# EXPLORING THE MEASUREMENT OF FORESTS WITH FULL WAVEFORM LIDAR THROUGH MONTE-CARLO RAY TRACING

*Steven Hancock[1,2], Mathias Disney[1], Philip Lewis[1] Jan-Peter Muller[2].*

1, Department of Geography, University College London, Gower street, London, WC1E 6BT. United Kingdom and NERC Centre for Terrestrial Carbon Dynamics.
2, Mullard Space Science Laboratory, University College London, RH5 6NT, United Kingdom.

**KEY WORDS:** Forestry, structure, lidar, vegetation, simulation, 3D modelling, satellite.

**ABSTRACT:**
This paper presents results from two simulation studies which attempt to measure forest height with full waveform lidar. Monte-Carlo ray tracing is used to simulate a full waveform lidar response over explicitly represented 3D forest models. Gaussian decomposition and multi-spectral edge detection are used to estimate tree top and ground positions over a range of forest ages, stand densities and ground slopes. The error of the height estimates are precisely quantified by comparison with the 3D model height. This paper discusses the development and testing of the inversion of tree height from simulations of lidar data, assuming a fixed set of lidar characteristics (corresponding to an LVIS-like instrument). It is shown that Gaussian decomposition performs reasonably well (mean 5.3m overestimate for <75% cover) for all but the densest canopies over flat ground. The potential of multi-spectral edge detection to separate ground and canopy returns from blurred waveform is demonstrated. The methods presented will be refined and extended to instrument-specific cases.

## Background

Carbon flux models are essential for understanding the complex processes involved in the Earth's climate (Woodward et al, 2004). These models need variables, such as biomass and leaf area at a range of scales and locations (Williams et al, 2005). Many areas are inaccessible and it would be prohibitively expensive to cover the world with airborne sensors. Space-borne remote sensing may be the solution.

Empirical relationships and physical models have been developed to relate biomass and leaf area to vegetation indices from passive optical instruments, such as MODIS (Myneni et al, 2002). These indices saturate at moderate canopy densities (LAI of 3 or 4, Lovell et al, 2003). Synthetic aperture radar suffers from similar saturation problems (Waring et al., 1995). This saturation would bias any global remotely sensed data assimilation scheme.

In contrast lidar derived vegetation parameters are less prone to saturation over forests as the light can fit through small gaps (Lefsky et al., 1999) and it allows a direct measurement of tree height. The capabilities of spaceborne lidars for measuring vegetation were demonstrated by the GLAS instrument aboard ICESat (Rosette et al., 2008).

### Simulation system

There has been a great deal of work on estimating forest parameters from full waveform lidar in recent years (Wagner et al, 2008, Reitberger et al, 2008). The results of these studies are promising; however positional uncertainty of remote measurements and the difficulty in seeing the tree top from ground level make validation difficult (Hyde et al, 2005).

Computer simulations allow "validation" as the true parameters of the virtual forest are known, unlike reality where there is always some uncertainty. This paper will use a Monte-Carlo ray tracer based upon the RAT library developed from "frat" (Lewis, 1999) to simulate a waveform lidar. Frat has been validated in the RAdiation transfer Model Inter-comparison exercises (RAMI) two and three, (Pinty et al. 2004, Widlowski et al, 2007). These exercises have not yet tested range resolved simulations but they have shown that the radiometry of a core set of explicit 3D models, including frat, agree to within 1% over vegetation for most cases.

Explicit geometric forest models, in which every needle is described (Disney et al, 2006) were used for the simulations. Individual Sitka spruce trees of different ages were created using the Treegrow model (Leersnijder, 1992). These trees were cloned at random locations, with a minimum separation, over a variety of slopes to form forests with a range of stand densities and different age mixes. The affect of different light regimes in different densities on tree structure was ignored. Figure 1 shows a ray traced image of a forest with an equal proportion of trees of each age on a $20^{o}$ slope. In addition to the simulated light returns the percentage coverage of each material in each range bin were recorded and used for validation of derived parameters. Spectra from the Prospect model (Jacquemond and Baret, 1990), measurements from the LOPEX dataset (Hosgood et al, 1995) and the model of Price (1990) were used for leaf, bark and soil spectra respectively. Modelled spectra were matched against OTTER field data. In a change to the method of Disney et al., (2006), needles were allowed to transmit light, trusting in the accuracy of Prospect in the absence of reliable transmittance data.

