms, the effect of species explained 19% (NEON as GEDI) and 16% (G-LiHT as GEDI) of the variation in PAI 11–12 m, compared to the 22% of variation explained by noiseless waveforms. In contrast, species did not account for a large proportion of the variance in day-cover waveforms (3% for G-LiHT as GEDI, and 11% for NEON as GEDI; Table 5).



**Figure 12.** ANOVA with simulated noise and spatial coverage of GEDI data. Boxplots of the quantiles of PAI change at 11–12 m are plotted by species for best (night-power; **a** and **c**) and worst (day-coverage; **b** and **d**) case scenarios of GEDI’s noise parameters and laser power. Species boxplots are colored by a significant difference from hemlock (magenta;  $p < 0.05$ ), as determined by Tukey–Kramer post-hoc tests.

**Table 5.** ANOVA results of the effect of species on PAI 11–12 m with GEDI’s noise and spatial coverage applied

|             | Day-Coverage Noise |           | Night-Power Noise |           | Noiseless |
|-------------|--------------------|-----------|-------------------|-----------|-----------|
|             | G-LiHT 2012        | NEON 2016 | G-LiHT 2012       | NEON 2016 |           |
| F-statistic | 0.43               | 1.47      | 6.79              | 5.39      | 15.22     |
| p-value     | 0.65               | 0.250     | 0.002             | 0.007     | <0.001    |
| $\eta^2$    | 0.03               | 0.11      | 0.16              | 0.19      | 0.22      |

## 4. Discussion

### 4.1. Overview

This study evaluates the ability of above-canopy lidar sensors, including the newly launched GEDI spaceborne lidar, to detect distinct changes in canopy structure brought about by an invasive insect infestation in a temperate New England forest. ALS datasets acquired with different lidar instruments, with their varying pulse densities and scan angles, were successfully compared at the scale of a large-footprint lidar instrument (~25 m diameter) using the GEDI simulator [32]. Waveform metrics were correlated with the condition of hemlock trees, measured as mortality, in stands infested with the hemlock woolly adelgid. The structural changes identified by this study, particularly the PAI loss in the midstory and the increased permeability of the canopy indicated by RH10 change, reveal the utility of lidar data for detecting the unique lower canopy impacts of the HWA disturbance.

The relationship between lidar metrics and forest condition held true when ALS and simulated GEDI waveforms were compared, provided that the acquisitions were with the power beam at night and that the waveforms being compared had comparable ground heights (within 1 m). The results show that this study provides a viable method for change detection and disturbance monitoring with real GEDI data for temperate field sites with spatially coincident lidar datasets.

#### 4.2. Structural Impacts of Hemlock Woolly Adelgid

Analysis with lidar waveforms confirmed that initial HWA impacts occur in the middle and lower canopies of hemlock-dominated forests. The loss of PAI at 11–12 m above ground and the drop in RH10 from 2012 to 2016 were both found to be significantly related to hemlock mortality.

The change in PAI 11–12 m successfully distinguished infested hemlock stands from those of other non-host tree species, including both deciduous and evergreen tree species. Vertical profiles of PAI change showed that the upper and lower portions of infested hemlock canopies are growing in a similar manner to the canopy of red oak, a healthy deciduous species. In contrast, the middle portion of hemlock canopies (40–70% of maximum height) was losing PAI, a signal of structural change that was unique to infested hemlock stands. Long-term observations in other infested hemlock plots in southern New England corroborate these results, documenting a foliar loss that starts in the lower and central portions of tree crowns before it reaches the top of the crowns [51].

A decline in RH10 was also strongly correlated with hemlock mortality in HWA infested stands. The negative change in RH10 over time indicates that more laser energy was able to pass through the canopy, and therefore trigger a ground return of higher amplitude. This finding supports the hypothesis that the HWA infestation increases canopy permeability through defoliation. The importance of RH10 as a disturbance metric corresponds with the findings of other studies with large-footprint lidar data [29] that have linked positive changes in RH metrics with canopy growth and negative changes in RH metrics with canopy loss.

