Geosci. Model Dev. Discuss., 4, 2327–2363, 2011
www.geosci-model-dev-discuss.net/4/2327/2011/
doi:10.5194/gmdd-4-2327-2011

Geoscientific Model Development Discussions

Vegetation height products between 60° S and 60° N from ICESat GLAS data

S. O. Los[1], J. A. B. Rosette[1, 2], N. Kljun[1], P. R. J. North[1], J. C. Suárez[2, 3], C. Hopkinson[4], R. A. Hill[5], L. Chasmer[6], E. van Gorsel[7], C. Mahoney[1], and J. A. J. Berni[7]

[1]Department of Geography, Swansea University, Singleton Park, Swansea SA2 8PP, UK
[2]Biospheric Sciences Branch, code 614.4, NASA/Goddard Space Flight Center, Greenbelt and University of Maryland College Park, Maryland 20771, USA
[3]Forest Research, Northern Research Station, Roslin, Midlothian, EH25 9SY, UK
[4]Applied Geomatics Research Group, Centre of Geographic Science, Nova Scotia Community College, Lawrencetown, Nova Scotia, B0S 1P0, Canada
[5]School of Applied Science, Bournemouth University, Poole, Dorset, BH12 5BB, UK
[6]Cold Regions Research Centre, Wilfrid Laurer University, Waterloo ON N2L 3C5, Canada
[7]CSIRO Marine and Atmospheric Research, Pye Laboratory, Canberra ACT 2601, Australia

Received: 23 August 2011 – Accepted: 14 September 2011 – Published: 19 September 2011

Correspondence to: S. O. Los (s.o.los@swansea.ac.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

We present a new method to obtain coarse resolution ($0.5^\circ \times 0.5^\circ$) vegetation height and vegetation-cover fraction data sets between 60° S and 60° N for use in climate models and ecological models. The data sets are derived from the Geoscience Laser Altimeter System (GLAS) on the Ice, Cloud and land Elevation Satellite (ICESat), which is the only LiDAR instrument that provides close to global coverage when all data collected for 2003–2009 are combined. Filters are applied to the GLAS data to identify and eliminate spurious observations, e.g. data that are affected by clouds, atmosphere and terrain and as such result in erroneous estimates of vegetation height or vegetation cover. GLAS vegetation height estimates are aggregated in histograms from 0 to 70 m in 0.5 m intervals. The GLAS vegetation height product is evaluated in four ways. First, unfiltered and filtered individual GLAS vegetation height measurements are compared with aircraft LiDAR measurements of the same from seven sites in the Americas, Europe, and Australia. Application of filters increases the correlation with aircraft data from $r = 0.36$ to $r = 0.67$ and decreases the root-mean-square error by a factor 3. Second, the global aggregated GLAS vegetation height product is tested for sensitivity towards the choice of data quality filters; areas with frequent cloud cover and areas with steep terrain are the most sensitive to the choice of thresholds for the filters. Thirdly, the GLAS global vegetation height product is compared with two other global vegetation height products and is believed to produce more realistic characteristics: dominant vegetation height for tropical forests between 30 and 60 m versus 20 and 40 m in existing products. Finally, the GLAS bare soil cover fraction is compared globally with the MODIS bare soil fraction ($r = 0.55$) and with the FASIR bare soil cover fraction estimates ($r = 0.58$); the correlation between GLAS and MODIS tree-cover fraction was ($r = 0.76$). The evaluation indicates that filters applied to the GLAS data are conservative and eliminate a large proportion of spurious data, while only in a minority of cases at the cost of removing reliable data as well. The present GLAS vegetation height product appears more realistic than previous data sets used for input to climate

GMDD
4, 2327–2363, 2011

models and ecological models and hence should significantly improve simulations that involve the land surface.

1 Introduction

Global biophysical parameters such as the fraction of photosynthetically active radiation (fAPAR) and leaf area index (LAI) are essential parameters in calculating fluxes in the global carbon cycle, water cycle and energy budget. They are closely linked to the amount of solar radiation absorbed and scattered by the vegetation canopy. Because these biophysical parameters influence the solar radiation reflected from the earth's surface, they can be estimated from data collected by passive optical radiometers that measure in visible and near-infrared wave bands. Examples of these sensors are the advanced very high resolution radiometer (AVHRR; August 1981–present), the Sea-viewing Wide Field-of-view Sensor (SeaWiFS; September 1997 – December 2010), the Système Pour l'Observation de la Terre – Vegetation instrument (SPOT-VGT; April 1998 – present), the Along Track Scanning Radiometer (ATSR-2 and AATSR; June 1995 – present) and the moderate resolution image spectrometer (MODIS; February 2000 – present); see e.g. (Sellers et al., 1996; Myneni et al., 2003; Gobron et al., 2005). However, these systems are not particularly suitable to obtain estimates of biophysical parameters linked to canopy structure – e.g., vegetation height, above ground biomass, canopy inflection point and stem diameter – although there are approaches that exploit indirect relationships between measurements such as the normalised difference vegetation index (NDVI) and biomass with some degree of success for particular biomes (Tucker et al., 1986; Prince, 1991; van der Werf et al., 2006). Knowledge of structural vegetation parameters is, for example, essential to assess the amount of carbon stored in vegetation, to improve modelling of light absorption and scattering through the canopy and of photosynthesis (Alton et al., 2005) and to model the wind profile at the surface which affects the exchange of water and carbon between the land and atmosphere (Sellers et al., 1996). A problem using passive optical sensors to infer

GMDD
4, 2327–2363, 2011

canopy structure is that different canopy structures can lead to the same spectral and bidirectional response; the inversion of biophysical parameters in these cases is a non-unique problem with more than one solution and this inhibits unambiguous estimation of canopy parameters.

An active optical sensor, such as the Geoscience Laser Altimeter System (GLAS) on the Ice, Cloud and land Elevation Satellite (ICESat) emits a light pulse of known intensity and duration (Zwally et al., 2002; Brenner et al., 2003). The pulse is transmitted, absorbed and scattered at various depths throughout the vegetation canopy by leaves and branches and the returned waveform therefore provides information on canopy structure and height (Drake et al., 2003; Lefsky et al., 2005; Rosette et al., 2008). Compared to active microwave (RADAR) instruments spaceborne LiDAR saturates at much higher biomass levels (Drake et al., 2003; Waring et al., 1995) but is also more sensitive to atmospheric interference by clouds, water vapour and aerosols (Spinhirne et al., 2005). Furthermore, interpretation of GLAS waveforms is not straightforward since the waveform is not only affected by the vegetation canopy, but also by other factors such as the occurrence of thin clouds and topography (Rosette et al., 2008; North et al., 2010; Rosette et al., 2010).

The objective of the present paper is to obtain a vegetation height and vegetation cover data set from the GLAS instrument for most of the land surface between 60° S and 60° N that can be used in global models. Filters are applied to the GLAS data to obtain representative estimates of vegetation height and vegetation cover fraction at $0.5^\circ \times 0.5^\circ$ resolution. Because the GLAS instrument is the only of its kind to provide consistent global coverage, there are very few options to test the data at the global scale. An evaluation of this GLAS vegetation height product is carried out (1) in Sect. 4.2 by comparing GLAS vegetation height estimates with measurements from airborne LiDAR; (2) in Sect. 4.3 by testing the sensitivity of estimated near-global vegetation height to the choice of filters; (3) in Sect. 4.4 by comparing the GLAS global vegetation height product with vegetation height estimated from land-cover classes (Sellers et al., 1996) and a different GLAS tree height product (Lefsky, 2010); (4) in

GMDD
4, 2327–2363, 2011

Sect. 4.5 by comparing GLAS estimates of tree-cover fraction and bare soil fraction with the MODIS vegetation-cover fraction estimates (Hansen et al., 2003, 2006) and the FASIR vegetation-cover fraction (Los et al., 2000).

2 Data

We used the ICESat GLAS land data (GLA14) product, release 31 (Zwally et al., 2008; Brenner et al., 2003). GLAS emits a pulse waveform in the 532 or 1064 nm bands which is 1 m wide (corresponding to a duration of 7 ns) between the points where the signal is half the size of the maximum amplitude. The returned waveform is measured for a duration equivalent to a length of about 82 m at 15 cm intervals for the Laser 1A and 2A periods, and for an equivalent length for 150 m for the other periods. The footprint size is an ellipse with dimension of 95 by 52 m for the Laser 1A to 2C periods and 61 by 47 m for the other periods. The returned waveform contains various peaks which are fitted by up to 6 Gaussians. The GLAS instrument collected data intermittently during 2003-2009, usually for 1 to 3 periods of about 1 month per year (Zwally et al., 2002; Harding et al., 2005). For the derivation of the filters we used data from the Laser 1A period; for testing the filters and assembling the global vegetation height data we used data from all laser periods.