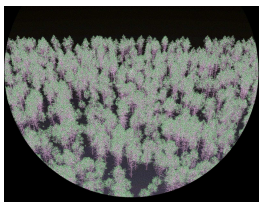

Figure 1, Simulated image of mixed age Sitka spruce forest model.

Using explicit 3D models is computationally expensive but avoids assuming that canopies behave as turbid media; an

assumption that ignores the heterogeneity of real trees. It is not clear how such an assumption would affect derived results, especially when parameters are derived using the same assumptions used to create the forests (Widlowski et al, 2005). Figure 2 shows a simulated waveform over a single thirty year old Sitka spruce tree model. The heterogeneity and subterranean echoes caused by multiple scattering are apparent. The ground is at a range of 1,200m.

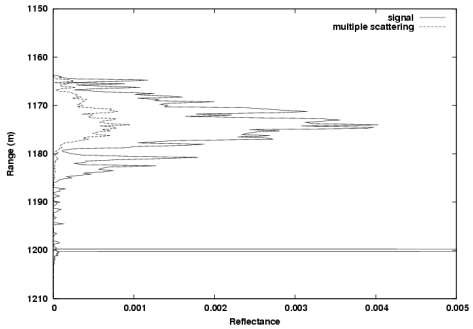


Figure 2. Simulated waveform from a single Sitka spruce tree.

Simulations were run with a resolution of 25cm (which can be coarsened for analysis), a range of wavelengths (including 532nm, 850nm, 1064nm, 1650nm and 2060nm), a 30m ground footprint (the optimum for forestry, Zwally et al, 2002) and with and without a temporal laser pulse (100ns is proposed for A-scope). Gaussian noise can be added before analysis. The width of the Gaussian is defined as a percentage of the maximum signal return. The methods will be developed for an infinitesimally short pulse before the extra complication of deconvolution is added. Systems such as GLAS have short pulses (around 2ns) which should not need any deconvolution unless the range sampling is significantly finer.

**Derivation of parameters**

Estimation of forest parameters from lidar relies on the ground returns being distinguishable from the canopy returns (Hofton et al, 2002). This can either be achieved with multiple first return scans (Koetz et al, 2007) or a single full waveform measurement (Zwally et al, 2002) to get a distribution of returns from throughout the canopy. Due to the speed of spaceborne platforms and the subsequent sparsity of sampling only full waveform lidar is suitable for measuring vegetation from space.

The standard method is to decompose the waveform into a set of Gaussians by non-linear regression (Hofton et al, 2000). The distribution of Gaussians can be used to classify cover type (Wagner et al, 2008, Reitberger et al, 2008) and (taking into account relative cross sections) can be used to derive vegetation height (Blair et al, 1999), estimate canopy cover (Lefsky et al, 2005) and through metrics derive other biophysical parameters such as leaf biomass and leaf area index (Lefsky et al, 1999).

Often there is no clear separation of ground and canopy returns, either due to dense understory, small separation of canopy and ground or topography. Attempts have been made to improve height estimates in these situations by using another data source to estimate the ground position (Rosette et al, 2008). Care must be taken that these ground elevation datasets give the true height (for example SRTM saturates over forests). Accurate datasets are not available globally.

# Method

This investigation explores methods that use only the waveform to estimate tree height (which can be linked to biomass through allometric relationships and stand counts). Other characteristics would need to be inferred with additional information and will not be investigated in this paper. Fusing lidar with hyper-spectral and multi-angular data would greatly help in the derivation of these biophysical parameters however the lidar waveform alone should provide the best height profile.

Before Gaussians are fitted to the simulated signal it is pre-processed in the following order;

- 5% Gaussian noise was added, as described above.
- The signal was pre-smoothed by convolution with a 3m Gaussian.
- Noise statistics are calculated from a known empty portion of signal (above canopy to avoid echoes).
- The signal was de-noised by subtracting a threshold of the mean noise plus (an arbitrary) three standard deviations
- The signal was post-smoothed with a 1m Gaussian.

