We would like to thank Reviewer # 1 for his/her timely review and for pointing out some minor inaccuracies. We would, however, respectfully disagree with the title of the review: 'The methodology requires major improvements' and the grounds on which this conclusion is based. We would like to take the opportunity to explain why we largely disagree with the conclusions reached by Reviewer # 1. Reviewer # 1 raises three main issues (comments made by the reviewer in bold):

1. **Why design the GLAS shot filtering from a small desert area using only L1A data. Since the mapping is global all desert areas could be used with all laser campaigns.**

The key question is whether or not the filtering method is adequate in identifying spurious data. It is less important whether or not the filters are derived from a sample of the data or from the full data set. In our opinion, deriving the filters from a subset and then successfully applying the filters to a much larger data is a good test to see if the filters are more generally applicable. Using the entire data set to derive the thresholds does not leave any data to test the method on. We therefore disagree with Reviewer # 1, we think that the chosen approach enhances to the credibility of the applicability of the filters, rather than diminishes it. We would be happy to include this point in the text.

2. **Computation of the bare ground fraction does not take into account the filtering procedure which has a major impact on the number of selected GLAS shots. (i.e. The selection process removes more forest shots than bare ground).**

The specific comment by the reviewer below is added here because it sheds further light on this comment.
In section 4.5, Filtered shots are used to compute bare ground cover fraction (p2330 line 20). The computation of bare soil fraction must be strongly driven by you filter selection method. More shots are removed in forest areas as indicated in table 2.

There appears to be a misunderstanding. The reviewer is correct that Table 2 indicates that more data are removed over tropical forests than deserts. Conditions for tropical forests and deserts are different (e.g. tropical forests have dense cloud cover) it is therefore not correct to draw the conclusion that the method eliminates more data from trees than from bare soil (e.g. bare soil patches in forests) for all conditions. The analysis of the sites (Fig. 2) demonstrates that both positive and negative outliers (points further away from the 1:1 line) are removed and that the removal of spurious data increases the correlation and decreases the RMSE and bias. The global analysis of GLAS and MODIS vegetation cover fraction shows that removing spurious data increases the correlation between the two. The *raw* GLAS data show a lower correlation with the MODIS tree cover fraction than the *filtered* GLAS data. Thus neither the site analysis nor the global analysis provide any evidence that the bare soil fraction is overestimated as a result of the filtering. We are therefore of the opinion that this criticism is unfounded.

3. **The comparison with Lefsky's map lacks quantitative analysis**

We agree that a spatial comparison with Lefsky's data would be useful and interesting. The comparison with Lefsky's data is limited in our paper, because his data are not publicly available. We therefore had to limit the evaluation to a comparison of height distributions per biome type (compare Lefsky's height distributions per biome type with our height distributions). The one difference that stands out from this comparison is that we measure much higher tree heights in the rainforests (peak in height distribution at 40 m) than Lefsky (peak at 25 m). We think Lefsky's estimates are too low: Sellers et al (1996) use a mean tree height of 35 m for tropical forests that is based on a review of the literature; Asner et al (BIOTROPICA 34(4): 483-492 2002) found a mean tree height of 45 m for the upper story; The Aircraft data from Peru indicate values up to 50 m and we have reports from people working in the field (Amazon LBA and Malaysia) as well that indicate that tree heights are on average much higher than 25 m. We agree that our comparison is limited but the limited conclusion we draw from this comparison is valid. We're happy to add references from the peer reviewed literature that report on tree height measurements in tropical forests.

Specific comments:

p2336: it is puzzling that a threshold of 8meters is selected. As mentioned in the paper, the bias due to presence of vegetation will be significant. Thus such threshold does not make sense once it is applied to forest areas.
The ICESAT/GLAS elevation provided in GLA14 is computed from the full waveform (not the last peaks) so that the 8 meters defined from desert areas may not be appropriate.

