7-2022

The Potential of Citizen Science Data to Complement Satellite and Airborne Lidar Tree Height Measurements: Lessons from The GLOBE Program

Josh Enterkine
Boise State University

Brian A. Campbell
NASA Goddard Space Flight Center

Holli Kohl
NASA Goddard Space Flight Center

Nancy F. Glenn
Boise State University

Kristen Weaver
NASA Goddard Space Flight Center

See next page for additional authors

Authors
Josh Enterkine, Brian A. Campbell, Holli Kohl, Nancy F. Glenn, Kristen Weaver, David Overoye, and Deanna Danke

IOP Publishing Environ. Res. Lett. 17 (2022) 075003 https://doi.org/10.1088/1748-9326/ac77a2

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
10 December 2021

REVISED
17 May 2022

ACCEPTED FOR PUBLICATION
10 June 2022

PUBLISHED
21 June 2022

LETTER

The potential of citizen science data to complement satellite and airborne lidar tree height measurements: lessons from The GLOBE Program

Josh Enterkine[1], **Brian A Campbell**[2,3,4,*], **Holli Kohl**[2,5], **Nancy F Glenn**[1], **Kristen Weaver**[2,5], **David Overoye**[6] **and Deanna Danke**[7]

[1] Department of Geoscience, Boise State University, Boise, ID, United States of America
[2] Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, MD, United States of America
[3] Laboratory for Atmospheres, NASA Wallops Flight Facility, Wallops Island, VA, United States of America
[4] Global Science and Technology, Inc., Greenbelt, MD, United States of America
[5] Science Systems and Applications, Inc., Lanham, MD, United States of America
[6] Science Systems and Applications, Inc., Pasadena, CA, United States of America
[7] Humanities and STEAM, Monsignor McClancy Memorial High School, Queens, NY, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: brian.a.campbell@nasa.gov

Keywords: citizen science, ICESat-2, NASA GLOBE Observer, tree height, The GLOBE Program, carbon storage

Abstract

The Global Learning and Observations to Benefit the Environment (GLOBE) Program is an international science, citizen science, and education program through which volunteers in participating countries collect environmental data in support of Earth system science. Using the program's software application, GLOBE Observer (GO), volunteers measure tree height and optional tree circumference, which may support the interpretation of NASA and other space-based satellite data such as tree height data from the Ice, Cloud, and land Elevation Satellite-2 (ICESat-2) and Global Ecosystem Dynamics Investigation instrument. This paper describes tree heights data collected through the GO application and identifies sources of error in data collection. We also illustrate how the ground-based citizen science data collected in the GO application can be used in conjunction with ICESat-2 tree height observations from two locations in the United States: Grand Mesa, Colorado, and Greenbelt, Maryland. Initial analyses indicate that data location accuracy and the scientific relevance of data density should be considered in order to align GLOBE tree height data with satellite-based data collections. These recommendations are intended to inform the improved implementation of citizen science environmental data collection in scientific work and to document a use case of the GLOBE Trees data for the science research community.

1. Introduction

1.1. Why measure tree height

Measuring tree height can be used to understand larger environmental conditions. The height of trees is correlated with available resources and stressors in the ecosystem, and thus is a widely used indicator of an environment's ability to grow trees (Koch *et al* 2004). Tree height is also used as an input for calculating aboveground biomass and for global land cover in calculations of carbon absorption (e.g. Hunter *et al* 2013, Sintayehu *et al* 2020).

Traditional methods of measuring tree height include field measurements with a hand-held clinometer, hypsometer, or measurement pole. As these methods are time-consuming and difficult to scale, satellite-based laser altimeters are ideally suited for mapping vegetation height globally (e.g. Li *et al* 2020, Potapov *et al* 2021). The NASA Ice, Cloud, and land Elevation Satellite-2 (ICESat-2), launched on 15 September 2018, carries the Advanced Topographic Laser Altimeter System (ATLAS) instrument designed to measure surface elevations, including tree canopy height, by counting laser photons leaving and returning to the satellite (Campbell 2021). While the primary science objective of ICESat-2 is to measure the changing elevation of ice, ICESat-2's ATLAS instrument can also measure the elevation changes of any surface or near-surface object (Markus *et al* 2017). Recent applications of ATLAS data to measure tree and canopy height have demonstrated its potential (Neuenschwander and Pitts 2019, Sun *et al* 2020). An essential component in understanding ICESat-2's performance in measuring tree height is the comparison to ground-based measurements. As an example, a recent study (Nandy *et al* 2020), found a 1.1 m root mean square error (a measure of the differences between values predicted by a model and the values observed), between field-measured and ICESat-2 tree heights for trees that ranged from 14.2 to 32.4 m. This error information is critical to utilizing ICESat-2 data for aboveground biomass estimates or global land cover uses (Narine *et al* 2020).

As such, citizen science may offer a valuable source of tree and vegetation height data for validating satellite-based measurements. By distributing data collection to many volunteers, citizen science can scale across broad geographic regions, and may be capable of generating a large enough sample data set to support the development of satellite-based methods of calculating tree height. Previous studies have demonstrated the potential value and application of citizen science data in ecology including tree inventories, while considering data quality (e.g. Crall *et al* 2011, Molinier *et al* 2016, Roman *et al* 2017, Dujardin *et al* 2022).