Waveform metrics were also able to model the condition of hemlock stands in the ForestGEO plot. The strong correlation ( $R^2 = 0.60$ ) between hemlock mortality and the change in PAI 11–12 m and RH10 further suggests that the HWA infestation produces unique structural impacts that could be further exploited to monitor the insect infestation at a regional scale. While the models of hemlock mortality derived here may not be accurate enough for large-scale prediction (RMSE = 0.08), they demonstrated that waveform variables can capture the progression of structural changes that correspond with the severity of a disturbance.

While the most important indicators of the HWA disturbance in lidar waveforms were found in the midstory and near the ground, this study also showed that plots with infested hemlocks continued to display some growth in the overstory. Potentially, this growth may be led by less-impacted trees that are taking advantage of the new resources made available by the HWA infestation. During the initial stages of the HWA infestation, the healthiest foliage remains at the top of hemlock trees, and conditions could still be adequate for canopy growth. Growth in the upper canopy could also be indicative of a ramping up of production in response to the stress of the insect, as has been observed in individual hemlock trees [13]. Another possible explanation is that the other tree species mixed in with infested hemlock stands, such as eastern white pine trees (*Pinus strobus*), are growing rapidly and taking a more dominant position in the canopy. Whatever the ecological driver may be, this finding makes an argument for monitoring studies to rely on measurements from the entire profile of forest canopies, rather than solely on top-of-canopy metrics.

#### 4.3. Comparing ALS Datasets with the GEDI Simulator

This study compared ALS datasets that were collected with different sensors and flight acquisition parameters using the GEDI simulator. After testing for potential biases between datasets, we concluded that the observed changes in waveform metrics were of ecological origin, and not artifacts of the difference between sensors. This conclusion is supported by previous research on simulated lidar data [32], which found insignificant biases in simulated waveforms from different sources of ALS data. This study also tested for potential biases in these particular datasets that could

have caused the observed changes in simulated waveforms at Harvard Forest. When the input ALS datasets were subset to footprints with the most similar sets of scan angles and point densities, the PAI loss in hemlock dominated footprints was still found to be significantly different from that of any other species (Appendix A). This finding demonstrates that any biases that may have been caused by differences in scan angle and point density were not the cause of the observed changes in PAI and RH metrics.

In addition to scan angle and point density, the two datasets also differed in other potentially important ways, such as sensor wavelength. The 1550 nm wavelength of the G-LiHT lidar sensor would be expected to be less sensitive to vegetation than would be the NEON sensor, with a 1064 nm wavelength. Based on wavelength alone and given that all other flight parameters are held equal, it could be expected that the G-LiHT sensor would obtain fewer canopy returns as compared to NEON, which would be more sensitive to vegetation and thus have a lower threshold for producing canopy returns. However, in direct opposition to these expectations, NEON data observed less foliage in 2016 than did G-LiHT in 2012. Any bias introduced by the difference in sensor wavelength would have been expected to dampen, rather than amplify, the changes in mid-canopy foliage that were observed in this study. Following this logic, we conclude that we see the impacts of HWA in simulated GEDI data in spite of, not because of, the potential biases caused by differences in sensor wavelength.

Furthermore, there is strong ecological evidence to support the conclusion that differences between simulated waveforms from 2012 to 2016 were the result of HWA impacts, and not sensor artifacts. The observed changes in waveform metrics were exclusive to infested hemlock stands, even when compared with stands of other evergreen trees, such as white pine and red pine. In addition, phenological differences between the datasets cannot explain the changes in waveform metrics, as both datasets were collected during the leaf-on season (June and August). Lastly, the mid-canopy changes captured by the simulated lidar waveforms are fully supported by field-based ecological studies, which found that HWA impacted the inner layers and lower canopy of hemlock trees before spreading to the top [51].

While this study is limited in its capacity to further evaluate potential sensor-based biases, a strong body of evidence supports the conclusion that simulated GEDI waveform metrics captured the impacts of HWA from 2012 to 2016 in the ForestGEO plot. A large-scale validation study, building off of the work of Hancock et al. [32], could provide more insight into any biases that might arise between datasets from different sensors and acquisition times. Future studies could also draw on a longer time series of ALS data from a single sensor over the ForestGEO plot, such as by using the continued data collections planned by NEON at Harvard Forest, to follow up these findings and lengthen the time series that is established by this initial study.