The empty tails are cropped from the signal to constrain the Gaussian decomposition. The positions and amplitudes of all turning points are recorded along with the width of peaks. If more features than the number of Gaussians to be fitted are found (due to heterogeneous or noisy signals) the Gaussians with the largest cross sections are used first. If too few are found (skewed Gaussians for example) the extra Gaussians are evenly spaced in the gaps. An implementation of the Levenberg-Marquardt method was used to minimise the root mean square difference between the fitted Gaussians and original signal (Press et al, 1994). It has been found that the best fits are achieved when the x and y axes are rescaled to between 0 and 100.

The fitted Gaussians and the pre-processed signal were analysed to derive biophysical parameters. The centre of the last Gaussian is taken as the ground position if the energy contained within is more than an arbitrary percentage (1%) of the total energy. This should avoid any Gaussians fitted to noise or subterranean echoes caused by multiple scattering. If the Levenberg-Marquardt method fails to find a solution or the derived parameters are unrealistic the fitting is repeated with one less Gaussian.

The tree top is calculated form the pre-processed signal. Taking it as the point at which the signal rises above the noise threshold will always lead to an underestimate of height. Data assimilation schemes such as the Kalman filter rely on unbiased observations (Williams et al, 2005). For this reason it may be better to try to estimate a point that could be an over or under estimate. The first point at which the signal drops to the mean noise level before it rises above the noise threshold would seem to be a sensible, unbiased estimate of tree top position. Figure 3 shows a histogram of the signal start position error with and without tracking back from the noise threshold to the mean noise value. One hundred simulated waveforms were used with ten thousand separate sets of random noise added to give one million estimates. A negative error means a premature signal start; this was common in both methods.

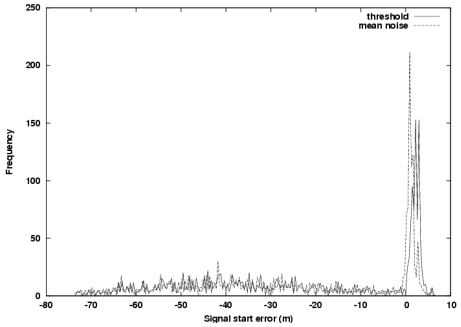

Figure 3. Signal start error histogram.

The non-tracking method gave a mean error of -22.99m compared to the tracking method's error of -23.41m. However, the tracking method has a modal error of 0.91m compared to the non-tracking method modal error of 2.23m. The tracking method error distribution looks less biased than the non-tracking; however the long tail of premature triggerings have distorted the mean. Ground height is unlikely to vary more than 30m over a 30m footprint in forests and the maximum height of trees can be estimated, therefore a gradual increase of the noise threshold to ensure the signal start is not more than the footprint size plus maximum tree height with a tolerance from the last return may eliminate the long tail.

These methods perform reasonably well when the ground return contains significant energy and is distinguishable form the canopy return. In very dense canopies (>90% coverage) little signal reaches the ground. In such cases a proportion of the measurements will fail; this proportion needs to be quantified. This canopy cover is not uncommon for evergreen broadleaf forests (Hofton et al, 2002). Accuracy will be sensitive to the ratio of the ground signal to noise.

Figure 4 shows true tree height against derived height for flat ground with a range of densities and ages with 5% noise added. The five different age classes are visible up the y axis. The actual tree height was calculated from the material information, taking the range difference between the first leaf return and the mean position of the soil returns. This may be slightly different to the actual tree height (also recorded) because of the sampling of the ray tracer but is the best estimate that can possibly be derived from the signal.

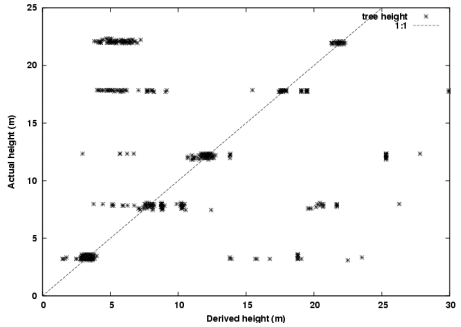

Figure 4. True against derived tree height with 1:1 line.