The Global Learning and Observations to Benefit the Environment (GLOBE) Program's software application (app) for smartphones and other mobile devices, GLOBE Observer (GO), engages volunteers worldwide in the collection of environmental data including tree height estimates. This paper describes the methods used to collect tree height measurements through the Trees tool in the GO app, hereafter GO Trees, and identifies sources of error in data collection. We also discuss how the ground-based citizen science data collected using GO Trees can be used in conjunction with space-based (ICESat-2) tree height observations at two locations in the United States. We make recommendations for citizen science data collection, specifically for measuring accurate tree heights, and for correlating this data to satellite-based observations.

1.2. The GLOBE Program

Founded in 1994 as a science and education program, The GLOBE Program is an international science, citizen science, and education program that invites volunteers of all ages to monitor the environment in support of Earth system science (Finarelli 1998, Butler and MacGregor 2003, GLOBE 2019). GLOBE is implemented through bilateral agreements between the United States and the governments of 126 participating countries as of late 2021. GLOBE data include more than 50 environmental parameters and are freely accessible through The GLOBE Program's website, www.globe.gov.

Launched in 2016, GO is the smartphone app of The GLOBE Program. Volunteers receive tiered access to GO depending on training and certification. Within most GLOBE countries, universal access is granted to collect environmental data for five protocols that have app-based mechanisms or tools for data collection. The app may not be used in nor are data accepted from countries that have not signed a formal GLOBE agreement. For the most part, the protocols—clouds, mosquito habitats, land cover, trees, and eclipse—do not require equipment beyond the app tools to collect meaningful data. The eclipse protocol requires a thermometer and trees has an optional tree circumference observation that requires a tape measure (Amos *et al* 2020). Training for these protocols is provided within the app.

In-app tools for each of the protocols guide volunteers through environmental data collection in which users take geotagged photographs and answer questions to document and classify clouds (Colón Robles *et al* 2020), eclipse conditions (Dodson *et al* 2019), mosquito habitats (Low *et al* 2021), or land cover (Kohl *et al* 2021). GO Trees is unique among the other tools because tagged photos are not the primary data collected, but rather it uses a smartphone's gyroscope and user input to calculate and estimate tree height.

1.3. The Trees tool

The Trees tool in the GO app provides a mechanism for citizen science-based estimates of tree height by tapping into the phone's gyroscope to turn the device into a handheld clinometer. A user downloads the app, creates an account, and follows the guides through the observation process. During account creation, the citizen scientist enters their height from which the individual's average stride length and height of the phone at eye level are estimated. Stride length (L) is defined as $L = 0.413 \times \textit{person's height}$. A user may independently measure actual average stride length and enter that value in place of the estimated stride length. The height of the phone at eye level is estimated as user height minus 10.14 cm. This estimate may also be manually updated with a measured value.

After account setup, the citizen scientist must complete a simple training, which includes instructions on how to measure stride length, review tree observation techniques, and assess accuracy before unlocking the rest of the tool to collect data. To measure a tree, the user confirms the date and time and records surface conditions around the tree (dry vs wet ground, leaves on trees, raining, snow/ice on the ground). The user is then guided to select an individual tree of any species that is at least 5 m tall and for which the base and crown are clearly visible and identifiable, and with a clear, unobstructed walking path to the tree. Standing 7–25 m away from the tree, the user taps the '*measure tree base*' button. The app opens the camera with a dotted line and instructions overlaid on the screen. Holding the phone at eye height, the user aligns the dotted line with the base of the tree and taps the screen to capture the angle to the base of the tree as measured by the phone's internal gyroscope. Without moving the phone vertically, the user tilts the camera to measure the angle to the crown of the tree, and then photographs the tree.

The user walks to the base of the tree in a straight line, records the number of steps taken to reach the tree (to determine distance), then confirms the latitude and longitude location of the tree. The app estimates location accuracy, and the user is requested to refresh the location or to manually enter the tree's location on a map to improve accuracy. The app then uses the above information to calculate an estimate of the tree's height. If a measuring tape is available, an optional step includes measuring the circumference of the tree at breast height (to obtain the commonly measured diameter at breast height). In the first three years of data collected via the GO Trees tool (March 2019 through March 2022), approximately 30% of tree observations included circumference measurements. As this is an optional secondary measurement in the app, and not comparable to satellite data, further discussion of circumference data is not included in this paper. Tree species data are not collected in this protocol.

The final screen of the tool provides a summary of the values entered. Each field may be edited (see figure 1). The user may edit the estimated distance to the tree with a measured value to improve height accuracy.

Tree height is calculated as follows (GLOBE 2019):

Figure 2 shows the geometry associated with the GO Trees tool. User inputs provide the angles 'A' and 'B', the camera height 'h' and the distance to the tree 'd'. The tree height 'T' is the sum of '$t1$' and '$t2$'.

$$T = t1 + t2$$

$$t1 = f \times \mathrm{Tan}\,(B)$$

$$t2 = f \times \mathrm{Tan}\,(A)$$

$$\text{and } f = d \times \mathrm{Sin}\,(F)$$

$$\text{So } T = f \times (\mathrm{Tan}\,(B) + \mathrm{Tan}\,(A))$$

Using basic geometry principles, we can see:

$$F = D + E$$

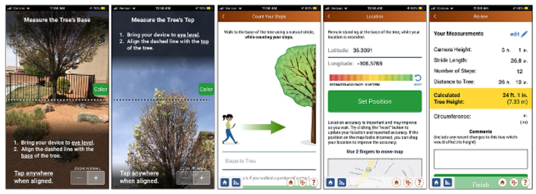

Figure 1. On-screen instructions guide citizen scientists to record the angle to the base and crown of the tree, estimate the distance to the tree, and record its location in GO Trees.