#### *4.4. Toward Change Detection and Disturbance Monitoring with GEDI*

While we focus on exploring the utility of GEDI and ALS for monitoring structural changes associated with HWA, the general method applied in this study could be expanded to other disturbances in temperate or tropical forests around the world. Wherever ALS data are available, GEDI waveforms can be geolocated and ALS waveforms can be simulated for comparison with real GEDI data. Structural change along the entire vertical profile of the canopy can be calculated for individual GEDI waveforms between time points: the instance of GEDI acquisition and that of the ALS data. Given additional field data on forest condition and canopy structure, a model can be trained to classify forest condition for a wide area.

This simulation study had the advantage of control over the noise parameters, sampling coverage, and location of the GEDI waveforms. Real GEDI data is estimated to have beam sensitivities between 0.92 and 0.996. Results of noise analyses showed that GEDI waveforms with beam sensitivities of greater than 0.96 and data acquisitions made at night with the power laser were most suitable for monitoring HWA in forests with comparable structure (mean canopy cover of 0.90). However, night-power waveforms had different capabilities depending on which input ALS data was degraded with noise. For instance, noisy G-LiHT and noisy NEON datasets produced slightly different quality thresholds when analyzing the effect of species on PAI 11–12 m change with

ANOVA (0.94 and 0.96 beam sensitivity). This finding demonstrates that future disturbance studies will need to develop their own quality thresholds to fit the structural complexity of the target site and the species of interest.

Forests with different structures and species compositions will vary in their sensitivity to noise, but hemlock forests seem to be particularly affected. Added noise caused hemlock stands to accumulate ground errors at a faster rate than those of other species. This difference is likely due to the unique structural properties of hemlock stands, including their dense canopy cover. Future studies could adjust for these errors by using the true ground elevation from coincident ALS datasets.

A potential limitation of this study is its reliance on absolute PAI metrics processed at 1 m vertical resolution. In contrast, the GEDI PAI product (L2B) is being released to users at 5 m height intervals [40]. In light of this, future studies should consider how varying vertical resolutions of PAI differ in their ability to capture structural changes related to disturbances. In addition, monitoring studies should test the utility of normalized or proportional metrics for capturing disturbances across a wider region. When expanding to a wider region with a greater variation of canopy structure, it may be beneficial to convert absolute metrics, such as aboveground heights and PAI change, into normalized metrics, such as percentiles of maximum height and proportional PAI change.

The sparse spatial coverage of GEDI also posed an issue for the regression model used in this study, since only eight footprints overlapped with the western portion of the ForestGEO plot that was sampled for mortality in 2016. Future studies will be able to use field data from the entire 35-hectare ForestGEO plot as a result of the recent field surveys conducted in 2018 and 2019. In general, future studies with GEDI data would benefit from drawing on large field plots, such as the entire 35-hectare ForestGEO plot and others in the CTFS network [47], and focusing on sites with the highest density of GEDI acquisitions, such as those locations currently identified for GEDI's validation studies.

## 5. Conclusions

By linking the change in simulated GEDI waveforms with the deteriorating condition of hemlock stands in a temperate New England forest, this study demonstrates how a time series of lidar waveforms can capture structural change and classify disturbance. The GEDI Simulator [32] successfully enabled an analysis of structural change between two ALS datasets using simulated GEDI data. Lower canopy PAI and RH metrics from the simulated GEDI waveforms emerged as significant predictors of the severity of the HWA infestation. These structure–condition relationships held true even when the quality of GEDI waveforms was degraded with realistic noise and the sparse coverage of GEDI acquisitions was applied. In addition, the waveform metrics identified growth in the upper canopies of hemlock stands, revealing that infested plots are undergoing a complex shift in the vertical distribution of leaf matter, with loss in the lower canopy and continued growth in the upper canopy.

Initial GEDI data is just becoming available, and data will continue to be collected and released over the course of the next 2 years. Following the methods of this study, a time series of ALS data and/or GEDI data can be compared for other forested sites and disturbances. The findings of this study open up new opportunities for ecological research and disturbance monitoring in temperate and tropical forests around the world.

