First validation of GEDI canopy heights in African savannas

Remote Sensing of Environment

First Validation of GEDI Canopy Heights in African Savannas

--Manuscript Draft--

Manuscript Number:	RSE-D-22-00262R1
Article Type:	Research Paper
Section/Category:	Vegetation structural parameters
Keywords:	lidar; GEDI; vegetation structure; savannas; canopy height; validation; Africa.
Corresponding Author:	Konrad Wessels, Ph.D. George Mason University Fairfax, Virginia UNITED STATES
First Author:	Xiaoxuan Li
Order of Authors:	Xiaoxuan Li
	Konrad Wessels, Ph.D.
	John Armston, Ph.D.
	Steven Hancock, Ph.D.
	Renaud Mathieu, Ph.D.
	Russell Main, Ph.D.
	Laven Naidoo, Ph.D.
	Barend Erasmus, Ph.D.
	Robert Scholes, Ph.D.
Abstract:	Savannas have complex, discontinuous woody vegetation structures that vary greatly in vertical and spatial arrangement and change due to climatic, ecological and management impacts. While airborne laser scanning (ALS) data have provided detailed information on vertical vegetation structure and is widely used in ecological studies, it is lacking in availability and repeat frequency. Although the Global Ecosystem Dynamics Investigation (GEDI) waveform Light Detection and Ranging (LiDAR) sensor and algorithms were optimized for measuring dense forests, it was anticipated that GEDI metrics could provide useful characterization of lower stature, sparse savannas structures. This study provided the first baseline validation of Version 2 GEDI (L2A) relative height 98 (RH98) by comparing the on-orbit GEDI-RH98 orb to the simulated GEDI-RH98 sim derived from ALS data across diverse savanna vegetation. It furthermore determined the influence of various factors on error, e.g. algorithm setting group (SGs), beam type, day vs. night, beam sensitivity, and vegetation phenology. After applying quality flags, 22813 GEDI footprints were analyzed across 11 sites. SGs 4-6 that are aimed at dense forests had much larger errors than SGs 1-3. The phenological conditions at the time of GEDI data acquisition had a very large influence on the error of RH98 orb . During leaf-on conditions for savanna vegetation with RH98 sim < 15 m, RH98 orb was very accurate with R^2 = 0.61, mean bias = -0.55 m, %bias = -11.1%, RMSE = 1.64 m and %RMSE = 29.8%. In leaf-off conditions where RH98 sim < 15 m, RH98 orb was less accurate with R^2 = 0.43, mean bias = -1.47 m, %bias = -26.5%, RMSE = 2.03 m and %RMSE = 40.9%. During leaf-off conditions, the GEDI LiDAR signal at the start of the waveform may be weaker as it interacts with denuded branches and may be truncated as noise, leading to a large negative height bias. Therefore, assessments of deciduous vegetation structures should be conducted during leaf-on periods. In leaf-on conditions, GEDI's RH98 orb was very accurate between canopy heights of 3 and 7 m, with a mean bias of -0.79 m (-10%). The bias of RH98 orb was not influenced by canopy cover. Due to the GEDI LiDAR pulse width of 15.6 ns, the GEDI-RH98 data product cannot reliably estimate canopy heights of shrubs below 2.34 m and will require more complex deconvolution of the waveform. GEDI's RH98 accurately estimates the canopy height of trees between 3 and 15 m allowing assessment of canopy heights over vast savanna areas.

Suggested Reviewers:	Niall Hanan, Ph.D. New Mexico State University nhanan@ad.nmsu.edu
	Jan Van Aardt, Ph.D. Rochester Institute of Technology vanaardt@cis.rit.edu
	Nancy Glen, Ph.D. Boise State University nancyglenn@boisestate.edu
	Shaun Levick, Ph.D. Commonwealth Scientific and Industrial Research Organisation Shaun.Levick@csiro.au

Dear Editor,

Thank you for accepting our manuscript with major revisions. We believe it is now ready for publication in RSE and that it will appeal to readers who are interested in using GEDI data for characterizing vegetation structure in biomes other than dense tropical forests. Reviewer #3 provided several comments which we addressed that improved the manuscript. The review was generally positive, but the reviewer ended up rejecting the manuscript due to a perceived lack of novelty. We now present clear evidence of novelty in the rebuttal and with additions to the text to ensure we communicate this to the reader. It is evident that reviewer #3 is eager to see results on the ecological application of GEDI, and while we share this sentiment and are actively engaged in such research, we argue that a paper on the validation of GEDI in sparse, short stature savannas is fully justified at this point in time. We furthermore present evidence and an explanation of the lower limits of canopy height estimation with GEDI (as it relates to LiDAR pulse width) which is very important to ecological studies involving shrubs and bush encroachment in savannas. We presented sufficient evidence to demonstrate novelty in line with the RSE's stated scope and objectives as it relates to remote sensing. It is furthermore evident that the Reviewer#3 did not fully understand or appreciate the value of simulating the GEDI waveforms from ALS point clouds for validation, as demonstrated by several references to the canopy height model (CHM) comparisons, which were not used in this manuscript. This may have influenced the reviewer's general impression of the manuscript.

Below is a summary of all the changes made to address comments by reviewer #3:

1. Additional details on settings of gediSimulator and CHM generation were provided
2. The relative influence of remaining 0-2 m random geolocation error was simulated. These simulations and sensitivity analysis took considerable time, and the details are explained in supplementary material and rebuttal. In summary, the results suggest that random geolocation errors of 0-2 m GEDI contribute 14% of absolute bias of RH98 < 8 m and 20.7% of RMSE of the RH98. This is a valuable addition to the results, especially in savannas with very high spatial diversity in vegetation height that may vary over short distances.
3. Figure 1 of the study area was completely revised and much improved.
4. The suggested comparison of evergreen sites in dry and wet season was not possible as the evergreen sites do not have a distinct dry season as illustrated by their climate zones characteristics. We therefore are unable to tease out seasonal effects of moisture conditions on the validation which the reviewer seems to hint at.
5. The explanation on the potential influence of phenological condition of ALS data was improved with additional text.
6. Minor edits were addressed
7. As suggested by the reviewer we produced a very useful visualization of ALS point cloud data inside multiple GEDI footprints and added it to supplementary material.
8. Finally, as requested, we discussed the limitations of GEDI which essentially samples the landscape and does not produce high resolution continuous surfaces. We briefly explain the appropriate use of GEDI data to address specific regional ecological questions, which is the subject of our on-going research.

Reviewer #4 was very positive and suggested only minor edits which were all implemented.

Overall, we believe that the additions improved the paper, that the novelty is now clearly stated and that it is ready for publication.

Kind regards

Konrad Wessels

Rebuttal and itemized reply to Reviewer's comments for RSE-D-22-00262 [220802-022158]

Reviewer #3 provided several comments that we addressed and that improved the manuscript. Reviewer #4 provided only minor edits, which were all addressed.

Below is a summary of all the changes made to address comments by reviewer #3:

1. Additional details on settings of gediSimulator and CHM generation were provided
2. The relative influence of remaining 0-2 m random geolocation error was simulated. These simulations and sensitivity analysis took considerable time, and the details are explained in supplementary material and rebuttal. In summary, the results suggest that random geolocation errors of 0-2m GEDI contribute 14% of absolute bias of RH98 < 8 m and 20.7% of RMSE of the RH98. This is a valuable addition to the results, especially in savannas with very high spatial diversity in vegetation height that may vary over short distances.
3. Figure 1 of the study area was completely revised and much improved.
4. The suggested comparison of evergreen sites in dry and wet season was not possible as the evergreen sites do not have a distinct dry season as illustrated by their climate zone characteristics. We therefore are unable to tease out seasonal effects of moisture conditions on the validation which the reviewer seems to hint at.
5. The explanation on the potential influence of phenological condition of ALS data was improved with additional text.
6. Minor edits were addressed
7. As suggested by reviewer we produced a very useful visualization of ALS point cloud data inside multiple GEDI footprints and added it to supplementary material.
8. Finally, as requested, we discussed the limitations of GEDI which essentially samples the landscape and does not produce high resolution continuous surfaces. We briefly explain the appropriate use of GEDI data to address specific regional ecological questions, which is the subject of our on-going research.

Reviewer #4 was very positive and suggested only minor edits that were all implemented.

Overall, we believe that the additions constituted minor revisions which improved the paper, that the novelty is now more evident and that it is ready for publication.

After each numbered comment by a reviewer, our reply is indicated by ">>>". Newly inserted text is given in Bold along with the approximate page number (Page + page number) and line number (Line + line number) in the edited manuscript (including track changes).

Reviewer #3:
This paper compared the heights of savanna vegetation measured with the relatively new GEDI satellite, with heights derived from well-established ALS. Vegetation height is particularly relevant in savannas, because ecological processes affect tree size which is not necessarily detected with established woody cover products. The authors found good agreement between canopy top (RH98) measured with GEDI, and RH98 modelled with ALS derived CHM. Importantly, they found that the accuracy was seasonal - with much stronger agreements in leaf-on conditions compared to leaf-off conditions.

This paper contributes to an important need in savanna vegetation monitoring: scaling-up vegetation cover estimates to better manage these dynamic systems. This was a very well written paper with a clear scope and interesting results. The methods and results were clearly presented and were generally easy to follow.

>>We thank the reviewer for these insights and comments on the quality of the paper.

The major weakness of the paper is a **lack of novelty - it appears to be applying existing tools (developed for forests) in a new environment.** A potential, significant expansion of the paper would be to flesh out the **ecological insights** perceivable from this new dataset. Given that the major weakness of ALS is considered to be "very limited aerial coverage, precluding regional monitoring of vegetation structure" (Line 80), what interpretations can be **made at broader scales that can't be seen at the smaller scales of ALS**? Does ALS under- or over-estimate cover, lead to incorrect interpretation of processes due to a lack of spatial coverage? **Are there any other insights that can be gleaned from the regional GEDI coverage? (touch in Discussion)**

>>The novelty of the research was not sufficiently explained in the paper, but we believe that the below rationale and added text exhibits sufficient novelty for publication in RSE at this point in time. GEDI was launched relatively recently and is the first space-based laser specifically designed for measuring vegetation structure. As with any other new sensor, it is common practice that the reliability of GEDI's on-orbit measurements first be validated before it can be relied upon for environmental insights (e.g. regional ecological studies), which is a subject of our on-going research under a NASA-funded project. The reviewer rightly points out that ecological application of GEDI across savannas would be very interesting, but this was not the topic of this paper. We would argue that a study on the validation of GEDI canopy height measurements in short, sparse savannas is fully justified at this point in time and should precede regional ecological studies based on GEDI, especially since scientists do not yet know how reliable the GEDI measurements are in savannas. Our study's novelty is based on (i) its overall timing given the fairly recent launch of GEDI (ii) the first validation study on date of submission (Feb 2022) to simulate GEDI waveforms from ALS for direct validation of GEDI's on-orbit relative height measurements (only one other paper, Wang et al. 2022, has been published recently using this approach - May 2022), (iii) the study provides detailed insights into the causes of the lower limit

of detectability of short stature shrubs that are essential to ecological studies in savannas, and (iv) the validation of GEDI canopy height measurements in savannas, despite the fact that GEDI sensor and algorithms were designed for dense, tall forests, as stated in the mission objectives.

While a few recent papers compared GEDI RH98 to canopy height models derived from ALS canopy height models (CHM), such comparisons are not direct validation of GEDI height metrics, since they essentially compare different structural metrics that may confound interpretation. We, therefore, simulated RH98 from ALS point cloud data to provide a more comprehensive and reliable validation of the on-orbit GEDI RH98. Using the simulation approach for validation of GEDI is novel in itself as it enabled the relative importance of different measurement errors (i.e. signal start detection, geolocation) to be assessed. It is also, to our knowledge, only the second paper to do so (after Wang et al. May 2022), irrespective of the biome of application, and the first to be focused specifically on savanna woody vegetation. The claim by the reviewer that we are "*applying existing tools (developed for forests) in a new environment"* is therefore not completely justified. However, we acknowledge that we should have provided more text in the manuscript to highlight the novelty of the study, which is addressed below. The paper furthermore makes an important contribution by validating and explaining the potential limits of detectability of very low shrub vegetation (< 3 m) by GEDI due to its pulse width, which is especially relevant to ecological insights in savannas.

The following text (in bold) was therefore added to highlight the novelty and importance of the paper.

Line 93-100: The GEDI pulse length (FWHM 15.6 ns) is short enough to vertically discriminate canopy and ground returns in forested ecosystems, however the ability to characterize short stature, discontinuous vegetation has not been a design requirement for GEDI **and therefore needs to be evaluated before applying it to regional ecological studies**. While the GEDI instrument design, algorithm calibration and validation were focused on the measurement of tall, dense, and continuous vegetation typical of tropical and temperate forests, it is anticipated that it could also provide useful, accurate characterization of lower stature, discontinuous savanna vegetation. **This was the first study to our knowledge to validate GEDI canopy height measurements in savanna vegetation.**

Line 111-114: The present study, therefore, **followed the novel approach of first simulating GEDI waveforms from the ALS point cloud data** before extracting of ground elevation and RH metrics (Hancock et al., 2019; Silva et al., 2018) to generate an ALS-based reference dataset before performing an error assessment **(Wang et al. 2022)**.

Line 117: This will provide the **first baseline assessment** of the performance of GEDI-RH98 products in African savannas, which will enable the identification of research priorities to expand the applicability of GEDI canopy height products to non-forest vegetation.

Line 545-547: **This study provided crucial insight into the lower limits of canopy height detection that impacts the measurement of shrub cover which is essential to ecological studies and biomass estimation in savannas** (Mograbi et al., 2015; O'Connor et al., 2014; Stevens et al., 2016; Venter et al., 2018) and may require different approaches to waveform

interpretation and estimation of other GEDI metrics, e.g., canopy cover fraction, plant area index.

Line 560: Conclusion: This study provided the **first** baseline validation of Version 2 GEDI RH98 canopy height estimates in African savannas **by simulating the GEDI waveform from ALS data**. The results provide valuable insights into the accuracy and precision of GEDI-RH98 across highly heterogeneous, short stature vegetation.

Existing text highlighting novelty:

Line 19-22: Although the Global Ecosystem Dynamics Investigation (GEDI) waveform Light Detection and Ranging (LiDAR) sensor and algorithms were **optimized for measuring dense forests**, it was anticipated that GEDI metrics could provide useful characterization of lower stature, sparse savannas structures. This study provided the **first baseline validation** of Version 2 GEDI (L2A) relative height 98 (RH98) by comparing the on-orbit GEDI-RH98orb to the simulated GEDI-RH98sim derived from ALS data **across diverse savanna vegetation**.

Finally, on the topic of novelty, the Aims and Scope of the RSE journal states that "*Original Research Articles describe important* ***significant new results or methods that will advance the science or application of remote sensing****. The* ***main contribution should be the remote sensing component****,* ***rather than investigation of an environmental problem*** *in which remote sensing data or techniques do not play a major role*." Given all the above evidence, we argue that our methods and results are sufficiently novel to justify publication in RSE at this point in time, especially considering that the GEDI sensor was actually designed for dense, tall forests. Our study does however raise an important caveat that low shrub vegetation below 3m cannot be accurately characterized with current GEDI Level 2 data products, which is very important in savannas.
This validation lays the foundation for regional ecological application in savannas, which ourselves and other researchers can now pursue with confidence.

Further comments below (Reviewer #3).

1. There is limited detail in some of the methods steps.

i.e. "The gediSimulator software was used to simulate GEDI full-waveform from the ALS data, collocate these with GEDI waveforms…" While I am not fully familiar with the software, I assume there would be parameter tuning that may be of consequence and worth mentioning.

>>> To emphasize the settings of the gediSimulator we added clear numbering, additional description and one additional setting (v). No additional fine-tuning was required beyond these settings.

Additions to Section 2.3 Data Processing, Page 13, Line 243-254: The gediSimulator software (Hancock et al., 2019) is comprised of collocateWaves and gediMetric programs which were used to simulate a GEDI full-waveform from the ALS data, collocate these with recorded GEDI waveforms (Fig. S1), and create a series of simulated vegetation metrics (RH, canopy cover, etc.) (Hancock et al., 2019) (Fig. 2). The collocateWaves program was implemented to simulate GEDI waveforms from ALS data and collocate the simulated

GEDI waveforms with recorded GEDI waveforms (Fig. S1). The parameter settings in the collocateWaves program were as follows, (i) the minimum GEDI beam sensitivity was set to 0.9 (ii) the minimum ALS point density was set to 3 points per/m2, (iii) the on-orbit GEDI pulse shapes were derived from the GEDI Level 1B data for each beam and were used instead of the default Gaussian shape, (iv) the search window used to match ALS-derived and GEDI waveform was set to 20 m x 5 m to handle heterogeneous vegetated areas, (v) the simulated GEDI waveforms were generated without Gaussian noise. To resolve the systematic geolocation errors, the collocateWaves program was applied independently to all GEDI tracks. The simulated GEDI RH metrics generated in collocateWaves program were then extracted from the simulated GEDI waveforms by applying gediMetric program (Hancock et al., 2019).

1b. Another potentially important example: "Digital terrain models (DTM), digital surface models (DSM), and canopy height models (CHM) were generated from the ALS data at 1 m resolution, using lidR package (Roussel et al., 2018) in R." The settings used to generate the CHM could have an impact on the comparison. For example, if the local maximum (highest point) is used, this will produce a CHM that is generally higher than if other approaches are used (i.e. 80th quantile, smoothing etc). It would be worth, at minimum, documenting the workflow used in lidR. Better still, investigating the impact of these settings on the final result would be interesting - if smoothing is used, CHM will probably be lower, and the negative bias persistent in the comparison may be reduced.

>>> Firstly, note that the ALS CHMs were **not** used in any direct comparisons with the GEDI data during validation, including the bias calculation the reviewer pointed out. The CHMs were only used to generate the descriptive statistics on cover, slope and mean height of the ALS sites given in Table 1. To clarify this, the following was added Line 170-172: **"Note that in contrast to other recent studies (Liu et al., 2021; Potapov et al., 2021), the CHMs were not used in any of the comparisons with $RH98_{orb}$, but were only used to generate the vegetation structure metrics (cover and mean height) of sites in Table 1."**
>>> We included the following detail on the settings used to generate the CHM as requested by reviewer:
Line 167-169: **The Grid_canopy (lidR package) function was used to generate CHMs at 1m resolution using Digital Surface Model Algorithm (p2r) without any smoothing or filtering.**

2. Line 418-420. It would be interesting to test the impact of random errors on the relationship between ALS modelled and GEDI height based on random error. A sensitivity analysis in this regard would add to the paper - by determining how much the strength of the relationship relies on geometric accuracy.

>>> The comment refers to the following text in the Discussion: "**This study only accounted for systematic geolocation error of the GEDI tracks, while random geolocation errors for individual footprints remain in the data, contributing to the overall error in on-orbit and simulated GEDI comparisons.**"

We consulted with members of the GEDI Science Team on this topic. The random jitter on the GEDI instruments is believed to be 10% of the beam divergence which translates to an estimated

0-2 m geolocation error, but this has not been sufficiently quantified or published. In general, this random geolocation error is believed to be very small in comparison with other sources of error, but we suspect that the impact on our results might be non-trivial, especially in diverse savannas. We devised methods to simulate the impact of the random geolocation error and calculated its relative contribution to the overall error estimates. The methods and the results (including 3 figures) require significant text and space. Given that this is already a long manuscript with a large number of figures, we decided to present this work in the Supplementary Material (SM) and only summarize the findings in the Discussion (Line 452-460), especially since this is not a core result. The methods and detailed results are nevertheless interesting and thus included in the SM. In summary, the results suggest that random geolocation errors of 0-2 m GEDI contribute 14% of absolute bias of RH98 < 8 m and 20.6% of RMSE of the RH98. Therefore 14-20% of the error can be attributed to random, 0-2 m mismatches in location of the simulated and on-orbit RH98, rather than error in vertical canopy height estimates.

Please allow us to first recap how geolocation error is addressed. Version 2 GEDI data products have a geolocation uncertainty of ~11 m (1 sigma). The collocateWaves program in gediSimulator was used to correct the systematic error component of this geolocation uncertainty for each GEDI track. The collocation process moves the search window of 20 m x 20 m and matches the on-orbit to the simulated GEDI waveforms with the highest Pearson's correlation coefficient. The corrected coordinate offsets can therefore be determined and applied to each on-orbit GEDI tracks based on the corrected footprint locations. Even though the gediSimulator can solve this systematic geolocation error of each GEDI track, random geolocation errors (estimated to be 0-2 m) of individual GEDI footprints remain, contributing to overall uncertainty in the comparison between on-orbit and simulated GEDI metrics. The reviewer was asking if we could quantify the contribution of this remaining random geolocation error to the overall error and we used the gediSimulator to add random geolocation error and determine its contribution to the overall estimated error.

>>> All of the following text and figures were added to the Supplementary Material.

Simulating the contribution of random geolocation error to $RH98_{orb}$ error

The gediSimulator can only estimate the systematic geolocation error of each GEDI track, while random geolocation errors (0-2 m) of individual GEDI footprints remain, contributing to overall uncertainty in the comparison between on-orbit and simulated GEDI RH98. The relative contribution of the random geolocation error to the overall error of $RH98_{orb}$ was estimated as described below.

First, an ensemble of RH98 estimates were simulated within the extend of the random geolocation error. This was computed by: (i) 100 leaf-on GEDI footprints were selected from all savanna sites; (ii) for each footprint, 100 offset values ranging from -2 to 2 m were randomly and uniformly generated in x and y direction and added to the corrected GEDI footprint coordinates to generate 100 relocated coordinates for each GEDI footprint; (iii) the gediSimulator (gediRat program) was used to simulate GEDI waveforms from the ALS data for both the corrected and relocated GEDI footprint coordinates; and (iv) these simulated waveforms were then used as inputs in

gediSimulator programs (gediMetric) to produce relocated simulated GEDI RH98 ($GEDI\ RH98_{Sim\_R}$).

To quantify the difference between the corrected ($GEDI\ RH98_{Sim\_C}$) and relocated ($GEDI\ RH98_{Sim\_R}$) simulated GEDI RH98 at each footprint, the mean absolute difference (MAD) between these two RH98's was calculated using equation (1):

$$MAD = \frac{\sum_{i=0}^{n=100} |GEDI\ RH98_{Sim\_R} - GEDI\ RH98_{Sim\_C}|}{n} \quad (1)$$

The distribution of the MAD values of 100 random relocations at each simulated RH98 for 100 GEDI footprints is given in Fig. S5:

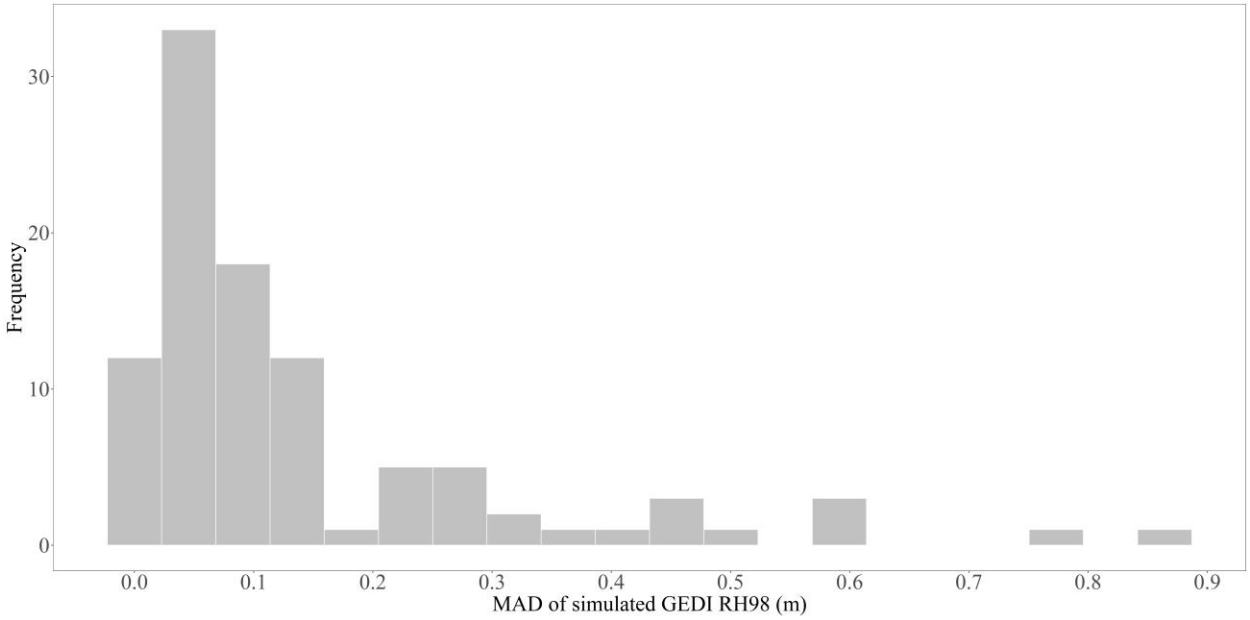


Fig. S5. Mean absolute difference (MAD) between relocated and corrected simulated GEDI RH98 for 100 random relocations of each of 100 GEDI footprints selected at random from all study sites (leaf-on only data).

Overall, 58% of the total MAD values are below 0.1 m, and the mean MAD values was 0.14 m (STD = 0.17). The majority of MAD data points (2 sigma, > 95%) fell within the range between 0 to 0.31 m. The percentage contribution of the MAD of random geolocation error to the bias of GEDI RH98 of individual footprints was estimated using the equation (2):

$$MAD_P = \frac{MAD}{|GEDI\ RH98_{Obs} - GEDI\ RH98_{Sim\_C}|} \times 100 \quad (2)$$

Where the MAD_P refers to the percentage contribution of random geolocation error (MAD) to the absolute bias of GEDI RH98 per GEDI footprint. $GEDI\ RH98_{Obs}$ denotes the on-orbit GEDI RH98. The MAD_P values were plotted for various RH98 height bins in Fig. S6:

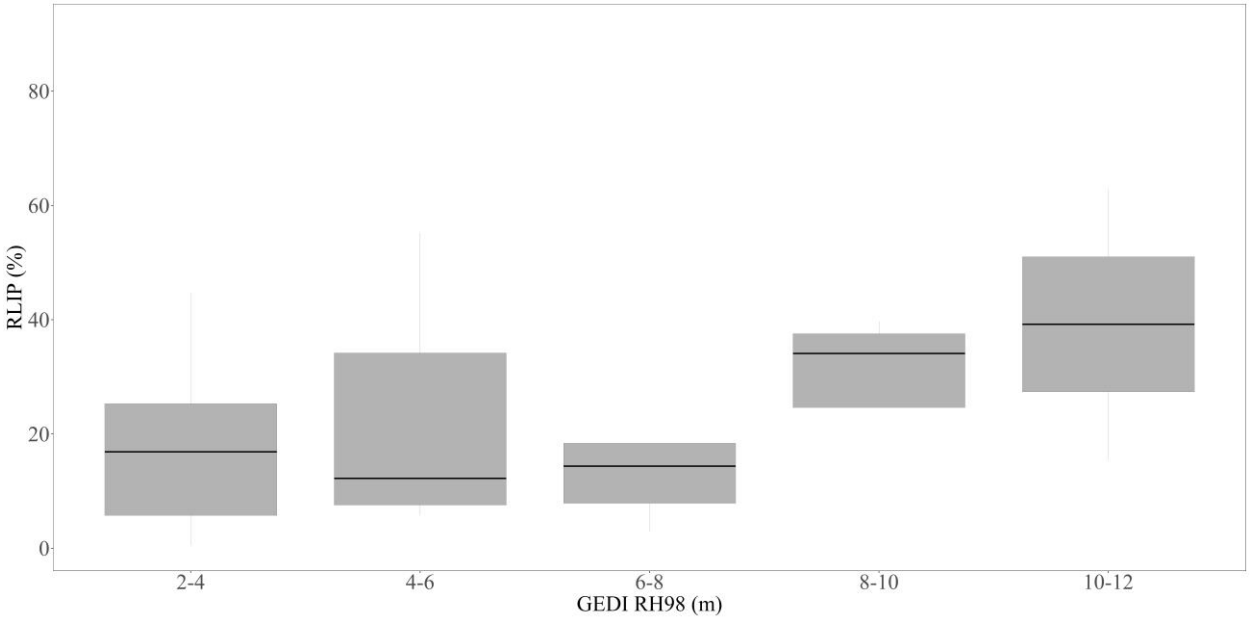


Fig. S6. MAD_P values across various RH98 height bins for 100 relocated GEDI footprints.

The average MAD_P value below 8m was approximately 14% up to 8 m, which included 92.9% of all data. Over the entire range of heights, the mean RLIP was 23.6% (STD = 22.5%). Three GEDI footprints had high RLIP values of > 40%, which could be caused by individual tall savanna trees (RH98 was > 7 m) along the edge of the GEDI footprint that may have a larger or smaller influence on the RH98 depending on the direction random relocation.