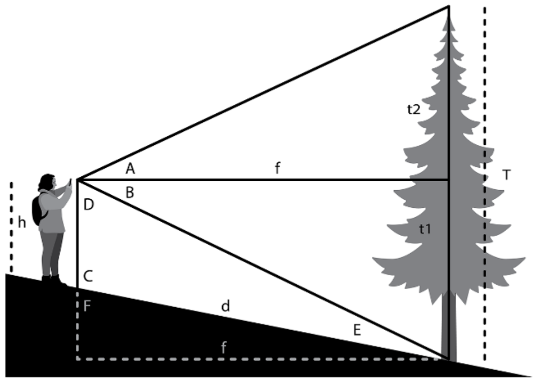

Figure 2. Geometry used in GO Trees to calculate tree height.

$$D = 90 - B$$

and using the law of sines, we can show:

$$E = \arcsin\left(h \times \frac{\mathrm{Sin}\,(D)}{d}\right).$$

Substituting $D + E$ for F, we get:

$$f = d \times \mathrm{Sin}\left(90 - B + \arcsin\left(h \times \frac{\mathrm{Sin}\,(90 - B)}{d}\right)\right).$$

So tree height, T becomes:

$$T = d \times \mathrm{Sin}\left(90 - B + \arcsin\left(h \times \frac{\mathrm{Sin}\,(90 - B)}{d}\right)\right) \times \mathrm{Tan}\,(B) + \mathrm{Tan}\,(A)).$$

Given that:
n: number of steps
L: stride length (m)
H_c: camera height (m)
H_{tree}: full tree height (m)
h_1: upper tree height (m)
h_2: lower tree height (m)
α: app-measured top angle (degrees)
β: app-measured bottom angle (degrees)

We calculate:

$$\varphi = \text{ radians } (90 - \beta)$$

$$\theta = \text{radians}\,(90 - a)$$

$$\lambda = \arctan\left(\frac{H_c}{nL}\right) - \varphi$$

$$h_1 = \tan(\theta)$$

$$h_2 = \tan(\varphi)$$

$$d = nL\cos(\lambda)$$

$$H_{\text{tree}} = d(h_1 + h_2)\,.$$

Certain constraints are placed on user inputs to avoid unreasonable numbers for characteristics such as height (0.61–2.44 m, the height of a small child to that of a very tall adult) and angles (angle to tree base [angle D] $\geqslant 15°$, angle to tree top [angle $A+B+D$] $\leqslant 165°$). In addition, the user must be at least 7.62 m and no more than 22.86 m away from the tree. Based on these constraints, the maximum tree height a 2.44 m tall person (camera height 2.34 m) could measure at the maximum distance is 90.66 m, and at the minimum distance is 31.78 m. The maximum tree height a 0.61 m tall person (camera height 0.51 m) could measure at the maximum distance is 88.83 m, and at the minimum distance is 29.95 m.

GO Trees was released on 26 March 2019. By the end of 2021, more than 40 000 GO Trees measurements have been submitted directly from the app. These data are available to view and download at www.globe.gov/globe-data.

2. Methods

2.1. Potential errors in GO Trees measurements

We assessed sources of error from user input in the GO Trees data. Previous studies have suggested the need for data quality assessments of citizen science data (e.g. Crall *et al* 2011). After launching the tool, we assessed variability in user-submitted height estimates through an analysis of a set of repeat measurements made by multiple users over a short period of time.

2.1.1. GO tree height accuracy testing

Because GO Trees offers an approach for measuring tree height that is significantly different from traditional field measurement techniques, we tested the accuracy across a range of angles and distances. In development of GO Trees, we assessed the sensitivity of each user input variable to error by comparing measurements of A, B, and d for eight trained users where both d and T_{tree} were known independently (table 1). Each of the eight users performed three supervised measurements, with angles and distances recorded and calculations compared to the known values.

For the two angles recorded, the standard deviation was calculated. The bottom angle recorded had the largest average standard deviation across users of 0.27 degrees (Kuhlmann *et al* 2021) estimated angle error measurements due to sensor differences across a large number of phones had a mean deviation of 0.05 degrees. As a result, a component of what is being attributed to user error may include sensor inaccuracies.

The true distance to the object was compared with the estimated distance, d, and an error in stride length was determined for each test user.

To assess variability across volunteers and the potential impact of user error on height measurements, we identified a cluster of data in which 76 unique users measured two trees using GO Trees. Photos and locations confirm that the group (a Queens, NY, high school Algebra class) measured two trees, A and B, between 6 and 19 October 2021. All students started their observations at approximately the same locations 16.4 m (tree A) and 16.8 m (tree B) away from the trees. We removed data submitted by duplicate user identification numbers to eliminate potential error in H_c introduced by a student using another's device without adjusting the user height. The dataset included six anomalously low height estimates indicative of user error, and these were also removed.