The following statistics from the GLA14 product are used: the start of signal, the 6 Gaussians fitted to the waveform data, the area under the 6 Gaussians, the amplitude for the 6 Gaussians, the topographic elevation estimated from the GLAS instrument and adjustment factors to relate the GLAS reference elevation to the Shuttle RADAR Topographic Mission (SRTM) digital elevation model (DEM) version 4.1, i.e., the geoid deviation from the ellipsoid used for the TOPEX/Poseidon mission and the saturation elevation correction (Table 1). A subset of the GLA14 data was organised in $5^\circ \times 5^\circ$ tiles to conform to the tiles of the SRTM version 4.1 data (Rodriguez et al., 2005; Jarvis et al., 2008). Table 1 shows a list of parameters retained. Data without geo-location, i.e., missing latitude and longitude values, are removed, as are data for

GMDD
4, 2327–2363, 2011

which no saturation elevation adjustment is possible (GLAS quality flag i_satElevCorr > 2). Data below 60° S and above 60° N are not analyzed because no SRTM data are available to test data for a correct estimation of elevation or for slope effects (Sect. 3.1).

The interpolated SRTM DEM version 4.1 distributed by the Consultative Group for International Agriculture Research – Consortium for Spatial Information (CGIAR-CSI) (Jarvis et al., 2008; Rodriguez et al., 2005) was used to compare with the GLAS elevation and to obtain an indication of the slope. The CGIAR-CSI data were used rather than the SRTM DEM data included in the GLA14 product because the agreement with GLAS elevation was closer.

The MODIS continuous fractional cover data (Hansen et al., 2003, 2006) and FASIR vegetation-cover fraction (Los et al., 2000) were compared with vegetation-cover fraction estimates from the GLAS data.

Aircraft LiDAR measurements of vegetation height from Canada, Peru, the United Kingdom, the Netherlands, Germany and Australia were used to test the GLAS vegetation height estimates. These globally distributed validation test sites incorporate boreal, temperate and tropical vegetation; managed and natural woodland and varied canopy cover (e.g., part cover in the case of the Australian sites and near complete closure for the Peru site). The product is thus evaluated using a range of conditions including those known to be problematic for GLAS.

The Canadian sites, the former southern BOREAS study sites in Saskatchewan, consisted of fairly homogeneous forested areas and flat topography with an aspen stand (*Populus tremuloides Michx.*), a black spruce stand (*Picea mariana Mill.*) and a jack pine site (*Pinus banksiana Lamb.*; Barr et al., 2006; Kljun et al., 2007). The Peru site is located in the Tambopata National Reserve and consists of dense mature forest, regenerating forest, part flood plain and wetland, in an area of flat topography (Hill et al., 2011). The UK sites are the Glen Affric and Aberfoyle sites both measured by the UK Forest Research. Glen Affric (Suárez et al., 2008) is an area of ancient woodland, it contains one of the largest ancient Caledonian pinewoods in Scotland. Common species are Scots pine (*Pinus Sylvestris*, Juniper (*Juniperus communis*), birch (*Betula*

GMDD
4, 2327–2363, 2011

pubescens), and aspen (*Populus tremula*). The Aberfoyle site (Suárez, 2010) is a silviculture area where trees are planted and clearfelled in rotations of 40–60 years. The dominant species is Sitka spruce (*Picea sitchensis (Bong.) Carr.*). The Netherlands Loobos site is a Scots pine (*Pinus Sylvestris*) stand covering about 89 % of the areas that is planted on flat, sandy terrain with some open areas (Dolman et al., 2002). The German Tharandt site is a mixed forest stand with trees of different ages consisting of mainly spruce (*Picea abies*) with scattered pine (*Pinus Sylvestris*) and European Larch (*Larix decidua*) on undulating terrain (Grünwald and Bonhofer, 2007). The Australian data were collected 7 km East of Tumbarumba to coincide with the GLAS measurements. The area is located in Bago State Forest, New South Wales and consisted of mainly eucalyptus trees (*Eucalyptus delegatensis R. T. Baker* and *Eucalyptus dalrympleana Maiden*) in relatively complex terrain (Leuning et al., 2005).

3 Method

To obtain consistent, realistic estimates of vegetation height, it is essential to identify and remove spurious observations from the GLAS data while retaining a sufficient proportion of good quality data. As a first step, GLAS data from a desert site are explored. Vegetation height estimates for deserts should as a general rule be low; high values therefore indicate problems in the GLAS data. The occurrence of spurious, high vegetation height values is compared with other measures such as slope, the difference between elevation measured by GLAS and the elevation indicated by a DEM and the strength of the GLAS signal. By changing thresholds on these measures, spurious GLAS data can be identified and eliminated.

GLAS data from a 5° × 5° tile between 20° N – 25° N and 0° – 5° E are analysed; this tile covers a desert area with the northern part located in Algeria. Data collected over 41 days in February 2003 and March 2003 during the Laser 1A operations period are investigated. The location of the data is shown in Fig. 1a. The elevation measured by the GLAS instrument and the elevation in the SRTM DEM version 4.1 data (Rodriguez

GMDD
4, 2327–2363, 2011

et al., 2005; Jarvis et al., 2008) are compared in Fig. 1b as a function of latitude. Topographic elevation is calculated from the GLAS data using:

$$h = h_e + \Delta h_e - \Delta h_g + \Delta h_l \quad (1)$$

with

h = topographic elevation
h_e = GLAS elevation; i_elev
Δh_e = Saturation elevation correction; i_satElevCorr
Δh_g = Height of the geoid above the TOPEX/Poseidon ellipsoid; i_gdHt
Δh_l = Difference WGS84 and TOPEX/Poseidon ellipsoid
$= \Delta r_a(\cos\phi)^2 + \Delta r_b(\sin\phi)^2$

with

Δr_a = Difference radius of WGS84 and TOPEX/Poseidon ellipsoids at equator (0.7 m)
Δr_b = Difference radius for meridian (0.713682 m).
ϕ = Latitude

Parameter names i_elev, i_satElevCorr, i_gdHt indicate records of the GLAS data (Table 1); a further description of these records can be found in Zwally et al. (2002) and the GLAS on-line documentation provided by the National Snow and Ice Data Center (http://nsidc.org/).

Vegetation height was estimated according to Rosette et al. (2008):

$$h_V = 1.06(r_1 - r_{A_{1,2}}) \quad (2)$$

GMDD
4, 2327–2363, 2011

with

$$h_V = \text{vegetation height}$$
$$r_1 = \text{signal start (i\_SigbegOff)}$$
$$r_{A_{1,2}} = \text{the centroid range increment, i\_gpCntRngOff for max amplitude between Gaussians 1 and 2}$$

Vegetation height as a function of latitude is shown in Fig. 1c. High vegetation height estimates are found in areas where the topography changes rapidly; note that, e.g. the spikes in vegetation height in Fig. 1c occur in the same location as the spikes in topography in Fig. 1b. Thus a first inspection of the data indicates that a large proportion of high vegetation height values are spurious.

3.1 Data filters

The tests below are intended to detect and eliminate spurious values, e.g. high vegetation height values over deserts, from the GLAS data. Where possible, thresholds for the data filters rely on error estimates from the peer reviewed literature. In cases where no estimates are available, the thresholds rely on visual interpretation of the data. A test of the filtered GLAS product is carried out in Sect. 4.2 and a sensitivity analysis of the filters in Sects. 4.3 and 4.5.

3.1.1 Slope test (Fig. 1d)

Slopes affect the GLAS waveform; the waveform from a slope without vegetation can look similar to that of a vegetation canopy over a flat surface (North et al., 2010; Rosette et al., 2010). Using the SRTM DEM 4.1 data, the slope was calculated as the maximum of the 8 slopes between the grid cell for which the GLAS measurement was collected and its 8 surrounding neighbours. The grid cell size of the SRTM DEM 4.1 data is 90 m; thus in areas with variations in terrain at shorter lengths the SRTM slope

GMDD
4, 2327–2363, 2011

will underestimate the topographic variations within the 50 to 60 m footprint most commonly produced by GLAS. Grid cells with a slope exceeding 10° (17 %) were removed from further analysis. Based on theoretical grounds and analysis of the desert data, a threshold of a 10° slope appears a reasonable compromise between retaining a sufficient proportion of the signal and avoiding erroneous values (Nelson et al., 2009; North et al., 2010; Rosette et al., 2010). Figure 1d indicates that for a slope <17 % both realistic low values and spurious high values are collected; whereas for a slope >17 % only spurious high values for vegetation height are found.

3.1.2 Elevation test (Fig. 1e)

The GLAS topographic height (Eq. 1) is compared with the SRTM DEM version 4.1. It is assumed that large differences between the SRTM DEM version 4.1 data and the GLAS elevation indicate problems in either data set. For the area shown in Fig. 1a, the root mean square error (RMSE) between GLAS and the SRTM 4.1 DEM data was about 3.7 m for February 2003 only and was 4.2 m for data of February and March 2003 combined. The 95 % confidence interval of the SRTM data globally is estimated at approximately 8 m; it varies for different continents between 7 m to 8.8 m with the exception of New Zealand where the RMSE was about 12 m (Rodriguez et al., 2005; Jarvis et al., 2008). The errors in SRTM elevation include an error for geo-location (i.e. no adjustment for geo-location was made). Based on Rodriguez et al. (2005) and our analysis of the Sahara desert we set a threshold at 8 m, approximately the 95 % confidence interval; data are deemed spurious and are eliminated when the difference between the GLAS elevation and SRTM DEM version 4.1 data is larger than 8 m (Fig. 1e). In cases where dense canopy exists the RADAR data may not reach the ground and the filter may be too conservative.