The RMSEs of the relocated simulated vs. corrected simulated GEDI RH98 were also tested. **Fig. S7** plots RMSEs vs the absolute Euclidean distance of the x and y offsets. As expected, there was an increase in RMSE with increased relocation distance from 0 to 2 m. The %RMSE of relocated and corrected simulated GEDI RH98 relative to corrected simulated and on-orbit GEDI RH98 ($\%RMSE_P$) was also calculated using the equation (3) and plotted in Figure S7 below.

$$\%RMSE_P = \frac{RMSE_{(Sim\_R,Sim\_C)}}{RMSE_{(Sim\_C,Obs)}} \times 100 \quad (3)$$

Where $RMSE_{(Sim\_R,Sim\_C)}$ represents the RMSE between relocated and corrected simulated GEDI RH98, and $RMSE_{(Sim\_C,Obs)}$ represents the RMSE between on-orbit and corrected simulated GEDI RH98. The mean %RRMSE was 20.7% (STD = 6.8%).

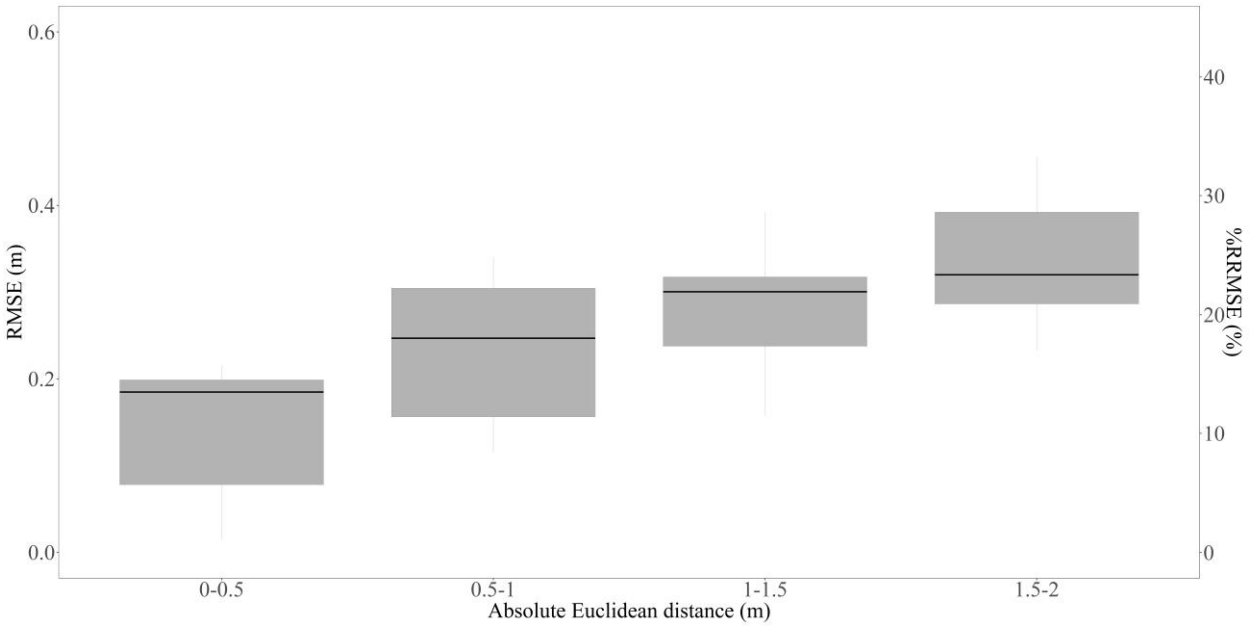


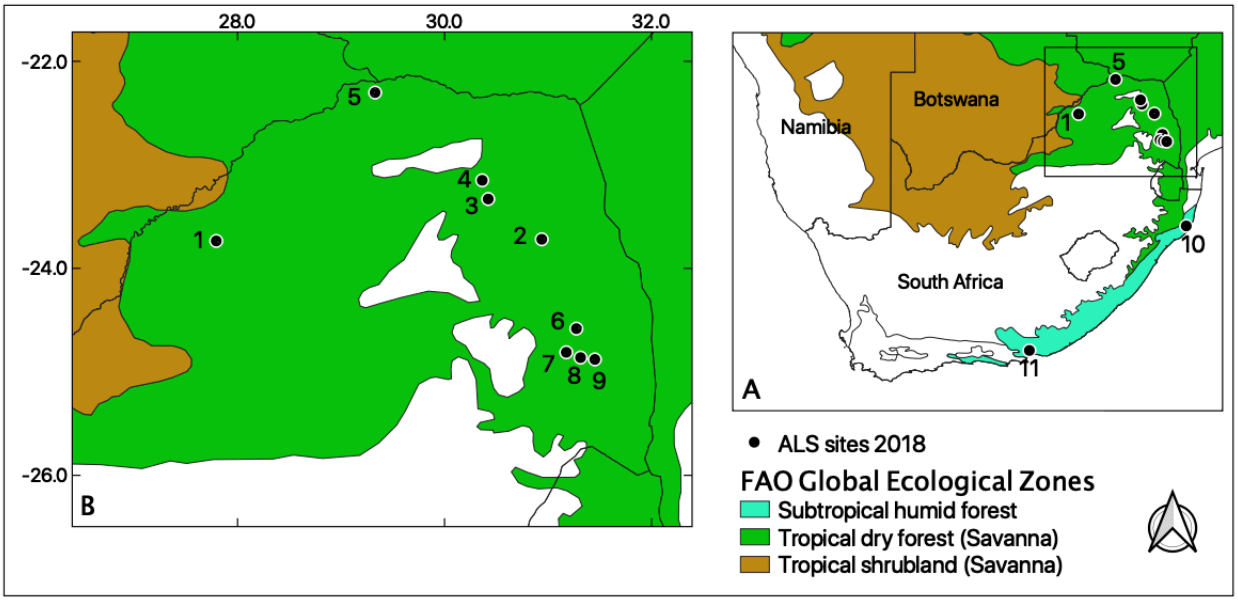
Fig. S7. RMSE and %RRMSE of the relocated simulated vs. corrected simulated GEDI RH98 against the absolute Euclidean distance of relocation.

In summary, the results suggest that random geolocation errors of 0-2 m GEDI contribute 14% of absolute bias of RH98 < 8 m and 20.7% of RMSE of the RH98 in this study.

>>> The following text was added to the Discussion (Line 452-460)
The present study only accounted for systematic geolocation error of the GEDI tracks, however random geolocation errors remain for individual footprints, conservatively estimated to be within 2 m, contributing to the overall error in on-orbit and simulated GEDI comparisons. **These random geolocation errors have not been formally quantified by the GEDI Science Team, so were conservatively estimated to be within 2 m (~10% of the footprint diameter). To quantify the contribution of this random geolocation error, 100 random offsets were generated by sampling from a random uniform distribution between 0 and 2 m and added to the 100 leaf-on GEDI footprint coordinates across all sites, before simulating the waveforms. The difference between the relocated and corrected RH98, along with various error metrics were calculated (for details see Supplementary Material). For RH98, the results suggest that these random geolocation errors contribute 14% of absolute bias below 8 m and 20.7% of the overall RMSE. Therefore, a minor but non-trivial component of the reported uncertainty can be attributed to random geolocation errors, rather than error in vertical canopy height estimates.**

3. Figure 1. It is unclear what CRS is being used, and the colors of the pop-out make the figure cluttered and difficult to follow. Would suggest removing red arrows. Also not entirely clear which map the scale belongs to, although would assume it's the centre one.

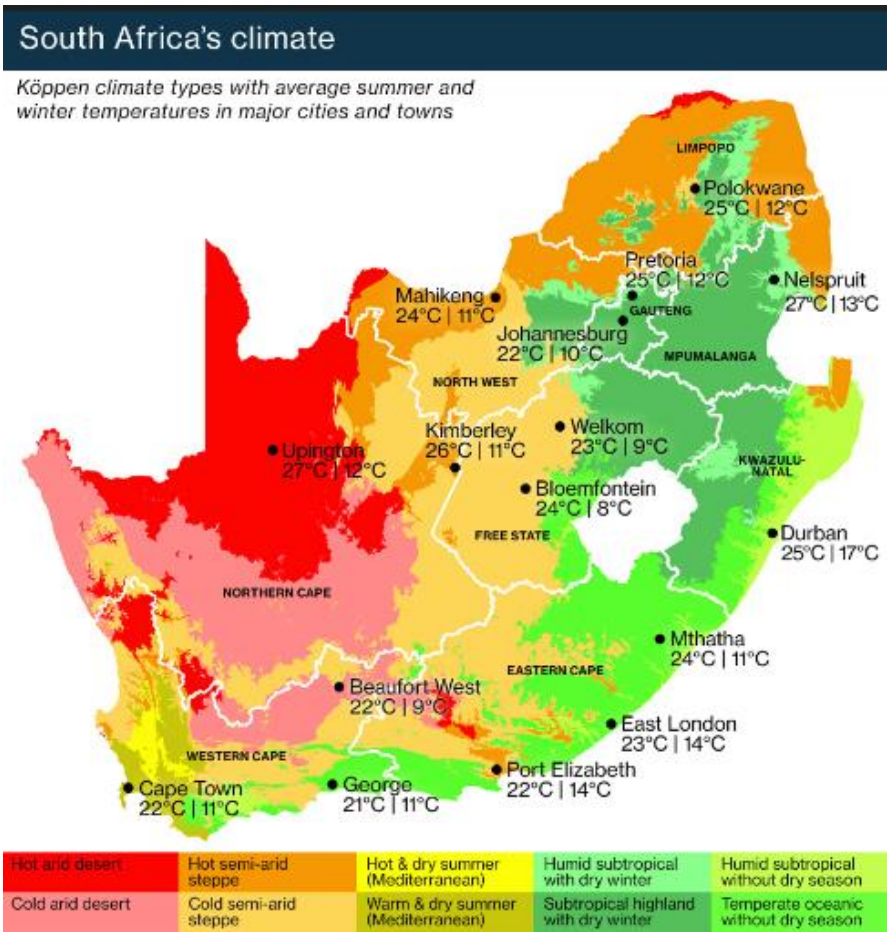
>>> Figure 1 (Page 8) was completely changed as suggested. We added the relevant FAO Global Ecological zones and captions updated accordingly.



New Figure 1 of study area included in manuscript.

4. A clear comparison between the evergreen sites and the deciduous sites for wet and dry seasons would help to confirm that it was in-fact phenology (as opposed to some seasonal impact) that drove the differences. The information is there in table 5, but not easy to determine whether the evergreen site maintains relative accuracy in the dry season, when the other sites are in leaf-off conditions.

>>> In our study area the deciduous woodlands have a very predictable phenology based on season. The phenological conditions were inferred from the date of the GEDI acquisitions (as explained in Section 2.4). Therefore, we cannot separate the influence of phenology from seasonality in the savanna sites. The reviewer therefore suggests that we use the evergreen sites to investigate if leaf on/off or dry vs. wet seasonal conditions drive any differences in the results (Table 5). While this is an interesting question, we found it hard to implement for several reasons. Firstly, the evergreen (non-savanna) sites do not have a distinct dry season as indicated by their climate zone names Dukuduku in the "Humid Subtropical without dry season" and Addo in the "Temperate Oceanic without Dry season" (see figure of Climate types below). The data can therefore not confidently be split into dry or wet seasons based on the date of GEDI acquisitions. Furthermore, the Dukuduku evergreen forest site we only have 3 cloud-free GEDI orbits of useful data that do not allow to do a potential clear dry vs. wet comparison, especially since the spatial coverage and beam-types also differ. The Addo study site has a bimodal rainfall pattern, and it is very hard to split the data into dry and wet. Therefore, it was unfortunately not possible to implement the investigation into the potential impact of seasonal wetness on the results. Furthermore, if the factors such as surface moisture potential had an influence on the results, it would be very difficult to know which orbits had higher surface moisture than others in the absence of a clear dry season.



Climate zones of South Africa. Note that the evergreen ALS sites fall within zones "without a dry season"

5. I find the claim that there was no impact on the phenological conditions during ALS acquisition to be questionable. "the bias of ALS Leaf-on vs. GEDI Leaf-off and ALS leaf-off vs. GEDI leaf-off was -1.48 and -1.46m, respectively (Table S1).
The bias does seem similar for GEDI whether the ALS is collected with leaf on or off, but the R2 and RMSE are considerably different. Eg, using the same comparison as above, the R2 is 0.68 vs 0.45.

>>> See Table S1 included below.
We provide additional interpretations of the potential impact of the phenological conditions of the ALS data. Firstly, the comparison of R^2 of ALS Leaf-on vs. GEDI Leaf-off and ALS leaf-off vs. GEDI leaf-off is actually 0.45 vs. 0.41 respectively, and not 0.68 vs. 0.45 as suggested by the reviewer (Table S1 below). To best investigate the influence of ALS data phenology one needs to compare the ALS leaf-on and GEDI leaf-on vs ALS leaf-off and GEDI leaf-on, however these data were not available to this study.

We added the following text (Page 15, Line 285-291): **The ALS leaf-on and GEDI leaf-on combination had a higher R^2 of 0.68 and lower RMSE of 1.15m compared to ALS leaf-off and GEDI leaf-on combination with R^2 of 0.56 and RMSE of 1.65 m. This indicates that there is a slightly weaker relationship between RH98sim and $RH98_{orb}$ when the $RH98_{sim}$ was based on leaf-off ALS data. However, the bias did not show much difference between ALS leaf-on or off. This indicated that since waveforms were simulated without noise, the discrete return ALS LiDAR and therefore simulated signal start was able to detect canopy tops irrespective of any reduction of canopy cover resulting from leaf-off conditions in these four sites.**

Table S1. Statistical summary of comparison between GEDI-$RH98_{orb}$ vs. GEDI-$RH98_{sim}$ for combined analysis grouped by ALS and GEDI phenological status.

Scenario	ALS (Leaf-on) + GEDI (Leaf-on)	ALS (Leaf-on) + GEDI (Leaf-off)	ALS (Leaf-off) + GEDI (Leaf-on)	ALS (Leaf-off) + GEDI (Leaf-off)
R^2	0.68	0.45	0.56	0.41
Bias (m)	-0.49	-1.48	-0.58	-1.46
%bias (%)	-21.56	-16.72	-17.64	-17.1
RMSE (m)	1.15	2.33	1.65	2.2
%RMSE (%)	26.38	41.51	30.98	40.48

6. Line 265, should read "in comparison to" (not of)
>> Done in Line 283.

7. Line 320, Hancock should be outside brackets to fit in sentence
>> Done in Line 345.

8. Line 344, remove "almost"
>> Done in Line 371.

9. Figure 10, remove square brackets, be clear that these are bin ranges. For example, 0-2.35 rather than (0,2.35),
>>> Done in Page 24.

10. Equation 1, define variables.
>>> The below text was added Line 212-213.

$$threshold = mean + x * \sigma \quad (1)$$

where "mean" represents the mean noise level, sigma is the standard deviation of the noise of the smoothed waveform, and x is a predetermined multiplier named "*Preprocessing threshold"* which was set to 4 sigma.

11. Maps showing examples of a CHM, modelled GEDI footprints and actual GEDI footprints would add to the clarity of the paper.

>> This was a great idea by the reviewer and the figure will help to visualize the data comparison approach, as well as GEDI footprint across the landscape. We produced the image below and included it in supplementary material (Fig. S4). Note that our image shows the ALS point clouds within each GEDI footprint, not CHM as suggested, because we did not use CHMs in any comparisons.

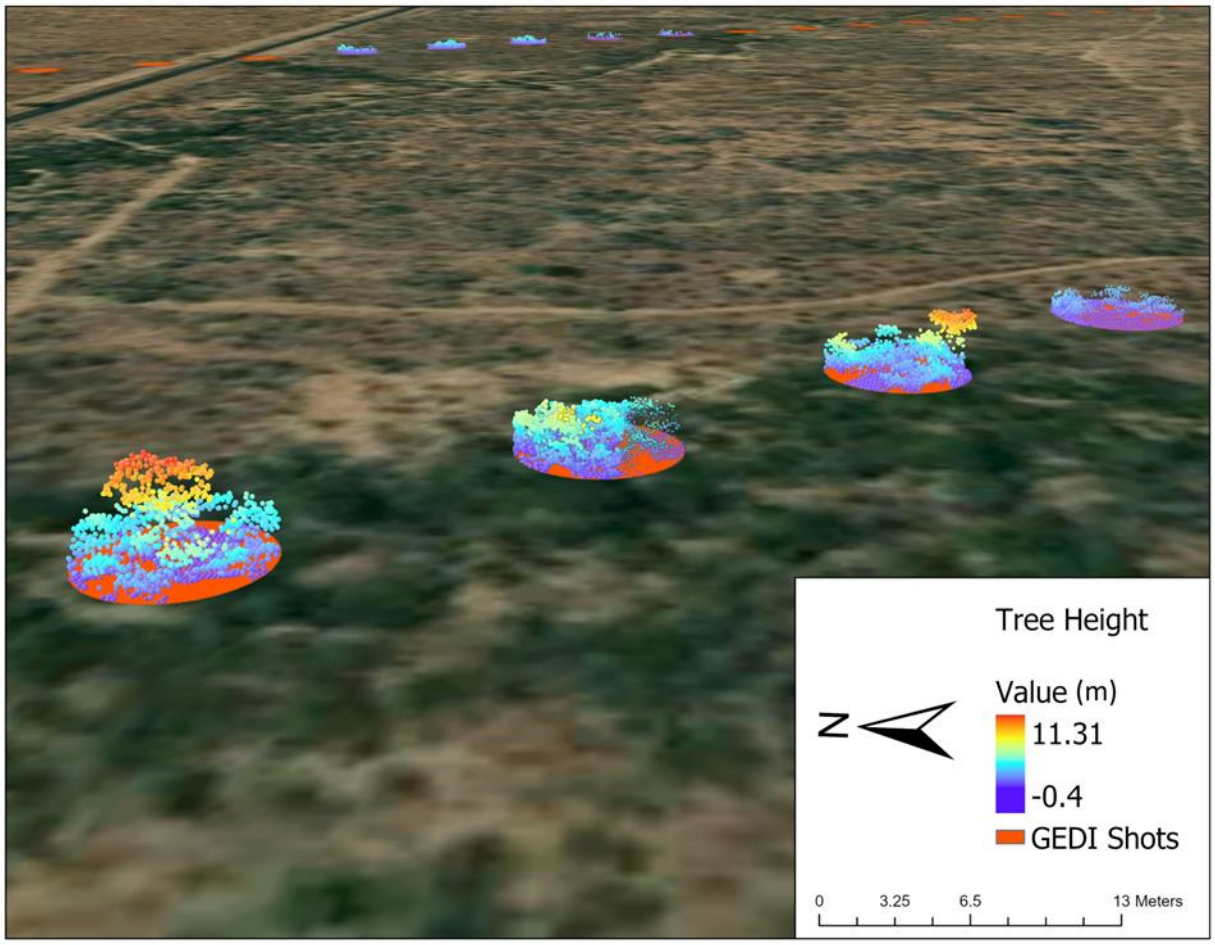


Fig. S4. ALS point clouds inside 25 m GEDI footprints used for simulating the individual GEDI waveforms. The GEDI footprints are spaced 60 m apart along one track and the track are 600 m apart. Backdrop: High resolution optical imagery of study area. (Google, © 2020 Maxar Technologies)

12. Don't discuss some of the limitations of using satellite LiDAR - including large footprint size, non-contiguous spatial coverage.

>>> We added the following text to the end of the Discussion to point out limitations of GEDI and its potential appropriate future use (Page 31, Line 550-558).

GEDI is a sampling mission with footprints separated 60 m along track and 600 m across track (Dubayah et al., 2020; Fig. S4), and therefore does not provide continuous measurements akin to a conventional high resolution ALS CHM used in small area ecological studies. These GEDI footprint samples (L2 data), as well as the L3 and L4B gridded products (1 km x 1 km) can, however, provide unique broad-area metrics of woody vegetation structure and aboveground biomass estimates (Duncanson et al. 2022; Dubayah et al., 2022) for regional ecological studies investigating, e.g. the drivers of structure changes (Sankaran 2008), impact of fire regimes (Smit et al, 2010, 2016) and carbon storage of savannas (Ross 2021). The research community is actively engaged in using GEDI, ICESat2 and the combination of the two space-based LiDAR sensors to address these pressing ecological questions.

>>>Furthermore, the study provided crucial insights into the lower limit of detectability of short stature shrubs that are essential to ecological studies in Savannas.

The following text was added (in bold) to Line 545-547:
Moreover, the measurement of shrub vegetation may be approaching GEDI's limits of detectability, which is determined not only by the pulse duration, but by multiple factors including cover, target reflectivity, topographic relief, and noise (Adam et al., 2020; Liu et al., 2021). **This study provided crucial insight into the lower limits of canopy height detection that impacts the measurement of shrub cover which is essential to ecological studies and biomass estimation in savannas** (Mograbi et al., 2015; O'Connor et al., 2014; Stevens et al., 2016; Venter et al., 2018) and may require different approaches to waveform interpretation and estimation of other GEDI metrics, e.g., canopy cover fraction, plant area index.

Reviewer #4:
Savanna is a special ecosystem in the Africa which provides shelter to various animals. Accurately obtaining the structure information of savanna ecosystem is very important for understanding its ecosystem process and making management policy. This study mainly focusses on accuracy of canopy height provided by the GEDI (L2A). With the large coverage of airborne lidar data, this study explores the accuracy of GEDI' height with different factors, such as GEDI algorithm setting, beam types, phenology. Compared to other research, this study is conducted in a low tree canopy cover area which would be important complement for understanding the performance of GEDI under different ecosystems. The experiment is well designed, and manuscript is well written. I think this paper would be accepted by revised following minor comments.

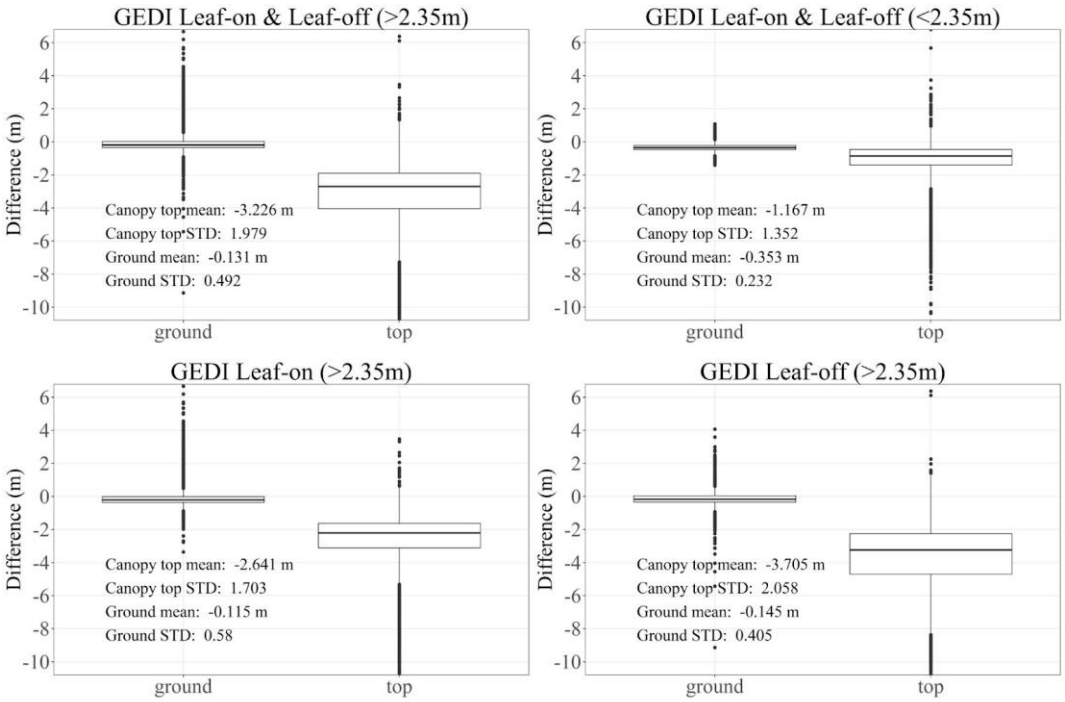
Comments:
1) Line 134: The effects of ALS canopy cover were not reported in this section.
>> Study Area: Canopy cover derived from ALS CHMs was reported in the table 1, but its effect was only discussed later in the manuscript - section 3.3, 3.5 and figure 11 e. We thus did not make any changes.

2) It would be better to make a significant difference test to comparison in Figure 6 and Figure 9.
>> We performed the Welch's Two Sample t-test in R (package: stats) on paired samples in both Figure 6 and 9. The following text was inserted in the manuscript:

Line 346-350:
"A Welch's Two Sample t-test in R (package: stats) (Team, 2013) was conducted on 100 randomly sampled data to investigate if the sensitivity values of the beam types differed significantly. The result indicated that the coverage beams and power beams had significantly different sensitivity values, while the coverage day and coverage night or power day and power night did not. There was no difference in bias for power vs. coverage beam or day vs. night acquisitions with a mean bias between 0 m and 0.5 m (data not shown)."
Line 373-376:

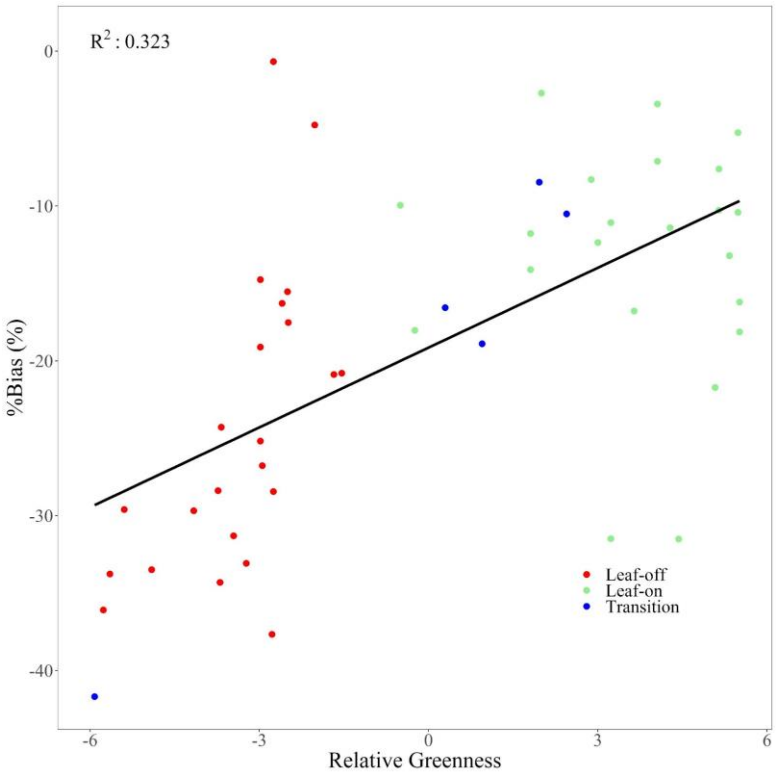
"A Welch's Two Sample t-test (Team, 2013) was implemented on 100 random samples of data each group and demonstrated a significant difference ($p < 0.05$) between the canopy top estimates across all scenarios. However, the mean values of ground estimates were not significantly different."

In addition, we added mean and std numbers to figure 9 which is shown below:

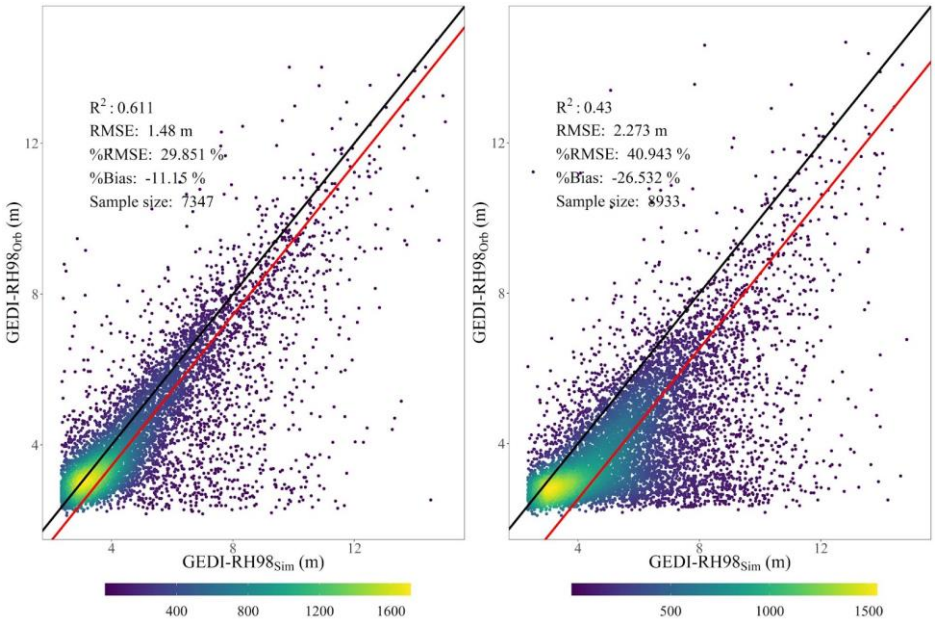


3) Table 4: GEDI footprints should be the number of GEDI footprints
>> Done - changed to "Samples size" in Table 4 and Table 5.