2.2. Considering GO tree height with ICESat-2 and airborne lidar

In order to inform the collection of tree height observations for comparison to satellite data, we performed two investigations. We compared GO Trees height data with ICESat-2 observations for an area in Greenbelt, MD. The second investigation used tree height data collected from airborne lidar data to compare to ICESat-2 observations in Grand Mesa, CO. The more dense and precise observations of airborne lidar data are ideal for comparison with space-based altimeters such as ICESat-2 (e.g. Ilangakoon *et al* 2021, Malambo and Popescu 2021). Comparing airborne lidar-measured tree heights with measurements from space-based lidar will enhance understanding of such tree measurements and enable us to make recommendations for GO Tree observers and researchers who are interested in using GO Tree observations with ICESat-2.

Several processed data products relevant to canopy heights are available from ICESat-2, including medium and high confidence along-track geolocated photons (ATL03) and along-track elevation profiles of terrain and canopy heights (ATL08). The ATL03 photons are comparable to traditional lidar returns in regard to their positional precision and density and are classified by signal vs background and in the case of terrestrial returns may include noise,

Table 1. Measurement and tree height errors for the test group. Standard deviations and averages are across 3 measurements per user.

	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	Average (abs)	Standard deviation
Bottom angle (β) stdev (degrees)	0.29	0.29	0.39	0.01	0.23	0.25	0.45	0.21	0.27	
Top angle (α) stdev (degrees)	0.35	0.13	0.27	0.23	0.14	0.047			0.20	
Average distance to tree error (m)	0.037	−0.305	1.982	1.494	−0.061	0.823	1.341	−0.534	0.823	0.939
Stride length average error per stride (m)	0.003	0.014	−0.098	−0.073	0.003	−0.012	−0.064	0.024	0.037	0.046
Average height error (m)	0.329	0.091	1.098	0.701	−0.061	−0.335	0.732	−0.305	0.457	0.524
% Error in tree height	4%	1%	14%	9%	−1%	−4%	9%	−4%	6%	7%

ground, and vegetation. The ICESat-2 data have a 6.5 m geolocation knowledge mission requirement (Neuenschwander and Magruder 2019). The ALT08 provides ground and a canopy surface posted at variable length scales relative to signal level, for each beam presented along-track. In cases with fewer than 50 signal photons in a 100 m segment, height is considered invalid and not provided. As this type of measurement averages vegetation height over a distance, we did not evaluate the relationship between ATL08 and GO- or airborne lidar-measured tree heights.

2.2.1. Greenbelt, Maryland

To collect citizen science comparison data, we identified an ICESat-2 track (Track 103, path gt1l, collected on 5 April 2019) through wooded public land on level ground in Greenbelt, MD. GO project staff took 21 tree measurements with GO Trees between April and December 2019. The trees measured were primarily oak, sweet gum, and Virginia pine. We compared these measurements to ATL03 data from the ICESat-2 track.

2.2.2. Grand Mesa, Colorado

Our second study area, Grand Mesa, was chosen based on available data from the NASA SnowEx campaign (Currier *et al* 2019). The study area is relatively flat terrain. The tree species within the ICESat-2 transect include Engelmann spruce, subalpine fir, and lodgepole pine.

Airborne lidar data were collected for the study area in September 2016, with an average density of approximately 3 points m^{-2}. The horizontal and vertical accuracy of the airborne data is 30 and 8 cm, respectively (Hojatimalekshah *et al* in review).

We extracted the ground surface and all above ground canopy heights to compare with ICESat-2 data and clipped the point cloud to 10 m on either side of the ICESat-2 track. We utilized ATL03 data from Track 1156, path gt2l, collected on 10 June 2020.

3. Results

3.1. GO Trees user variable error assessment

During development testing, all test users, except one, measured tree height to an accuracy better than 10% (the design requirement) with a standard deviation across users of 7%.

Based on these results and to extrapolate performance across a wide range of situations, we used the angle standard deviation of 0.27 degrees, and a step standard deviation of 0.046 m and calculated what the measurement error would be for 1×, 2× and 3× the standard deviation for a tree of 15 m tall, for a user at distances of 25, 50, 75, 100 and 200 steps away. We also calculated the impact a change in camera height of 3, 5, and 8 cm away from the calculated camera height (H_c) would have on the overall error.

The drawing (figure 3) below illustrates this set of cases.

Table 2 presents the results as percent error for the estimated tree height with each variable at 1×, 2× and 3× the standard deviation determined in table 1 for the distances of 25, 50, 75, 100 and 200 steps. For example, column two indicates that with a stride delta of 4.57 cm, the error in tree height measurement ranges from 5% to 6% as the user's distance increases from 25 to 200 steps.

The following may be interpreted from these results:

(a) The impact of user distance from the tree is moderate up to 100 steps. As the user reaches 200 steps, angle differences cause greater inaccuracies. For example, the top angle delta of 0.54 degrees causes a 2% error at 25 steps but increases to 11% at 200 steps. To reduce error, GO Trees warns the user if they exceed 50 steps.

(b) For all angles measured, and for camera height errors of up to 3 sigma (based on a test case),

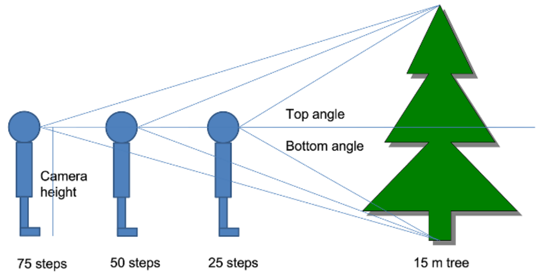

Figure 3. Same height tree observed from different distances.