GMDD
4, 2327–2363, 2011

3.1.3 Area under first Gaussian test (Fig. 1f)

Estimates of vegetation height in the present paper use the difference between the start of signal and the centroid range increment of the first or second Gaussian (Rosette et al., 2008). The returned waveform will always have a measurable width even in cases where no vegetation is present because of the duration of the emitted signal, the atmospheric attenuation of the signal and the reflection of the signal from a surface that is rarely completely flat. The implication is that for bare soil a small difference between the signal start and the centre of the first Gaussian is found and this translates into an equivalent estimate of vegetation height. In Fig. 1f the estimated vegetation height is plotted as a function of the area under the first Gaussian (in units of V × ns; i.e., Volt times nano second) to obtain an indication of the magnitude of the effect. Figure 1f shows that as the area under the first Gaussian increases, the estimate for the minimum vegetation height increases. It is assumed that the 5 % values of the height distributions (per interval on the x-axis) provide an indication of the magnitude of the effect. A line is fitted and the estimated vegetation height (Eq. 2) is subsequently adjusted according to:

$$h_{0.05} = a + bA \tag{3}$$

with A the area under the first Gaussian (V ns) and fitted coefficients $a = 1.91$ and $b = 0.11$. This value for $h_{0.05}$ is subtracted from all GLAS vegetation height estimates.

Figure 1f reveals a second potential problem; for low values of the area under the first Gaussian, the spread in estimated vegetation height is large. The higher values in this interval are likely unrealistic. A likely cause is that low values for the area under the first Gaussian indicate weak signal strengths, possibly caused by attenuation of the signal in the atmosphere or by low energy emitted. The latter problem occurred frequently during the last two years of the ICESat mission (Lefsky, 2010). A threshold is applied to eliminate values with low first Gaussian areas. Because a low area under the first Gaussian can also occur for vegetation with a dense canopy, the threshold cannot be

GMDD
4, 2327–2363, 2011

too large so as not to eliminate values from tall, dense vegetation. As a compromise a value of 1 V ns was selected.

3.1.4 Amplitude of First Gaussian test (Fig. 1g)

A low amplitude of the first Gaussian indicates a data quality problem similar to the low area under the first Gaussian. The ability to separate the true returned waveform start and end from the background noise is reduced. A test was implemented to eliminate data with low amplitude (Fig. 1g) here set at 0.05 V. Figure 1g indicates a number of outliers over the entire range of amplitudes. A second test was applied to eliminate the highest 0.1 % of values per amplitude interval of 0.1 V; these values appear as outliers in Fig. 1g.

3.1.5 Sigma test (Fig. 1h)

Gaussians with a large spread (range between the 5 % and 95 % values over 80 m or so) are unlikely to be from vegetation which only in exceptional cases reaches these heights. A test was applied to all Gaussians to remove waveforms with high sigma values. The threshold for the sigma test was calculated as the > 99.9 % value; this test eliminates the data with the highest 0.1 % sigma values. The thresholds for this test were calculated from frequency distributions of the unfiltered data.

3.1.6 Neighbour test (Fig. 1i)

Finally, data were removed where the along-track neighbour on either side failed any of the above tests.

3.1.7 Choice of filters

The choice of thresholds for some of the data filters is somewhat arbitrary. The scatter plots (Fig. 1) indicate that a large proportion of spurious data is removed but some

GMDD
4, 2327–2363, 2011

spurious values are likely still to be present (Fig. 1i). The discussion in the next section and Table 2 provide further indications as to how much data are removed by the filters. If thresholds are adjusted, a larger proportion of spurious values is removed, but this may be at the cost of removing too many reliable data. Prior to a potential adjustment of the thresholds, the filtered vegetation height values are evaluated in Sect. 4.

4 Analysis of the GLAS data filters

Raw, unfiltered GLAS data were organised in $5^\circ \times 5^\circ$ tiles similar to the SRTM DEM v 4.1 tiles. A selection of statistics from the GLA14 record was retained and a number of measures were added (Table 1). The filters and adjustments discussed in Sect. 3 were applied to the tiled GLAS data; data that did not pass the filters were removed. An estimate of vegetation height (Field height; Eq. 2) adjusted for the area under the first Gaussian (Eq. 3) was added.

4.1 Application of filters to a temperate and a tropical area

As a first step the filters are applied to data collected over 2003 for two $5^\circ \times 5^\circ$ tiles, one located in western Europe and the other in the Amazon, with fairly dense vegetation cover. The same filters were applied as in Sect. 4 and their effect compared to their improvement on data of the desert tile. The amount of data removed by each filter for the three tiles is summarised in Table 2. For the desert tile about 9 % of the data is removed; the filters with the most impact are the elevation test, the area under the first Gaussian test and the neighbour test. For the tile that covers part of western Europe most of the spurious data are removed by the slope test; a majority of data removed by this test is because of missing SRTM DEM values over the sea. The elevation test, area under the first Gaussian test and neighbour test each remove approximately 3 % of the data. For tropical forests the largest amount of data, about 35 %, is removed by the area under the first Gaussian test. About 10 % is removed by the amplitude

GMDD
4, 2327–2363, 2011

test and neighbours test each. The large effect of the area under the first Gaussian test may indicate problems with the ground return of the waveform for dense vegetation canopies. Therefore, in Sect. 4.3 it is investigated how much the canopy height changes in response to changing the filters.

4.2 Comparison with airborne LiDAR

Filtered GLAS vegetation height estimates obtained for all Laser periods (2003–2009) were compared with airborne LiDAR measurements of vegetation height for 9 sites (Sect. 2): the former southern old aspen, old black spruce and old jack pine BOREAS sites in Canada, a tropical forest site in Tambopata near Puerto Maldonado, Peru, the Loobos needle-leaf forest site in the Netherlands (Dolman et al., 2002), the Tharandt mixed forest site in Germany, the Glen Affric and Aberfoyle sites measured by Forest Research in the UK, and a transect 7 km East of the Tumbarumba site in Australia. Airborne LiDAR data were collected at a point density of 0.25 m, 0.5 m or 1 m. LiDAR point data were sampled to a 50 m resolution by either taking the 99.9 % values of vegetation height in a 50 by 50 m grid cell (the Glen Affric, Aberfoyle, Tumbarumba, Loobos and Tharandt sites) or the maximum value (BOREAS and Peru sites). The Peru data were matched with the centres of the GLAS footprint; reported GLAS footprint dimensions and azimuth for each laser campaign (http://nsidc.org/data/icesat/laser_op_periods.html) were used to extract coincident subsets of the airborne LiDAR data. Vegetation height estimated from the GLAS waveforms and the airborne LiDAR point clouds could then be directly compared. Aircraft data were mapped to a universal transverse Mercator (UTM) projection. Latitude and longitude were calculated for the centres of all grid cells, and data were compared if the distance (in the horizontal plane) between the centre of the 50 by 50 m grid cell and the centre of the GLAS footprint was less than 20 m. The comparison was carried out for unfiltered GLAS data, using the difference of start of signal and end of signal to indicate vegetation height, and for GLAS data with the filters of Sect. 3.1 applied and field height calculated with Eqs. (2) and (3).

GMDD
4, 2327–2363, 2011

Figure 2 and Table 3 summarise the results of the comparison. Overall, application of the filters led to a significant improvement in the agreement between the GLAS data and aircraft data. All correlations between GLAS data and aircraft data, with the exception of the Tharandt site, increased. The root-mean-square error decreased significantly in all cases with the exception of the Tharandt site, and in one case (Glen Affric) by a factor 10. The bias decreased for most cases, only for the Peru data the bias became larger. A possible reason for the outliers in GLAS versus aircraft vegetation height scatterplots is the spatial variability in the scene. The major axis of the GLAS footprint can be larger than 50 m; and may incorporate a response of a tree within an adjacent 50 m grid cell. Anecdotal evidence for this effect can be found at the Glen Affric site, where the one outlier could be caused by a small number of trees standing adjacent to the validation grid cell. The Tharandt site, which is the most problematic in that it has a large variability in tree type as well as in tree height, shows an improvement in values close to the 1:1 line, but contains various outliers that remain in the data. There is reason to assume that these outliers are related to small differences in footprint size in combination with a large variability in tree height (below). The overall improvement is demonstrated when all data (without Peru; not included because information from surrounding grid cells was missing) are combined (Fig. 3a); the correlation increases from 0.36 to $r = 0.67$ and the RMSE decreases from 22.7 to 8.1 (Table 3).

Differences in vegetation height estimated from the GLAS instrument and aircraft LiDAR can be caused by errors in either instrument, registration errors, differences in the size of the footprint and land-cover change between times of measurement. If the errors are caused by registration errors it is expected that errors would increase as a function of distance between the centre points of the GLAS waveforms and the 50 m grid cells derived from the aircraft. Figure 3b shows the absolute difference between the height measurements as a function of distance of the centres of the GLAS waveforms and the 50 m grid cells derived from aircraft. There is no statistically significant relationship between the two, therefore, registration errors are an unlikely source

GMDD
4, 2327–2363, 2011

of error. This is further supported by the fact that geo-location knowledge of GLAS footprints for the data release used for this work is sub-metre for the majority of laser campaigns. The maximum error does increase with distance, however. This increase is caused by high spatial variability in tree height. Differences in size of the footprint, resulting in different areas measured may result in an error if the variation in vegetation height is large over short distances. This is likely to be an important factor, as the difference in vegetation height increases significantly as a function of spatial variability, expressed as the standard deviation in vegetation height for a 3×3 grid cell window (Fig. 3c).