4) Figure 7 and Figure 8: please change to "R2"
>> Done – changed to "R^2" in Fig. 7 and Fig. 8.



Updated Fig 7.



Updated Fig. 8.

4b) and the description of "R2" should be provided in the section 2.5.
>> We believe that R^2 is sufficiently understood by the readers and does not require a description.

Other updates:

>>> Reference was made to the recently published Wang 2022 in the Discussion Line 506-507: **Wang et al., (2022) compared the on-orbit GEDI RH100 with the simulated GEDI RH100 from ALS datasets across the US, with an R2 of 0.83, bias of -0.6 m and RMSE of 3.09 m.**

First Validation of GEDI Canopy Heights in African Savannas

Xiaoxuan Li [a], Konrad Wessels [a,*], John Armston [b], Steven Hancock [c], Renaud Mathieu [d], Russell Main [e], Laven Naidoo [e], Barend Erasmus [f], and Robert Scholes [g]

a Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, United States

b Department of Geographical Sciences, University of Maryland College Park, College Park, MD, United States

c School of Geosciences, University of Edinburgh, Edinburgh, EH9 3FF, United Kingdom

d IRRI South Asia Research Center, International Rice Research Institute, Varanasi, India

e Precision Agriculture Group, Advanced Agriculture and Food Cluster, Council for Scientific and Industrial Research, Pretoria, South Africa

f Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

g Global Change Institute, University of Witwatersrand, Johannesburg, South Africa

* *Corresponding Author: kwessel4@gmu.edu*

Abstract:

Savannas have complex, discontinuous woody vegetation structures that vary greatly in vertical and spatial arrangement and change due to climatic, ecological and management impacts. While airborne laser scanning (ALS) data have provided detailed information on vertical vegetation structure and is widely used in ecological studies, it is lacking in availability and repeat frequency. Although the Global Ecosystem Dynamics Investigation (GEDI) waveform Light Detection and Ranging (LiDAR) sensor and algorithms were optimized for measuring dense forests, it was anticipated that GEDI metrics could provide useful characterization of lower stature, sparse savannas structures. This study provided the first baseline validation of Version 2 GEDI (L2A) relative height 98 (RH98) by comparing the on-orbit GEDI-$RH98_{orb}$ to the simulated GEDI-$RH98_{sim}$ derived from ALS data across diverse savanna vegetation. It furthermore determined the influence of various factors on error, e.g. algorithm setting group (SGs), beam type, day vs. night, beam sensitivity, and vegetation phenology. After applying quality flags, 22813 GEDI footprints were analyzed across 11 sites. SGs 4-6 that are aimed at dense forests had much larger errors than SGs 1-3. The phenological conditions at the time of GEDI data acquisition had a very large influence on the error of $RH98_{orb}$. During leaf-on conditions for savanna vegetation with $RH98_{sim}$ < 15 m, $RH98_{orb}$ was very accurate with R^2 = 0.61, mean bias = -0.55 m, %bias = -11.1%, RMSE = 1.64 m and %RMSE = 29.8%. In leaf-off conditions where $RH98_{sim}$ < 15 m, $RH98_{orb}$ was less accurate with R^2 = 0.43, mean bias = -1.47 m, %bias = -26.5%, RMSE = 2.03 m and %RMSE = 40.9%. During leaf-off conditions, the GEDI LiDAR signal at the start of the waveform may be weaker as it interacts with denuded branches and may be truncated as noise, leading to a large negative height bias. Therefore, assessments of deciduous vegetation structures should be conducted during leaf-on periods. In leaf-on conditions, GEDI's $RH98_{orb}$ was very accurate between canopy heights of 3 and 7 m, with a mean bias of -0.79 m (-10%). The bias of $RH98_{orb}$ was not influenced by canopy cover. Due to the GEDI LiDAR pulse width of 15.6 ns, the GEDI-RH98 data product cannot reliably estimate canopy heights of shrubs below 2.34 m and will require more complex deconvolution of the waveform. GEDI's RH98 accurately estimates the canopy height of trees between 3 and 15 m allowing assessment of canopy heights over vast savanna areas.

Keywords

LiDAR, GEDI, vegetation structure, Savannas, canopy height, validation, Africa.

1. Introduction

Savannas are defined as "communities or landscapes with a continuous grass layer and a discontinuous tree layer" (Scholes and Archer, 1997). Savannas cover more than 20% of the Earth's surface and account for the third-largest above ground carbon stock, after Tropical Wet and Tropical Moist climate regions (Scharlemann et al., 2014). Although the carbon density of savannas and woodlands is lower than that of dense forests, in Africa 52% of the total above-ground carbon is stored in savannas and woodlands (Bouvet et al., 2018). Savannas furthermore contain a very large proportion of the world's human population and the majority of its livestock and wildlife (Hill and Hanan, 2010). An estimated 150 million people in the developing world are directly dependent on local savanna ecosystem services (McNicol et al., 2018), e.g. livestock grazing, fuelwood as primary source of household energy, especially in India and Africa (Behera and Gupta, 2015; Wessels et al., 2013), timber and non-timber products (Shackleton and Shackleton, 2004; Twine et al., 2003).

Tree-grass dynamics in African savannas are complex, under the influence of highly variable precipitation, and major disturbances such as fire and domestic or wild herbivores (Sankaran et al., 2008, 2005). Human activities directly alter the woody components through deforestation and degradation. For instance, deforestation rates in African savanna woodlands are reported to be higher than in tropical rain forests (Ciais et al., 2011) and carbon losses are 3-6 times higher than previously estimated (McNicol et al., 2018). On the other hand, African savannas are also experiencing rapid woody encroachment (Stevens et al., 2016; Venter et al., 2018) in lower height classes, possibly due to increases in CO_2 globally (Bond and Midgley, 2012; Ratnam et al., 2016), and locally due to fire suppression, livestock overgrazing, historical loss of browsing wildlife, as well as long-term interactions between fire and rainfall episodes (Joubert et al., 2013, 2008; O'Connor et al., 2014). Furthermore, simultaneous and spatially coincident gains of shrub cover (< 3 m) and losses of large trees (> 5 m) due to high elephant density, humans and fire, have been reported in South Africa, with large impacts on wildlife biodiversity and livestock grazing in rural rangelands (Asner et al., 2016; Davies et al., 2018; Dean et al., 1999; Levick et al., 2015, 2009; Mograbi et al., 2017, 2015; Smit et al., 2016, 2010; Smit and Prins, 2015; Wessels et al., 2011). Spatial data on woody vegetation structure, specifically canopy height, are therefore essential to understand and manage savannas for a wide variety of land uses.

Three-dimensional vegetation structure is defined as the vertical configuration of above-ground vegetation and its horizontal, leaf to landscape-scale variations (Brokaw and Lent, 1999) including tree and canopy height, canopy cover, leaf area density profile and stem diameter. Savannas are complex and structurally diverse ecosystems, in terms of the varying proportions and spatial distributions of trees, shrubs and grass, as well as the architectures of trees, and shrubs at varying stages of growth (Scholes, 1997). Most woody species in southern African savannas are deciduous and display strong phenological cycles with woody vegetation losing their leaves during the dry winter period (Archibald and Scholes, 2007). The woody vegetation structure of savannas has been predominantly measured with discrete return ALS data, which provides detailed data on the proportions of woody canopies at various heights (Fisher et al., 2015, 2014; Levick et al., 2009), the understory structure (Fisher et al., 2015; Mograbi et al., 2015) and their changes through time (Mograbi et al., 2017). The typical ALS products used to assess structure include (i) point clouds with density of 1-15 points / m^2 (ii) canopy height models with of 1-2m resolution and (iii) voxel data of 1 m^3 (Fisher et al., 2015). While the ALS data provide detailed information that are suited to characterize the structure at a local, landscape scales (Camarretta et al., 2019; Lefsky et al., 2002; Wulder et al., 2012), it is expensive, infrequently repeated to provide change information and often has very limited aerial coverage, precluding regional monitoring of vegetation structure. Space-based LiDAR sensors, such as GEDI may be able to address this observation gap by providing near-global coverage and frequent observations of vertical structures of global savannas, in particular woody vegetation height. The overall goal of the GEDI mission is to characterize ecosystem structure and its change due to climate and land use (Dubayah et al., 2020). The GEDI instrument is a geodetic-class laser altimeter and waveform LiDAR with a 25m footprint that is optimized for measurements of vertical structure. Mounted on the International Space Station, GEDI measurements have a near-global coverage (within 51.6° N & S), which has sampled 4% of Earth's land surface during the initial two-year mission. GEDI provides critical datasets of woody structural metrics including canopy height, canopy cover, plant area index and vertical foliage profiles, topography, as well as footprint-level and gridded above-ground biomass. The GEDI instrument was designed to measure vertical canopy profiles in conditions of up to 95% and 98% canopy cover for the coverage and power beams, respectively (Dubayah et al., 2020). Its design was optimized for the measurement of dense forests, following more than two decades of research using airborne large footprint waveform LiDAR, such as that from the Land, Vegetation and Ice Sensor (LVIS) (Drake et al., 2002a; Drake et al., 2002b; Dubayah et al., 2010; Hyde et al., 2005). The GEDI pulse length (FWHM 15.6 ns) is short enough to vertically

discriminate canopy and ground returns in forested ecosystems, however the ability to characterize short stature, discontinuous vegetation has not been a design requirement for GEDI and therefore needs to be evaluated before applying it to regional ecological studies. While the GEDI instrument design, algorithm calibration and validation were focused on the measurement of tall, dense, and continuous vegetation typical of tropical and temperate forests, it is anticipated that it could also provide useful, accurate characterization of lower stature, discontinuous savanna vegetation. This was the first study to our knowledge to validate GEDI canopy height measurements in savanna vegetation.

Relative Height (RH) metrics derived from full-waveform LiDAR systems, such as GEDI, are considered "LiDAR perceived metrics" which estimate the height at which a particular quantile returned energy is reached relative to the elevation of the lowest waveform mode. i.e. ground level (Hofton and Blair, 2019). This paper focused on RH98, which represents the 98th quantile of the returned energy distribution. RH98 is representative of the top of the canopy, or the near-highest vegetation in the footprint, but contains fewer outliers and is less sensitive to noise than RH100 (Blair and Hofton, 1999). A few recent studies have directly compared GEDI RH metrics to ALS canopy height models (CHM) by calculating percentiles of CHM values within GEDI footprints (Liu et al., 2021; Potapov et al., 2021). While these studies provided essential early insights into the global and continental-scale reliability of GEDI metrics with very promising results, such comparisons to ALS CHM-derived height metrics alone are not the most appropriate validation approaches, as they essentially compare different structural metrics which may confound interpretation. The present study, therefore, followed the novel approach of first simulating~~, simulated~~ GEDI waveforms from the ALS point cloud data before ~~, followed by the~~ extracting~~on~~ of ground elevation and RH metrics (Hancock et al., 2019; Silva et al., 2018) to generate an ALS-based reference dataset before performing an error assessment (Wang et al., 2022)~~ (Roy et al., 2021)~~.

The aim of this study was to provide an independent validation of GEDI canopy height metrics in diverse savannas by comparing the relative height (RH) metrics of Release 2 on-orbit GEDI data to simulated GEDI-RH98 derived from reference ALS datasets. This will provide the first ~~a~~ baseline assessment of the performance of GEDI-RH98 products in African savannas, which will enable the identification of research priorities to expand the applicability of GEDI canopy height products to non-forest vegetation. The specific objectives of the study were to: (i) assess the accuracy of on-orbit GEDI relative height metric (RH98) by comparison to the reference estimates simulated from

ALS, and (ii) determine the influence of various factors on this error, e.g. algorithm setting group (SGs), beam type, day vs. night, sensitivity, and vegetation phenology.

2. Materials and Methods

2.1. Study area

Nine study sites were distributed across South Africa representing various savanna vegetation types that fall within the Tropical Dry Forest (FAO, 2012) (Table 1) (Fig. 1). The sites target specific vegetation types (Dayaram et al., 2019), but often include some adjacent transformed cover types such as subsistence agriculture (Table 1). Ireagh, Justicia, Agincourt and Welverdiendt sites are located in the Lowveld of Mpumalanga Province, next to Kruger National Park and include Granite Lowveld and Legogote Sour Bushveld vegetation, as well as some subsistence cultivation (Dayaram et al., 2019). The D'Nyala site is situated within the D'Nyala Nature Reserve in the far western extent of the Limpopo Province. The reserve contains several Bushveld vegetation types (Table 1), with scattered large Nyala (*Xanthocercis zambesiaca*) and Baobab (*Adansonia digitata*) trees up to 20 m in height. The Venetia site is located at the northern tip of South Africa, in the Musina Mopane Bushveld that is dominated by mopane trees (*Colophospermum mopane*). The three Limpopo sites are also situated within the Limpopo Province, just west of the Kruger National Park. These sites are primarily dominated by Mopane trees of the Tsende Mopaneveld and Granite Lowveld vegetation types with sporadic pockets of Gravelotte Bushveld and Tzaneen Sour Bushveld (Dayaram et al., 2019). All nine sites mentioned above have deciduous vegetation that lose their leaves during the dry winters (Archibald and Scholes, 2007).

Two additional evergreen, non-savanna sites classified as Subtropical humid forests (FAO, 2012), were included to provide a wider range of canopy height and cover for the site-specific assessments. Their data was not included in the combined savanna-specific analyses. The Addo site transects part of the Addo Elephant National Park in the Eastern Cape Province. The site's thicket vegetation is dominated by very dense, evergreen, stout (maximum height 3.6m), succulent shrubs and geophytes, such as Spekboom trees (*Portulacaria afra*) which is a major source of food for the large elephant population (>600). The Dukuduku site is mainly covered by evergreen, tall lowland coastal forest located along the northern coast of the Kwa-Zulu Natal Province. The area is a flat coastal plain with dense

forest growth supported by the high, year-round rainfall. It is the last remaining lowland indigenous forest in South Africa and is dominated by numerous endemic forest tree species (e.g. *Syzygium cordatum*, *Cussonia zuluensis* and *Ficus natalensis*). The site includes surrounding grasslands, exotic tree plantations, and subsistence cultivation, providing very diverse vegetation structures. The thicket vegetation of Addo and the coastal forest vegetation of Dukuduku are juxtaposed to savanna vegetation both in location and along a continuum of vegetation structure. The topography of all the sites is flat or gently undulating and slope data derived from ALS DEM did not vary much within study sites (standard deviation (STD) < 4 degrees) (Table 1) and therefore it was assumed that slope did not influence the results.

Table 1. Properties of study sites and airborne LiDAR (ALS) data acquired.

Site#	Site name	Vegetation type (Dayaram et al. 2019)/ transformed cover	Deciduous Savannas /Evergreen	Mean annual rainfall (mm)	Mean annual temp (°C)	Mean (STD) ALS vegetation height (m)	ALS canopy cover (%)	Date of ALS	Area (km^2)	Mean (STD) Slope (degree)
1	D'Nyala	Roodeberg, Waterberg Mountain & Limpopo Sweet Bushveld	Deciduous Savannas	375	21	3.9 (2.3)	74.3	March 2018	53.26	2.36 (3.69)
2.3.4	Limpopo	Tsende Mopaneveld, Granite Lowveld, Gravelotte Bushveld & Tzaneen Sour Bushveld	Deciduous Savannas	613	27	3.3 (2.1)	51.5	March/ April 2018	163.32	1.7 (1.28)
5	Venetia	Musina Mopane Bushveld & Limpopo Ridge Bushveld	Deciduous Savannas	368	22.8	2.5 (1.1)	41.6	March 2018	56.31	1.97 (1.75)
6	Welverdiendt	Granite lowveld & Legogote Sour Bushveld / subsistence cultivation	Deciduous Savannas	353	25	3.7 (2.0)	44.2	June 2018	126.75	1.74 (1.68)
7	Agincourt		Deciduous Savannas	353	25	4.2 (2.4)	41.9	May 2018	35.88	5.14 (4.84)
8	Ireagh		Deciduous Savannas	687	21.3	3.2 (2.3)	32.3	June 2018	65.08	2.43 (2.09)
9	Justicia		Deciduous Savannas	550	25	3.4 (2.5)	28.9	June 2018	81.25	2.55 (1.48)
10	Dukuduku	Maputaland Coastal belt & Northern Coastal Forest mix, Subtropical Alluvial / exotic tree plantations, subsistence cultivation	Evergreen	967	21.7	9.3 (5.6)	70.0	June 2018	141.38	1.84 (2.41)
11	Addo	Sundays mesic & Valley thicket, Grassridge Bontveld & Albany Alluvial	Evergreen	388	18.4	2.3 (0.88)	66.3	March 2018	109.84	2.63 (3.05)

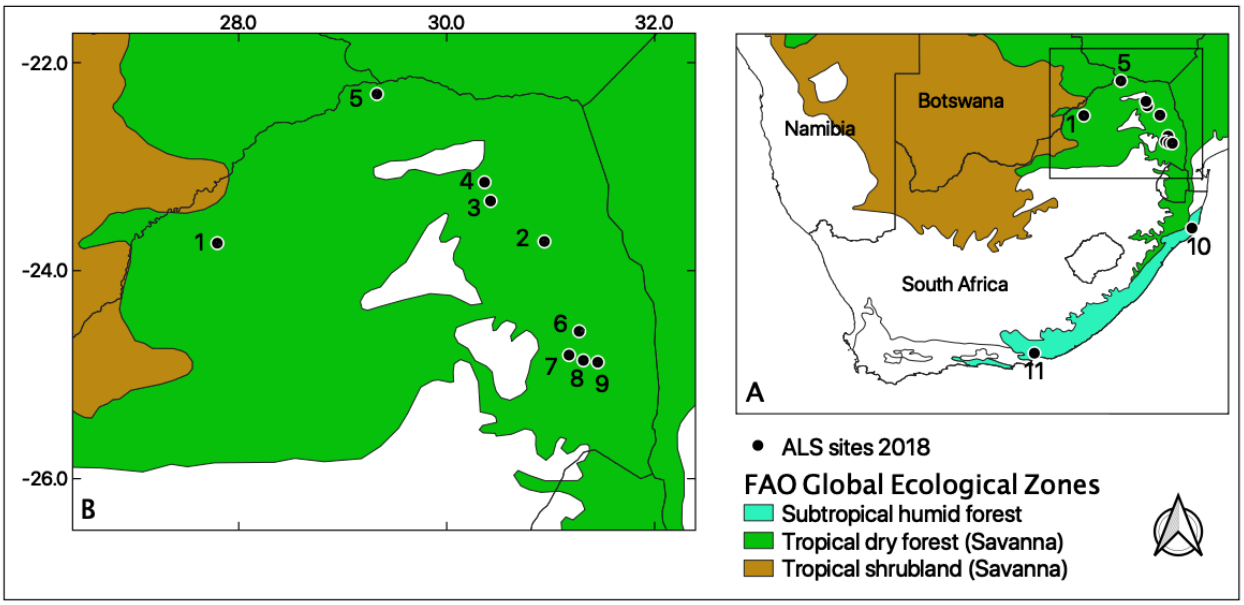


Fig. 1. (A) Study area in South Africa with relevant FAO Global Ecological Zones. (B) Zoom-in of northern South Africa with detailed location of individual sites covered by airborne LiDAR data (ALS): (1) D'Nyala, (2, 3, 4) Limpopo, (5) Venetia, (6) Welverdiendt, (7) Agincourt, (8) Ireagh, (9) Justicia, (10) Dukuduku, (11) Addo (B). (2-column fitting image)

2.2. Data

2.2.1 ALS data

Approximately 800 km^2 of discrete ALS data were recorded between March and June of 2018 (Table 1) over the 11 study sites (Fig. 1). For four of the sites the ALS data was collected during May and June, the leaf-off period, while other sites were collected during leaf-on conditions (March) (Table 1). The airborne LiDAR datasets were collected from a fixed wing aircraft (700 m above ground) using Optech ALTM M300 (13SEN327) and Optech Gemini (09SEN258) sensors at a pulse repetition rate of 150 kHz. Acquisitions were planned to have 25 % overlap between flight lines and resulted in an average laser spot spacing of 0.4 m and an average point density of 8.6 points / m^2. Digital terrain models (DTM), digital surface models (DSM), and canopy height models (CHM) were generated from the ALS data at 1 m resolution, using lidR package (Roussel et al., 2018) in R. The Grid_canopy (lidR package) function was used to generate CHMs at 1m resolution using DSM Algorithm (p2r) without any smoothing or filtering. Within each study site, the canopy cover variable was determined by calculating the percentage of pixels with CHM greater than 1.5 m (Table 1). Note that in contrast to other recent studies (Liu et al., 2021; Potapov et al., 2021), the CHMs were not used in any of the comparisons with $RH98_{orb}$, but were only used to generate the vegetation structure metrics (cover and mean height) of sites in Table 1.

2.2.2 GEDI data

GEDI was launched in December 2018 and attached to the International Space Station's Japanese Experiment Module-Exposed Facility. As a waveform LiDAR, GEDI is specifically designed to measure vegetation structural metrics and estimate biomass under different environmental conditions (Dubayah et al., 2020). There are three lasers mounted on the GEDI platform, two of which operate at full power and another one was split up into two coverage beams. These four beams are optically dithered to four power ground tracks (BEAM0101, BEAM0110, BEAM1000, and BEAM1011) and four coverage ground tracks (BEAM0000, BEAM0001, BEAM0010, and BEAM0011). This design produces simultaneously eight parallel tracks of GEDI footprints, with a nominal diameter of 25 m and spacing of 60 m along track and 600 m across track.

The study used Version 2 GEDI data that became available in April 2021. Compared to Version 1 GEDI data, key improvements in Version 2 data include (i) improved geolocation of orbital segments, and (ii) algorithm setting group (SG) selection per laser shot (Beck et al., 2021). This study quantified the systematic geolocation error in Version 2 and compared the error of different SGs in savannas.

The input data and processing workflow is given in Fig. 2. The GEDI datasets used in this study are subsets of attributes stored in GEDI L1B, L2A, and L2B data products, which were downloaded from the NASA Land Processes Distributed Active Archive Center (DAAC) using the LP DAAC download tool (https://git.earthdata.nasa.gov/projects/LPDUR/repos/daac_data_download_r/browse). Version 2 GEDI data have a known geolocation error of approximately 10 m (1-sigma) (Beck et al., 2021). This error is dominated by a systematic component that may be removed by maximizing the correlation between on-orbit and simulated GEDI waveforms to determine the horizontal and vertical offset between the GEDI and ALS datasets (Hancock et al., 2019; Hofton et al., 2000). This geolocation error was corrected using the collocateWaves program (see section 2.3).

In this study, 40 GEDI orbits intersected 9 deciduous savannas study sites, providing a total of 22813 useful GEDI footprints (samples) in 58 test cases (each orbit intersecting one site is one test case) during the seventeen months of the GEDI mission, between April 2019 and September 2020, which was all the data released at time of writing. Most of these test cases had more than 100 footprints of one orbit within a site. The most relevant variables used are GEDI L2B quality flags, GEDI L1B waveforms, GEDI L2A vegetation height metrics calculated by six algorithm SGs, and

geolocation attributes such as latitudes, longitudes, and orbits. Quality flags were used to remove erroneous and low-quality returns. Only GEDI footprints with quality flag = 1, were considered as "good quality" data when specific requirements were met in terms of energy, sensitivity, amplitude, and real-time surface tracking quality (Hofton and Blair, 2019).

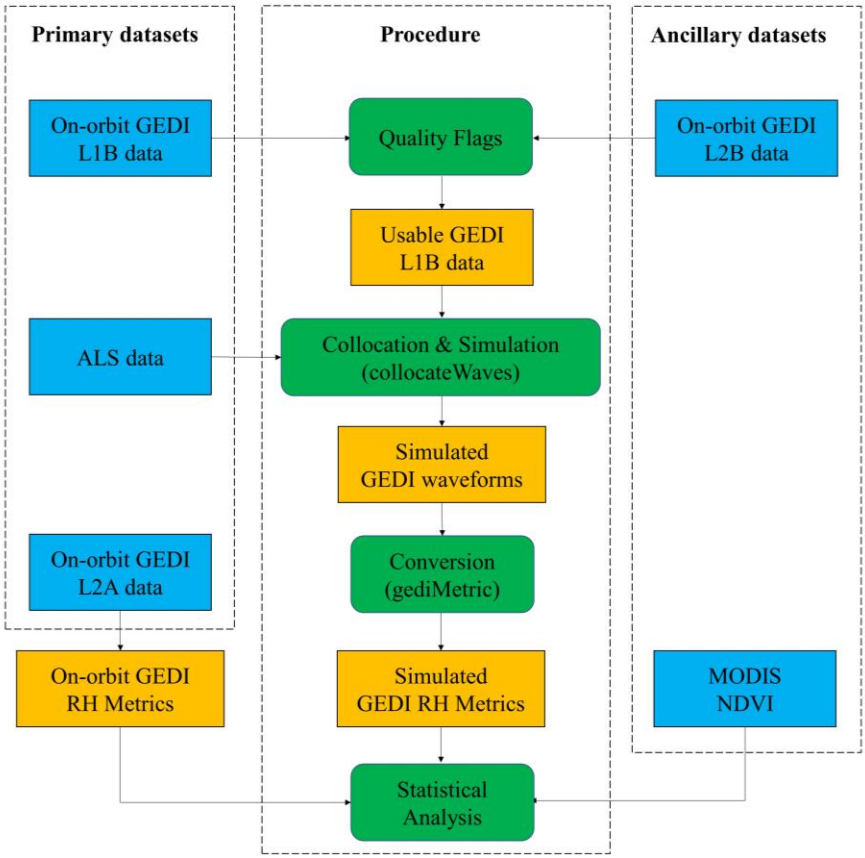


Fig. 2. Workflow of GEDI and ALS data processing and analysis. (1.5-column fitting image)

2.2.3 GEDI waveform processing and algorithm setting groups

In order to interpret the results of the six SGs (Table 2), it is useful to review waveform processing and algorithm parameters as described in the GEDI Algorithm Theoretical Basis Documents (ATBD) (Hofton and Blair, 2019). Equation (1) is used to determine the highest detectable return (canopy top, named "*toploc*") or the lowest detectable return (ground, named "*botloc*") within the GEDI waveform extent (between the start "*searchstart"* and end positions "*searchend"*), where x representing either the '*Front threshold*', or the '*Back threshold*'. The "*toploc*" or

"*botloc*" are detected where two adjacent waveform intensities are above the '*Front threshold*', or the '*Back threshold*' (Fig. 3).

$$threshold = mean + x * \sigma \quad (1)$$

where "*mean*" represents the mean noise level, sigma is the standard deviation of the noise of the smoothed waveform, and x is a predetermined multiplier named "*Preprocessing threshold*" which was set to 4 sigma. Next, a Gaussian filter with different smoothing widths in ns (*Smooth width zcross*) is applied to identify distinctive canopy and ground modes within the signal-only section of the GEDI waveform (Table 2). Finally, the waveform digitizer counts within the section of the smoothed waveform between "*toploc*" and "*botloc*" are summed to create the cumulative energy metrics from 0 to 1 at 1% interval, where metrics such as the RH98 can be calculated relative to the lowest waveform (ground) mode.

Six waveform SGs are implemented with different smoothing width (*Smooth width zcross*) and threshold ('*Front threshold*', '*Back threshold*') settings to allow for the detection of the ground and vegetation structures under different conditions (Hofton and Blair, 2019) (Table 2). Version 2 GEDI data automatically selects the most appropriate algorithm setting group (SG S) for individual laser shots depending on plant functional type, geographic region and laser return energy (Beck et al., 2021). For instance, SG 1 (which was selected in the majority of cases) (Beck et al., 2021), uses conservative settings that may result in the ground mode not being distinguished from noise in dense canopies, or potentially a canopy mode not being detected in very sparse canopies.