Table 2. % error in tree height measurement for varying distance and measurement errors.

	Stride delta (cm)			Top angle (α) delta (degrees)			Bottom angle (β) delta (degrees)			Camera height delta (cm)		
Steps	4.57	9.15	13.72	0.27	0.54	0.81	0.27	0.54	.081	2.44	5.18	7.62
25	5%	11%	16%	1%	2%	2%	−1%	−1%	−2%	0%	0%	0%
50	5%	11%	16%	1%	2%	4%	−1%	−2%	−3%	0%	−1%	−1%
75	6%	12%	18%	2%	4%	6%	−1%	−3%	−4%	1%	0%	−1%
100	6%	12%	18%	2%	4%	6%	−3%	−5%	−8%	−1%	−2%	−3%
200	6%	12%	18%	6%	11%	16%	−3%	−8%	13%	2%	−1%	−3%

Table 3. % error in user input variables at varying distance to the tree and tree heights.

		Stride delta (cm)			α delta (degrees)			β delta (degrees)			Camera height delta (cm)		
Tree height	Steps	4.57	9.15	13.72	0.27	0.54	.081	2.44	5.18	7.62	0.27	0.54	.081
9	25	6%	11%	17%	1%	2%	3%	−1%	−2%	−3%	0%	0%	−1%
18	50	6%	12%	17%	1%	2%	3%	−1%	−2%	−3%	0%	−1%	−1%
27	75	6%	12%	18%	1%	2%	3%	−1%	−2%	−3%	0%	−1%	−1%
35	100	6%	12%	18%	1%	2%	3%	−1%	−2%	−3%	0%	−1%	−1%
70	200	6%	12%	18%	1%	2%	3%	−1%	−2%	−3%	0%	−1%	−1%

and up to a maximum of 100 steps, all scenarios indicate a maximum error of less than 10%.

(c) GO Trees is most sensitive to stride length. A one sigma stride length error remains within the 10% range, 2 sigma just exceeds, and 3 sigma shows significant errors. These can be mitigated by asking the user to validate the stride length estimated for them by the app.

(d) This analysis considers input errors individually but does not evaluate how combined errors (e.g. stride error plus camera height error) will affect overall height estimates. In practice, errors will most likely be combined for either improved or decreased accuracy.

Table 3 showcases results from this analysis indicate:

(a) Percent errors stay consistent as tree height increases and distance from tree increases.

(b) For all angle measurement errors and camera height errors, the tree height measurement remains very small (<3%).

(c) The sensitivity to stride length is again the largest factor, with the 2 sigma value just exceeding the goal of 10% error (12%).

3.2. GO Trees citizen science height measurement consistency field accuracy

To assess the consistency with which unsupervised volunteers used the app to measure tree height, we identified two trees measured by 76 unique users (see figure 4). For tree A, the average height measured across 38 users was 10.08 m with a standard deviation from the average of 1.42 m or 14% from the average. Tree B had an average height across 38 users of 14.92 m with a standard deviation of 2.27 m or 15% from the average.

As described above, testing among experienced, supervised volunteers showed no clear bias in the app with height errors below 10%. If we assume that the app measurement is unbiased and the average height of each tree is close to the actual height, a spread from the average greater than 10% may be attributed to user error. Fifty percent of the GLOBE tree A

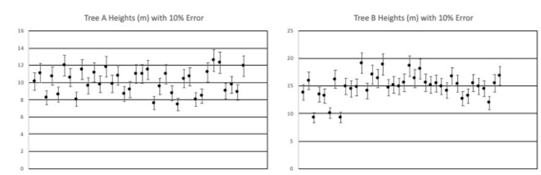

Figure 4. Tree heights, with 10% errors, reported for repeat GO measurements of two trees.

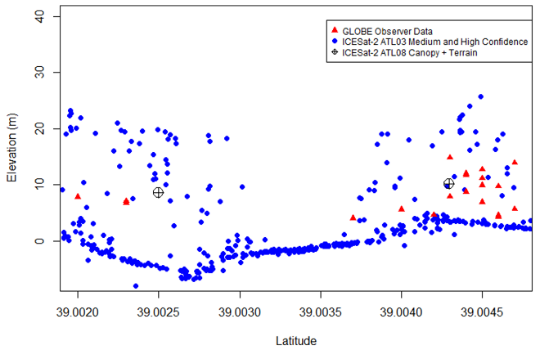

Figure 5. Transect showing GO Trees data (red) overlaid with ICESat-2 ATL03 medium and high confidence returns (blue), and ATL08 canopy heights (black).

measurements fell beyond 10% of the average height. Thirty-four percent of the measurements of tree B fell outside of the 10% range. The most common user errors are movement of the camera to different heights while measuring the angles to the top and base of the tree and altering pace length when walking to the tree. We recommend additional data collection of trees of known height by unsupervised volunteers to further test consistency.

3.3. Greenbelt, Maryland

We next compared GO Trees data to an ICESat-2 track in Greenbelt, Maryland, USA (see figure 5). Figure 6 plots the GO Trees data against the ATL03 (ATL08 data included for reference). While the GO height data are relatively sparse ($n = 21$) in comparison to the ICESat-2 data, the trend in elevations is similar.