Overall the comparison with the aircraft data indicates a dramatic improvement in the estimates of vegetation height when the filters are applied to the GLAS data. A large amount of error, expressed as the RMSE in Table 3 is caused by high spatial variability in combination with a difference in what the GLAS waveform measures and what is represented by the 50 m aircraft grid cell. The RMSE values in Table 3 are likely too high, an error estimate more resistant to outliers is the 68 % value of the distances in Fig. 3b (≈ 4.9 m). This value is only marginally higher than the RMSE of the elevation measured by GLAS (4 m).

4.3 Sensitivity of vegetation height estimates to application of filters

The screened GLAS data are aggregated into frequency distributions from 0 to 70 m in 0.5 m intervals for each $0.5^\circ \times 0.5^\circ$ land-surface cell between 60° S and 60° N. The 90 % vegetation height value was determined from the histograms. The sensitivity of the 90 % vegetation height value to the choice of data filters is explored. Thresholds for three filters are varied simultaneously by a factor $k = 1,2,3$, producing increased severity of the filters:

$$
\begin{aligned}
&(\theta < 10^\circ/k) \\
&\&(A_1 > k \times 1\text{Vns}) \\
&\&(S_1 > k \times 0.05\text{V})
\end{aligned}
\tag{4}
$$

GMDD
4, 2327–2363, 2011

where θ is the slope, A_1 the area of the first Gaussian (V ns) and S_1 amplitude of the first Gaussian (V). Figure 4 compares the cumulative distributions of vegetation height per Simple Biosphere model (SiB) vegetation cover type (Loveland et al., 2001) for a filter factor $k = 1$ versus $k = 2$ in twelve quantile-quantile plots and Fig. 5 shows the same comparison but for a filter factor $k = 2$ versus $k = 3$. The quantile-quantile plots of vegetation height for a filter factor $k = 1$ versus $k = 2$ vary only slightly for most biomes, indicating that the choice of filters does not affect the height distributions much at the biome level. The exceptions are mostly in the shorter vegetation classes: for the shrubs and bare soil, and to a lesser extent for ground cover and shrubs and tundra. For these classes the larger height estimates for the filter factor $k = 2$ are somewhat lower. Changing the filter factor from $k = 2$ to $k = 3$ affects the broad-leaf deciduous class; for most other classes the height distributions are similar. Thus at the biome level, application of filters does not change the height distribution much.

The effect of application of the filters for a specific locale is investigated by looking at the sensitivity global distribution of 90 % values of the height frequency distributions per $0.5^\circ \times 0.5^\circ$ cell. The 90 % of the height distributions globally for a filter factor $k = 3$ are shown in Fig. 6a. The values range from over 40 m in tropical forests to 0 m in deserts. The effect of the filter factors $k = 1$ and $k = 3$ is shown spatially as a change in difference in the 90 % value for filter factor $k = 1$ and $k = 3$ in Fig. 6b. Most areas do not show a significant change. In some areas, mostly in the tropical forests, vegetation increases in height by up to 4 m if $k = 3$ is used. In some other, mostly mountainous areas, the vegetation decreases in height by at most 4 m.

4.4 Global vegetation height evaluation

Histograms of the 90 % value of the globally retrieved vegetation height distributions (filter $k = 3$ to conform with Fig. 6) are shown per SiB biome type (Sellers et al., 1996) in Fig. 7. Where in previous work one vegetation height per biome was used, e.g., to obtain an estimate of surface roughness (Sellers et al., 1996), we find a wider, more realistic, distribution of vegetation heights per biome. There is good agreement be-

GMDD
4, 2327–2363, 2011

tween vegetation cover types 1 to 6 (dominated by trees) and the occurrence of tall vegetation and cover types 7–12 (shrubs, grasses, tundra , agriculture, bare soil) and the occurrence of mostly short vegetation. The exception is agriculture and to a lesser extent tundra. It is likely, however, that these classes do contain a minority proportion of tall vegetation.

Lefsky (2010) derives vegetation height for forests and woodlands at 1 km resolution by merging the MODIS land-cover product (Friedl et al., 2010) with ICESat GLAS measurements. The MOD12Q1 product he uses is different from the SiB classification scheme used in the present paper. Nevertheless, for the more or less comparable tropical forest class Lefsky (2010) derives height intervals different from the present results; his tropical and sub-tropical moist broadleaf height estimates range between 10 and 30 m with a peak at 25 m, whereas our estimates for broad-leaf evergreen forest show a dominant peak between 30 and 60 m with a peak at 40 m (Fig. 7a). Height estimates for other tall vegetation classes have a similar range to the estimates by Lefsky (2010), differences can to some extent be attributed to differences in class definitions. The much larger heights for the tropical broadleaf evergreen class found in the present study indicates a substantial improvement in height estimates for this biome.

4.5 Comparison of GLAS cover fraction with MODIS data

The University of Maryland (UMD) MODIS continuous field land-cover product provides the percentage cover for three classes: bare soil, trees and other vegetation (Hansen et al., 2003, 2006). The Fourier Adjusted, Solar and sensor zenith angle corrected, interpolated and reconstructed (FASIR) vegetation-cover fraction (Los et al., 2000) can be used to calculate the bare soil fraction as well: $f_b = 1 - f_V$, with f_V the vegetation-cover (all vegetation) fraction. From the GLAS height estimates a bare-cover fraction and a tree-cover fraction can be estimated and these can be compared with the MODIS continuous fields and the FASIR bare soil fraction. Bare soil fraction can be calculated as the fraction of GLAS measurements within each $0.5^\circ \times 0.5^\circ$ cell heights below a set threshold. This threshold is likely to be higher than some value above zero, otherwise

GMDD
4, 2327–2363, 2011

small unevenness of the soil topography may appear as low estimates of vegetation height. The bare soil fraction was calculated from the $0.5^\circ \times 0.5^\circ$ degree GLAS height frequency distributions as the proportion of footprints below a height threshold, starting at 0 m and moving up at increments of 0.5 m:

$$f_{b,z} = \frac{\sum n_{h \leq z}}{N} \tag{5}$$

with $\sum n_{h \leq z}$ being the number of observations for a height interval smaller than z m with z varying from 0 to 70 m in 0.5 m intervals and N the total number of observations per grid cell. Similarly, tree-cover fraction for each grid cell was calculated using the fraction of observations above a height threshold:

$$f_{t,z} = \frac{\sum n_{h \geq z}}{N} \tag{6}$$

with $\sum n_{h \geq z}$ being the number of observations for a height interval larger than or equal to z m.

The GLAS bare soil fraction and tree-cover fraction are compared with the MODIS bare soil and tree-cover fraction sampled to $0.5^\circ \times 0.5^\circ$ resolution. Bare soil fraction and tree-cover fraction were estimated from the raw GLAS data and the filtered GLAS data ($k = 1,2,3$). For the 4 versions of GLAS bare soil fraction and tree-cover fraction, a coefficient of correlation with the MODIS data for land data between 60° S and 60° N was calculated for every height interval z. The correlation as a function of the threshold height is shown in Fig. 8a for bare soil and in Fig. 8b for tree-cover. The highest agreement was obtained for $k = 2$; the GLAS bare soil fraction using a threshold height $z = 1$ m resulted in the highest correlation ($r = 0.553$), for tree-cover fraction this was $k = 1$ at 7.5 m ($r = 0.768$); the difference with $k = 2$ at 7 m was very small ($r = 0.759$). In all cases, estimates of tree height fraction and bare soil fraction using filters were in much closer agreement with the MODIS data compared to estimates from the raw data (Fig. 8). Filter $k = 2$ appears an acceptable compromise between retaining sufficient

GMDD
4, 2327–2363, 2011

high quality data to obtain reasonable height estimates and removing the bulk of spurious data.

The maximum correlations between the GLAS bare soil fraction and the FASIR bare soil fraction are slightly higher than the correlations with the MODIS bare soil fraction (Fig. 8a). The FASIR vegetation-cover fraction is estimated as a continuous parameter but from data with a coarser resolution (≈ 50 km) than the MODIS data (1 km). The difference is not large, but could indicate that use of a continuous parameter (f_{APAR}) from which vegetation cover fraction was calculated (Los et al., 2000) may be more appropriate than using a discrete parameter (land-cover class).

5 Discussion and conclusion

The present study describes the estimation of a global vegetation height data set from the ICESat GLAS instrument. The spatial extent of the data is limited to the spatial coverage of the SRTM DEM data between 60° S and 60° N. Filters were designed to eliminate spurious data; some of the filters (slope, elevation) were based on literature, whereas other filters (the area under the first Gaussian, peak of the first Gaussian, neighbour test) were based on a visual analysis of desert data. The filters are not optimised using an objective minimization criterion such as least squares, because of the large volumes of data that need to be handled and the limited amount of validation data. As more data sets from air campaigns become available, optimisation of the thresholds can be further improved. However the product has been thoroughly tested for a range of vegetation types and conditions found globally, including those known to be challenging for the GLAS instrument.