The 8 m is approximately the 95 % uncertainty range in the SRTM DEM. The uncertainty is from Rodriguez et al from their uncertainty analysis for all continents (not just deserts). The threshold was confirmed in our desert analysis. As stated in our paper we would assume the SRTM DEM to be less accurate over forests. Because Radar saturates over dense vegetation it will only penetrate about halfway into dense canopy. This measure more or less agrees with d_elev, which is generally located at the centroid of the waveform. The elevation test is principally intended to eliminate cloud contaminated data and it may be possible to relax the 8m uncertainty range over dense forests to acknowledge the greater uncertainty in the SRTM and d_elev and not to reject GLAS data that could be otherwise be included in the analysis. Analysis of the Peru data (dense forest) and Tumbarumba data (dense forest) indicate the method still works, despite this potential limitation. The global analysis indicates tall trees in tropical forest, the average pretty much in line with what's known from the literature. Thus, although we flagged this as a potential point of concern in the paper, it does not seem to lead to gross inaccuracies and as a result, we did not see a need to relax the 8m threshold over dense forests.

p2331: add GLA14:

Thanks for pointing this out, we will add this.

section 3.1.2In the GLA14 product there are spurious shots with much larger DEM error. Any relationship with tree height?
The x-y plots from the test sites (Fig 2) indicate that data removed by the filters (including the elevation filter) have a poor relationship with tree height.

Section 3.1.3 You use 1 or Gaussians. How is the area of the 2nd changing?

For this test only the area under the first Gaussian is considered.

Conclusions on line26: How can you reach such conclusion given the test is over the desert?

It is not a conclusion, it is an explanation as to how this threshold is obtained. For the desert we can set the threshold much higher, but we need to be aware that a very high threshold may eliminate too many data for other areas (dense vegetation) since the magnitude of the first Gaussian will be affected by vegetation cover. We therefore derive a threshold for the desert data that is as low as possible, below this threshold a large proportion of data is likely to be subject to error. We later increase this threshold in the sensitivity analysis to see how its variation affects the selection of data.

section 3.1.5 It lookks like tht spread of heights is larger for low values...explain.

Fig 1.h: It is not the spread of heights versus sigma, it is amplitude versus sigma. When the amplitude is large the sigma is small and vice versa. This is because the total pulse strength (area) increases as a function of both amplitude and sigma.

section 4.1 The filters are applied sequentially. What happens when the order is changed?

The current order of filters is chosen to remove the most unreliable data first (e.g., clouds (elevation test) and slope). The neighbour test removes the neighbours of eliminated data and has to come last. Several other tests are based on absolute numbers (test to eliminate weak signals identified by amplitude and area under the first Gaussian) so their order in the sequence does not matter. The amplitude vs sigma test is derived from the raw data and is therefore independent of order as well. Thresholds based on a particular % of a distribution (the outlier test) may be affected by the order if a disproportionate amount is removed from one part of the distribution. If one swaps the filters one may want to adjust the thresholds to get a similar result. We're not sure what the benefit is of swapping the order of the filters.

section 4.2 line 17: wasn't that done for other sites? line 25: Why did you sample the airborne lidar data to a 50m grid cell that is not centered on the GLAS shots? Why not use the all airborne shots within the GLAS shot like you haved done for the Peru site. In that case Figure 3b would bot be necessary.

One of the co-authors (Craig Mahoney) has looked into this as part of his PhD research. It makes very little difference what you choose. We think Fig 3b is a nice illustration of the uncertainty in the comparison caused by spatial variation, we would like to leave it in.

p2341 line 16: Doesn't the use of "max tree height" least sensitive to variability? As long as the tallest tree is somewhere within the footprint, one should obtain the same height. please discuss.