It is not certain which tree corresponds to which ICESat-2 observation, if any, given the geolocation error of ICESat-2 returns (within 6.5 m) and GO Trees data, which ranged from 5 to 20 m for most measurements. These geolocation errors may account for perceived differences in tree heights; however, the observed correlation in height distributions indicates that citizen science data may have potential to serve as comparison data for ICESat-2.

3.4. Grand Mesa, Colorado

To better assess correlation of GO Trees height measurement data with ICESat-2 data, we matched ICESat-2 data with an independent airborne lidar

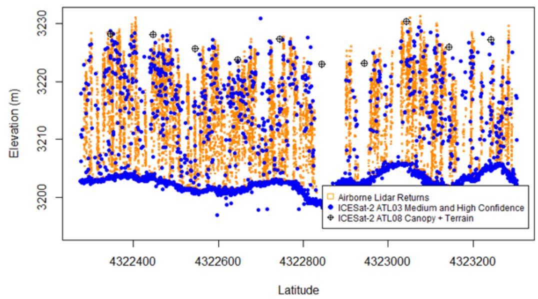

Figure 6. Transect showing airborne lidar returns (orange) overlaid with ICESat-2 ATL03 medium and high confidence returns (blue), and ATL08 canopy heights (black).

dataset with high-accuracy canopy and ground elevation measurements. We hypothesize that the airborne lidar would elicit an understanding of ICESat-2 canopy measurements, especially if ATL03 returns are representative of the height of a tree or a coarser measurement of canopy structure. The comparison for the example transect indicates that the ICESat-2 data captures similar canopy structure to the airborne lidar canopy, as well as similarly captures the ground surface. A Welch two-sample t-test was used to compare the distributions of measured elevations and found that the sample means were within 1.06 m (for canopy returns), and 0.83 m for (ground returns).

For examining the potential impacts of ICESat-2 locational uncertainty and the potential that the return was measuring parts of the non-top canopy heights that were not the top, we did two comparisons. First, when comparing the ICESat-2 returns with airborne lidar returns from a much smaller swath (1 m, representing the surface below the coordinates of the ATL03 data), the relationship of the canopy elevation of canopy returns was no longer significant; however, the t-test for ground returns was still statistically significant. This indicates that the ICESat-2 returns for canopies likely are measurements of tree canopies from within the 6.5 m locational geolocation uncertainty and within the approximately 13 m footprints, and may be possibly representing a fair an amount of non-top canopy.

The second comparison method examined the locational uncertainties of canopy measurements by comparing ICESat-2 and data with airborne lidar data returns at increasingly coarser scales in order to emulate the locational uncertainty of the GO Trees observations locations. Comparing the ICESat-2 canopy returns with rasterized versions of the airborne canopy returns at 1, 2, and 5 m pixel sizes, t-tests indicated that the relationship between the canopy elevations was significant at the 1 and 2 m scales (p-values <0.05) but not the 5 m scale (p-value $= 0.15$). This suggests that when the uncertainty of canopy height measurements is greater than 2.5 m (Nyquist rate), measurements that they may no longer correlate well with correspond to the ICESat-2 measured canopy values.

The relative horizontal uncertainty and pulse footprint size of ICESat-2 returns may be best mitigated by many GO Trees observations beneath ICESat-2 tracks, and by measuring distinct trees, notable in that they stand out by height or distance from other trees. The results of the spatial resolution and positional accuracy indicate that if the horizontal positional uncertainty of the GO Trees observations are greater than 2.5 m or the treetops are within 2.5 m from one-another, it is not advisable to use for ICESat-2 analysis comparing individual tree heights.

An analysis of the GO Trees data from the first three years of the app shows that 17.8% of the data points for which accuracy information is available, have a location accuracy of 10 m or better, while another 12.4% are between 11 and 20 m. The majority of the data points, 61.9%, have an accuracy between 21 and 100 m. As the location accuracy is stored with each observation, users of the data are able to select the data points with accuracy ranges appropriate to their needs. As an example, in the case of the data from Greenbelt, Maryland described above, 60% of the data with accuracy information available had an accuracy of 10 m or better, and fully 80% had an accuracy of 20 m or better.

4. Conclusions

4.1. Tree height accuracy testing conclusion

Within the GO Trees method of estimating tree height, stride length is the greatest source of error. A one sigma error in stride length still results in good performance (within 10%), while a 2 sigma stride length error provides adequate performance (around 10%). Any further increase in stride length error results in poor height estimates. Errors in stride length have a greater impact on the height estimate the farther the user is from the tree, and for this reason, users must be closer than 50 steps. Citizen scientists can significantly improve this accuracy by validating their stride length or measuring the actual distance to the tree rather than using the value that is calculated by the app.

In the sample of 8 test users with 3 repeat measurements against a known object of 7.9 m, 7 of 8 test users measured height within 10% accuracy with a standard deviation across all 24 measurements of 7%. The average of the 24 measurements demonstrated a 4% error in tree height.

Among citizen scientists in the field, app-based estimates of tree height are generally less accurate than other means but perform at their best for taller trees: we saw a 32% increase in accuracy in repeat measurements of a 14.9 m tree versus a 10.1 m tree. For these reasons, it is recommended that citizen scientists focus on measuring taller, more prominent trees.

In fortuitous harmony with other findings of this paper, GO Trees data are typically not collected in dense canopy, as accurate measurements require that the base and top of the tree are visible, and that the observer is able to walk to the tree in a straight line with no obstacles.