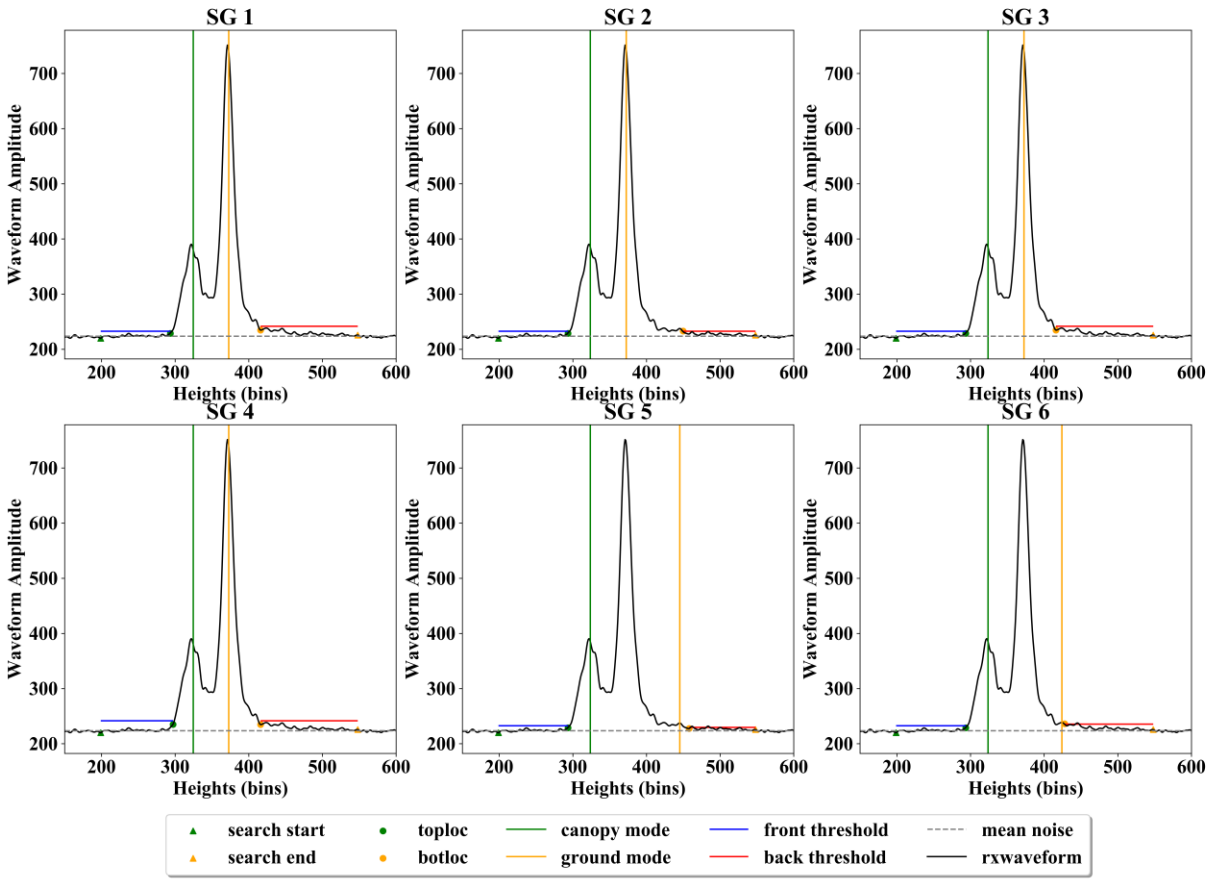


Fig. 3. Example of one GEDI L1B waveform (GEDI orbit: 7433, waveform ID: 54731) and parameters used by different algorithm setting groups (SGs) (Table 2) to identify canopy mode (yellow line) and ground mode (green line). (2-column fitting image)

Table 2. Parameters of different GEDI algorithm setting groups (SGs).

SG	Smooth width zcross (ns)	Front threshold	Back threshold
1	6.5	3	6
2	3.5	3	3
3	3.5	3	6
4	6.5	6	6
5	3.5	3	2
6	3.5	3	4

The effects of instrument acquisition factors, e.g. time of acquisition (day vs. night), and beam type (coverage vs. full power) on RH98 errors, were investigated. These factors all influence beam sensitivity, a waveform metric that indicates the maximum canopy cover through which a GEDI waveform can detect the ground (Hancock et al., 2019). The sensitivity is important because it is determined by laser pulse energy, background solar illumination, surface

reflectance, and atmospheric transmission, and can potentially influence the variability or errors of RH metrics (Dubayah et al., 2020; Hancock et al., 2019). Solar elevation defines when GEDI data products were acquired, with daytime defined by positive solar elevation values, and nighttime by negative values (Beck et al., 2021). Under average conditions, the GEDI power beams are expected to have beam sensitivities of 99% during the nighttime, and 94% during the day. The GEDI coverage beams are expected to have beam sensitivities of 96% during the nighttime and 92% during the day (Hancock et al., 2019).

2.3 Data processing

To help remove erroneous or cloud-contaminated land surface waveforms, GEDI L2A and L2B quality flags were first extracted from GEDI L2B datasets and used to filter out low-quality GEDI L1B data (Fig. 2). The subset of high-quality GEDI L1B waveforms was then intersected with the ALS datasets in Table 1. The gediSimulator software (Hancock et al., 2019) is comprised of collocateWaves and gediMetric programs which were~~as~~ used to simulate a GEDI full-waveform from the ALS data, collocate these with recorded GEDI waveforms (Fig. S1), and create a series of simulated vegetation metrics (RH, canopy cover, etc.) (Hancock et al., 2019) (Fig. 2). ~~-~~The collocateWaves program was used to simulate GEDI waveforms from ALS data and collocate the simulated GEDI waveforms with recorded GEDI waveforms (Fig. S1). The parameter settings in the collocateWaves program were as follows, (i) the minimum GEDI beam sensitivity was set to 0.9 (ii) the minimum ALS point density was set to 3 points per/m^2, (iii) the on-orbit GEDI pulse shapes were derived from the GEDI Level 1B data for each beam and were used instead of the default Gaussian shape, (iv) the search window used to match ALS-derived and GEDI waveform was set to 20 m x 5 m to handle heterogeneous vegetated areas, (v) the simulated GEDI waveforms were generated without Gaussian noise. To resolve the systematic geolocation errors, the collocateWaves program was applied independently to all GEDI tracks. ~~The collocation procedure was applied to each GEDI track independently and only used GEDI waveforms with beam sensitivity > 0.9 and simulated waveforms with an ALS point density >= 3 points / m^2. To define the transmitted pulse shape for simulations, average pulse shapes were calculated from the GEDI Level 1B data for each beam.~~ The simulated GEDI RH metrics generated in collocateWaves program were then extracted from the simulated GEDI waveforms by applying gediMetric program (Hancock et al., 2019).

Given the pulse width of 15.6 ns (FWHM) of the GEDI laser, GEDI's on-orbit and simulated RH98 both have a theoretical minimum height of retrieval equal to half the pulse width, i.e. 2.34 m, which may be above the maximum

height of short woody vegetation in savannas (e.g. shrubs < 2.34 m). Since RH98 is the 98th quantile of the returned energy, its value will therefore be ~ 2.34 m even over flat, bare ground containing no woody vegetation. Variations in topographic relief within the GEDI footprint would broaden the returned waveform and increase the RH98 independently of vegetation structure (Harding and Carabajal, 2005). Therefore, variations in woody vegetation below 2.34 m have a very minor influence on the received waveform shape and the derived RH98. Hence, comparisons of on-orbit and simulated RH98 below 2.34 m ($RH98_{sim}$ < 2.34 m) are typically not sensitive to variation in vegetation structure, and these data points were excluded from the analyses a priori.

2.4 Vegetation phenology

The influence of vegetation phenology on the RH98 error estimates was investigated by considering, (i) a relative greenness index and (ii) a phenology status classification based on the date of the GEDI acquisition. The normalized difference vegetation index (NDVI) of the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MOD13A3, 1 km) time series (Didan, 2015) was used to calculate the relative greenness of each test case (all pixels within the site) on the acquisition date of the $GEDI_{orb}$ data, according to the following equation (Peters et al., 2002):

$$Greenness_{relative} = (NDVI_{GEDI} - NDVI_{mean}) \div NDVI_{std} \quad (2)$$

where $NDVI_{GEDI}$ is the average NDVI of the pixels on the date of a specific test case, and $NDVI_{mean}$ and $NDVI_{std}$ are the mean and standard deviation of 20-year MODIS NDVI (2000 to 2020) of all the same pixels in the site, respectively. The relative greenness values ranged between -6 and +6. The phenological status of the woody vegetation on the date of GEDI acquisition was classified as, (1) leaf-on; (2) leaf-off; and (3) transition from leaf-on to leaf-off or from leaf-off to leaf-on. Each test case was classified as follows: Leaf-on (Nov. to Apr.), Transition (May and Oct.), and Leaf-off (Jun. to Sep.)

The potential influence of the phenological conditions during the ALS acquisition was tested but found to have little impact on the results. There were only minor differences between the metrics of GEDI RH98 accuracy in comparison ~~of~~ to ALS leaf-on and leaf-off data with GEDI data acquired in either leaf-on or leaf-off conditions (Table S1). For example, the bias of ALS Leaf-on vs. GEDI Leaf-off and ALS leaf-off vs. GEDI leaf-off was -1.48

and -1.46m, respectively ~~(Table S1)~~. The ALS leaf-on and GEDI leaf-on combination had a higher R^2 of 0.68 and lower RMSE of 1.15 m compared to ALS leaf-off and GEDI leaf-on combination with R^2 of 0.56 and RMSE of 1.65 m. This indicates that there is a slightly weaker relationship between $RH98_{sim}$ and $RH98_{orb}$ when the $RH98_{sim}$ was based on leaf-off ALS data. However, the bias did not show much difference between ALS leaf-on or off. This indicated that since waveforms were simulated without noise, the discrete return ALS LiDAR and therefore simulated signal start was able to detect canopy tops irrespective of any reduction of canopy cover resulting from leaf-off conditions in these four sites. ~~This was not unexpected since comparisons were undertaken using waveforms simulated without noise, so the discrete return ALS LiDAR and therefore simulated signal start was able to accurately detect canopy tops irrespective of leaf-off conditions in four sites. All the ALS data were therefore treated as equivalent in terms of phenology for the rest of the analysis.~~

2.5 Statistical comparison of on-orbit and simulated GEDI RH98

To investigate the impact of other potential factors, e.g., SGs, vegetation phenology, sensitivity, beam type and acquisition time, the datasets were split accordingly during statistical analysis. The SGs were compared in terms of the R^2 between the recorded, on-orbit GEDI RH98 (named GEDI-$RH98_{orb}$ hereafter) and the simulated GEDI RH98 (name GEDI-$RH98_{sim}$ hereafter), as well as the bias within 1m bins of $RH98_{sim}$. The error of GEDI-$RH98_{orb}$ was determined by comparing it to the GEDI-$RH98_{sim}$. ΔRH98 was the difference between the GEDI-$RH98_{orb}$ and GEDI-$RH98_{sim}$ (equation 3) (Neuenschwander et al., 2020):

$$\Delta RH98 = GEDI_{RH98_{orb}} - GEDI_{RH98_{sim}} \quad (3)$$

Other error measures, included mean bias, root mean squared error (RMSE), relative RMSE (%RMSE), and relative bias (%bias) (equation 4-7):

$$Meanbias = \frac{\sum(\Delta RH98)}{n} = \frac{\sum_{i=1}^{n} Bias_i}{n} \quad (4)$$

$$RMSE = \sqrt{\frac{(\sum_{i=1}^{n} Bias_i)^2}{n}} \quad (5)$$

$$\%RMSE = \frac{RMSE}{mean\left(GEDI_{RH98_{sim}}\right)} \times 100 \quad (6)$$

$$\%bias = \frac{Meanbias}{mean\left(GEDI_{RH98_{sim}}\right)} \times 100 \quad (7)$$

3. Results

3.1. GEDI geolocation accuracy

After co-locating the GEDI and ALS data using the collocateWaves program in gediSimulator, the coordinate offsets of all GEDI footprints provided estimates of systematic geolocation errors in Euclidian distance. The geolocation of GEDI tracks was accurately correlated with the ALS-derived GEDI waveforms. The mean geolocation error of the Version 2 GEDI data was 12 m. These results correspond with the expected geolocation (horizontal) accuracy of 10.3 m for Version 2 (Beck et al., 2021).

3.2. GEDI algorithm setting groups

The R^2 of the relationship between GEDI-$RH98_{orb}$ and GEDI-$RH98_{sim}$ (Table 3) for each SG indicates that SG 4-6 performed much poorer than SG 1-3, with relatively low R^2 values below 0.4. SG 5 had the lowest R^2 of 0.129, while SG S had the highest R^2 of 0.48. SG S is based on the prediction of the best SG from SGs 1-6 for individual laser shots depending on plant functional type, geographic region, and laser return energy.

The %bias of GEDI-$RH98_{orb}$ was calculated for 1 m intervals of GEDI-$RH98_{sim}$ (Fig. 4). Compared to the other SGs, SG 4 had the most pronounced negative bias of -20% to -45%, where $RH98_{orb}$ was consistently underestimating $RH98_{sim}$ above 3 m (Fig. 4d). This may be attributed to the high front threshold of SG 4 (6 sigma) that leads to higher fraction of the low amplitude canopy top signal being excluded, resulting in an underestimation of the canopy top (Fig. 5e). SG 5 significantly overestimated RH98 in low height bins below 8 m (Fig. 4e) due mainly to the small back threshold value (2 sigma) compared to SG 3 (6 sigma), SG 6 (4 sigma), and SG 2 (3 sigma) (Table 2) (Adam et al., 2020; Beck et al., 2021; Fayad et al., 2021). This is often caused by the lowest detected mode being noise rather than the ground leading to an underestimation of ground elevation (Beck et al., 2021) (Fig. 5a). SG 3 and SG 1 have

similar threshold settings but different Gaussian smoothing widths. Some samples of SG 1, with the larger smoothing width (6.5 ns), had much higher lowest detected modes than SG 3 (Fig. 5c, 6f). This suggests that SG 1's lowest detected modes were convolved by its larger Gaussian smoothing width of 6.5 ns, resulting in a reduced number of detected modes that, in turn, often caused the highest detected mode to be misclassified as the ground. SG 1 and SG S results were similar because SG 1 was selected by SG S in ~92% of data. Due to the overall best performance of SG S all results reported in the remainder of this study are based on the SG S.

Table 3. R^2 values of the linear models of $GEDI\text{-}RH98_{orb}$ and $GEDI\text{-}RH98_{sim}$ for each algorithm setting group.

Setting group	R^2
1	0.479
2	0.448
3	0.455
4	0.341
5	0.129
6	0.361
S	0.48

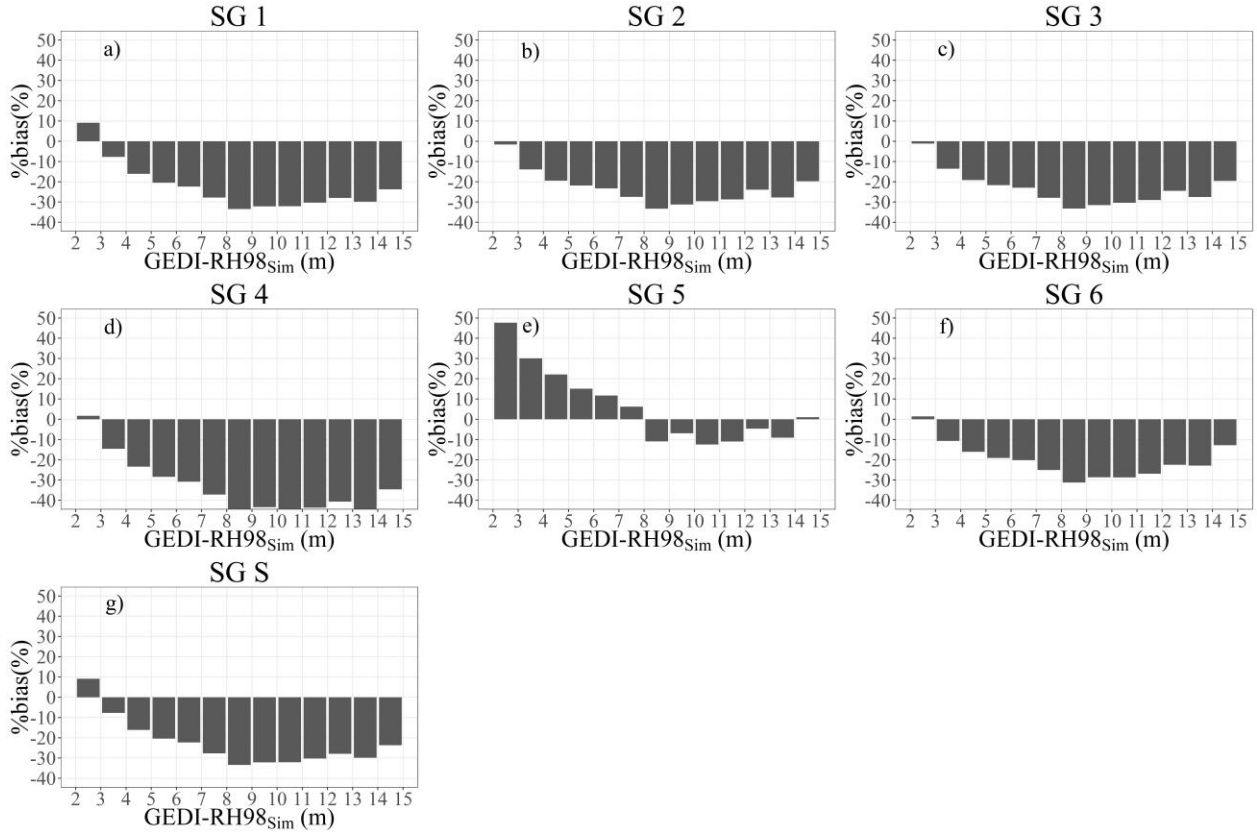


Fig. 4. The % bias of various GEDI algorithm setting groups (a-g) within 1m bins of $RH98_{sim}$. (2-column fitting image)

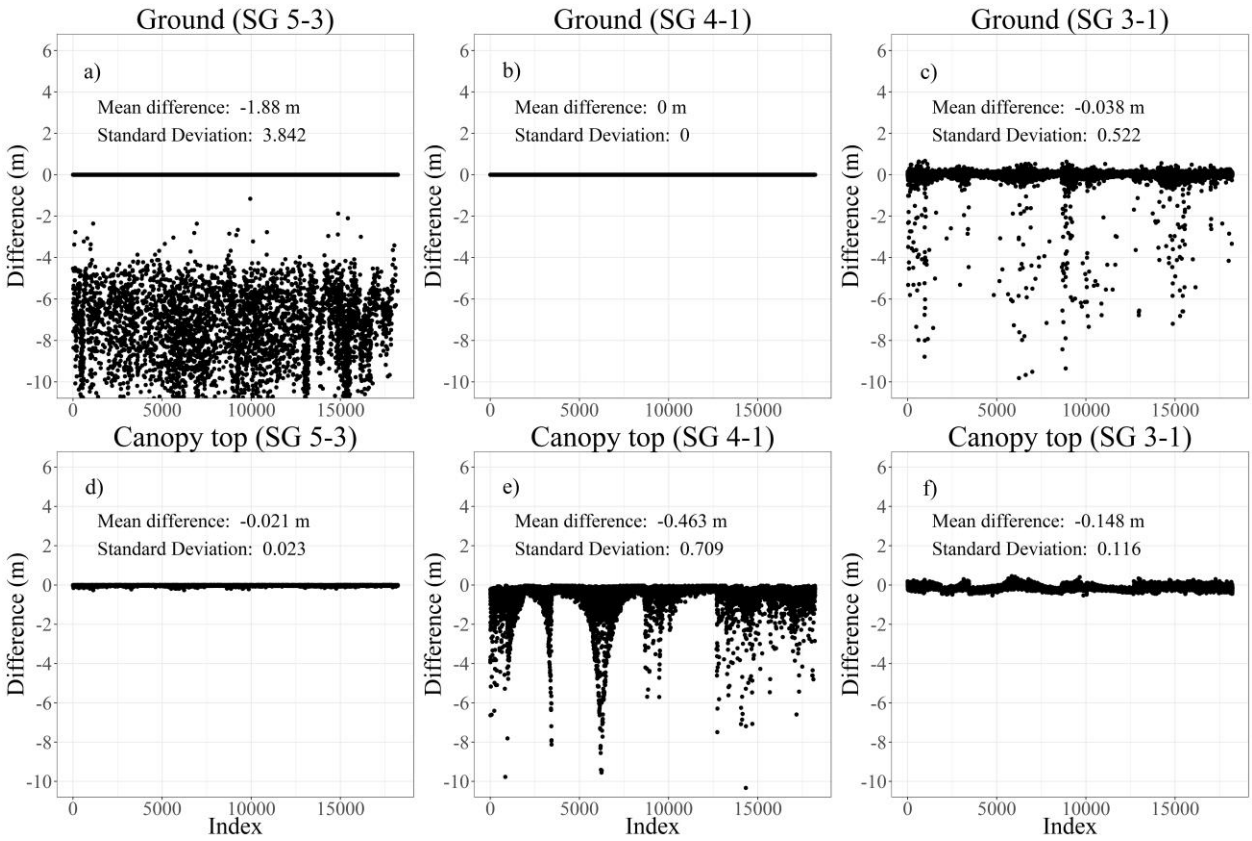


Fig. 5. The difference of ground and canopy top estimations between GEDI algorithm setting groups to illustrate the impact of different parameter settings. Back threshold (SG 5: 2 sigma, SG 3: 6 sigma) (a, d); front threshold (SG 4: 6 sigma, SG 1: 3 sigma) (b, e); and second Gaussian smoothing width (SG 3: 3.5 sigma, SG 1: 6.5 sigma) (c, f). (2-column fitting image)

3.3. GEDI sensitivity ~~and ALS canopy cover~~

The number of power and coverage beam GEDI footprints collected in nighttime conditions were 6301 and 6436, respectively. However, the daytime counts were much less, 1482 and 1602 for power and coverage beams, respectively. The large number of good quality GEDI footprints collected during the night can be explained by the lower background noise compared to the daytime acquisitions. The results are consistent with the reduced sensitivity of the day and night coverage beam (92%, 93%) compared to the day and night power beam (96%) (Fig. 6). The sensitivity of the night power beams were approximately 97% and did not reach the 99% stated by ~~(~~Hancock et al., 2019~~)~~, presumably due to the influence of atmospheric transmission and the relatively high front and back thresholds for SG 1 (Hancock et al., 2019). A Welch's Two Sample t-test in R (package: stats) (Team, 2013) was conducted on 100 randomly sampled data to investigate if the sensitivity values of the beam types differed significantly. The result indicated that the coverage beams and power beams had significantly different sensitivity values, while the coverage

day and coverage night or power day and power night did not. There was no difference in bias for power vs. coverage beam or day vs. night acquisitions with a mean bias between 0 m and 0.5 m (data not shown).

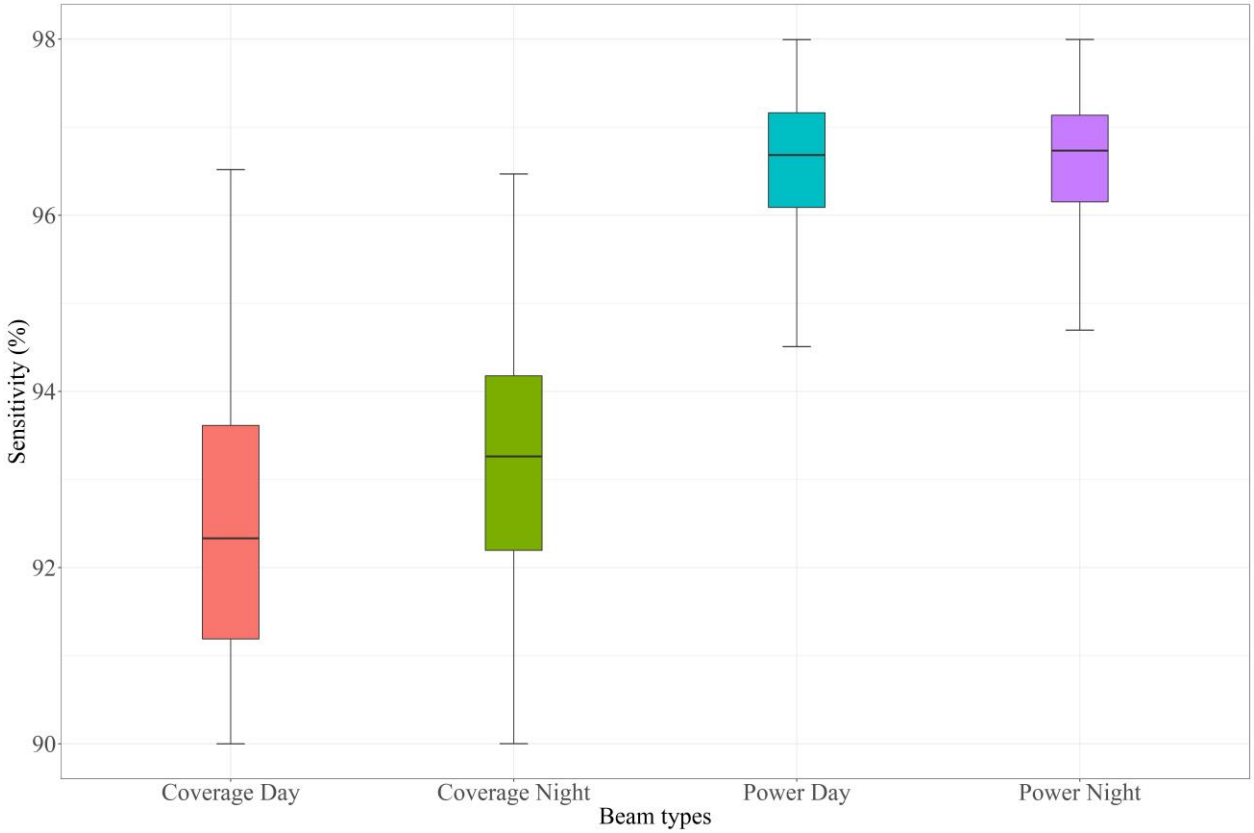


Fig. 6. Sensitivity of GEDI beam types (coverage vs. power) combined with time of acquisition (day vs. night). (Single fitting image)

3.4. Influence of phenology on GEDI-RH98$_{orb}$ error

The samples (footprints) of all test cases were pooled together in a single, combined analysis without considering phenological conditions. Overall, there was a positive relationship between GEDI-RH98$_{orb}$ and GEDI-RH98$_{sim}$, with R^2 of 0.48 (Table 4), and the GEDI-RH98$_{orb}$ observations underestimated GEDI-RH98$_{sim}$, with a bias of -1.02 m, %bias = -19.48%, RMSE = 1.93 m, %RMSE of 36.7%.

After splitting the data based on phenological conditions, the leaf-on and leaf-off test cases had an R^2 of 0.61 and 0.43, respectively (Fig. 8, Table 4). For leaf-on test cases, the mean bias was -0.55 m (bold red line in Fig. 8a). Leaf-off test cases, on the other hand, underestimated GEDI-RH98$_{sim}$ with a mean bias of -1.47 m. Leaf-on and leaf-off test cases had RMSE of 1.48 m and 2.27 m, respectively, indicating that leaf-off conditions lead to nearly 34% higher RMSE (Fig. 8b). The error measures of leaf-on condition were all lower than that of leaf-off condition, which

indicates that for the selected algorithm setting in the Version 2 product, GEDI estimates are more likely to detect the canopy top in leaf-on conditions. The impact of transitional phenological conditions (spring or autumn) could not be sufficiently assessed, as there were only five such cases. There was a weak positive relationship between the bias of test cases and relative greenness (R^2 = 0.32) (Fig. 7), suggesting again that the GEDI-RH98$_{sim}$ was more representative of the canopy top when the trees had full, green canopies, than during drier, leaf-off conditions when the LiDAR pulse may be penetrating the denuded canopy to interact with lower parts of the trees.

When examining the difference between the GEDI and ALS derived ground and canopy top estimates separately, the ground bias was ~~almost~~ close to 0 regardless of phenological impacts (Fig. 9c, 9d). The canopy top estimates were underestimated in both leaf-on and leaf-off conditions, but~~and in~~ leaf-on conditions the mean bias of canopy top estimates (-2.64 m) was about 1 m smaller than leaf-off conditions (-3.7 m) (Fig. 9c, 9d). A Welch's Two Sample t-test (Team, 2013) was implemented on 100 random samples of data each group and demonstrated a significant difference ($p < 0.05$) between the canopy top estimates across all scenarios. However, the mean values of ground estimates were not significantly different.

Table 4. Summary of comparison between GEDI-RH98$_{orb}$ vs. GEDI-RH98$_{sim}$ for all savanna sites and test cases under leaf-on and leaf-off conditions.