**Author Contributions:** Conceptualization, P.B., S.H., Z.L., and C.S.; Methodology, P.B., S.H., D.O., L.D., J.A., H.T., A.E., and Z.L.; Software, P.B. and S.H.; Formal analysis, P.B., I.P., and S.H.; Investigation, P.B., I.P., K.K., B.C., and D.O.; Resources, K.K., B.C., D.O., and C.S.; Data curation, P.B. and S.H.; Writing—original draft preparation, P.B., S.H., I.P., A.E., D.O., and C.S.; Writing—review and editing, P.B., L.D., J.A., H.T., B.C., K.K., Z.L., A.E., and D.O.; Visualization, P.B. and S.H.; Supervision, C.S.; Project administration, C.S.; Funding acquisition, C.S. and P.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received funding from the NSF Integrative Graduate Education and Research Traineeship (IGERT) award, number 1249946. In addition, this work was supported by NASA MODIS, grant number 80NSSC18K0642, and NASA VIIRS, grant number 80NSSC18K0621.

**Acknowledgments:** We would like to acknowledge Ralph Dubayah (UMD, PI of GEDI) and the entire GEDI Science Team for their support of this work, which began well before the launch of GEDI. In particular, we thank Scott Luthcke at the NASA Goddard Space Flight Center for providing the pre-launch simulated GEDI acquisition locations for the Harvard Forest region. We also heartily thank Audrey Barker-Plotkin and all of the staff at Harvard Forest, as well as the field crews funded by the Smithsonian Institute for collecting and maintaining detailed field data in the CTFS ForestGEO plot.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Appendix A

### Evaluating Potential Sensor Bias

Working with simulated data from two different airborne lidar instruments allowed for further analysis of sensor bias in this study. The effects of the scan angle and the pulse density distributions of the input ALS datasets upon simulated waveform metrics were examined.

Scan angle and point density distributions differed greatly between both sets of input data. G-LiHT ALS data had a greater range of scan angles per footprint (max scan angle of 36 degrees in G-LiHT vs. 18 degrees in NEON), as well as a higher average point density per footprint (29.6 points per m<sup>2</sup>) compared to NEON data (6.8 points per m<sup>2</sup>; Table 1).

To mitigate potential biases caused by the different scan angle and point density distributions in the input ALS datasets, the PAI 11–12 m change was compared by species (hemlock, red oak, red maple, and white pine) for a subset of footprints with the most similar distributions of input data. When the data were subset to 157 footprints (one-third of the dataset) with similar scan angle distributions, an ANOVA with Tukey–Kramer post-hoc tests still found significant differences between hemlock and red oak, red maple, and white pine ( $p < 0.05$ ). When the data were subset to those with similar point density distributions, PAI 11–12 m changes in hemlock footprints were also significantly different from those of red oak, red maple, and white pine ( $p < 0.05$ ). These results reinforce the findings of this study, showing that the changes observed in waveform variables are not the result of sensor differences, but rather, have ecological origins.