In lieu of an objective criterion for determining optimum data quality, the GLAS vegetation height product was evaluated in four ways. Individual GLAS vegetation height measurements were compared with airborne LiDAR measurements from nine sites. For global aggregates of GLAS vegetation height distributions, a sensitivity test and comparisons with other data products were carried out. Both the analysis of the site

GMDD
4, 2327–2363, 2011

data and the analysis with other global data sets showed that the GLAS vegetation products improved substantially using data filters. The filters remove 10 % of values over deserts and 75 % of values over tropical forests; over test sites about 75 % of values was removed. Application of filters increases the correlation with aircraft data from $r = 0.36$ to $r = 0.67$ and decreases the root-mean-square error from 23 m to 8 m or even to 5 m when the 68 % value of the error distribution is used. This latter value is comparable in magnitude to the RMSE in elevation, around 4 m, indicating that the proposed method approaches the maximum accuracy that can be obtained with the GLAS instrument.

The sensitivity test of filter thresholds showed that only a minority of cells was sensitive to varying threshold values; for tropical forests application of more progressive filters increased estimates of vegetation height; whereas at the same time over mountainous areas estimates of vegetation height decreased. Comparison of GLAS bare soil cover and tree cover estimates with comparable MODIS products showed small differences between filters used; the median option, filter $k = 2$ provided a good compromise between removing spurious data and retaining good data. A future area of research will be to optimise thresholds further; this would require availability of a larger number of test sites. Optimizing filters per continent, region or land-cover class may be a further useful refinement.

Vegetation height histograms per $0.5^\circ \times 0.5^\circ$ cell show more realistic values than existing products. For example, vegetation height derived by biome uses only one average value, the GLAS data indicate that a large variation in vegetation height exists within land-cover classes. The latter is more realistic. Compared to the tree height product of Lefsky (2010), 10–30 m with a peak at 25 m for tropical forest, our estimate of the coressponding 90 % height values is twice as large: a range up to 60 m with 40 m heights being the most frequently occurring. Measuring tree height from waveform LiDAR in tropical forests is notoriously difficult to determine due to the difficulty in identifying the ground return. Further improvements can be expected if ground elevation can be estimated with higher certainty. This is challenging for a large footprint

GMDD
4, 2327–2363, 2011

LiDAR such as GLAS. A future satellite waveform sensor, producing a smaller footprint, would improve the capability of detecting the ground for sloped and vegetated surfaces.

The GLAS vegetation height data show remaining problems over bare soil ($r = 0.553$ for a height threshold of 1 m). However this offers a significant improvement on observations of other authors of estimated vegetation heights of several metres for bare soil. Combining an NDVI-based bare soil estimate or land-cover classification-based bare soil estimate with the GLAS estimates should improve the overall product further. Compared to calculating the bare soil fraction, measuring of the tree cover fraction is more straightforward and correlation with the MODIS product is higher than for bare soil ($r = 0.759$ for a 7 m height threshold).

Only a small percentage of each $0.5^\circ \times 0.5^\circ$ grid cell is sampled by the GLAS instrument. This can lead to uncertainties as to how representative the sample average is for the grid cell average. MacDonald and Hall (1980) found that crop yield for large areas could be estimated well with only a small percentage of land sampled. The limited sensitivity of the GLAS $0.5^\circ \times 0.5^\circ$ vegetation height estimates to varying the data quality filters is further indication that reasonable estimates are obtained.

The GLAS vegetation height distributions derived in the present paper are a first attempt to obtain near-global estimates of vegetation height for all biomes. Despite some limitations, the present product makes a substantial improvement over existing products.

Acknowledgements. The MODIS global vegetation continuous fields, MOD44B, were obtained from the Global Land Cover Facility, http://www.landcover.org/, the ICESat GLAS data were obtained from the National Snow and Ice Data Center (NSIDC), http://nsidc.org/, and the interpolated SRTM-DEM version 4.1 data were obtained from the Consultative Group for International Agriculture Research – Consortium for Spatial Information (CGIAR-CSI), http://www.cgiar-csi.org/. Airborne LiDAR data for the UK were provided by the UK government Forestry Commission research agency, Forest Research. The LiDAR data of Peru were acquired by Dr Bryan Mark of Ohio State University. We are also grateful to Doreen Boyd of the University of Nottingham who is part of the team working on this data set. Airborne LiDAR data from the Canadian sites were obtained with support from NERC (grant NE/G000360/1, PI

GMDD
4, 2327–2363, 2011

NK), from the Netherlands and German sites with support from NERC/ARSF grant EU10-01 (PI NK) and from Australia with support from NCEO EO mission support 2009 (PI NK). Special thanks to the Applied Geomatics Research Group (AGRG) in Nova Scotia, the NERC Airborne Research and Survey Facility (ARSF) and Airborne Research Australia (ARA) for carrying out the airborne campaigns.

References

Alton, P. B., North, P., Kaduk, J., and Los, S. O.: Radiative transfer modelling of direct and diffuse sunlight in a Siberian pine forest, J. Geophys. Res., 110, D23209, doi:10.1029/2005JD006060, 2005. 2329

Barr, A. G., Morgenstern, K., Black, T. A., McCaughey, J. H., and Nesic, Z.: Surface energy balance closure by eddy covariance method above three boreal forest stands and implications for the measurement of CO_2 fluxes, Agric. Forest Meteorol., 140, 322–337, 2006. 2332

Brenner, A. C., Zwally, H. J., Bentley, C. R., Csatuo, B. M., Harding, D. J., Hofton, M. A., Minster, J. B., Roberts, L. A., Saba, J. L., Thomas, R. H., and Yi, D.: Geoscience laser altimeter system algorithm theoretical basis document: Derivation of range and range distributions from laser pulse waveform analysis, ATBD v 4.1, available at: http://www.csr.utexas.edu/glas/, 2003. 2330, 2331

Dolman, A. J., Moors, E. J., and Elbers, J. A.: The carbon uptake of a mid-latitude pine forest growing on sandy soil, Agric. Forest Meteorol., 157–170, 2002. 2333, 2340

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B., and Hofton, M.: Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: factors affecting the generality of relationships, Global Ecol. Biogeogr., 12, 147–159, 2003. 2330

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X. M.: MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sens. Environ., 114, 168–182, 2010. 2344

Gobron, N., Pinty, B., Taberner, M., Mélin, F., Verstraete, M. M., and Widlowski, J.: Monitoring the photosynthetic activity vegetation from remote sensing data, Adv. Space Res., 38, 2196–2202, doi:10.1016/j.asr.2003.07.079, 2005. 2329

Grünwald, T. and Bonhofer, C.: A decade of carbon, water and energy flux measurements of an old spruce forest at the Archer Station Tharandt, Tellus, 59B, 387–396, 2007. 2333

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., and Sohlberg, R. A.: Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm, Earth Interact., 7, 1–15, 2003. 2331, 2332, 2344

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., and Sohlberg, R. A.: Vegetation continuous fields MOD44b, 2001 percent tree cover, collection 4, Tech. Rep., University of Maryland, College Park, Maryland, 2006. 2331, 2332, 2344

Harding, D. J. and Carabajal, C. C.: ICESat waveform measurements of within-footprint topographic relief and vegetation structure, Geophys. Res. Lett, 32, L21S10, doi:10.1029/2005GL023471, 2005. 2331

Hill, R. A., Boyd, D. S., and Hopkinson, C.: Relationship between canopy height and Landsat ETM+ response in lowland Amazonian rainforest, Remote Sens. Lett., 2, doi:10.1080/01431161.2010.510810, 203–212, 2011. 2332

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, A.: Hole-filled SRTM for the globe version 4, available from the CGIAR-CSI SRTM 90 m database, http://srtm.csi.cgiar.org, 2008. 2331, 2332, 2334, 2336

Lefsky, M., Harding, D., Keller, M., Cohen, W., Carabajal, C., Del Bom Espirito-Santo, F., Hunter, M., and De Oliveira Jr, R.: Estimates of forest canopy height and aboveground biomass using ICESat, Geophys. Res. Lett., 32, L22S02, doi:10.1029/2005GL023971, 2005. 2330

Lefsky, M. A.: A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System, Geophys. Res. Lett., 37, L15401, doi:10.1029/2010GL043622, 2010. 2330, 2337, 2344, 2347

Kljun, N., Black, T. A., Griffs, T. J., Barr, A. G., Gaumont-Guay, D., Morgenstern, K., McCaughey, J. H., and Nesic, Z.: Response of net ecosystem productivity of three boreal forest stands to drought, Ecosyst., 10, doi:10.1007/s10021-007-9088-x, 1039–1055, 2007. 2332

Leuning, R., Cleugh, H. A., Zegelin, S. J., and Hughes, D.: Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates, Agric. Forest Meteorol., 129, doi:10.1016/j.agrformet.2004.12.004, 151–173, 2005. 2333