Spatial variability does affect the accuracy of the comparison, given that there is a probability that the top of a tree can be in one footprint (either GLAS or aircraft), but not in the other. This probability increases if the spatial variation in the tree height increases. Moreover, the maximum canopy height is most variable in both small and large footprint LiDAR. The energy distribution within the LiDAR footprint is Gaussian. For large footprint LiDAR, a single tall tree towards the edge of the footprint may not product a sufficient energy return to trigger the threshold of the start of the waveform. For small footprint systems, the distribution of footprints is commonly such that it does not hit the very top of a tree crown (particularly prone in conifers)

p2342 line 2: sub-meter GLAS location accuracy. I have strong doubts. Can you prove that? line6-8: This argument is far from being convincing when Figure 3C shows a correlation of 0.25.

The reported location accuracy is based on a document provided by the NSIDC. It can be obtained from http://nsidc.org/data/icesat/docs/. The horizontal geolocation accuracy mean and st. dev. (m) are given for various campaigns and are in the order of 0.4 m +/- 3 m, only for 4 campaigns is the mean error above 1 m. We can add this information to the manuscript.
line 6-8: We're reporting the results of the statistical test (significant at $p \ll 0.01$). The figure indicates that as spatial variability increases the probability increases that the aircraft and GLAS data differ. The figure indicates that it is still possible to hit a tree of similar height if spatial variability is high, it just becomes more unlikely. In general, the magnitude of the correlation is related to, but not always the same as, the significance of the correlation.

p: 2343 line 5: what about k=0 or 0.5?
It is unlikely that k=0 or k=0.5 will lead to better results. k=0 represents the raw data screened only for elevation differences and would incorporate data from steep slopes that are spurious. k=0.5 would incorporate data from slopes <= 34 % and would incorporate a substantial amount of spurious data as well. Another argument against this suggestion is that since k=2 appears better than k=1, it does not make sense to investigate k< 1 (k=0.5 or k=0).

p: 2344 line 6-17. Lefsky's map is not 1km resolution. Large scale segments are used. This comparison should be done quantitatively (simple subtraction could be performed.

Thanks for pointing this out, we'll correct this to 5 km^2 in the updated version of our manuscript. Lefsky's data was not available to us. I agree a spatial comparison would be interesting and useful. Based on the comparison of height distributions per biome type between our and Lefsky's data we would expect greater similarity in the boreal and temperate forests, and large discrepancies in the tropical forest. The spatial comparison will provide further detail but will not change this conclusion.

In section 4.5, Filtered shots are used to compute bare ground cover fraction (p2330 line 20). The computation of bare soil fraction must be strongly driven by you filter selection method. More shots are removed in forest areas as indicated in table 2.

We addressed this point in our comment regarding main criticism 2.

p:2348 line19: The paper does not show the substantial improvement over existing products as no quantitative analysis is provided.

We compared our present product with two other vegetation height products.

1. Vegetation height estimates from Sellers et al 1996 (1 average height value per land-cover type obtained from the literature). Our averages are not that different from those by Sellers et al, with the exception of short vegetation types and especially agriculture (contains patches of tall trees, not considered by Sellers et al). Our product shows height variation within land-cover type, is tested on site data and compares well with the MODIS tree cover (and bare soil) product. Our present data is therefore an improvement.

2. The tree height product by Lefsky. The first improvement is that our product estimates vegetation height for all biomes, not just forest biomes. Thus we have better coverage, be it at lower spatial resolution. Moreover, our estimates for tropical forests are higher and we think more realistic (see previous response).

Our estimates of tree height are similar for boreal forests and temperate forests where there are fewer problems with the GLAS data.
We think the difference in overall averages for tropical forests + by providing height estimates for all biomes is sufficient evidence that our data set is a substantial improvement over Lefsky's data.

technical comments: Caption of Table 1: there is no "double" line.
Thanks for pointing this out, we will revise this to read the last three parameters instead.

Concluding remarks

We are happy to make various minor corrections that the reviewer suggests. We do disagree with the major issues raised by the reviewer and are of the opinion that we produced sufficient evidence in our paper and in this document to address the concerns expressed by reviewer 1.