## References

1. Oliver, C.D. Forest Development in North America Following Major Disturbances. *For. Ecol. Manag.* **1980**, *3*, 153–168.
2. Orwig, D.A.; Foster, D.R. Forest Response to the Introduced Hemlock Woolly Adelgid in Southern New England, USA. *J. Torrey Bot. Soc.* **1998**, *125*, 60–73.
3. Orwig, D.A.; Barker Plotkin, A.A.; Davidson, E.A.; Lux, H.; Savage, K.E.; Ellison, A.M. Foundation species loss affects vegetation structure more than ecosystem function in a northeastern USA forest. *PeerJ*. **2013**, *1*, e41.
4. Parker, B.L.; Skinner, M.; Gouli, S.; Ashikaga, T.; Teillon, H.B. Low lethal temperature for hemlock woolly adelgid (Homoptera: Adelgidae). *Environ. Entomol.* **1999**, *28*, 1085–1091.
5. McClure, M.S.; Cheah, C.A.S.-J. Important Mortality Factors in the Life Cycle of Hemlock Woolly Adelgid, *Adelges tsugae* Annand (Homoptera: Adelgidae) in the Northeastern United States. In Proceedings of the Hemlock Woolly Adelgid Symposium, East Brunswick, NJ, USA, 5–7 February 2002; Onken, B., Reardon, R., Lashomb, J., Eds.; 13–22.
6. Parker, B.L.; Skinner, M.; Gouli, S.; Ashikaga, T.; Teillon, H.B. Survival of Hemlock Woolly Adelgid (Homoptera: Adelgidae) at Low Temperatures. *For. Sci.* **1998**, *44*, 414–420.
7. Skinner, M.; Parker, B.L.; Gouli, S.; Ashikaga, T. Regional Responses of Hemlock Woolly Adelgid (Homoptera: Adelgidae) to Low Temperatures. *Environ. Entomol.* **2003**, *32*, 523–528.
8. Paradis, A.; Elkinton, J.; Hayhoe, K.; Buonaccorsi, J. Role of winter temperature and climate change on the survival and future range expansion of the hemlock woolly adelgid (*Adelges tsugae*) in eastern North America. *Mitig. Adapt. Strateg. Glob. Chang.* **2008**, *13*, 541–554.
9. Dukes, J.S.; Pontius, J.; Orwig, D.; Garnas, J.R.; Rodgers, V.L.; Brazee, N.; Cooke, B.; Theoharides, K.A.;

- Stange, E.E.; Harrington, R.; et al. Responses of insect pests, pathogens, and invasive plant species to climate change in the forests of northeastern North America: What can we predict? *Can. J. For. Res.* **2009**, *39*, 231–248.
10. Trotter, R.T.; Shields, K.S. Variation in Winter Survival of the Invasive Hemlock Woolly Adelgid (Hemiptera: Adelgidae) Across the Eastern United States. *Environ. Entomol.* **2009**, *38*, 577–587.
  11. Elkinton, J.S.; Lombardo, J.A.; Roehrig, A.D.; McAvoy, T.J.; Mayfield, A.; Whitmore, M. Induction of cold hardiness in an invasive herbivore: The case of hemlock woolly adelgid (Hemiptera: Adelgidae). *Environ. Entomol.* **2017**, *46*, 118–124.
  12. Dubayah, R.O.; Blair, J.B.; Goetz, S.J.; Fatoyinbo, L.; Hansen, M.C.; Healey, S.P.; Hofton, M.; Hurtt, G.; Kellner, J.R.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Sci. Remote Sens.* **2020**, *1*, 100002.
  13. Domec, J.-C.; Rivera, L.N.; King, J.S.; Peszlen, I.; Hain, F.; Smith, B.; Frampton, J. Hemlock woolly adelgid (*Adelges tsugae*) infestation affects water and carbon relations of eastern hemlock (*Tsuga canadensis*) and Carolina hemlock (*Tsuga caroliniana*). *New Phytol.* **2013**, *199*, 452–463.