Los, S. O., Collatz, G. J., Sellers, P. J., Malmström, C. M., Pollack, N. H., DeFries, R. S., Bounoua, L., Parris, M. T., Tucker, C. J., and Dazlich, D. A.: A global 9-yr biophysical land

surface dataset from NOAA AVHRR data, J. Hydrometeorol., 1 (2), 183–199, 2000. 2331, 2332, 2344, 2346

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, J., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data, Int. J. Remote Sens., 21, 1303–1330, 2001. 2343, 2362

MacDonald, R. B. and Hall, F. G.: Global crop forecasting, Science, 208 (4445), 670–679, 1980. 2348

Myneni, R., Knyazikhin, Y., Glassy, J., Votava, P., and Shabanov, N.: User's guide FPAR LAI (ESDT: MOD15A2) 8-day composite NASA MODIS land algorithm, Tech. rep., Terra MODIS Land Team. http://cybele.bu.edu/, 2003. 2329

Nelson, R., Ranson, K. J., Sun, G., Kimes, D. S., Kharuk, V., and Montesano, P.: Estimating Siberian timber volume using MODIS and ICESat/GLAS, Remote. Sens. Environ., 113, 691–701, 2009. 2336

North, P. R. J., Rosette, J. A. B., Suárez, J. C., and Los, S. O.: A Monte Carlo radiative transfer model of satellite waveform LiDAR. Int. J. Remote Sens., 31, 1343–1358, 2010. 2330, 2335, 2336

Prince, S. D.: Satellite remote sensing of primary production: comparison of results for Sahelian grasslands 1981–1988, Int. J. Remote Sens., 12, 1301–1311, 1991. 2329

Rodriguez, E., Morris, C. S., Belz, J. E., Chapin, E. C., Martin, J. M., Daffer, W., and Hensley, S.: An assessment of the SRTM topographic products, Tech. Rep. JPL D-31639, NASA Jet Propulsion Laboratory, Pasadena, California, 2005. 2331, 2332, 2333, 2336

Rosette, J. A. B., North, P. R. J., and Suárez, J. C.: Vegetation height estimates for a mixed temperate forest using satellite laser altimetry, Int. J. Remote Sens., 29 (5), 1475–1493, 2008. 2330, 2334, 2337, 2356

Rosette, J. A. B., North, P. R. J., Suárez, J. C., and Los, S. O.: Uncertainty within satellite LiDAR estimations of vegetation and topography, Int. J. Remote Sens., 31, 1325–1342, 2010. 2330, 2335, 2336

Sellers, P. J., Los, S. O., Tucker, C. J., Justice, C. O., Dazlich, D. A., Collatz, G. J., and Randall, D. A.: A revised land-surface parameterization (SiB2) for atmospheric GCMs. part 2: The generation of global fields of terrestrial biophysical parameters from satellite data, J. Climate, 9, 706–737, 1996. 2329, 2330, 2343, 2362

Spinhirne, J. D., Palm, S. P., Hart, W. D., Hlavka, D. L., and Welton, E. J.: Cloud and aerosol measurements from GLAS: Overview and initial results, Geophys. Res. Lett., 32, L22S03,

GMDD
4, 2327–2363, 2011

doi:10.1029/2005GL023507, 2005. 2330
Suárez, J. C.: An analysis of the consequences of stand variability in Sitka spruce plantations in Britain using a combinations of airborne LiDAR analysis and models, PhD Thesis, University of Sheffield, Sheffield, UK, 2010. 2333
Suárez, J., Rosette, J., Nicoll, B., and Gardiner, B.: A practical application of airborne LiDAR for forestry managemen in Scotland, in: Proc. Silvilaser 2008, edited by: Hill, R. A., Rosette, J., and Suárez, J., 8th Int. Conf. on LiDAR applications in forest assessment and inventory, Edinburgh, UK, ISBN 978-0-85538-774-7, 2008. 2332
Tucker, C. J., Justice, C. O., and Prince, S. D.: Monitoring the grasslands of the Sahel 1984–1985, Int. J. Remote Sens., 7, 1571–1582, 1986. 2329
van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423–3441, doi:10.5194/acp-6-3423-2006, 2006. 2329
Waring, R. H., Way, J., Hunt Jr., E. R., Morrissey, L., Ranson, K. J., Weishampel, J. F., Oren, R., and Franklin, S. E.: Imaging RADAR for ecosystem studies, BioScience, 45, 715–723, 1995. 2330
Zwally, H. J., Schutz, R., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J., and Thomas, R.: ICESat's laser measurements of polar ice, atmosphere, ocean, and land, J. Geodyn., 34 (3–4), 405–445, 2002. 2330, 2331, 2334
Zwally, H. J., Schutz, R., Bentley, C., Bufton, J., Herring, T., Minster, J., Spinhirne, J., and Thomas, R.: GLAS/ICESat L2 Global Land Surface Altimetry Data (GLA14), release V31, http://nsidc.org/data/gla14.html, 2008. 2331

GMDD
4, 2327–2363, 2011

Table 1. List of GLAS parameters retained and of parameters added (below second double line).

GLA14 code	Description
i_lat	Latitude
i_lon	Longitude
i_elev	Waveform reference elevation
i_SolAng	Solar incidence angle
i_gdHt	Geoid height
i_DEM_elv	DEM elevation
i_sigBegOff	Signal begin range increment
i_ldRngOff	Land range offset
i_SigEndOff	Signal end range offset
i_gpCntRngOff	Centroid range increment for up to six peaks
i_maxSmAmp	Peak amplitude of smoothed received echo
i_numPk	Number of peaks found in the return
i_Gamp	Amplitude of up to six Gaussians
i_Garea	Area under up to six Gaussians
i_satElevCorr	Saturation Elevation Correction
i_satCorrFlg	Saturation Correction Flag
i_FRir_cldtop	Full Resolution 1064 Cloud Top
Field	Vegetation (field) height (m)
slope	Maximum of slope with 8 surrounding cells (%)
jday03	Days since 1 January 2003 (=1)

GMDD
4, 2327–2363, 2011

Table 2. Cumulative percentage of data removed by subsequent filters (Sect. 3.1) for 3 test tiles (Analysis on data collected for 2003 only).

	20°–25° N, 0°–5° E (Algeria)	50°–55° N, 0°–5° E (W. Europe)	5° S–0°, 65°–60° W South America
Dominant land cover	Bare soil	Agriculture	Broad leaf evergreen
Missing data	0.00 %	0.00 %	0.00 %
Slope > 10°	1.33 %	56.25 %	0.99 %
Difference $h > 8$ m	2.93 %	59.2 %	11.54 %
Area Gaussian 1 > 1 V ns	5.49 %	62.4 %	46.41 %
Amplitude Gaussian 1 > 0.05 V	6.00 %	63.0 %	57.4 %
Outlier test (> 99.9 %)	6.10 %	63.1 %	57.5 %
Sigma test (Gaussian 1–6; > 99.9 %)	6.11 %	63.1 %	57.5 %
Neighbour test	9.16 %	66.8 %	76.1 %

GMDD
4, 2327–2363, 2011

Table 3. Summary statistics comparing estimates of vegetation height from GLAS data with aircraft LiDAR measurements. Columns under "Raw" show statistics with no filter applied to the GLAS data and the vegetation height estimated from the difference between the beginning and end of signal. Columns under "Filtered" show the statistics with a filter $k = 1$ applied to the GLAS data (Sect. 3.1); "n" indicates the number of observations where the centres of the aircraft laser shots and the GLAS laser shots were located within 20 m; "r" is the coefficient of correlation, "RMSE" is the root mean square error and "bias" is the average difference between GLAS and aircraft measurements; see also Fig. 2.

	Raw				Filtered			
	n	r	RMSE	bias	n	r	RMSE	bias
Boreas (CDN)	228	0.39	13.9	8.1	142	0.80	4.2	−0.5
Loobos (NL)	62	0.64	7.1	−0.4	36	0.81	5.6	−4.0
Tharandt (D)	127	0.54	11.9	4.0	46	0.34	15.2	−1.7
Tambopata (PE)	648	0.32	15.1	−3.9	27	0.72	9.9	−6.5
Tumbarumba (AUS)	420	0.39	15.5	−1.6	10	0.91	9.5	−5.8
Glen Affric (GB)	61	0.13	42.3	24.4	8	0.89	4.1	0.4
Aberfoyle (GB)	190	0.16	39.0	24.8	17	0.40	12.1	3.5
Combined (-Peru)	1088	0.36	22.7	7.2	256	0.67	8.1	−0.4