4.2. Using GO Trees data with ICESat-2

Hence, we suggest that for the purposes of supporting spaceborne elevation data with citizen-science observations, GO Trees measurements are beneficial: (a) when able to be measured with elevational precision and horizontal accuracy, that tree height measurements be taken of trees within a stand of similar-height trees, and (b) when able to be measured with elevational precision but without decent horizontal accuracy (toward the higher end of the average 15 m; Kohl *et al* 2021), that the citizen-scientist measure prominent or otherwise identifiable trees such as those that are significantly higher than their neighbors, or those that stand alone.

ICESat-2 has the ability to measure the canopy tops of individual trees to dense stands of trees. GLOBE volunteers may not be able to use the GO app to measure tree heights in areas with dense patches of trees because they lack a clear line of sight and walking path to the tree from the initial point of observation.

We suggest that ATL03 is more comparable (than ATL08) to tree height measurements of high geopositional accuracy, as it is more directly comparable to single-tree measurements as ATL08 is a derived product along ICESat-2 tracks (Neuenschwander and Pitts 2019). However, ALT08 may also be appropriate when larger-area measurements of similar age-class (cohort) trees are measured, regardless of horizontal accuracy.

Based on our research, our specific, ranked recommendations for maximum impact and applicability of the GO Trees data with other global-scale measurements such as ICESat-2 include:

(a) Align the relative geopositional accuracies (e.g. GLOBE tree height and beneath ICESat-2 tracks).
(b) That citizen scientists maximize height accuracy by following the instructions in GO Trees. Specifically, volunteers should see both the tree top and bottom, use a normal gait to walk in a straight, obstacle-free line from the observation point to the tree or measure the distance with a tape measure, select trees taller than 5 m to measure, and hold the camera at a steady height to measure the angles to the top and bottom of the tree.

(c) Such measurements observe prominent or identifiable trees (such as large, isolated trees, or trees that stand visibly above surrounding trees). Ideally, such measurements would have a location uncertainty within 2.5 m.
(d) If geolocation is less reliable or there are no trees meeting criteria (c), that accurate measurements of multiple trees within a larger homogeneous cohort be taken; such measurements may be comparable with both ICESat-2's ATL03 and ATL08 data.

We note that these observations are a result of significant experience with GO data (Dodson *et al* 2019, Amos *et al* 2020, Colón Robles *et al* 2020, Kohl *et al* 2021, Low *et al* 2021) and analyses between airborne lidar, ICESat-2, and GO Trees data limited to this study. Additionally, while smart phone GPS measurements are improving with most being at best accurate to within 4 m, an analysis of GO data found an average location accuracy of 14.5 m. Given the 13 m ICESat-2 photon pulse footprint, this means that GO Trees measurements within the footprint may not be resolvable or may be misleading. Similar results have been found in other studies (Li *et al* 2020, Nandy *et al* 2020), but additional studies should expand the types and amount of tree cover.

From a larger perspective, GLOBE tree height data, along with tree circumference data, may allow NASA and other scientists to extrapolate biomass which can inform calculations of the carbon that trees and forests either absorb or release into the atmosphere. The potential combination of GLOBE citizen-science tree height data and ICESat-2 tree canopy data can help build a much more robust dataset for scientific research of biomass and carbon sequestration by trees, as well as other globally scalable observations such as forest recovery, reaction to climatic or environmental stress, or other large processes.

While location accuracy within the app may limit one-to-one comparisons, ultimately the goal would be to match GLOBE data and other height data, particularly data from space-based missions and instruments. To this end, a new geofencing function is being added to the GO app that can direct users to measure tree heights at specified locations. The app has a similar time and location-based function that alerts users to take a cloud observation when a satellite is overhead (Hayden *et al* 2019), resulting in nearly one million citizen science observations matched to satellite data. The new geofencing function notifies users of locations where data are desired, providing a mechanism for collecting data based on a statistical sample.

Perhaps our most fundamental recommendation, in the interest of making citizen-science data collection more approachable to all, is that one takes a moment to appreciate a favorite or special tree through contemplative mensuration.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

The authors would like to thank NASA and The GLOBE Program and all the citizen scientists who have taken vital tree height observations with the GLOBE Observer. We would especially like to thank the 130 9th grade Algebra students at the Monsignor McClancy Memorial High School in Queens, New York for their exemplary tree height observations and Dr Christopher Shuman for his help on this paper. GLOBE Observer funding falls under NASA Award Number NNX16AE28A for the NASA Earth Science Education Collaborative. The GLOBE Program is funded by NASA with support from the National Oceanic and Atmospheric Administration, the National Science Foundation, and the US Department of State.