Topography blurs the ground and canopy returns together (Harding and Carabajal, 2005), even without noise. Figure 5 shows a return from an old growth forest on a 30° slope. The ground and canopy returns cannot be reliably separated by Gaussian decomposition. Undertory will have the same effect.

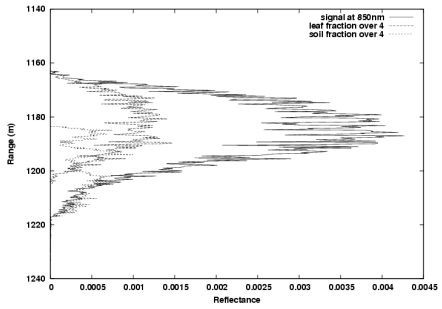

Figure 5. Topographically blurred waveform.

**Multi-spectral lidar**

Two wavelengths with different canopy to ground reflectances ratios should allow the ground to be distinguished. Figure 6 shows the ratio of leaf to soil reflectances against wavelength for the spectra used. A canopy also includes bark and this, being a similar colour to soil will bring the ratios closer to unity. There is still a contrast between the visible and near infra-red. Wavelengths of 650nm and 800nm have been used for the initial trials as they show a large contrast, although an instrument using 1064nm and 532nm will be easier to build. Hyper-spectral lidar may also be available one day (Kaasalainen et al, 2007).

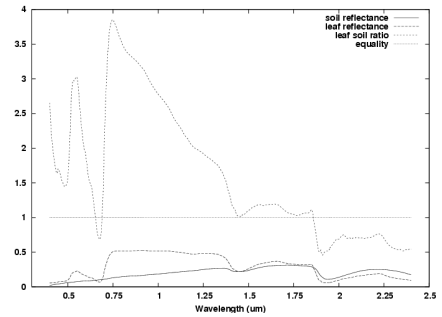

Figure 6. Leaf and soil spectra and their ratio.

In the absence of noise the ground occurs at the point of the maximum ratio of 550nm over 850nm. Noise complicates the issue, unsurprisingly.

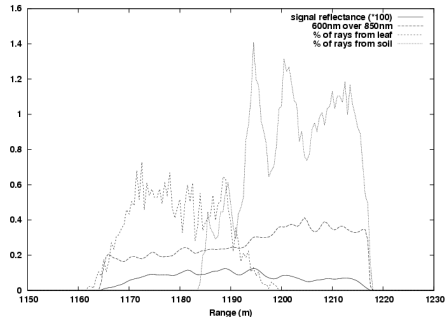

Figure 7. Spectral ratio showing edges due to ground start, with noise alongside material contributions.

If the signals are blurred by topography or understory the ground should be visible as a change in the ratio of visible to near infra-red reflectance. This can be found with traditional edge detection methods.

The signal was de-noised, taking care to preserve the energy ratio between the bands, assuming that any noise bias does not distort the ratio. The ratio of visible to near infra-red was smoothed with Gaussians of decreasing width so that the smallest Gaussian to leave a single crossing point of the second differential (corresponding to a maximum gradient) was used. Using the smallest possible Gaussian ensured the best localisation of the edge. To avoid the large gradients of the signal start and end only a window between the first and last crossing points of the third differential (after smoothing) was used. The position of the maximum gradient was taken as the start, the last return above noise as the end and the average taken as the centre of the ground return respectively. Figure 8 shows the mean error in derived ground position for one hundred simulated waveforms with 5% noise added over forests with different canopy covers on a 30$^o$ slope. Error bars show the standard deviations.

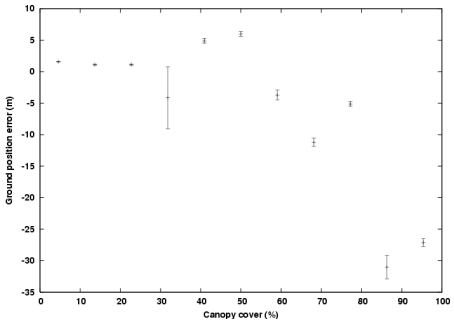

Figure 8. Ground position error against canopy cover

The method performs well provided there is a reasonable ground return. For very dense canopies the ground return is less than the noise threshold and so would be hard to see, even without topography.