GMDD
4, 2327–2363, 2011

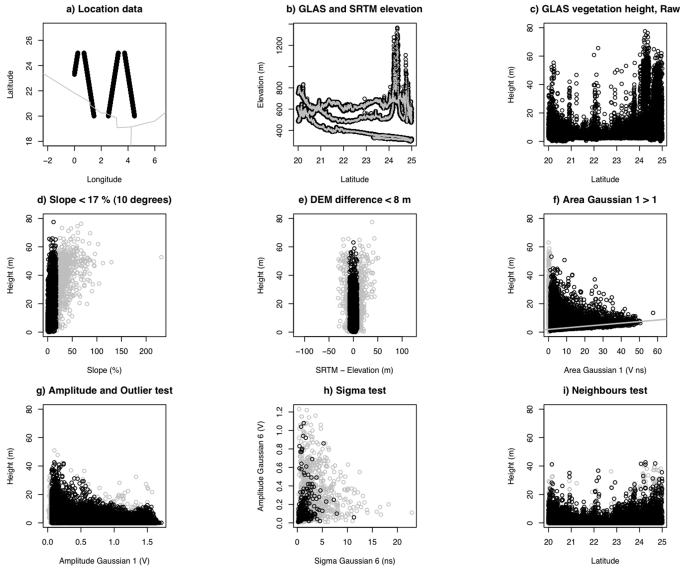

Fig. 1. **(a)** Location of the GLAS data collected between 20°–25° N and 0°–5° E prior to April 2003. **(b)** Elevation as a function of latitude for the measurements shown under a; black circles are GLAS elevation measurements; they are overlain by grey dots (SRTM 4.1 values). **(c)** Vegetation height estimated from the GLAS data after Rosette et al. (2008); no filter was applied. **(d)** Estimated vegetation height as a function of slope. The slope was calculated as the maximum of the slope in 8 directions calculated from the 90 m SRTM version 4.1 data. Grey values show data for slope ≥ 17%; black values are for slopes <17 %. **(e)** Vegetation height as a function of the difference between the GLAS reference elevation and the SRTM version 4.1 elevation. Grey circles show values that passed the 17 % slope filter in d; black circles show the data with a difference in DEM <8 m. 1.f) Vegetation height as a function of the Area of the first Gaussian; black circles pass the test, line indicates the best fit through the 5 % values per equal area interval of 10 V ns. **(g)** Amplitude test; threshold at 5 V, top 0.1 % of highest values per Amplitude interval are removed, **(h)** values with a very high signal width (sigma) are removed (grey values), **(i)** remaining values after Neighbour test is applied (compare with **(c)**).

GMDD
4, 2327–2363, 2011

Vegetation height between 60° S and 60° N from GLAS

S. O. Los et al.

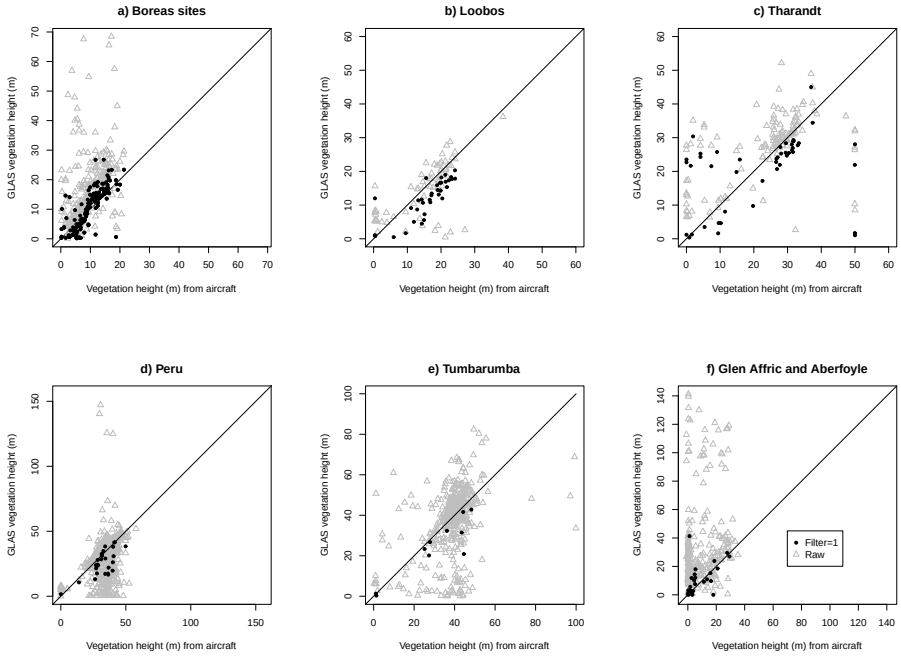

Fig. 2. Comparison of GLAS vegetation height retrievals with vegetation height measurements from aircraft LiDAR averaged to a 50 m by 50 m grid. Distance between the centre of the GLAS shot and centre of the 50 m grid cell is less than 20 m. Vegetation height from GLAS is estimated both from the raw data (grey triangles) and the filtered data ($k = 1$; black dots). Statistics are shown in Table 3. **(a)** Former Boreas sites (Canada), **(b)** Loobos site (the Netherlands) **(c)** Tambopata (Peru), **(d)** Tharandt (Germany), **(e)** E of Tumbarumba (Australia), **(f)** Glen Affric and Aberfoyle (UK).

GMDD
4, 2327–2363, 2011

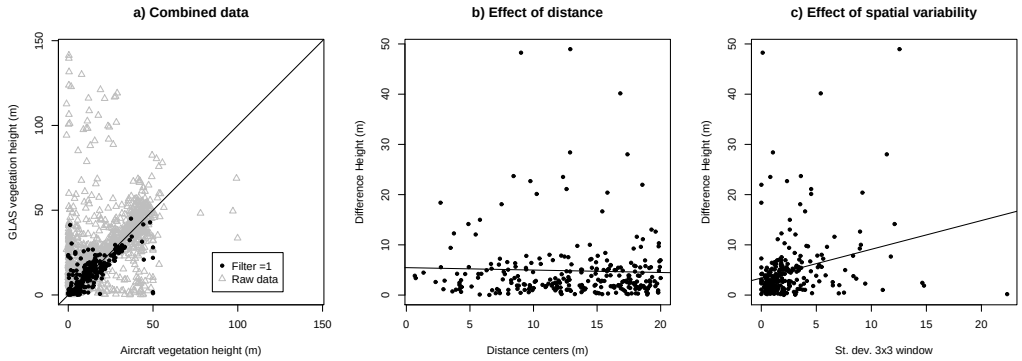

Fig. 3. (a) Combined aircraft data and GLAS data (-Peru) of Fig. 2. See Table 3 for statistics. **(b)** Difference in GLAS (Filter $k = 1$) and aircraft vegetation height estimates as a function of distance between the centre of the GLAS pulse and the centre of the aircraft 50 m by 50 m grid cell. The slope of the regression line is not statistically significant. The maximum error does increase with distance, however. **(c)** Variation in difference between the GLAS and aircraft vegetation height (absolute values) as a function of the spatial variability in vegetation height in the aircraft measurements (standard deviation of a 3 by 3 window around the centre of the 50 m grid cell). The slope of the regression line is statistically significant; ($p \ll 0.01$), the coefficient of correlation is $r = 0.25$.

GMDD
4, 2327–2363, 2011

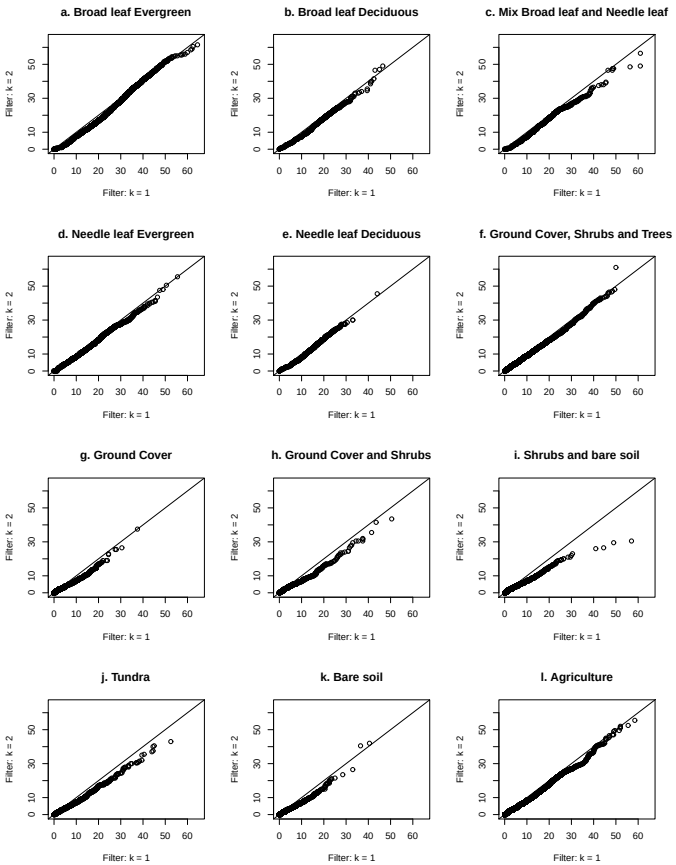

Fig. 4. Quantile-quantile plots for probability distributions of vegetation height using filtered data with $k = 1$ (x-axis) or $k = 2$ (y-axis).