Grouping	R^2	Bias (m)	%bias (%)	RMSE (m)	%RMSE (%)	Sample size~~GEDI Footprints~~
Combined analysis, all test cases	0.48	-1.02	-19.48	1.93	36.7	22813
Combined analysis, leaf-on	0.61	-0.55	-11.15	1.48	29.85	7347
Combined analysis, leaf-off	0.43	-1.47	-26.53	2.27	40.94	8933

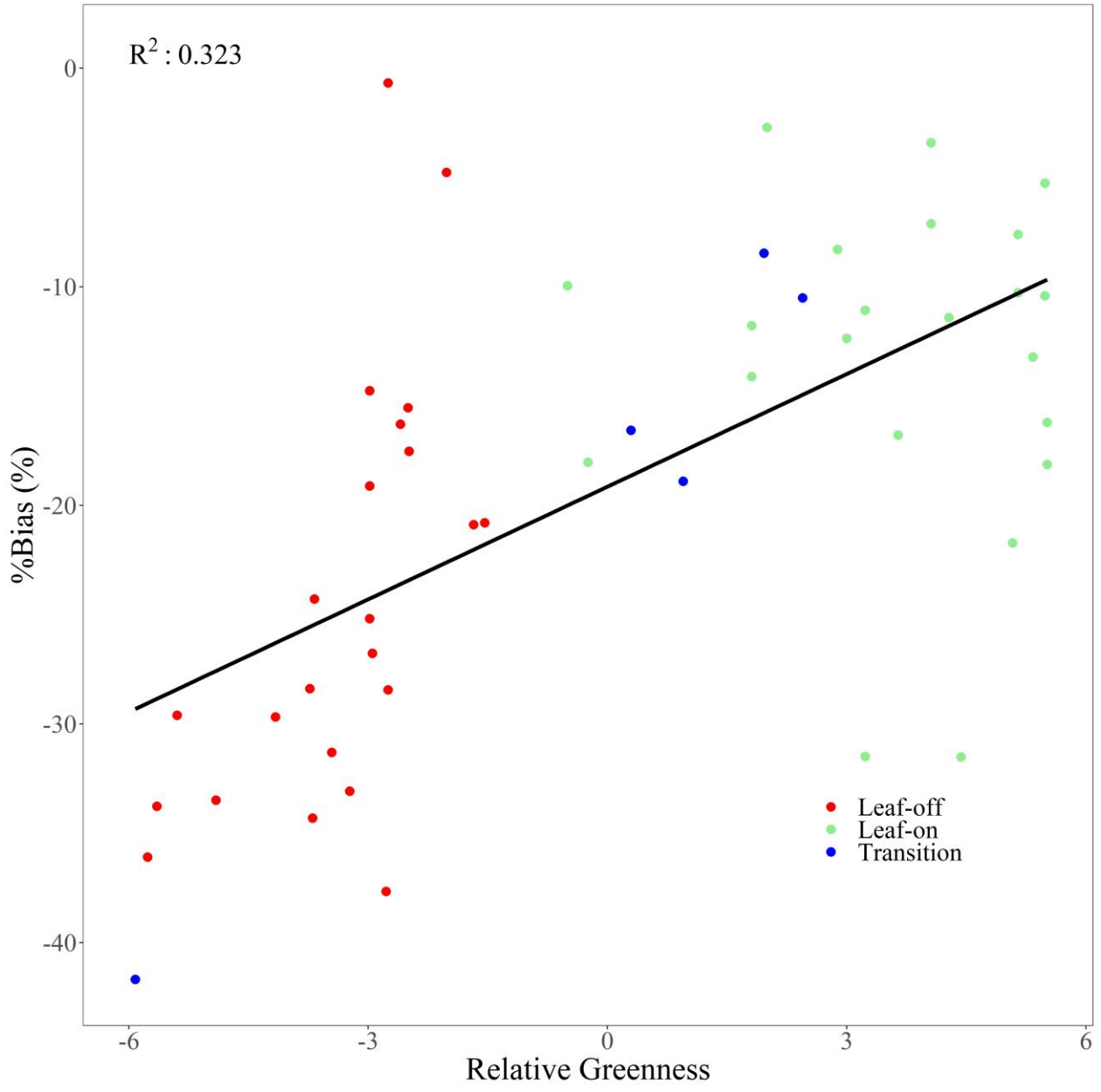


Fig. 7. The bias of test cases vs. MODIS relative greenness on the date of $GEDI_{orb}$ acquisition. A reduction in negative bias in RH98orb with increased greenness is evident. (Single fitting image)

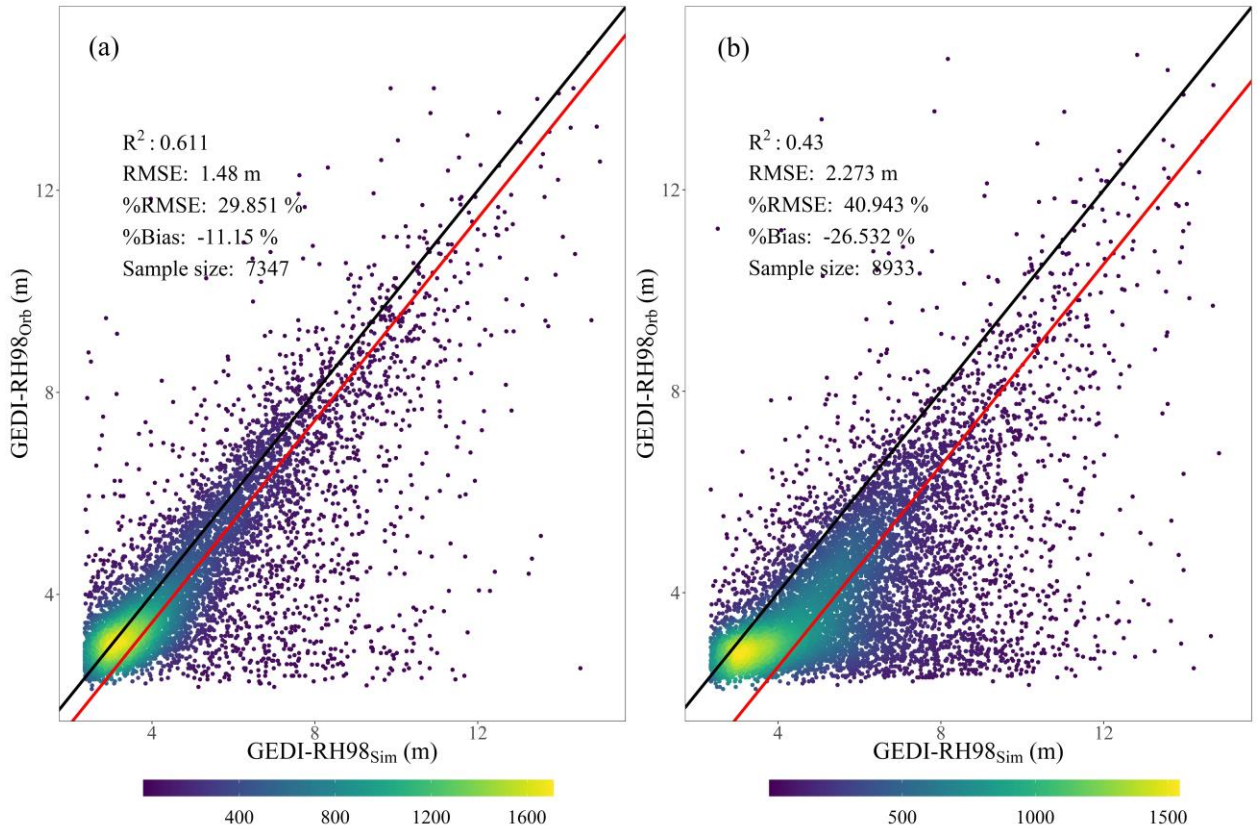


Fig. 8. Density plots of $RH98_{sim}$ vs. $RH98_{orb}$ below 15 m for all test cases with leaf-on (a) and leaf-off (b) conditions. The black line is the diagonal reference line (slope = 1, intercept = 0) and the red line represents the mean bias. (2-column fitting image)

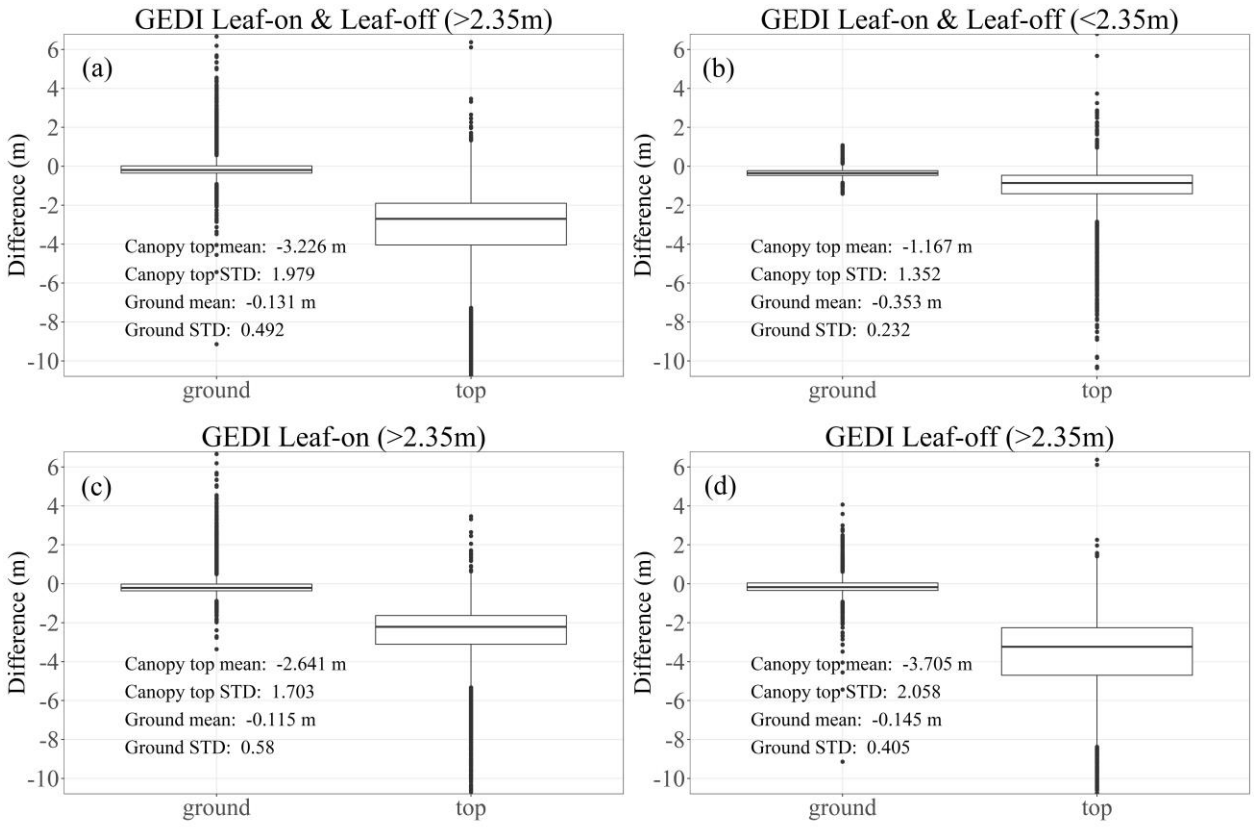


Fig. 9. The difference between GEDI and ALS derived ground and canopy top elevations for four scenarios: leaf-on & leaf-off (>2.35m) (a), leaf-on & leaf-off (<2.35m) (b), leaf-on (>2.35m) (c), and leaf-off (>2.35m) (d). The RH98 bias was mainly due to the underestimation of canopy top elevation. (1.5-column fitting image)

3.5. GEDI-RH98$_{orb}$ Bias distribution with RH98 and canopy cover

Fig. 10 gives the relative frequency of $RH98_{sim}$ values across all sites as context to the bias distribution. The bias of all individual GEDI footprints was investigated across RH98 height bins to determine any trends, stratified according to leaf-on and leaf-off conditions (Fig. 11a, 11b). For both leaf-on and leaf-off conditions, the variability of the bias generally increased with the $RH98_{sim}$ heights. Under leaf-on conditions, in the 2.35-3 m and 3-4 m $RH98_{Sim}$ bins, containing 35% of samples (Fig. 10), the mean bias (and %bias) was +0.31 m (+11.6%) and –0.17 m (-4.9%) respectively. The negative bias remained around 0.8 m (-10%), up to 7 m, which includes approximately 80% of the data. Above 7 m the bias peaked at -2.7 m (–20%) at $RH98_{sim}$ of 12-13 m and then steadily reduced to –2.5 m (-10%) at 15 m. The bias above 15 m was not reported as these heights represented only 0.16% of the samples.

The bias under leaf-off conditions was almost double that of leaf-on above 4 m and peaked at -4.6 m (-40%) for $RH98_{sim}$ 8-9 m (Fig. 11). During leaf-off conditions, severe underestimations of up to –6 m (~-30%) occurred at higher RH98 between 10 and 15 m. The canopy cover (calculated from ALS data) did not have a significant

influence on the bias of $RH98_{orb}$ (leaf-on) which remained close to a mean of –0.5m between 20-90% cover and increased slightly to –0.7 between 90-100% cover (Fig. 11e).

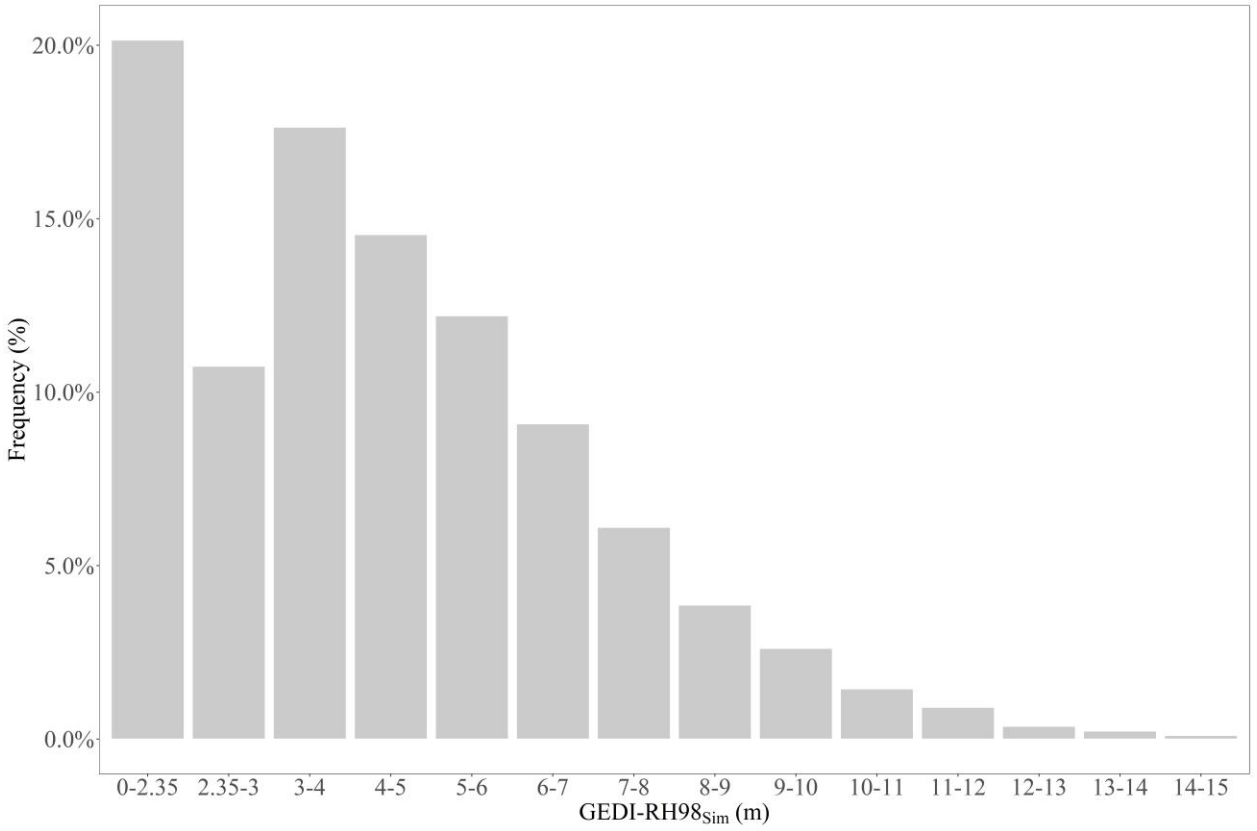
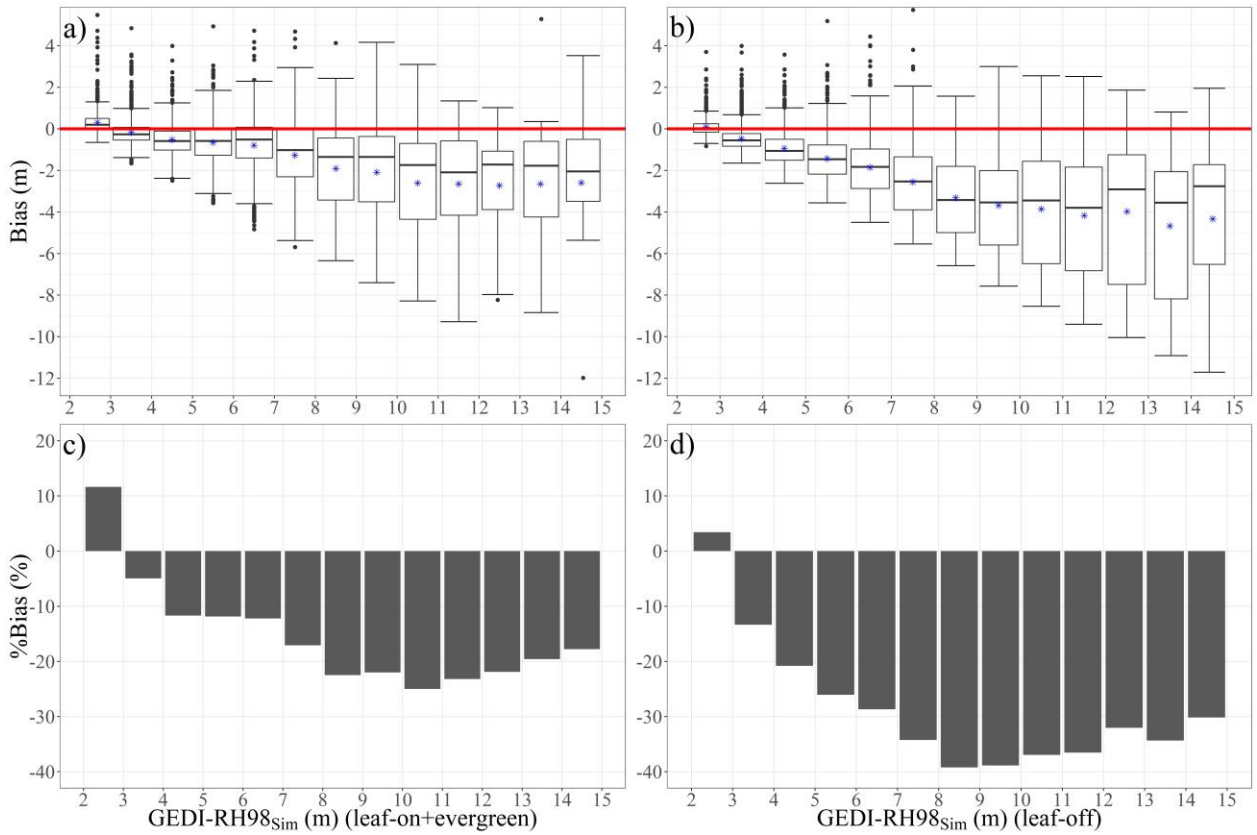
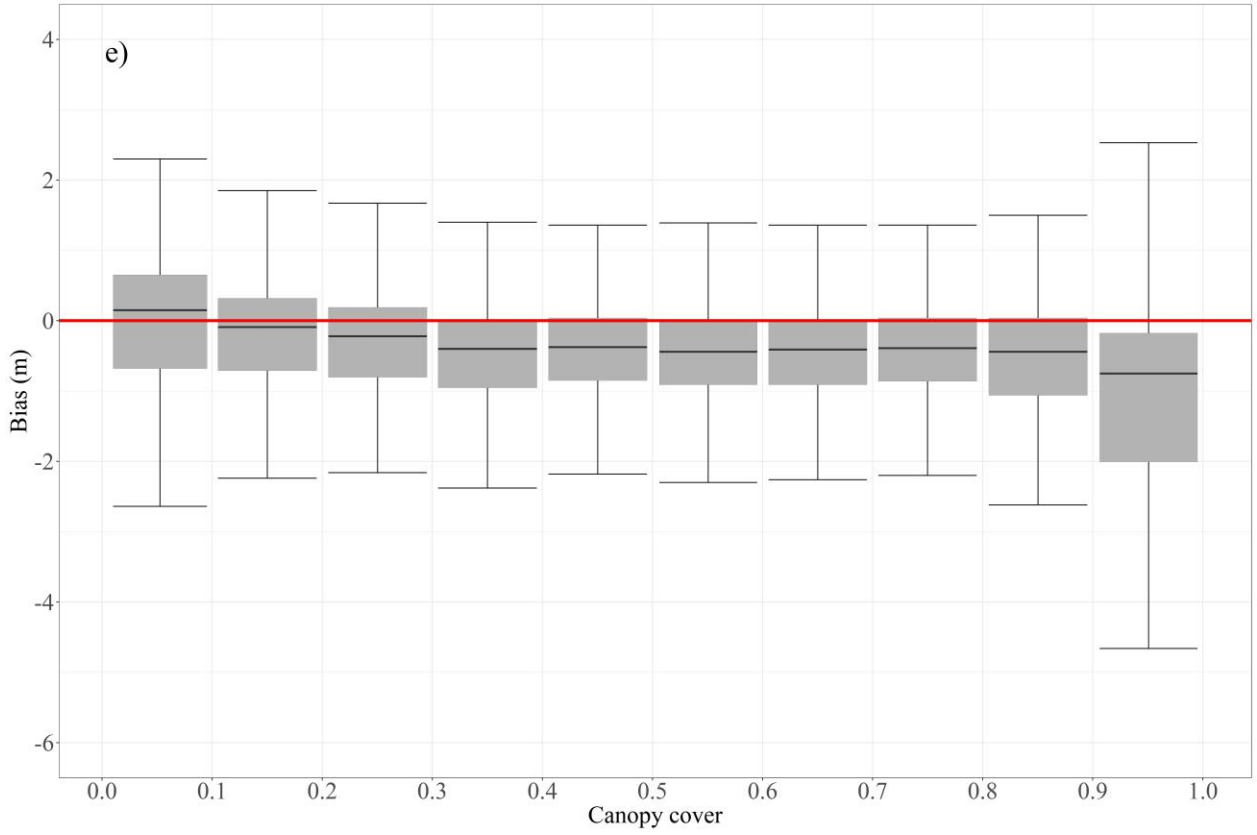


Fig. 10. Histogram of the relative frequency of $RH98_{sim}$ values of all GEDI footprints across all study sites (0-15 m). (single fitting image)


a)
b)
c)
d)
Bias (m)
%Bias (%)
GEDI-RH98Sim (m) (leaf-on+evergreen)
GEDI-RH98Sim (m) (leaf-off)

407


e)
Bias (m)
Canopy cover

408

Fig. 11. Bias (a, b), and %bias (c, d) vs. GEDI-RH98$_{sim}$ (0-15 m) for leaf-on (left) and leaf-off (right) test cases. Mean Bias values are indicated with blue asterisks. The lowest bin range is 2.35 – 3 m. RH98 bias (leaf-on) vs. airborne LiDAR (ALS) canopy cover (e). (2-column fitting image)

3.6. Error assessment of RH98 for individual study sites

The error of individual sites and their test cases was examined to determine how accurately GEDI can characterize the canopy height of smaller study areas with contrasting vegetation structures (Table 5). There was some variability in errors between test cases of the same site. This could partially be attributed to variability in the specific area of a site that was sampled by each GEDI track and the varied number of good quality data samples (footprints) intersecting the site (Table 5). For all the individual savanna sites (notably D'Nyala, Agincourt, Venetia) the leaf-on test cases had much lower RMSE, smaller negative bias, and higher R^2 compared to the leaf-off test cases of the same site (Table 5). The leaf-on test cases of all the savanna sites had R^2 ranging from 0.43 to 0.71, %bias from -12% to -2.7%, RMSE from 0.76 to 2.1m and %RMSE from 6% to 34%.

The Addo site with evergreen thicket vegetation had a low average ALS CHM height of 3.9 m, but a high average ALS canopy cover of 66% (according to ALS CHM) (Table 1). It produced a relationship with an R^2 of 0.42, mean bias of –0.08 m, %bias of -2.78%, and %RMSE of 6.75% (Fig. S2). This indicates that GEDI RH98 was fairly accurate in this dense, short stature vegetation above 2.35 m. However, it must be noted that within the Addo site a very large proportion of vegetation is below 2.35 m (16% of RH98$_{sim}$ < 2.35m), that cannot be accurately estimated due to the long GEDI pulse width (see Discussion for details). In contrast to the Addo site, the tall, evergreen forest of Dukuduku with an average ALS CHM of 9.7 m and 70% canopy cover, had an R^2 of 0.521, mean bias of –0.91 m, %bias of -12.5%, and %RMSE of 43.6% (Fig. S3). Overall, the error of GEDI RH98 did not show specific patterns related to sites or vegetation types and provided reliable estimates of canopy heights within individual sites during leaf-on conditions**.**

Table 5. Summary of comparison between GEDI-RH98$_{orb}$ vs. GEDI-RH98$_{sim}$ for each study site grouped by phenological status.

Study sites	**Mean R^2**	**Mean bias (m)**	**Mean %bias[1] (%)**	**Mean RMSE (m)**	**Mean %RMSE[2] (%)**	~~**GEDI footprints**~~**Sample size**
Addo (evergreen)	0.428	-0.08	-2.78	0.203	6.75	2133
Dukuduku (evergreen)	0.521	-0.91	-12.55	3.11	43.63	1262
Agincourt	0.486	-0.53	-8.93	1.852	31.30	479

(leaf-on)						
Agincourt (leaf-off)	0.167	-2.1	-36.08	3.089	51.08	337
D'Nyala (leaf-on)	0.712	-0.168	-2.7	1.413	22.82	86
D'Nyala (leaf-off)	0.157	-1.905	-28.43	2.72	40.69	896
Ireagh (leaf-on)	0.512	-0.571	-11.72	1.61	33.51	780
Ireagh (leaf-off)	0.352	-1.503	-31.54	2.295	47.92	736
Venetia (leaf-on)	0.611	-0.459	-11.94	0.76	19.81	2038
Venetia (leaf-off)	0.373	-0.831	-21.57	1.162	30.21	592
Limpopo (leaf-on)	0.438	-0.901	-14.09	2.202	34.37	677
Limpopo (leaf-off)	0.476	-1.514	-25.94	2.313	40.47	2330
Justicia (leaf-on)	0.495	-1.165	-21.88	2.183	41.06	1322
Justicia (leaf-off)	0.492	-1.387	-26.98	2.276	44.28	1319
Welverdiendt (leaf-on)	0.536	-0.608	-12.35	1.544	31.738	1965
Welverdiendt (leaf-off)	0.417	-1.221	-22.11	1.836	33.35	2723

4. Discussion

This study provided a baseline validation of Version 2 GEDI RH98 in predominantly sparse (30-60% canopy cover), short stature savannas vegetation and to our knowledge, was the first to do so at the time of writing. This study compared the on-orbit GEDI RH98 with the simulated GEDI RH98 derived from ALS point cloud using the gediSimulator (Hancock et al., 2019), and as such allows for a direct and rigorous comparison (Roy et al., 2021). Recent studies that have calculated the quantiles of canopy heights from a horizontal aggregation of equally weighted ALS CHM values within a GEDI footprint have not directly accounted for factors such as geolocation error, laser pulse shape and the distribution of energy within the footprint (Liu et al., 2021; Potapov et al., 2021), thus confounding interpretation of differences. In contrast to ALS CHM's, the GEDI RH98 is the 98th quantile of the energy returned by the vertical profile of vegetation and includes a large fraction of energy returned from the ground, especially in sparse savannas. We recommend that studies aimed at validation of GEDI footprint elevation and height

data products consider the separation of differences resulting from waveform processing and definition of canopy structure.

The Version 2 GEDI data products have a much-improved geolocation accuracy since systemic errors have been minimized as part of standard processing (Beck et al., 2021). This study found Version 2 GEDI products contained a mean geolocation error of approximately 12 m. However, these results were limited to GEDI orbits intersecting the study areas, and this study was therefore not a comprehensive assessment of GEDI geolocation error, as conducted elsewhere (Roy et al., 2021). The present study only accounted for systematic geolocation error of the GEDI tracks, however relatively small random geolocation errors remain for individual footprints, conservatively estimated to be within 2 m, contributing to the overall error in on-orbit and simulated GEDI comparisons. These random geolocation errors have not been formally quantified by the GEDI Science Team, so were conservatively estimated to be within 2 m (~10% of the footprint diameter). To quantify the contribution of this random geolocation error, 100 random offsets were generated by sampling from a random uniform distribution between 0 and 2 m and added to the 100 leaf-on GEDI footprint coordinates across all sites, before simulating the waveforms. The difference between the relocated and corrected RH98, along with various error metrics were calculated (for details see Supplementary Material). For RH98, the results suggest that these random geolocation errors contribute 14% of absolute bias below 8 m and 20.7% of the overall RMSE. Therefore, a minor but non-trivial component of the reported uncertainty can be attributed to random geolocation errors, rather than error in vertical canopy height estimates. ~~In addition, t~~Thehis ~~s~~present ~~study only accounted for systematic geolocation error of the GEDI tracks, while~~ 0–2m ~~random geolocation errors for individual footprints remain in the data, contributing to the overall error in on-orbit and simulated GEDI comparisons.~~

Recent GEDI studies confirmed that the SGs using different algorithm parameter settings performed differently in various vegetation types, as were expected (Adam et al., 2020; Beck et al., 2021; Hofton et al., 2020) (Table 2). In the current study, SG 5 performed poorly because it has the lowest back threshold that potentially causes confusion between noise and the start of the ground return, leading to significant overestimation of RH98 (Fig. 5a). This was expected, as SG 5 was designed specifically for dense vegetation where there is difficulty in estimating the peak ground return because of the lower laser penetration through the canopy. SG 4 had the most pronounced negative bias of –20 to –45% for $RH98_{sim}$ above 3m, which was caused by the high front threshold of 6 sigma that leads to truncation of the canopy top signal (Fig 6e). Version 2 of GEDI algorithms automatically select the best SG for each

GEDI footprint and indicates the selected algorithm in GEDI L2A data products. Thus, SG S provided the best overall performance, slightly better than SG 1, which was the most frequently selected by the SG S dataset for our study.