If it can be assumed that the tallest trees are equally spaced across the laser footprint then the tree height will be the distance between the first signal return and first ground return. The forest is unlikely to be so homogeneous over a 30m footprint. The measure of canopy height need not be the tree top. Lefsky et al (1999) showed the possibilities of using parameters such as the median canopy height to estimate biomass and other biophysical parameters through allometrics. This can be calculated provided the energy returned from the canopy and ground can be separated. The ground start and end could be used to further constrain the Gaussian fitting then the median of the canopy calculated. How this relates to biomass will depend upon the forest's horizontal and vertical heterogeneity.

## Conclusions

This paper has briefly outlined work in progress on the development of algorithms for measuring tree height from above canopy waveform lidar. The standard method of Gaussian decomposition has been implemented, using realistic simulations to precisely quantify the technique's error (mean overestimate of 5.7m for <75% cover; refining of the method is needed).

A method for determining ground position of signals blurred by topography or understory with dual waveband lidar has been introduced and shown to work reliably for all but the densest canopy covers and the errors quantified (mean RMSE of 3.2m for <65% cover). Relatively few simulations were available (100 over sloping and 225 over flat forests); more are being processed with a wider range of densities, slopes and tree species for use in further investigations. These will explore the affect of noise, thresholds, range resolution, canopy cover and pulse length on derived tree height error.

A reliable method for determining a measure of canopy height over slopes is needed. The median canopy position may be the easiest to derive. Its sensitivity to heterogeneity and noise will be investigated.

The premature estimation of tree top suggests that the threshold used was not high enough. This could be increased to the mean noise value plus four or more times the standard deviation, but having this fixed would increase the chances of thresholding out weak ground returns. An adaptive threshold based upon estimates of maximum likely tree heights or signal shape (the tree top is unlikely to be isolated from the rest of the tree) may give a better solution and should be investigated.

**Acknowledgments**

This work was funded by the NCEOI (Natural environment research council Centre for Earth Observation and Instrumentation) for project SA-014-DJ-2007 (Hyperspectral Imaging LiDAR) and the Hollow waveguide technology for $CO_2$ and canopy measurements (A-scope) project through the Centre for Terrestrial Carbon Dynamics (CTCD).and the EPSRC, studentship no. GR/9054658

## References

Blair JB, Rabine DL, Hofton MA. 1999. The laser vegetation imaging sensor: a medium-altitude digitisation-only, airborne laser altimeter for mapping vegetation and topography. ISPRS journal of photogrammetry and remote sensing, 54, pp. 115-122.

Disney, M., Lewis, P., Saich, P., 2006. 3-D modelling of forest canopy structure for remote sensing simulations in the optical and microwave domains. Remote Sensing of Environment, 100, pp. 114-132.

Harding DJ, Carabajal CC, 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. Geophysical research letters, 32, L21S10.

Hofton MA, Minster JB, Blair JB. 2000. Decomposition of laser altimeter waveforms. IEEE transactions on geoscience and remote sensing, 38, pp. 1989-1996.

Hofton MA, Rocchio LE, Blair JB, Dubayah R, 2002. Validation of vegetation canopy lidar sub-canopy topography measurements for a dense tropical forest. Journal of geodynamics, 34 pp. 491-502.

Hosgood G, Jacquemond S, Andreoli G, Verdebout J, Pegrini G, Schmuck G, 1995. Leaf optical properties experiment 93 (LOPEX 93). Ispra, Italy; European commission, Joint Research Centre institute of remote sensing applications.

Hyde P, Dubayah R, Peterson B, Blair JB, Hofton M, Hunsacker 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, pp. 427-437.

Jacquemond and Baret, 1990. Prospect: A model of leaf optical properties spectra. Remote sensing of environment, 34, pp. 75-91.

Kaasalainen S, Lindroos T, Hyyppa J, 2007. Toward hyperspectral lidar: Measurement of spectral backscatter intensity with a supercontinuum laser source. IEEE geoscience and remote sensing letters, 4, pp. 211-215.

Koetz B, Sun G, Morsdorf F, Ranson KJ, Kneubühler M, Itten K, Allgöwer B. 2007. Fusion of imaging spectrometer and LIDAR data over combined radiative transfer models for forest canopy characterization. Remote Sensing of Environment, 106, pp. 449-459.