  14. Fitzpatrick, M.C.; Preisser, E.L.; Porter, A.; Elkinton, J.; Ellison, A.M. Modeling range dynamics in heterogeneous landscapes: Invasion of the hemlock woolly adelgid in eastern North America. *Ecol. Appl.* **2012**, *22*, 472–486.
  15. Ellison, A.; Orwig, D.; Fitzpatrick, M.; Preisser, E. The Past, Present, and Future of the Hemlock Woolly Adelgid (*Adelges tsugae*) and Its Ecological Interactions with Eastern Hemlock (*Tsuga canadensis*) Forests. *Insects* **2018**, *9*, 172.
  16. Pekel, J.F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422.
  17. Little, E.L.J. *Atlas of the United States Trees*, 1st ed.; U.S. Department of Agriculture, Forest Service: Washington, DC, USA: 1971.
  18. Carey, J.H. *Tsuga canadensis*. In *Fire Effects Information System*; U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory, Missoula, MT: 1993. Available online: <https://www.fs.fed.us/database/feis/plants/tree/tsucan/all.html> (accessed on 13 April 2020).
  19. Kellner, J.R.; Clark, D.B.; Hubbell, S.P. Pervasive canopy dynamics produce short-term stability in a tropical rain forest landscape. *Ecol. Lett.* **2009**, *12*, 155–164.
  20. Calders, K.; Schenkels, T.; Bartholomeus, H.; Armston, J.; Verbesselt, J.; Herold, M. Monitoring spring phenology with high temporal resolution terrestrial LiDAR measurements. *Agric. For. Meteorol.* **2015**, *203*, 158–168.
  21. Solberg, S.; Næsset, E.; Hanssen, K.H.; Christiansen, E. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sens. Environ.* **2006**, *102*, 364–376.
  22. Meng, R.; Dennison, P.E.; Zhao, F.; Shendryk, I.; Rickert, A.; Hanavan, R.P.; Cook, B.D.; Serbin, S.P. Mapping canopy defoliation by herbivorous insects at the individual tree level using bi-temporal airborne imaging spectroscopy and LiDAR measurements. *Remote Sens. Environ.* **2018**, *215*, 170–183.
  23. Tang, H.; Dubayah, R. Light-driven growth in Amazon evergreen forests explained by seasonal variations of vertical canopy structure. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2640–2644.
  24. Wulder, M.A.; White, J.C.; Nelson, R.F.; Næsset, E.; Ørka, H.O.; Coops, N.C.; Hilker, T.; Bater, C.W.; Gobakken, T. Lidar sampling for large-area forest characterization: A review. *Remote Sens. Environ.* **2012**, *121*, 196–209.
  25. Weishampel, J.F.; Drake, J.B.; Cooper, A.; Blair, J.B.; Hofton, M. Forest canopy recovery from the 1938 hurricane and subsequent salvage damage measured with airborne LiDAR. *Remote Sens. Environ.* **2007**, *109*, 142–153.
  26. Harding, D.; Lefsky, M.; Parker, G.; Blair, J. Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broadleaf forests. *Remote Sens. Environ.* **2001**, *76*, 283–297.
  27. Drake, J.B.; Dubayah, R.O.; Clark, D.B.; Knox, R.G.; Blair, J.B.; Hofton, M.A.; Chazdon, R.L.; Weishampel, J.F.; Prince, S.D. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sens. Environ.* **2002**, *79*, 305–319.
  28. Marselis, S.M.; Tang, H.; Armston, J.D.; Calders, K.; Labrière, N.; Dubayah, R. Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in Lopé National Park, Gabon. *Remote Sens. Environ.* **2018**, *216*, 626–634.
  29. Dubayah, R.O.; Sheldon, S.L.; Clark, D.B.; Hofton, M.A.; Blair, J.B.; Hurtt, G.C.; Chazdon, R.L. Estimation