ORCID iD

Brian A Campbell https://orcid.org/0000-0003-3477-7371

References

Amos H M *et al* 2020 GLOBE Observer data: 2016–2019 *Earth Space Sci.* **7** e2020EA001175
Butler D M and MacGregor I D 2003 GLOBE: science and education *J. Geosci. Educ.* **51** 9–20
Campbell B 2021 ICESat-2 and the trees around the GLOBE student research campaign: looking at Earth's tree height, one tree at a time *Acta Astronaut.* **182** 203–7
Colón Robles M, Amos H M, Dodson J B, Bouwman J, Rogerson T M, Bombosch A, Farmer L, Burdick A, Taylor J E and Chambers L H 2020 Clouds around the world: how a simple data challenge became a worldwide success *Bull. Am. Meteorol. Soc.* **101** E1201–13
Crall A W, Newman G J, Stohlgren T J, Holfelder K A, Graham J and Waller D M 2011 Assessing citizen science data quality: an invasive species case study *Conserv. Lett.* **4** 433–42
Currier W R *et al* 2019 Comparing aerial lidar observations with terrestrial lidar and snow-probe transects from NASA's 2017 SnowEx campaign *Water Resour. Res.* **55** 6285–94
Dodson J B, Colón Robles M, Taylor J E, DeFontes C C and Weaver K L 2019 Eclipse across America: citizen science observations of the 21 August 2017 total solar eclipse *J. Appl. Meteorol. Climatol.* **58** 2363–85
Dujardin S *et al* 2022 Mapping abundance distributions of allergenic tree species in urbanized landscapes: a nation-wide study for Belgium using forest inventory and citizen science data *Landscape Urban Plan.* **218** 104286
Finarelli M G 1998 GLOBE: a worldwide environmental science and education partnership *J. Sci. Educ. Technol.* **7** 77–84
Global Learning and Observations to Benefit the Environment (GLOBE) 2019 *Data User Guide, Version 1.0* (available at: www.globe.gov/globe-data/globe-data-user-guide) (Accessed 18 November 2021)
GLOBE 2019 About GLOBE *Global Learning and Observations to Benefit the Environment*

(available at: www.globe.gov/about/overview) (Accessed 18 November 2021)
Hayden L, Taylor J and Colon Robles M 2019 GLOBE: connecting to community of observers directly to NASA satellites *IEEE Trans. Geosci. Remote Sens.* **7** 98–99
Hojatimalekshah A, Gongora J, Enterkine J, Glenn N F, Caughlin T and Marshall H P Lidar and machine learning reveal forest structural controls on snowpack *Front. Ecol. Environ.* in review
Hunter M O, Keller M, Victoria D and Morton D C 2013 Tree height and tropical forest biomass estimation *Biogeosciences* **10** 8385–99
Ilangakoon N, Glenn N F, Schneider F D, Dashti H, Hancock S, Spaete L and Goulden T 2021 Airborne and spaceborne lidar reveal trends and patterns of functional diversity in a semi-arid ecosystem *Front. Remote Sens.* **2** 74330
Koch G *et al* 2004 The limits to tree height *Nature* **428** 851–4
Kohl H A *et al* 2021 GLOBE observer and the GO on a trail data challenge: a citizen science approach to generating a global land cover land use reference dataset *Front. Clim.* **3** 620497
Kuhlmann T, Garaizar P and Reips U 2021 Smartphone sensor accuracy varies from device to device in mobile research: the case of spatial orientation *Behav. Res. Methods* **53** 22–33
Li W, Niu Z, Shang R, Qin Y, Wang L and Chen H 2020 High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data *Int. J. Appl. Earth Obs. Geoinf.* **92** 102163
Liu A, Cheng X and Chen Z 2021 Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals *Remote Sens. Environ.* **264** 112571
Low R, Boger R, Nelson P and Kimura M 2021 GLOBE mosquito habitat mapper citizen science data 2017–2020 *GeoHealth* **5** e2021GH000436
Malambo L and Popescu S C 2021 Assessing the agreement of ICESat-2 terrain and canopy height with airborne lidar over US ecozones *Remote Sens. Environ.* **266** 112711
Markus T *et al* 2017 The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): science requirements, concept, and implementation *Remote Sens. Environ.* **190** 260–73
Molinier M, López-Sánchez C A, Toivanen T, Korpela I, Corral-Rivas J J, Tergujeff R and Häme T 2016 Relasphone—mobile and participative *in situ* forest biomass measurements supporting satellite image mapping *Remote Sens.* **8** 869
Nandy S, Srinet R and Padalia H 2020 Mapping forest height and aboveground biomass by integrating ICESat-2, Sentinel-1 and Sentinel-2 data using random forest algorithm in northwest Himalayan foothills of India *Geophys. Res. Lett.* **48** e2021GL093799
Narine L L, Popescu S C and Malambo L 2020 Using ICESat-2 to estimate and map forest aboveground biomass: a first example *Remote Sens.* **12** 1824
Neuenschwander A L and Magruder L A 2019 Canopy and terrain height retrievals with ICESat-2: a first look *Remote Sens.* **11** 1721
Neuenschwander A and Pitts K 2019 The ATL08 land and vegetation product for the ICESat-2 mission *Remote Sens. Environ.* **221** 247–59
Potapov P *et al* 2021 Mapping global forest canopy height through integration of GEDI and Landsat data *Remote Sens. Environ.* **253** 112165
Roman L A, Scharenbroch B C, Östberg J P, Mueller L S, Henning J G, Koeser A K, Sanders J R, Betz D R and Jordan R C 2017 Data quality in citizen science urban tree inventories *Urban For. Urban Greening* **22** 124–35
Sintayehu D W, Belayneh A and Dechassa N 2020 Aboveground carbon stock is related to land cover and woody species diversity in tropical ecosystems of Eastern Ethiopia *Ecol. Process.* **9** 1–10
Sun T, Qi J and Huang H 2020 Discovering forest height changes based on spaceborne lidar data of ICESat-1 in 2005 and ICESat-2 in 2019: a case study in the Beijing-Tianjin-Hebei region of China *For. Ecosyst.* **7** 1–12