The GEDI power beams have a stronger energy pulse than the coverage beams as they were designed to provide more stable GEDI ground estimates in forests with canopy cover above 95% (Duncanson et al., 2020; Hancock et al., 2019). Although the present study confirmed the higher sensitivity of the power beam compared to the coverage beam (Fig. 6), there was no difference in the bias and RMSE of $RH98_{orb}$ between the two beam types. Daytime GEDI acquisitions are impacted by background solar illumination noise, while nighttime acquisitions are not (Beck et al., 2021). This resulted in a much larger number of good quality nighttime than daytime footprints in the data. However, the time of acquisition (day vs. night) did not have any influence on the bias or RMSE of $RH98_{orb}$. Furthermore, because the minimum sensitivity was set to 0.9, the ability of the GEDI power beam to penetrate high canopy cover of up to 90% brought no advantage in the present study where ALS canopy cover seldom exceeded 70%.

The prevailing phenological conditions of vegetation at the time of GEDI data acquisition had a very large influence on the error of $RH98_{orb}$. During leaf-on conditions for savanna vegetation with $RH98_{sim} < 15$ m, $RH98_{orb}$ was very accurate with $R^2 = 0.61$, mean bias = -0.55 m, %bias = -11.1%, RMSE = 1.64 m and %RMSE = 29.8%. During leaf-off conditions, $RH98_{orb}$ was less accurate with $R^2 = 0.43$, mean bias = -1.47 m, %bias = -26.5%, RMSE = 2.03 m and %RMSE = 40.9% (Table 4). Therefore, in leaf-off conditions, the mean bias of GEDI-$RH98_{orb}$ was about three times higher than during leaf-on conditions. Leaf-off conditions had significant underestimations of up to –6 m (~30%) at higher RH98 between 10 and 15 m, suggesting that the height of tallest deciduous trees would be significantly underestimated (Fig. 11). The influence of phenology is further underscored by the positive relationship ($R^2 = 0.32$) between relative greenness and bias (Fig. 7). The results clearly indicate that GEDI is better able to detect the highest canopies when they are fully covered in green leaves. After the trees have shed their leaves, the LiDAR signal presumably penetrates further through the denuded branches before interacting with the lower strata of the trees. The signal at the start of the waveform may therefore be weaker and truncated as noise, leading to a large negative height bias. Therefore, assessments of deciduous vegetation structure should ideally be conducted during leaf-on periods.

Overall, the results for leaf-on conditions agreed with recent publications on initial comparisons of GEDI RH metrics and ALS CHMs. Potapov et al., (2021) compared GEDI-RH90 metrics with 90th percentile distribution of ALS-derived height metrics at a global scale, with a similar R^2 of ~0.7 and bias of –0.7 m. Adam et al., (2020) tested the relationship between ALS and GEDI-derived canopy height models (CHMs) and found that GEDI-derived CHM's underestimated canopy heights in temperate forests, with a median CHM-difference of 0.23m. More recently, Liu et al., (2021) compared GEDI "canopy heights" (RH98) and canopy heights (98th quantiles calculated from ALS CHM values within GEDI footprints), across a wide range of vegetation types, with an R^2 of 0.82, bias of -0.8 m and %RMSE of 30%. Wang et al., (2022) compared the on-orbit GEDI RH100 with the simulated GEDI RH100 from ALS datasets across the US, with an R2 of 0.83, bias of -0.6 m and RMSE of 3.09 m. Ilangakoon et al., (2021) recently observed that GEDI RH98 (version 1) significantly overestimated canopy heights of low (<5m), semi-arid vegetation and suggested further investigation of GEDI's performance in shrublands. In the only other published study to simulate GEDI waveform and metrics (RH95) from ALS data collected five years earlier, Roy et al., (2021) demonstrated that geolocation uncertainty of GEDI Version 1 footprints should account for significant error in canopy height estimation (best R^2 = 0.59 after geolocation correction) in secondary tropical forests. The best performance attained using ICESat-2 in boreal forests (strong beam/night/summer acquisitions), underestimate canopy height by 0.56 m, with %bias of 3.18%, RMSE% of 13.75%, for canopy cover 40-85% (Neuenschwander et al., 2020). However, the residual error of these ICESat-2 height estimates increased rapidly with canopy cover below 40%, reaching 1.7 m at 20% canopy cover. While the results of the present study and this ICESat-2 study were not directly comparable, future research should directly compare the accuracy of coincident GEDI and ICESat-2 height metrics in savannas.

The investigation of the bias within specific RH98 bins was essential to understanding GEDI's under and overestimations across various canopy heights and canopy cover conditions (Fig. 11). When considering only leaf-on conditions, in the lowest $RH98_{sim}$ bin of 2.35-3 m, the mean bias was +0.31 m (+11.6%). This overestimation could be related to the GEDI pulse width of 15.6 ns, the waveform smoothing width of 6.5 ns of SG 1, and the resulting minimum height value for RH98 of 2.34 m over flat, bare ground, that is discussed in more detail below. In the 3-4 m bin accounting for 25% of all the data (Fig. 10), the mean bias was only –0.17 m or -4.9%. The mean negative bias remained less than 0.79 m or -10%, up to 7 m, which suggest that GEDI's $RH98_{orb}$ is very accurate between 3 and 7 m. This range incorporates a very large portion of canopy heights in savannas and 80% of our study sites (Fig. 10).

The mean bias increased from -1.2 m (-17%) to –2.7 m (-20%) from 7 to 13 m (Fig. 11). This negative bias could be due to the absence of GEDI signal noise in GEDI simulator, leading to the underestimation of canopy top elevations (Fig. 9c). There was no relationship between canopy cover (calculated from ALS data) and bias of $RH98_{orb}$ (leaf-on) that remained close to a mean of –0.5m between 30-90% cover (Fig. 11e). Notably, below 30% cover, the bias decreased to –0.2 m (Fig. 11e). GEDI's RH98 estimates, therefore, remained consistently accurate even with sparse canopy cover below 30%, which is often prevalent in savannas. This was in contrast with the findings for ICESat-2 which reportedly does not capture enough canopy reflection for accurate canopy height estimates when canopy cover was < 40%, leading to a positive bias of up to 0.5 m at 20% canopy cover (Neuenschwander et al., 2020). Other GEDI studies suggest that canopy cover only starts reducing canopy height estimation accuracy when it exceeds 90% and GEDI cannot effectively detect the ground (Liu et al., 2021).

Given the GEDI pulse length of 15.6 ns, RH98 has a theoretical minimum value of half the pulse width of 2.34 m over flat, bare ground, which is similar to the height of shrubs (< 3 m) that cover substantial areas of savannas (O'Connor et al., 2014). In the present study, approximately 20% of GEDI footprints had $RH98_{sim}$ < 2.35 m and 13.6% between 2.35 m and 3 m, where characterizing the vegetation canopy height can be challenging (mean bias = -1.16 m) (Fig. 9b). Characterizing low vegetation over the predominantly flat savannas may therefore benefit from more intricate deconvolution of the waveform (McGlinchy et al., 2014; Neuenschwander, 2008) which is a topic of on-going research. Moreover, the measurement of shrub vegetation may be approaching GEDI's limits of detectability, which is determined not only by the pulse duration, but by multiple factors including cover, target reflectivity, topographic relief and noise (Adam et al., 2020; Liu et al., 2021). This study provided crucial insight into the lower limits of canopy height detection that impacts the measurement of shrub cover which is essential to ecological studies and biomass estimation in savannas~~Measuring shrub cover is, however, essential to ecological studies and rangeland management in savannas~~ (Mograbi et al., 2015; O'Connor et al., 2014; Stevens et al., 2016; Venter et al., 2018) and may require different approaches to waveform interpretation and estimation of other GEDI metrics, e.g., canopy cover fraction, plant area index. GEDI is a sampling mission with footprints separated 60 m along track and 600 m across track (Dubayah et al., 2020) (Fig. S4), and therefore does not provide continuous measurements akin to a conventional high resolution ALS CHM used in small area ecological studies. These GEDI footprint samples (L2 data), as well as the L3 and L4B gridded products (1 km x 1 km) can, however, provide unique broad-area metrics of woody vegetation structure and aboveground biomass estimates (Dubayah et al., 2022;

Duncanson et al., 2022) for regional ecological studies investigating, e.g. the drivers of structure changes (Sankaran et al., 2008), impact of fire regimes (Smit et al., 2016, 2010) and carbon storage of savannas (Ross et al., 2021). The research community is actively engaged in using GEDI, ICESat2 and the combination of the two space-based LiDAR sensors to address these pressing ecological questions.

5. Conclusions

This study provided the first~~a~~ baseline validation of Version 2 GEDI RH98 canopy height estimates in African savannas by simulating the GEDI waveform from ALS data. The results provide valuable insights into the accuracy and precision of GEDI-RH98 across highly heterogeneous, short stature vegetation. The time of acquisition (day vs. night) and beam type did not have any influence on the bias or RMSE of RH98. The RH98 accuracy and bias was not influenced by canopy cover and remained consistently accurate even with sparse canopy cover below 30%. Woody vegetation phenology had a large influence on results. During leaf-off conditions, the mean bias of GEDI-$RH98_{orb}$ was about three times higher and significantly underestimated high canopy heights (10-15 m) by an average of 3.75 m. Therefore, the use of GEDI data to assess deciduous vegetation structure should be limited to leaf-on periods. During leaf-on conditions for vegetation, typical of savannas ($RH98_{sim}$ < 15 m), $RH98_{orb}$ was accurate with R^2 = 0.61, mean bias = -0.55 m, %bias = -11.1%, RMSE = 1.48 m and %RMSE = 29.8%. Due to the challenge of differentiating the vegetation return from that of the ground with a pulse width of 15.6 ns, the Version 2 GEDI-RH98 data product cannot reliably be used for estimating canopy heights of shrubs below 2.34 m. However, algorithm development focused on savanna and shrubland ecosystems may extend the limits of detection for low stature vegetation. In leaf-on conditions, GEDI's $RH98_{orb}$ was very accurate between 3 and 7 m, with a mean negative bias of about -0.79 m or -10%, up to 7 m. The mean bias remained –2.5 m up to 15 m. Therefore, while shrubs below 3 m could not be accurately measured by GEDI's RH98, the canopy height of trees between 3 and 15 m were reliably estimated. While GEDI was designed for measurement of Earth's tropical and temperate forests and has limitations for characterizing the vertical structure of shrubs, RH98 can reliably characterize tree canopy heights across African and, potentially, global savannas.

Author's Responsibilities

XL processed all GEDI and ALS data, performed analysis and wrote the paper. KW conceptualized the research, secured ALS data acquisitions and wrote the paper. JA conceptualized the research, provided technical oversight and

wrote the paper. SH provided advanced technical input and edited the paper. R Mathieu conceptualized the research, secured ALS data acquisitions and edited the paper. R Main and LN processed the ALS data, provided study area expertise and edited paper. BE and RS secured ALS data acquisition, provided expertise on savanna ecology and edited the paper.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

The research was funded by NASA Carbon Monitoring System (Grant number 80NSSC21K0967), the CSIR Strategic Research Panel and the National Research Foundation (South Africa) Global Change Grand Challenge Grant (Grant number 110496). BE and RS was supported by the University of the Witwatersrand. XL was furthermore supported by GMU Presidential Scholarship. The funding sources were not directly involved in the research.

References

Adam, M., Urbazaev, M., Dubois, C., Schmullius, C., 2020. Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters. Remote Sensing 12, 3948.

Archibald, S., Scholes, R.J., 2007. Leaf green-up in a semi-arid African savanna - separating tree and grass responses to environmental cues. Journal of Vegetation Science 18, 583–594.

Asner, G.P., Vaughn, N., Smit, I.P.J., Levick, S., 2016. Ecosystem-scale effects of megafauna in African savannas. Ecography 39, 240–252. https://doi.org/10.1111/ecog.01640

Beck, J., Armston, J., Hofton, M., Luthcke, S., 2021. Global Ecosystem Dynamics Investigation (GEDI) Level 02 User Guide. Sioux Falls, South Dakota, USA: EROS Center, US Geological Survey.

Behera, S., Gupta, N., 2015. Utilization of vegetable waste for biomass production of some wild edible mushroom cultures 5.

Blair, J.B., Hofton, M.A., 1999. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. Geophysical Research Letters 26, 2509–2512. https://doi.org/10.1029/1999GL010484

Bond, W.J., Midgley, G.F., 2012. Carbon dioxide and the uneasy interactions of trees and savannah grasses. Philos. Trans. R. Soc. B-Biol. Sci. 367, 601–612. https://doi.org/10.1098/rstb.2011.0182

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., Asner, G.P., 2018. An above-ground biomass map of African savannahs and woodlands at 25m resolution derived from ALOS PALSAR. Remote Sens Environ 206, 156–173.

Brokaw, N.V.L., Lent, R.A., 1999. Vertical structure. In 'Maintaining Biodiversity in Forest Ecosystems'.(Eds I. Hunter and L. Malcom.) pp. 373–399. Cambridge University Press: Cambridge.

Camarretta, N., Harrison, P.A., Bailey, T., Potts, B., Lucieer, A., Davidson, N., Hunt, M., 2019. Monitoring forest structure to guide adaptive management of forest restoration: a review of remote sensing approaches. New Forests. https://doi.org/10.1007/s11056-019-09754-5

Ciais, P., Bombelli, A., Williams, M., Piao, S.L., Chave, J., Ryan, C.M., Henry, M., Brender, P., Valentini, R., 2011. The carbon balance of Africa: synthesis of recent research studies. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 369, 2038–2057. https://doi.org/10.1098/rsta.2010.0328

Davies, A.B., Gaylard, A., Asner, G.P., 2018. Megafaunal effects on vegetation structure throughout a densely wooded African landscape. Ecological Applications 28, 398–408. https://doi.org/10.1002/eap.1655

Dayaram, A., Harris, L.R., Grobler, B.A., Merwe, S. van der, Rebelo, A.G., Powrie, L.W., Vlok, J.H.J., Desmet, P.G., Qabaqaba, M., Hlahane, K.M., Skowno, A.L., 2019. Vegetation Map of South Africa, Lesotho and Swaziland 2018: A description of changes since 2006. Bothalia 49, 11. https://doi.org/10.4102/abc.v49i1.2452

Dean, W.R.J., Milton, S.J., Jeltsch, F., 1999. Large trees, fertile islands, and birds in arid savanna. Journal of Arid Environments 41, 61–78. https://doi.org/10.1006/jare.1998.0455

Didan, K., 2015. Mod13a3 modis/terra vegetation indices monthly l3 global 1km sin grid v006. NASA EOSDIS Land Processes DAAC.

Drake, Jason B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F., Prince, S., 2002. Estimation of tropical forest structural characteristics using large-footprint lidar. Remote Sensing of Environment 79, 305–319.

Drake, Jason B, Dubayah, R.O., Knox, R.G., Clark, D.B., Blair, J.B., 2002. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. Remote Sensing of Environment 81, 378–392. https://doi.org/10.1016/S0034-4257(02)00013-5

Dubayah, R., Armston, J., Healey, S., Bruening, J., Patterson, P., Kellner, J., Duncanson, L., Saarela, S., Göran, S., Yang, Z., Tang, H., Blair, B., Fatoyinbo, L., Goetz, S., Hancock, S., Hanson, M., Hofton, M., Hurtt, G., Luthcke, S., 2022. GEDI Launches a New Era of Biomass Inference from Space. Environmental Research Letters.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P., Qi, W., Silva, C., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Science of Remote Sensing 100002. https://doi.org/10.1016/j.srs.2020.100002

Dubayah, R.O., Sheldon, S.L., Clark, D.B., Hofton, M.A., Blair, J.B., Hurtt, G.C., Chazdon, R.L., 2010. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. Journal of Geophysical Research: Biogeosciences 115. https://doi.org/10.1029/2009JG000933

Duncanson, L., Kellner, J.R., Armston, J., Dubayah, R., Minor, D.M., Hancock, S., Healey, S.P., Patterson, P.L., Saarela, S., Marselis, S., Silva, C.E., Bruening, J., Goetz, S.J., Tang, H., Hofton, M., Blair, B., Luthcke, S., Fatoyinbo, L., Abernethy, K., Alonso, A., Andersen, H.-E., Aplin, P., Baker, T.R., Barbier, N., Bastin, J.F., Biber, P., Boeckx, P., Bogaert, J., Boschetti, L., Boucher, P.B., Boyd, D.S., Burslem, D.F.R.P., Calvo-Rodriguez, S., Chave, J., Chazdon, R.L., Clark, D.B., Clark, D.A., Cohen, W.B., Coomes, D.A., Corona, P., Cushman, K.C., Cutler, M.E.J., Dalling, J.W., Dalponte, M., Dash, J., de-Miguel, S., Deng, S., Ellis, P.W., Erasmus, B., Fekety, P.A., Fernandez-Landa, A., Ferraz, A., Fischer, R., Fisher, A.G., García-Abril, A., Gobakken, T., Hacker, J.M., Heurich, M., Hill, R.A., Hopkinson, C., Huang, H., Hubbell, S.P., Hudak, A.T., Huth, A., Imbach, B., Jeffery, K.J., Katoh, M., Kearsley, E., Kenfack, D., Kljun, N., Knapp, N., Král, K., Krůček, M., Labrière, N., Lewis, S.L., Longo, M., Lucas, R.M., Main, R., Manzanera, J.A., Martínez, R.V., Mathieu, R., Memiaghe, H., Meyer, V., Mendoza, A.M., Monerris, A., Montesano, P., Morsdorf, F., Næsset, E., Naidoo, L., Nilus, R., O'Brien, M., Orwig, D.A., Papathanassiou, K., Parker, G., Philipson, C., Phillips, O.L., Pisek, J., Poulsen, J.R., Pretzsch, H., Rüdiger, C., Saatchi, S., Sanchez-Azofeifa, A., Sanchez-Lopez, N., Scholes, R., Silva, C.A., Simard, M., Skidmore, A., Stereńczak, K., Tanase, M., Torresan, C., Valbuena, R., Verbeeck, H., Vrska, T., Wessels, K., White, J.C., White, L.J.T., Zahabu, E., Zgraggen, C., 2022. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. Remote Sensing of Environment 270, 112845. https://doi.org/10.1016/j.rse.2021.112845

Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., Silva, C.A., Armston, J., Luthcke, S.B., Hofton, M., Kellner, J.R., Dubayah, R., 2020. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. Remote Sensing of Environment 242, 111779. https://doi.org/10.1016/j.rse.2020.111779

FAO, 2012. Global ecological zones for FAO forest reporting: 2010 Update. FAO: Rome, Italy.

Fayad, I., Baghdadi, N.N., Alvares, C.A., Stape, J.L., Bailly, J.S., Scolforo, H.F., Zribi, M., Maire, G.L., 2021. Assessment of GEDI's LiDAR Data for the Estimation of Canopy Heights and Wood Volume of Eucalyptus Plantations in Brazil. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 7095–7110. https://doi.org/10.1109/JSTARS.2021.3092836

Fisher, J.T., Erasmus, B.F.N., Witkowski, E.T.F., van Aardt, J., Wessels, K.J., Asner, G.P., 2014. Savanna woody vegetation classification - now in 3-D. Appl Veg Sci 17, 172–184. https://doi.org/10.1111/avsc.12048

Fisher, J.T., Witkowski, E.T., Erasmus, B.F., Mograbi, P.J., Asner, G.P., Aardt, J.A., Wessels, K.J., Mathieu, R., 2015. What lies beneath: Detecting sub-canopy changes in savanna woodlands using a three-dimensional classification method. Applied Vegetation Science 18, 528–540.

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R., Dubayah, R., 2019. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. Earth and Space Science 6, 294–310. https://doi.org/10.1029/2018EA000506

Harding, D.J., Carabajal, C.C., 2005. ICESat waveform measurements of within- footprint topographic relief and vegetation vertical structure. Geophysical research letters 32.

Hill, M.J., Hanan, N.P., 2010. Ecosystem function in savannas: Measurement and modeling at landscape to global scales. CRC Press.

Hofton, M., Blair, J., Story, S., Tang, H., Silva, C., Armston, J., Dubayah, R., 2020. B029-03 - GEDI Measurements of Topography, Height and 3D Structure Measurements.

Hofton, M., Blair, J.B., 2019. Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products 44.

Hofton, M.A., Minster, J.B., Blair, J.B., 2000. Decomposition of laser altimeter waveforms. IEEE Trans. Geosci. Remote Sensing 38, 1989–1996. https://doi.org/10.1109/36.851780

Hyde, P., Dubayah, R., Peterson, B., Blair, J.B., Hofton, M., Hunsaker, C., Knox, R., Walker, W., 2005. Mapping forest structure for wildlife habitat analysis using waveform lidar: Validation of montane ecosystems. Remote Sensing of Environment 96, 427–437. https://doi.org/10.1016/j.rse.2005.03.005

Ilangakoon, N., Glenn, N.F., Schneider, F.D., Dashti, H., Hancock, S., Spaete, L., Goulden, T., 2021. Airborne and Spaceborne Lidar Reveal Trends and Patterns of Functional Diversity in a Semi-Arid Ecosystem. Frontiers in Remote Sensing 2.

Joubert, D.F., Rothauge, A., Smit, G.N., 2008. A conceptual model of vegetation dynamics in the semiarid Highland savanna of Namibia, with particular reference to bush thickening by *Acacia mellifera*. Journal of Arid Environments 72, 2201–2210. https://doi.org/10.1016/j.jaridenv.2008.07.004

Joubert, D.F., Smit, G.N., Hoffman, M.T., 2013. The influence of rainfall, competition and predation on seed production, germination and establishment of an encroaching *Acacia* in an arid Namibian savanna. Journal of Arid Environments 91, 7–13.

Lefsky, M.A., W. B. Cohen, G. G. Parker, D. J. Harding, 2002. Lidar remote sensing for ecosystem studies. BioScience 52, 19–30.

Levick, S.R., Asner, G.P., Kennedy-Bowdoin, T., Knapp, D.E., 2009. The relative influence of fire and herbivory on savanna three-dimensional vegetation structure. Biological conservation 142, 1693–1700.

Levick, S.R., Baldeck, C.A., Asner, G.P., 2015. Demographic legacies of fire history in an African savanna. Functional Ecology 29, 131–139. https://doi.org/10.1111/1365-2435.12306

Liu, A., Cheng, X., Chen, Z., 2021. Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. Remote Sensing of Environment 264, 112571. https://doi.org/10.1016/j.rse.2021.112571

McGlinchy, J., van Aardt, J.A.N., Erasmus, B., Asner, G.P., Mathieu, R., Wessels, K., Knapp, D., Kennedy-Bowdoin, T., Rhody, H., Kerekes, J.P., Ientilucci, E.J., Wu, J., Sarrazin, D., Cawse-Nicholson, K., 2014. Extracting Structural Vegetation Components From Small-Footprint Waveform Lidar for Biomass Estimation in Savanna Ecosystems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 480–490. https://doi.org/10.1109/JSTARS.2013.2274761

McNicol, I.M., Ryan, C.M., Mitchard, E.T.A., 2018. Carbon losses from deforestation and widespread degradation offset by extensive growth in African woodlands. Nature Communications 9, 3045. https://doi.org/10.1038/s41467-018-05386-z

Mograbi, P.J., Asner, G.P., Witkowski, E.T.F., Erasmus, B.F.N., Wessels, K.J., Mathieu, R., Vaughn, N.R., 2017. Humans and elephants as treefall drivers in African savannas. Ecography 40, 1274–1284. https://doi.org/10.1111/ecog.02549

Mograbi, P.J., Erasmus, B.F.N., Witkowski, E.T.F., Asner, G.P., Wessels, K.J., Mathieu, R., Knapp, D.E., Martin, R.E., Main, R., 2015. Biomass Increases Go under Cover: Woody Vegetation Dynamics in South African Rangelands. PLoS ONE 10, e0127093. https://doi.org/10.1371/journal.pone.0127093

Neuenschwander, A., Guenther, E., White, J.C., Duncanson, L., Montesano, P., 2020. Validation of ICESat-2 terrain and canopy heights in boreal forests. Remote Sensing of Environment 251, 112110. https://doi.org/10.1016/j.rse.2020.112110

Neuenschwander, A.L., 2008. Evaluation of waveform deconvolution and decomposition retrieval algorithms for ICESat/GLAS data. Canadian Journal of Remote Sensing 34, S240–S246.

O'Connor, T.G., Puttick, J.R., Hoffman, M.T., 2014. Bush encroachment in southern Africa: Changes and causes. African Journal of Range & Forage Science 31, 67–88.

Peters, A.J., Walter-Shea, E.A., Ji, L., Vina, A., Hayes, M., Svoboda, M.D., 2002. Drought monitoring with NDVI-based standardized vegetation index. Photogrammetric engineering and remote sensing 68, 71–75.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment 253, 112165.

Ratnam, J., Tomlinson, K.W., Rasquinha, D.N., Sankaran, M., 2016. Savannahs of Asia: antiquity, biogeography, and an uncertain future. Philosophical Transactions of the Royal Society B: Biological Sciences 371, 20150305. https://doi.org/10.1098/rstb.2015.0305

Ross, C.W., Hanan, N.P., Prihodko, L., Anchang, J., Ji, W., Yu, Q., 2021. Woody-biomass projections and drivers of change in sub-Saharan Africa. Nat. Clim. Chang. 11, 449–455. https://doi.org/10.1038/s41558-021-01034-5

Roussel, J.-R., Auty, D., De Boissieu, F., Meador, A.S., 2018. lidR: Airborne LiDAR data manipulation and visualization for forestry applications. R package version 1.

Roy, D.P., Kashongwe, H.B., Armston, J., 2021. The impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring. Science of Remote Sensing 4, 100024. https://doi.org/10.1016/j.srs.2021.100024

Sankaran, M., Hanan, N.P., Scholes, R.J., Ratnam, J., Augustine, D.J., Cade, B.S., Gignoux, J., Higgins, S.I., Le Roux, X., Ludwig, F., Ardo, J., Banyikwa, F., Bronn, A., Bucini, G., Caylor, K.K., Coughenour, M.B., Diouf, A., Ekaya, W., Feral, C.J., February, E.C., Frost, P.G.H., Hiernaux, P., Hrabar, H., Metzger, K.L., Prins, H.H.T., Ringrose, S., Sea, W., Tews, J., Worden, J., Zambatis, N., 2005. Determinants of woody cover in African savannas. Nature 438, 846–849.