GMDD
4, 2327–2363, 2011

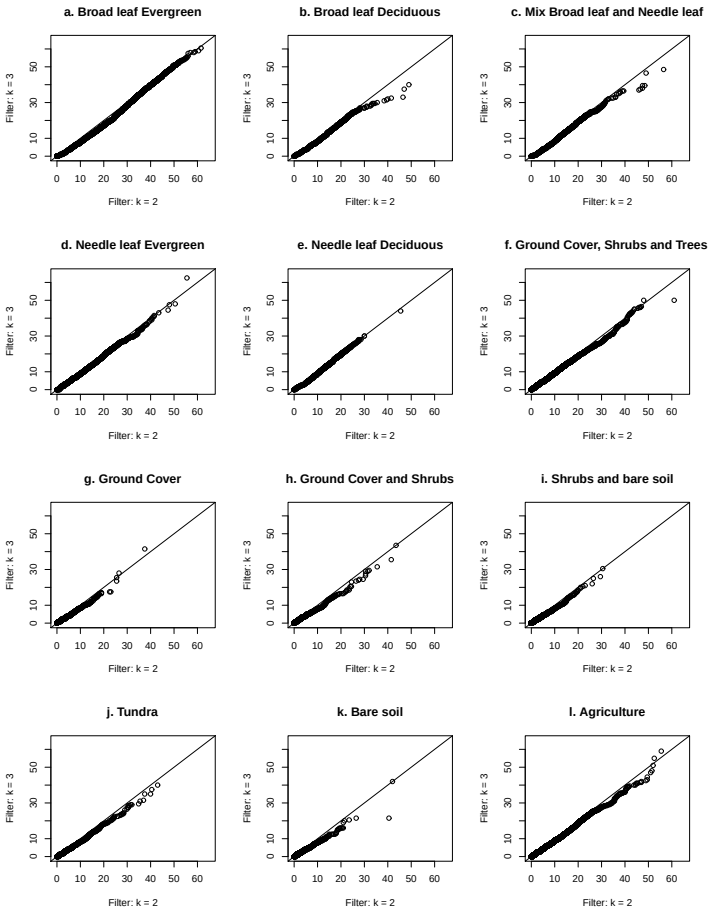

Fig. 5. Same as Fig. 4 but for filtered data with $k = 2$ (x-axis) and $k = 3$ (y-axis).

GMDD
4, 2327–2363, 2011

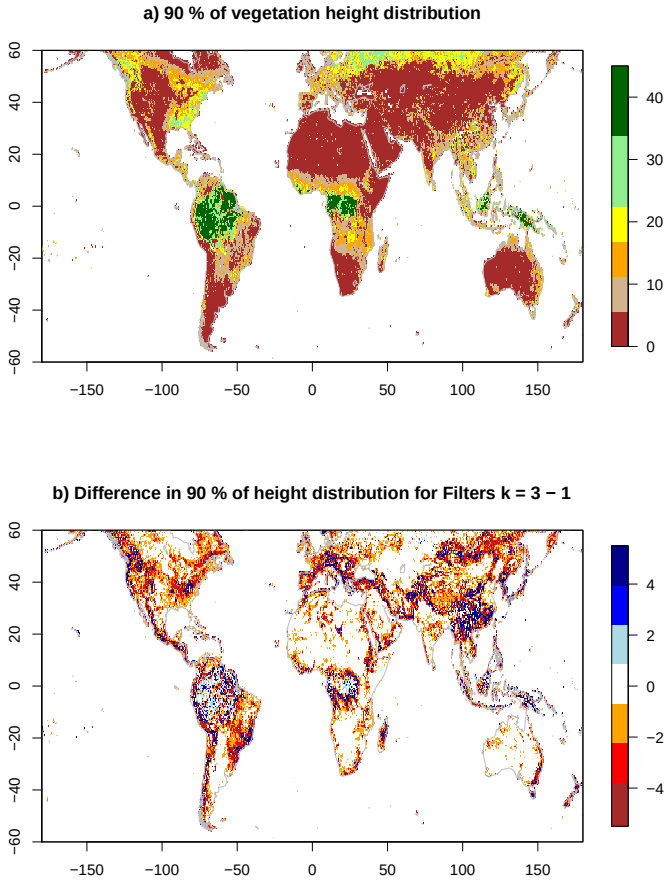

Fig. 6. **(a)** Spatial distribution of 90 % vegetation height in m for filtered data with $k = 3$; **(b)** difference in 90 % height in m for filtered data with $k = 3$ and $k = 1$ (filter $k = 3$ – filter $k = 1$); vegetation height in the tropics increases when a more conservative filter is used, whereas vegetation height in mountainous regions decreases at the same time.

GMDD
4, 2327–2363, 2011

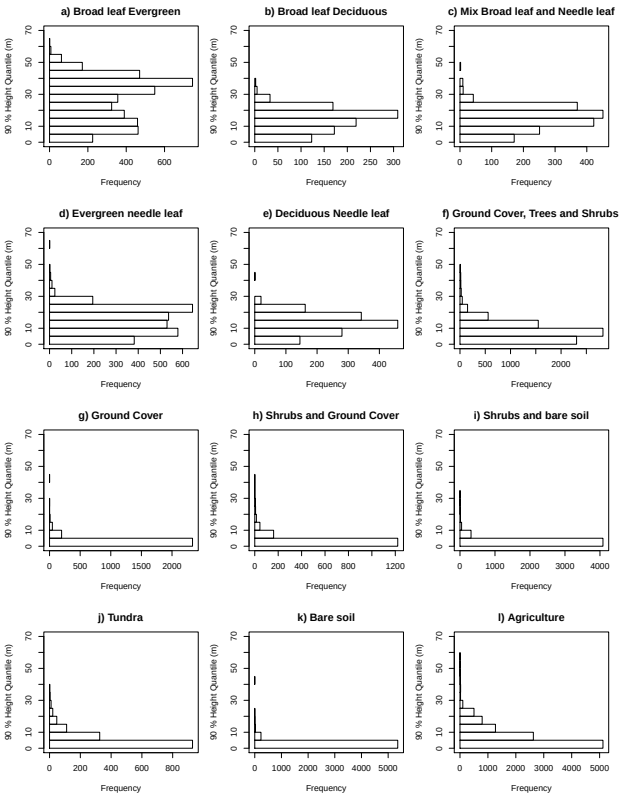

Fig. 7. Globally retrieved height frequency distributions by SiB vegetation class (Loveland et al., 2001) for Filter $k = 3$; height values for SiB biomes (Sellers et al., 1996) are given for comparison: broadleaf evergreen **(a)** = 35 m; broadleaf deciduous **(b)** and mixed broadleaf and needleleaf **(c)** = 20 m; evergreen needleleaf **(d)** and deciduous needleleaf **(e)** = 17 m; classes with a majority of ground cover (**(f)**, **(g)**, **(h)**, **(i)**) , bare soil **(k)** and agriculture **(l)** = 1 m; shrubs and bare soil = 0.5 m and tundra = 0.6 m

GMDD
4, 2327–2363, 2011

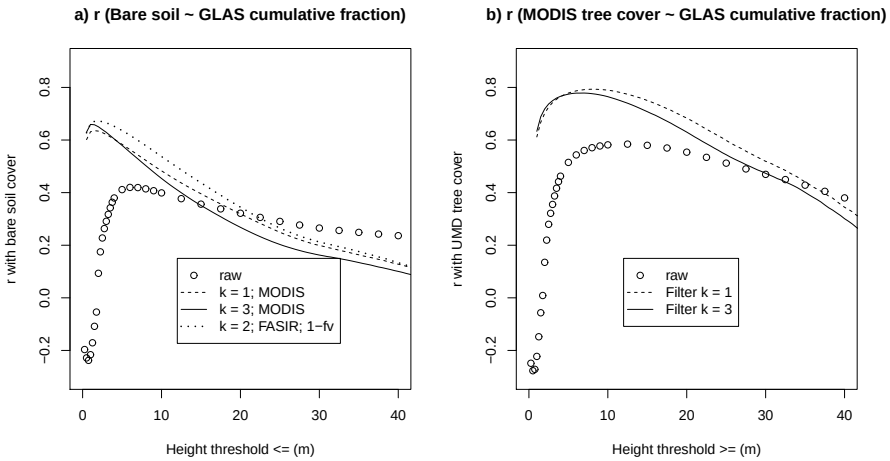

Fig. 8. **(a)** Coefficient of correlation between University of Maryland (UMD) MODIS bare soil fraction and GLAS bare soil fraction as a function of the height threshold used to identify bare soil. For bare soil estimated from raw data, the maximum $r = 0.42$ is at 6 m; for filter $k = 1$ the maximum $r = 0.549$ is at 1.5 m; for filter $k = 2$ the maximum $r = 0.553$ is at 1 m (line not shown); for filter $k = 3$ the maximum $r = 0.551$ is at 1 m. Maximum correlation with FASIR $1 - f_V$ $r = 0.58$ is at 2.0 to 2.5 m. **(b)** Coefficient of correlation between UMD MODIS tree-cover fraction and GLAS tree-cover fraction as a function of the height threshold used to identify trees. For raw data the maximum $r = 0.584$ is at 12.5 m; for filter $k = 1$ the maximum $r = 0.768$ is at 7.5 m; for filter $k = 2$ the maximum $r = 0.759$ is at 7 m (line not shown); for filter $k = 3$ the maximum $r = 0.743$ is at 6 m.

GMDD
4, 2327–2363, 2011