Leersnijder RP, 1992. PINOGRAM: A pine growth model. WAU dissertation 1499, Waganingen Agricultural university, The Netherlands.

Lefsky MA, Cohen WB, Acker SA, Parker GC, Spies TA, Harding D. 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir and western hemlock forests. Remote Sensing of Environment, 10, pp. 339-361.

Lefsky MA, Harding DJ, Keller M, Cohen WB, Carabajal CC, Espirito-Santo FDB, Hunter MO, de Oliveira Jr R. 2005. Estimates of forest canopy height and aboveground biomass using ICESat. Geophysical research letters, 32, L22S02.

Myneni RB, Hoffman S, Knyazikhin Y, Privette JL, Glassy J, Tian Y, Wang Y, Song X, Zhang Y, Smith GR, Lotsch A, Friedl M, Morisette JT, Votava P, Nemani RR, Running SW, 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote sensing of environment, 83, pp. 214-231.

Lewis, P. 1999. Three-dimensional plant modelling for remote sensing simulation studies using the Botanical Plant Modelling System. Agronomie, 19(3-4), pp. 185-210.

Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., Coops, N.C., 2003. "Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests." Canadian Journal of Remote Sensing 29(5), pp. 607-622.

Pinty B, Widlowski J-L, Taberner M, Gobron N, Verstraete MM, Disney M, Gascon F, Gastellu J-P, Jiang L, Kuusk A, Lewis P, Li X, Ni-Meister W, Nilson T, North P, Qin W, Su W, Tang S, Thompson R, Verhoef W, Wang H, Wang J, Yan G, Zang H. Radiation transfer model intercomparison (RAMI) exercise: results from the second phase. Journal of geophysical research, 109, D06210.

Press W.H., Tuekolsky S.A., Vetterling W.T. Flanery B.P. 1994. Numerical Recipes in C. Cambridge University Press, second edition (with reprints).

Price JC, 1990. On the information content of soil reflectance spectra. Remote sensing of environment, 33, pp. 113-121.

Reitberger J, Krzystek P, Stilla U. 2008. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. International journal of remote sensing, 29, pp. 1407-1431.

Rosette JAB, North PRJ, Suarez JC. 2008. Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. International journal of remote sensing, 29, pp. 1475-1493.

Wagner W, Hollaus M, Briese C, Ducic V. 2008. 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. International journal of remote sensing, 29, pp. 1433-1452.

Waring RH, Law B, Goulden ML, Bassow SL, McCreight RW, Wofsy SC, Bazzaz FA (1995) Scaling gross ecosystem production at Harvard Forest with remote sensing: a comparison of estimates from a constrained quantum-use efficiency model and eddy correlation. Plant Cell Environment, 18, pp. 1201–1213.

Widlowski J-L, Pinty B, Lavergne T, Verstraete MM, Gobron N. 2005. Using 1-D models to interpret the reflectance anisotrpy of 3-D canopy targets: Issues and caveats. IEEE transactions on geoscience and remote sensing, 4(9), pp. 2008-2017.

Widlowski J-L, Taberner M, Pinty B, Bruniquel-Pinel V, Disney M, Fernandes R, Gastellu-Etchegorry J-P, Gobron N, Kuusk A, Lavergne T, Leblanc S, Lewis PE, Martin E, Mõttus M, North PRJ, Qin W, Robustelli M, Rochdi N, Ruiloba R, Soler C, Thompson R, Verhoef W, Verstraete MM, Xie D. 2007. Third radiation transfer model intercomparison (RAMI) exercise: Documenting progress in canopy reflectance models. Journal of geophysical research, 112, d09111.

Williams M, Schwarz PA, Law BE, Irvine J, Kurpuis MR. 2005. An improved analysis of forest carbon dynamics using data assimilation. Global change biology, 11, pp. 89-105.

Woodward FI, Lomas MR, 2004. Vegetation dynamics – simulating responses to climate change. Biological reviews, 79, pp. 643-670.

Zwally HJ, Shutz B, 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, Thomas R. 2002. ICESat's laser measurements of polar ice, atmosphere, ocean and land. Journal of geodynamics, 34, pp. 405-445.