Sankaran, M., Ratnam, J., Hanan, N., 2008. Woody cover in African savannas: The role of resources, fire and herbivory. Glob. Ecol. Biogeogr. 17, 236–245.
Scharlemann, J.P., Tanner, E.V., Hiederer, R., Kapos, V., 2014. Global soil carbon: understanding and managing the largest terrestrial carbon pool. Carbon Management 5, 81–91. https://doi.org/10.4155/cmt.13.77
Scholes, R., Archer, S., 1997. Tree-grass interactions in savannas. Annual review of Ecology and Systematics 28, 517–544.
Scholes, R.J., 1997. Savanna, in: Cowling, R.M., Richardson, D.M., Pierce, S.M. (Eds.), Vegetation of Southern Africa. Cambridge University Press, Cambridge, UK, pp. 258–277.
Shackleton, C., Shackleton, S., 2004. The importance of non-timber forest products in rural livelihood security and as safety nets: a review of evidence from South Africa. South African Journal of Science 100, 658–664.
Silva, C.A., Saatchi, S., Garcia, M., Labrière, N., Klauberg, C., Ferraz, A., Meyer, V., Jeffery, K.J., Abernethy, K., White, L., Zhao, K., Lewis, S.L., Hudak, A.T., 2018. Comparison of Small- and Large-Footprint Lidar Characterization of Tropical Forest Aboveground Structure and Biomass: A Case Study From Central Gabon. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 3512–3526. https://doi.org/10.1109/JSTARS.2018.2816962
Smit, I.P.J., Asner, G.P., Govender, N., Kennedy-Bowdoin, T., Knapp, D.E., Jacobson, J., 2010. Effects of fire on woody vegetation structure in African savanna. Ecological Applications 20, 1865–1875.
Smit, I.P.J., Asner, G.P., Govender, N., Vaughn, N.R., Wilgen, B.W. van, 2016. An examination of the potential efficacy of high-intensity fires for reversing woody encroachment in savannas. Journal of Applied Ecology 53, 1623–1633. https://doi.org/10.1111/1365-2664.12738
Smit, I.P.J., Prins, H.H.T., 2015. Predicting the Effects of Woody Encroachment on Mammal Communities, Grazing Biomass and Fire Frequency in African Savannas. PLOS ONE 10, e0137857. https://doi.org/10.1371/journal.pone.0137857
Stevens, N., Erasmus, B.F.N., Archibald, S., Bond, W.J., 2016. Woody encroachment over 70 years in South African savannahs: Overgrazing, global change or extinction aftershock? Philosophical Transactions of the Royal Society B: Biological Sciences 371. https://doi.org/10.1098/rstb.2015.0437
Team, R.C., 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www. R-project. org/.
Twine, W., Moshe, D., Netshiluvhi, T., Siphugu, V., 2003. Consumption and direct-use values of savanna bio-resources used by rural households in Mametja, a semi-arid area of Limpopo province, South Africa. South African Journal of Science 99, 467–473.
Venter, Z.S., Cramer, M.D., Hawkins, H.-J., 2018. Drivers of woody plant encroachment over Africa. Nat Commun 9, 2272. https://doi.org/10.1038/s41467-018-04616-8
Wang, C., Elmore, A.J., Numata, I., Cochrane, M.A., Shaogang, L., Huang, J., Zhao, Y., Li, Y., 2022. Factors affecting relative height and ground elevation estimations of GEDI among forest types across the conterminous USA. GIScience & Remote Sensing 59, 975–999. https://doi.org/10.1080/15481603.2022.2085354
Wessels, K.J., Colgan, M.S., Erasmus, B.F.N., Asner, G.P., Twine, W.C., Mathieu, R., van Aardt, J.A.N., Fisher, J.T., Smit, I.P.J., 2013. Unsustainable fuelwood extraction from South African savannas. Environ. Res. Lett. 8, 014007. https://doi.org/10.1088/1748-9326/8/1/014007
Wessels, K.J., Mathieu, R., Erasmus, B.F.N., Asner, G.P., Smit, I.P.J., van Aardt, J.A.N., Main, R., Fisher, J., Marais, W., Kennedy-Bowdoin, T., Knapp, D.E., Emerson, R., Jacobson, J., 2011. Impact of communal land use and conservation on woody vegetation structure in the Lowveld savannas of South Africa. Forest Ecology and Management 261, 19–29. https://doi.org/10.1016/j.foreco.2010.09.012
Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T., 2012. Lidar sampling for large-area forest characterization: A review. Remote Sensing of Environment 121, 196–209. https://doi.org/10.1016/j.rse.2012.02.001

List of Figure Captions

Fig. 1. (A) Study area in South Africa with relevant FAO Global Ecological Zones. (B) Zoom-in of northern South Africa with detailed location of individual sites covered by airborne LiDAR data (ALS): (1) D'Nyala, (2, 3, 4) Limpopo, (5) Venetia, (6) Welverdiendt, (7) Agincourt, (8) Ireagh, (9) Justicia, (10) Dukuduku, (11) Addo (B). (2-column fitting image)

Fig. 2. Workflow of GEDI and ALS data processing and analysis. (1.5-column fitting image)

Fig. 3. Example of one GEDI L1B waveform (GEDI orbit: 7433, waveform ID: 54731) and parameters used by different algorithm setting groups (SGs) (Table 2) to identify canopy mode (yellow line) and ground mode (green line). (2-column fitting image)

Fig. 4. The % bias of various GEDI algorithm setting groups (a-g) within 1m bins of RH98sim. (2-column fitting image)

Fig. 5. The difference of ground and canopy top estimations between GEDI algorithm setting groups to illustrate the impact of different parameter settings. Back threshold (SG 5: 2 sigma, SG 3: 6 sigma) (a, d); front threshold (SG 4: 6 sigma, SG 1: 3 sigma) (b, e); and second Gaussian smoothing width (SG 3: 3.5 sigma, SG 1: 6.5 sigma) (c, f). (2-column fitting image)

Fig. 6. Sensitivity of GEDI beam types (coverage vs. power) combined with time of acquisition (day vs. night). (Single fitting image)

Fig. 7. The bias of test cases vs. MODIS relative greenness on the date of GEDIorb acquisition. A reduction in negative bias in $RH98_{orb}$ with increased greenness is evident. (Single fitting image)

Fig. 8. Density plots of RH98sim vs. RH98orb below 15 m for all test cases with leaf-on (a) and leaf-off (b) conditions. The black line is the diagonal reference line (slope = 1, intercept = 0) and the red line represents the mean bias. (2-column fitting image)

Fig. 9. The difference between GEDI and ALS derived ground and canopy top elevations for four scenarios: leaf-on & leaf-off (>2.35m) (a), leaf-on & leaf-off (<2.35m) (b), leaf-on (>2.35m) (c), and leaf-off (>2.35m) (d). The RH98 bias was mainly due to the underestimation of canopy top elevation. (1.5-column fitting image)

Fig. 10. Histogram of the relative frequency of RH98sim values of all GEDI footprints across all study sites (0-15 m). (single fitting image)

Fig. 11. Bias (a, b), and %bias (c, d) vs. GEDI-RH98sim (0-15 m) for leaf-on (left) and leaf-off (right) test cases. Mean Bias values are indicated with blue asterisks. The lowest bin range is 2.35 – 3 m. RH98 bias (leaf-on) vs. airborne LiDAR (ALS) canopy cover (e). (2-column fitting image)

- GEDI waveform LiDAR sensor and algorithms were designed for dense forests.
- First validation of GEDI canopy height estimates (Relative Height 98) in savannas.
- On-orbit GEDI-RH98 was compared to simulated RH98 from airborne LiDAR data.
- Leaf-on vs. leaf-off conditions had a very large influence on results.
- RH98 was very accurate: R^2 = 0.61, bias = -0.55 m, %bias = -11.1%, RMSE = 1.64 m.
- GEDI provides reliable canopy height estimates in savannas, for height classes >3m.

Table 1. Properties of study sites and airborne LiDAR (ALS) data acquired.

Site#	Site name	Vegetation type (Dayaram et al. 2019)/ transformed cover	Deciduous Savannas /Evergreen	Mean annual rainfall (mm)	Mean annual temp (°C)	Mean (STD) ALS vegetation height (m)	ALS canopy cover (%)	Date of ALS	Area (km^2)	Mean (STD) Slope (degree)
1	D'Nyala	Roodeberg, Waterberg Mountain & Limpopo Sweet Bushveld	Deciduous Savannas	375	21	3.9 (2.3)	74.3	March 2018	53.26	2.36 (3.69)
2.3.4	Limpopo	Tsende Mopaneveld, Granite Lowveld, Gravelotte Bushveld & Tzaneen Sour Bushveld	Deciduous Savannas	613	27	3.3 (2.1)	51.5	March/ April 2018	163.32	1.7 (1.28)
5	Venetia	Musina Mopane Bushveld & Limpopo Ridge Bushveld	Deciduous Savannas	368	22.8	2.5 (1.1)	41.6	March 2018	56.31	1.97 (1.75)
6	Welverdiendt	Granite lowveld & Legogote Sour Bushveld / subsistence cultivation	Deciduous Savannas	353	25	3.7 (2.0)	44.2	June 2018	126.75	1.74 (1.68)
7	Agincourt		Deciduous Savannas	353	25	4.2 (2.4)	41.9	May 2018	35.88	5.14 (4.84)
8	Ireagh		Deciduous Savannas	687	21.3	3.2 (2.3)	32.3	June 2018	65.08	2.43 (2.09)
9	Justicia		Deciduous Savannas	550	25	3.4 (2.5)	28.9	June 2018	81.25	2.55 (1.48)
10	Dukuduku	Maputaland Coastal belt & Northern Coastal Forest mix, Subtropical Alluvial / exotic tree plantations, subsistence cultivation	Evergreen	967	21.7	9.3 (5.6)	70.0	June 2018	141.38	1.84 (2.41)
11	Addo	Sundays mesic & Valley thicket, Grassridge Bontveld & Albany Alluvial	Evergreen	388	18.4	2.3 (0.88)	66.3	March 2018	109.84	2.63 (3.05)

Table 2. Parameters of different GEDI processing setting groups (SG).

SG	Smooth width zcross (ns)	Front threshold	Back threshold
1	6.5	3	6
2	3.5	3	3
3	3.5	3	6
4	6.5	6	6
5	3.5	3	2
6	3.5	3	4

Table 3. R^2 values of the linear models of GEDI-RH98$_{orb}$ and GEDI-RH98$_{sim}$ for each setting group.

Setting group	R^2
1	0.479
2	0.448
3	0.455
4	0.341
5	0.129
6	0.361
S	0.48

Table 4. Summary of comparison between GEDI-RH98$_{orb}$ vs. GEDI-RH98$_{sim}$ for all savanna sites and test cases under leaf-on and leaf-off conditions.

Grouping	Mean R^2 (STD)	Mean bias (m) (STD)	Mean %bias (%) (STD)	Mean RMSE (m) (STD)	Mean %RMSE (%) (STD)	Sample size
Combined analysis, all test cases	0.48	-1.02	-19.48	1.93	36.7	18215
Combined analysis, leaf-on	0.61	-0.55	-11.15	1.48	29.85	7347
Combined analysis, leaf-off	0.43	-1.47	-26.53	2.27	40.94	8933

Table 5. Summary of comparison between GEDI-RH98$_{orb}$ vs. GEDI-RH98$_{sim}$ for each study site grouped by phenological status.

Study sites	Mean R^2	Mean bias (m)	Mean %bias[1] (%)	Mean RMSE (m)	Mean %RMSE[2] (%)	Sample size
Addo (evergreen)	0.428	-0.08	-2.78	0.203	6.75	2133
Dukuduku (evergreen)	0.521	-0.91	-12.55	3.11	43.63	1262
Agincourt (leaf-on)	0.486	-0.53	-8.93	1.852	31.30	479
Agincourt (leaf-off)	0.167	-2.1	-36.08	3.089	51.08	337
D'Nyala (leaf-on)	0.712	-0.168	-2.7	1.413	22.82	86
D'Nyala (leaf-off)	0.157	-1.905	-28.43	2.72	40.69	896
Ireagh (leaf-on)	0.512	-0.571	-11.72	1.61	33.51	780
Ireagh (leaf-off)	0.352	-1.503	-31.54	2.295	47.92	736
Venetia (leaf-on)	0.611	-0.459	-11.94	0.76	19.81	2038
Venetia (leaf-off)	0.373	-0.831	-21.57	1.162	30.21	592
Limpopo (leaf-on)	0.438	-0.901	-14.09	2.202	34.37	677
Limpopo (leaf-off)	0.476	-1.514	-25.94	2.313	40.47	2330
Justicia (leaf-on)	0.495	-1.165	-21.88	2.183	41.06	1322
Justicia (leaf-off)	0.492	-1.387	-26.98	2.276	44.28	1319
Welverdiendt (leaf-on)	0.536	-0.608	-12.35	1.544	31.738	1965
Welverdiendt (leaf-off)	0.417	-1.221	-22.11	1.836	33.35	2723