- of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *J. Geophys. Res. Biogeosci.* **2010**, *115*, doi:10.1029/2009JG000933.
30. Huang, W.; Sun, G.; Dubayah, R.; Cook, B.; Montesano, P.; Ni, W.; Zhang, Z. Mapping biomass change after forest disturbance: Applying LiDAR footprint-derived models at key map scales. *Remote Sens. Environ.* **2013**, *134*, 319–332.
  31. Patterson, P.L.; Healey, S.P.; Ståhl, G.; Saarela, S.; Holm, S.; Andersen, H.-E.; Dubayah, R.O.; Duncanson, L.; Hancock, S.; Armston, J.; et al. Statistical properties of hybrid estimators proposed for GEDI-NASA's global ecosystem dynamics investigation. *Environ. Res. Lett.* **2019**, *14*, 065007.
  32. Hancock, S.; Armston, J.; Hofton, M.; Sun, X.; Tang, H.; Duncanson, L.I.; Kellner, J.R.; Dubayah, R. The GEDI simulator: A large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. *Earth Space Sci.* **2019**, *6*, 294–310.
  33. Blair, B.J.; Hofton, M.A. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophys. Res. Lett.* **1999**, *26*, 2509–2512.
  34. Orwig, D.; Foster, D.; Ellison, A. Harvard Forest CTFS-ForestGEO Mapped Forest Plot since 2014. 2015. Available online: <https://harvardforest1.fas.harvard.edu/exist/apps/datasets/showData.html?id=hf253> (accessed on 9 September 2019).
  35. Cook, B.; Corp, L.; Nelson, R.; Middleton, E.; Morton, D.; McCorkel, J.; Masek, J.; Ranson, K.; Ly, V.; Montesano, P. NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) Airborne Imager. *Remote Sens.* **2013**, *5*, 4045–4066.
  36. Kampe, T.; Johnson, B.R.; Kuester, M.; Keller, M. NEON: The first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *J. Appl. Remote Sens.* **2010**, *4*, 043510.
  37. Hancock, S.; Anderson, K.; Disney, M.; Gaston, K.J. Measurement of fine-spatial-resolution 3D vegetation structure with airborne waveform lidar: Calibration and validation with voxelised terrestrial lidar. *Remote Sens. Environ.* **2017**, *188*, 37–50.
  38. Blair, J.B.; Rabine, D.L.; Hofton, M.A. The Laser Vegetation Imaging Sensor: A medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS J. Photogramm. Remote Sens.* **1999**, *54*, 115–122.
  39. Hofton, M.A.; Minster, J.B.; Blair, J.B. Decomposition of laser altimeter waveforms. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 1989–1996.
  40. Tang, H.; Armston, J. Algorithm Theoretical Basis Document (ATBD) for GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics; Goddard Space Flight Center, Greenbelt, MD, USA: 2019. Available online: <https://gedi.umd.edu/data/documents/> (accessed on January 1 2020).
  41. Chen, J.M.; Rich, P.M.; Gower, S.T.; Norman, J.M.; Plummer, S. Leaf area index of boreal forests: Theory, techniques, and measurements. *J. Geophys. Res. Atmos.* **1997**, *102*, 29429–29443.
  42. Ni-Meister, W.; Jupp, D.L.; Dubayah, R. Modeling Lidar Waveforms in Heterogeneous and Discrete Canopies. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1943–1958.
  43. Armston, J.; Disney, M.; Lewis, P.; Scarth, P.; Phinn, S.; Lucas, R.; Bunting, P.; Goodwin, N. Direct retrieval of canopy gap probability using airborne waveform lidar. *Remote Sens. Environ.* **2013**, *134*, 24–38.
  44. Tang, H.; Dubayah, R.; Swatantran, A.; Hofton, M.; Sheldon, S.; Clark, D.B.; Blair, B. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sens. Environ.* **2012**, *124*, 242–250.
  45. Tang, H.; Brolly, M.; Zhao, F.; Strahler, A.H.; Schaaf, C.; Ganguly, S.; Zhang, G.; Dubayah, R. Deriving and validating Leaf Area Index (LAI) at multiple spatial scales through lidar remote sensing: A case study in Sierra National Forest, CA. *Remote Sens. Environ.* **2014**, *143*, 131–141.
  46. Hofton, M.; Blair, B.; Story, S.; Yi, D. Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products; Goddard Space Flight Center, Greenbelt, MD, USA: 2019. Available online: <https://gedi.umd.edu/data/documents/> (accessed on January 1 2020).
  47. Anderson-Teixeira, K.J.; Davies, S.J.; Bennett, A.C.; Gonzalez-Akre, E.B.; Muller-Landau, H.C.; Joseph Wright, S.; Abu Salim, K.; Almeyda Zambrano, A.M.; Alonso, A.; Baltzer, J.L.; et al. CTFS-ForestGEO: A worldwide network monitoring forests in an era of global change. *Glob. Chang. Biol.* **2015**, *21*, 528–549.
  48. Orwig, D.A.A.; Boucher, P.; Paynter, I.; Saenz, E.; Li, Z.; Schaaf, C. The potential to characterize ecological data with terrestrial laser scanning in Harvard Forest, MA. *Interface Focus* **2018**, *8*, 20170044.

49. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2009**, *33*, 1–22.
50. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288.
51. Orwig, D.A.; Foster, D.R.; Mauseel, D.L. Landscape patterns of hemlock decline in New England due to the introduced hemlock woolly adelgid. *J. Biogeogr.* **2002**, *29*, 1475–1487.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).