Figure 1

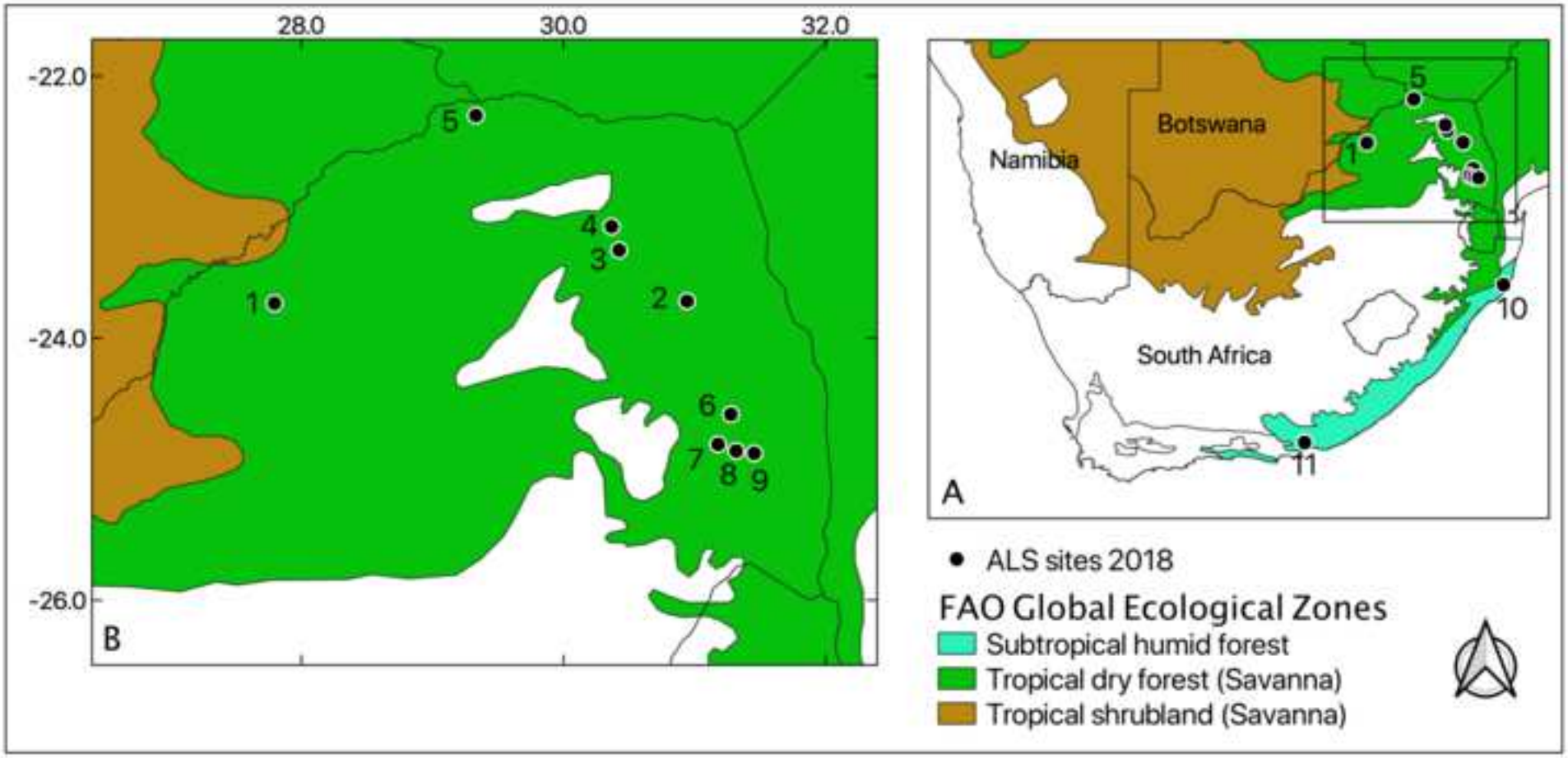


Figure 10

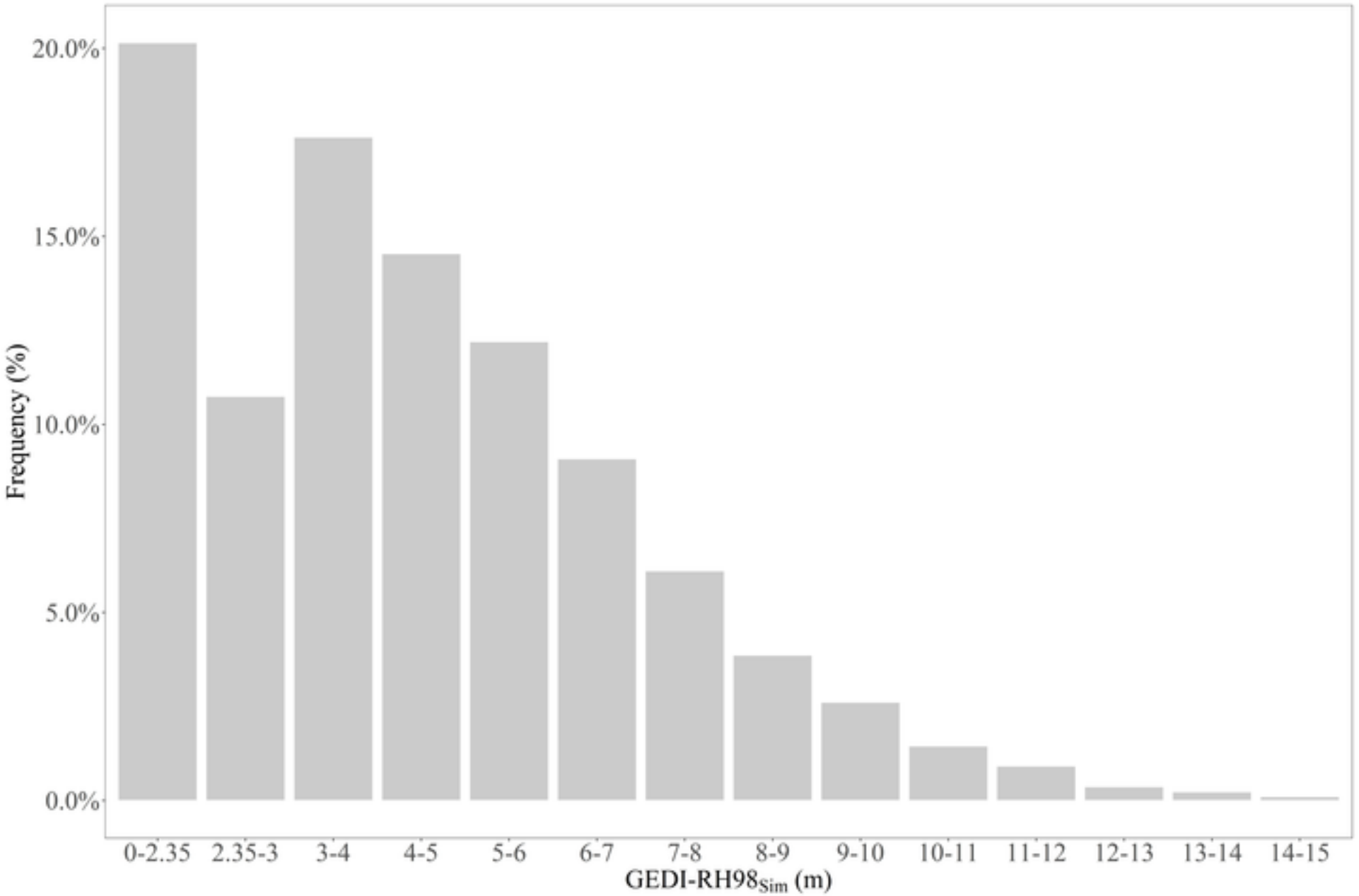


Figure 11.1

Click here to access/download;Figure;GEDI_RSE_Fig.11(a-d).jpg

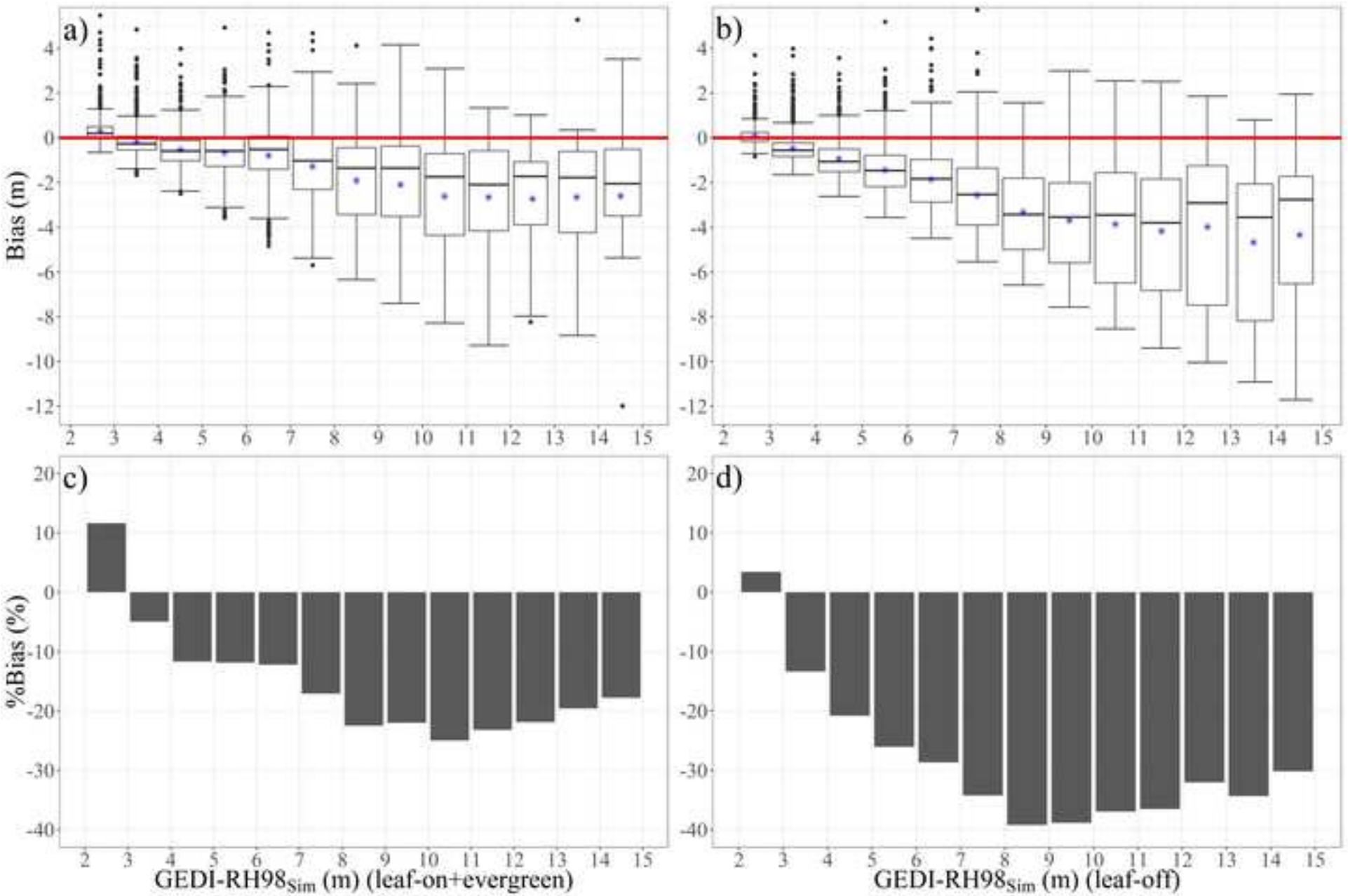


Figure 11.2

Click here to access/download;Figure;GEDI_RSE_Fig.11(e).jpg

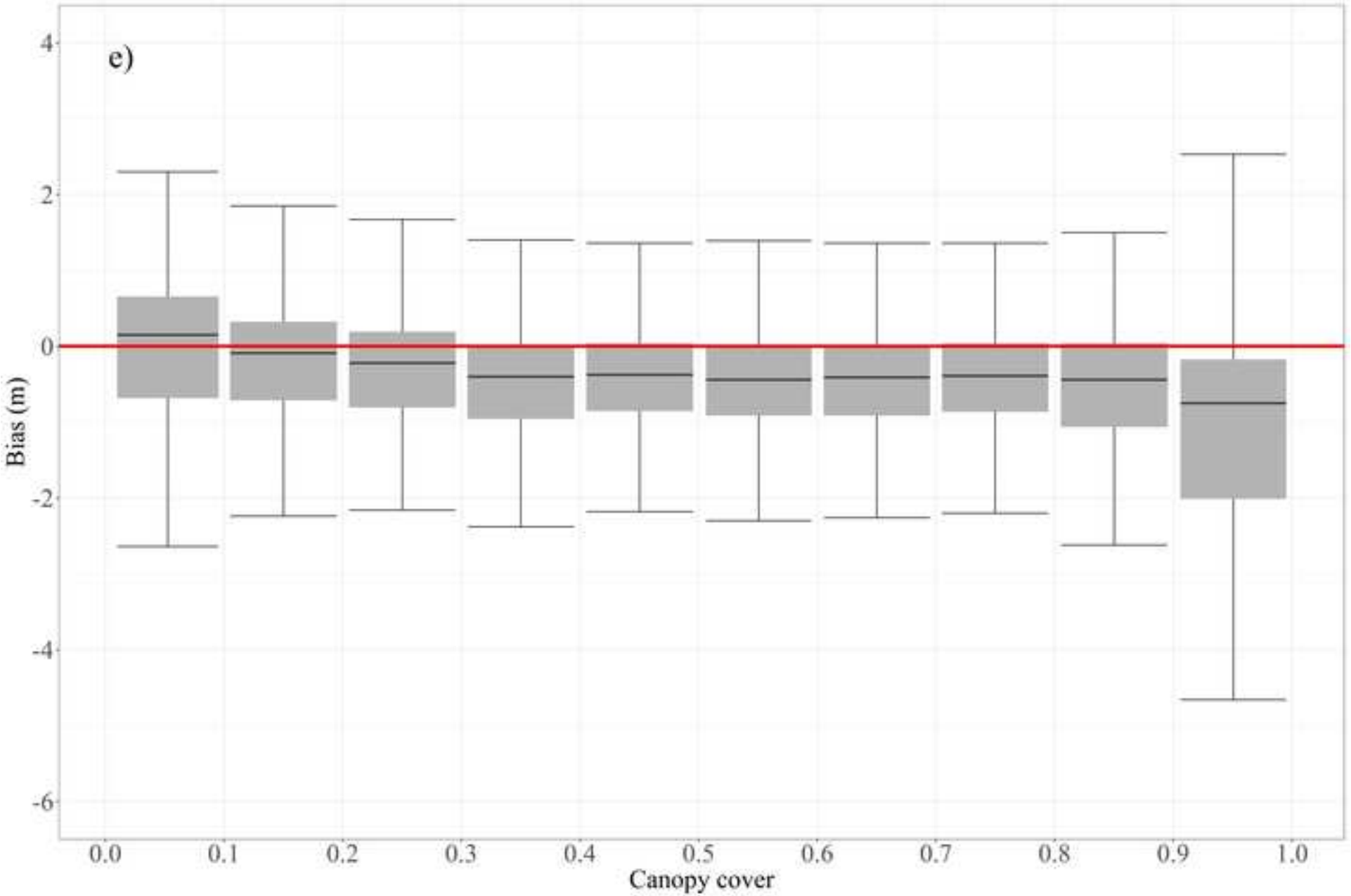


Figure 2

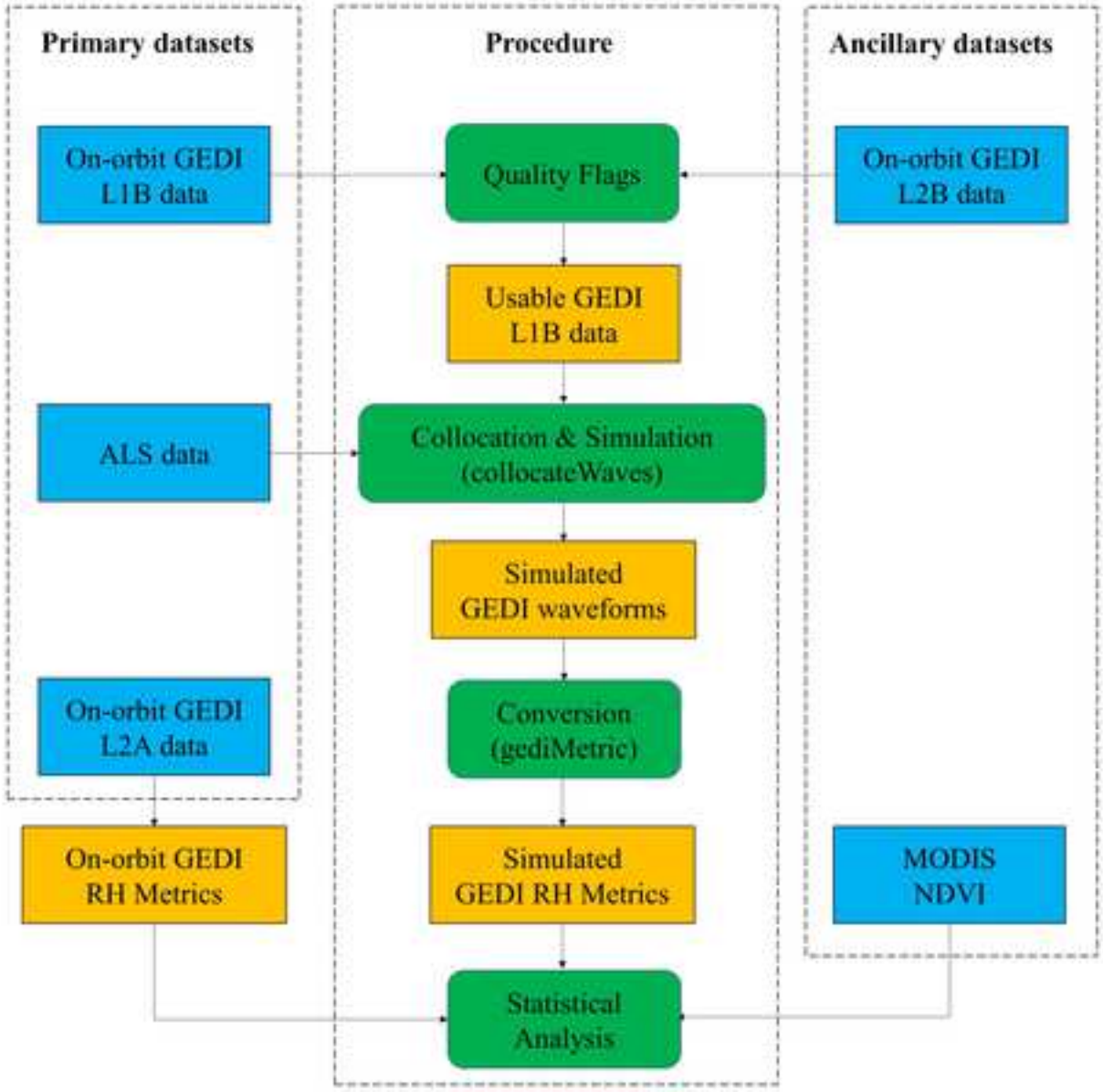


Figure 3

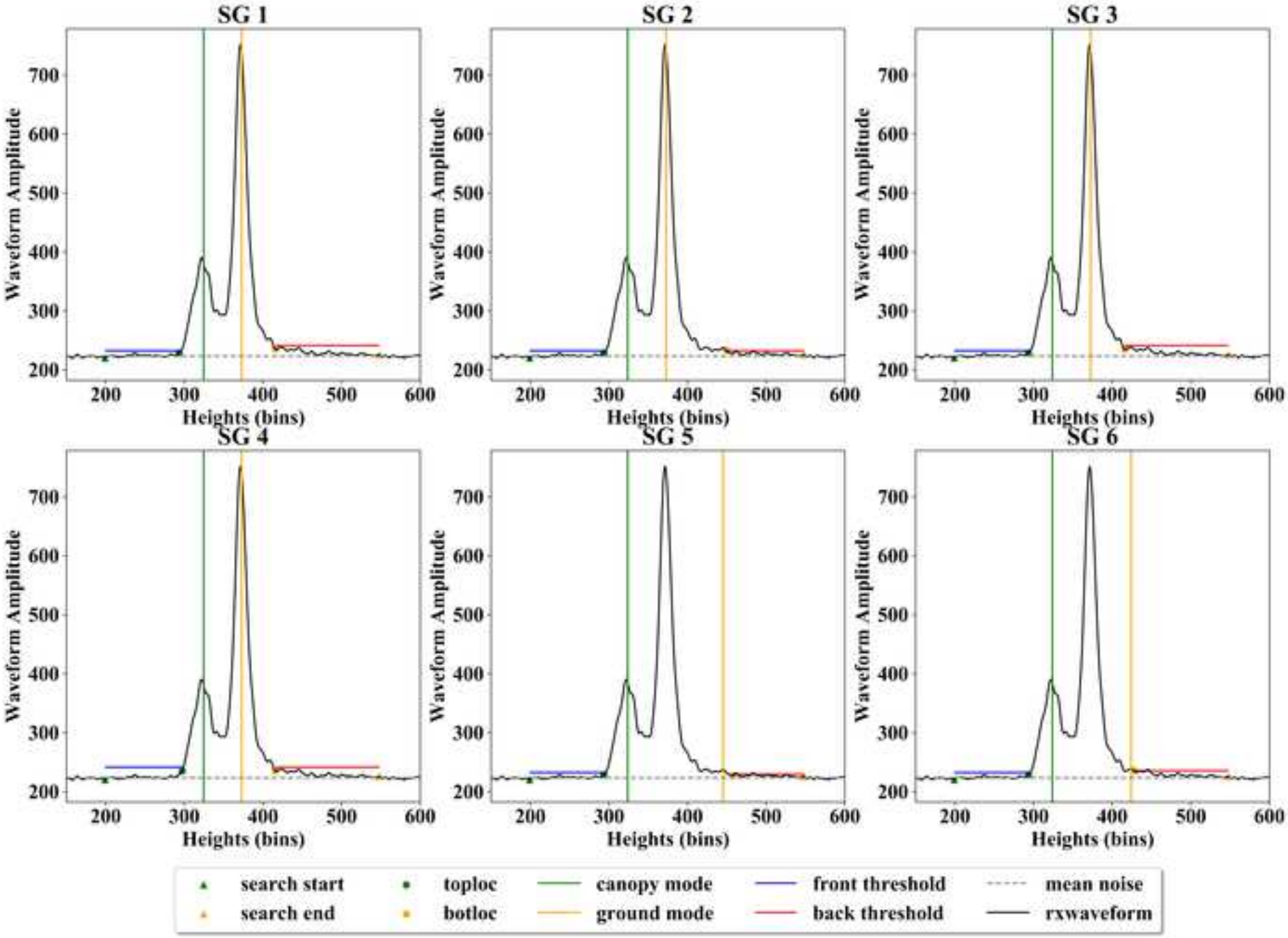


Figure 4

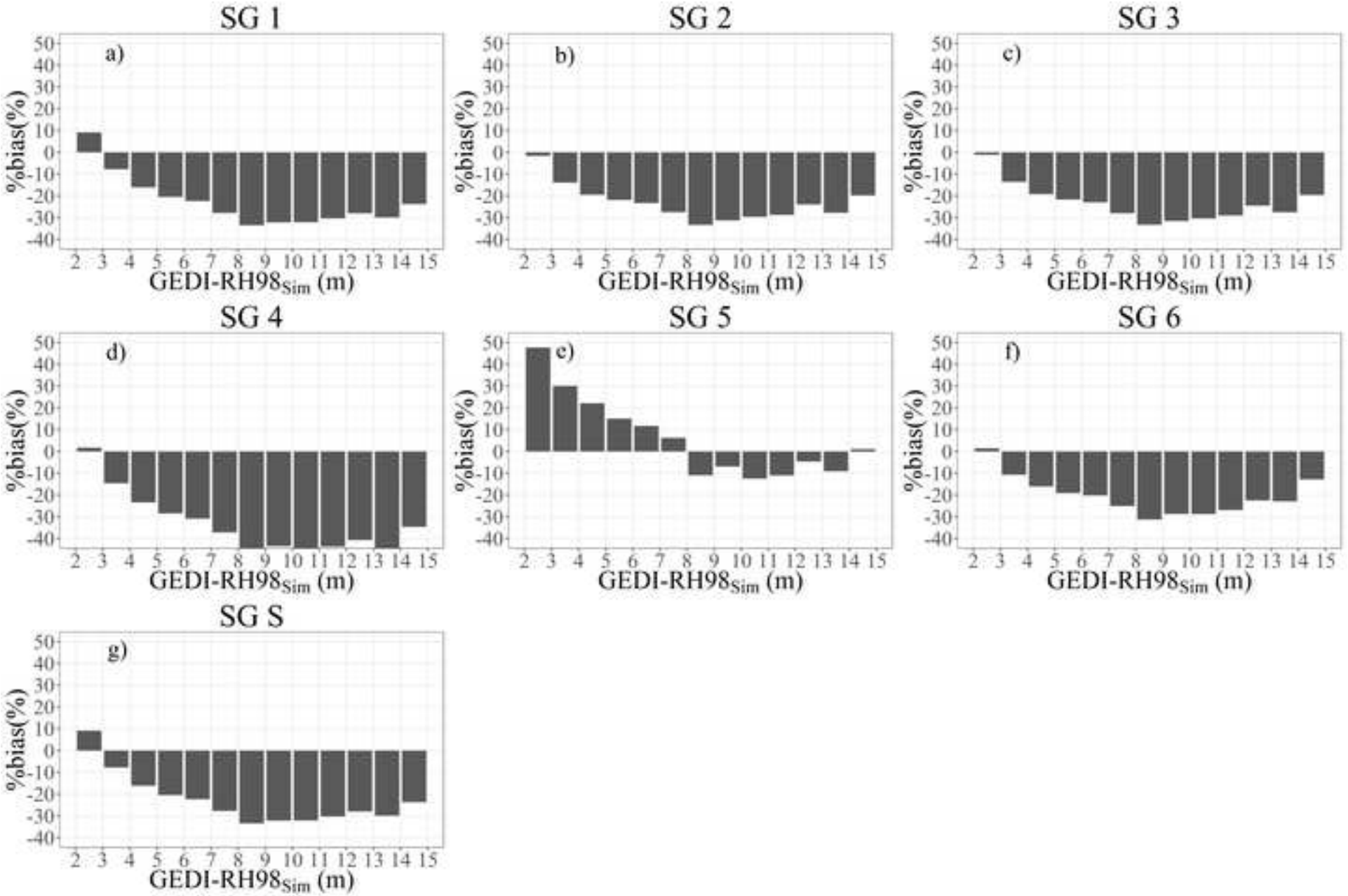


Figure 5

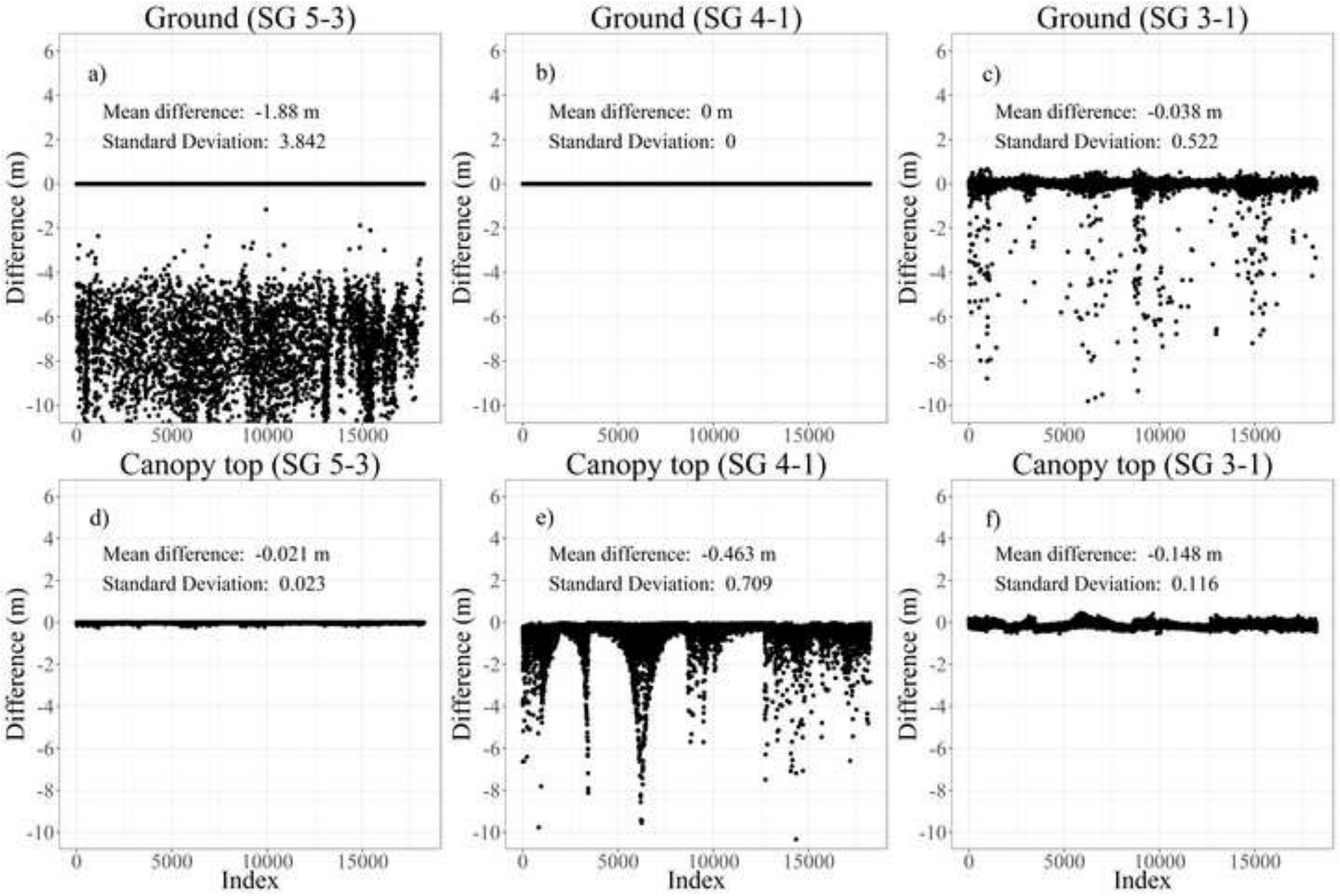


Figure 6

Click here to access/download;Figure;GEDI_RSE_Fig.6.jpg

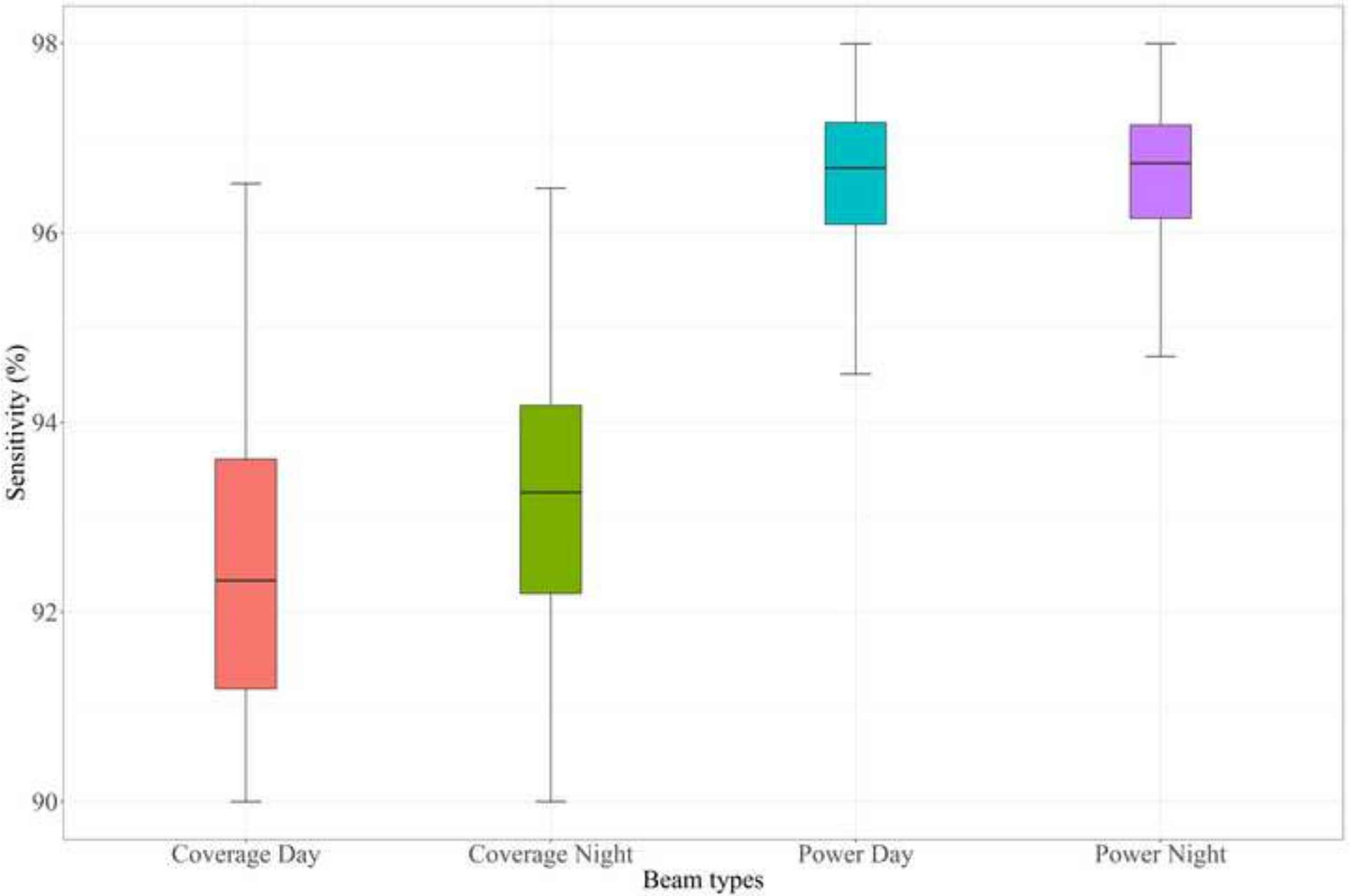


Figure 7

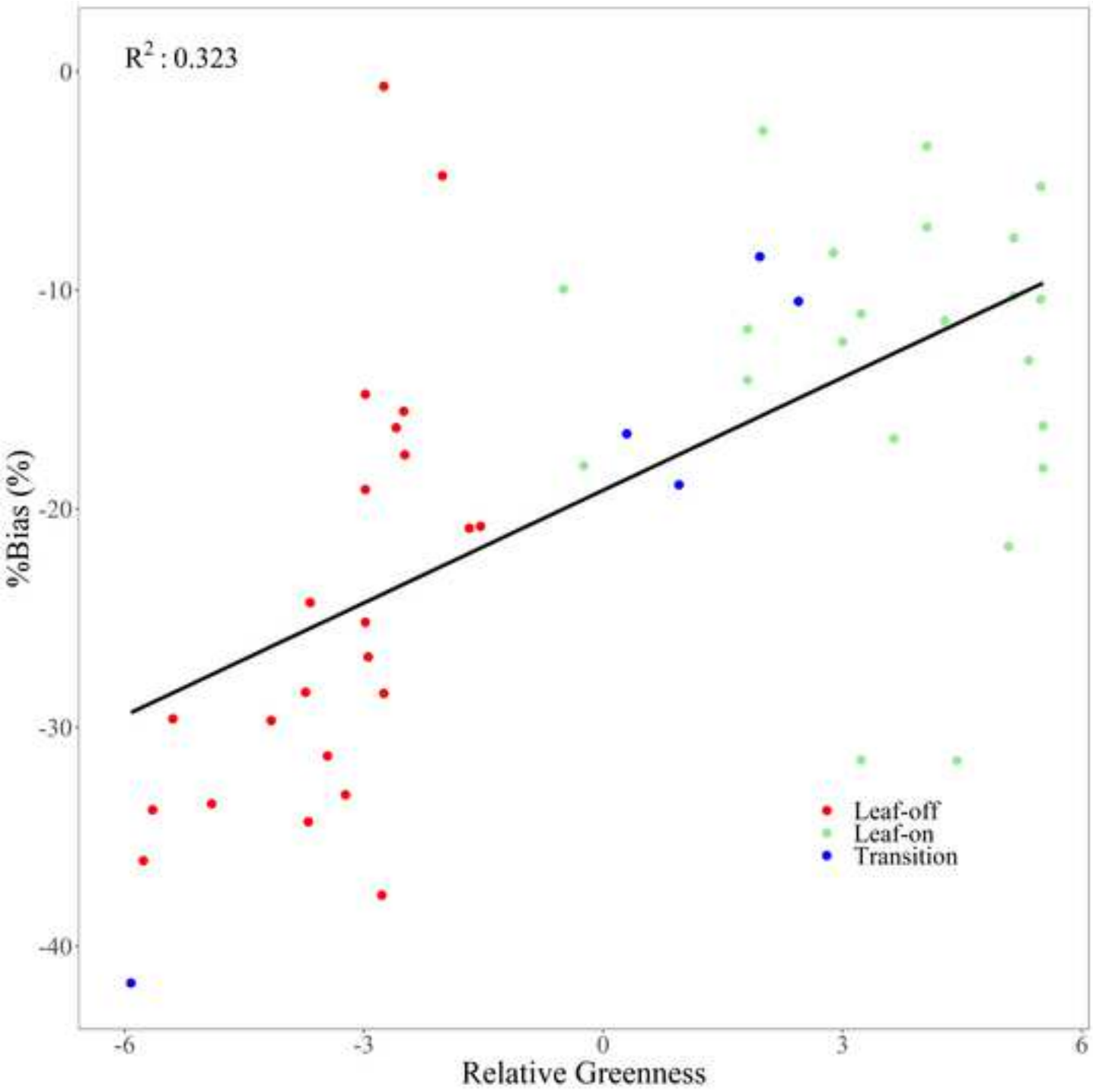


Figure 8

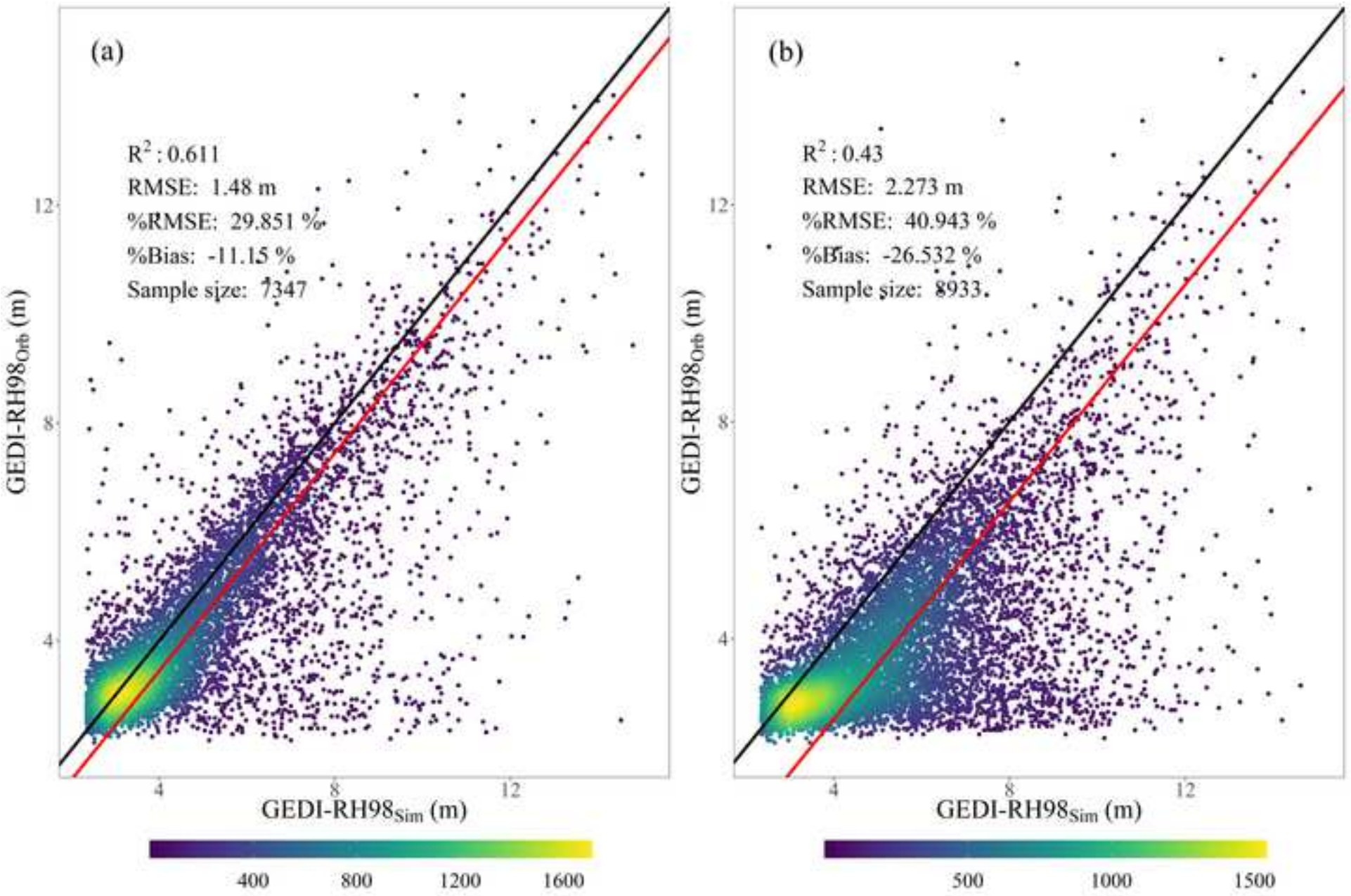
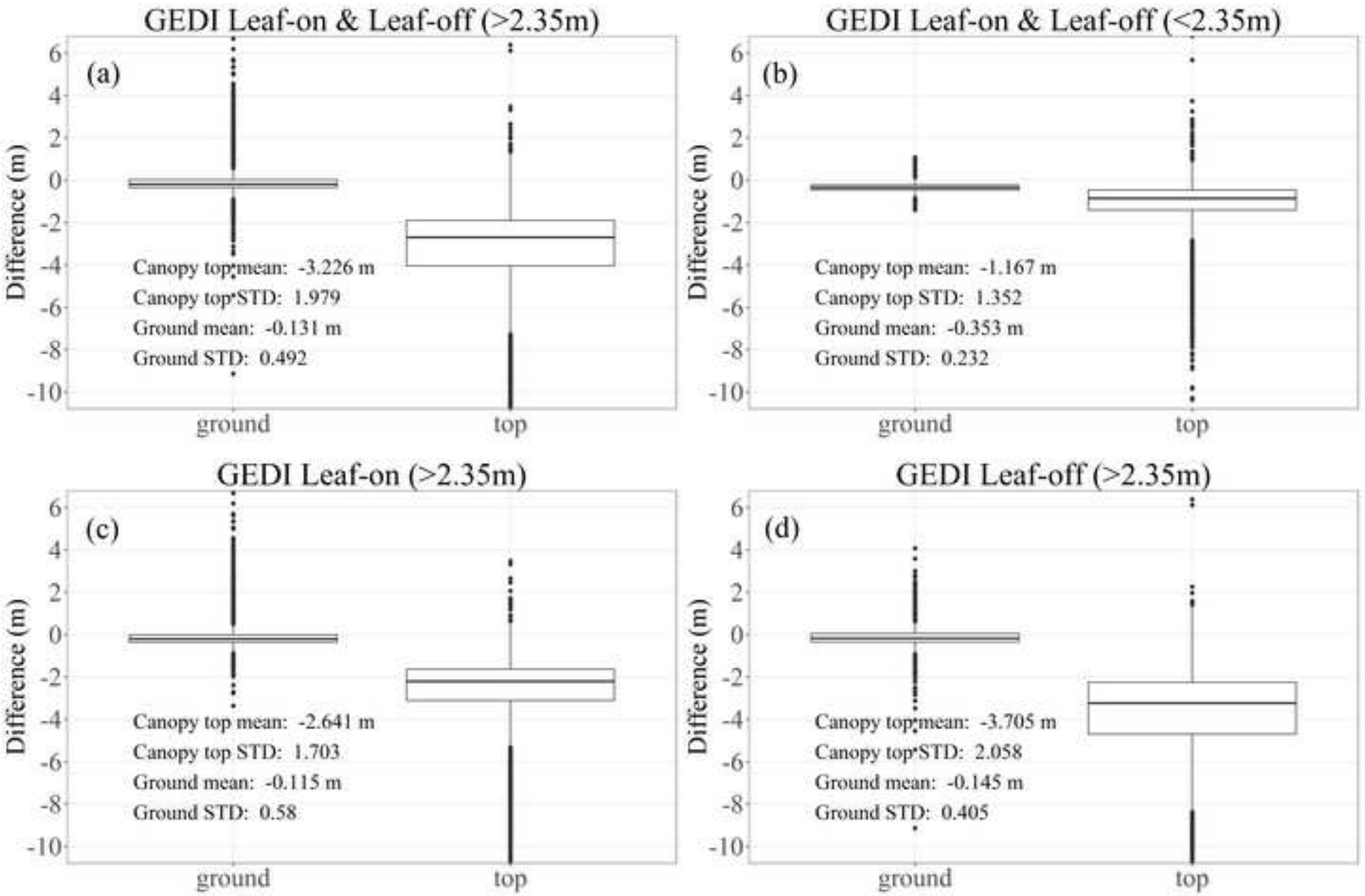


Figure 9

Click here to access/download;Figure;GEDI_RSE_Fig.9.jpg



Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Credit Author Statement

XL processed all GEDI and ALS data, performed analysis and wrote the paper. KW conceptualized the research, secured ALS data acquisitions and wrote the paper. JA conceptualized the research, provided technical oversight and wrote the paper. SH provided advanced technical input and edited the paper. R Mathieu conceptualized the research, secured ALS data acquisitions and edited the paper. R Main and LN processed the ALS data, provided study area expertise and edited paper. BE and RS secured ALS data acquisition, provided expertise on savanna ecology and edited the paper.