Waveform lidar over vegetation: An evaluation of inversion methods for estimating return energy

CrossMark

Steven Hancock [a,*], John Armston [b], Zhan Li [c], Rachel Gaulton [d], Philip Lewis [e], Mathias Disney [e], F. Mark Danson [f], Alan Strahler [c], Crystal Schaaf [g], Karen Anderson [a], Kevin J. Gaston [a]

[a] Environment and Sustainability Institute, University of Exeter, United Kingdom
[b] Remote Sensing Centre, Science Delivery, Department of Science, Information Technology and Innovation, 41 Boggo Road, QLD 4102, Australia and Joint Remote Sensing Research Program, School of Geography, Planning and Environmental Management, University of Queensland, St Lucia, QLD 4072, Australia
[c] Department of Earth and Environment, Boston University, United States
[d] School of Civil Engineering and Geosciences, Newcastle University, United Kingdom
[e] UCL Geography and NERC National Centre for Earth Observation, United Kingdom
[f] Ecosystems and Environment Research Centre, School of Environment & Life Sciences, University of Salford, United Kingdom
[g] School for the Environment, University of Massachusetts Boston, United States

ARTICLE INFO

ABSTRACT

Full waveform lidar has a unique capability to characterise vegetation in more detail than any other practical method. The reflectance, calculated from the energy of lidar returns, is a key parameter for a wide range of applications and so it is vital to extract it accurately. Fifteen separate methods have been proposed to extract return energy (the amount of light backscattered from a target), ranging from simple to mathematically complex, but the relative accuracies have not yet been assessed. This paper uses a simulator to compare all methods over a wide range of targets and lidar system parameters. For hard targets the simplest methods (windowed sum, peak and quadratic) gave the most consistent estimates. They did not have high accuracies, but low standard deviations show that they could be calibrated to give accurate energy. This may be why some commercial lidar developers use them, where the primary interest is in surveying solid objects. However, simulations showed that these methods are not appropriate over vegetation. The widely used Gaussian fitting performed well over hard targets (0.24% root mean square error, RMSE), as did the sum and spline methods (0.30% RMSE). Over vegetation, for large footprint (15 m) systems, Gaussian fitting performed the best (12.2% RMSE) followed closely by the sum and spline (both 12.7% RMSE). For smaller footprints (33 cm and 1 cm) over vegetation, the relative accuracies were reversed (0.56% RMSE for the sum and spline and 1.37% for Gaussian fitting). Gaussian fitting required heavy smoothing (convolution with an 8 m Gaussian) whereas none was needed for the sum and spline. These simpler methods were also more robust to noise and far less computationally expensive than Gaussian fitting. Therefore it was concluded that the sum and spline were the most accurate for extracting return energy from waveform lidar over vegetation, except for large footprint (15 m), where Gaussian fitting was slightly more accurate. These results suggest that small footprint ($\ll$ 15 m) lidar systems that use Gaussian fitting or proprietary algorithms may report inaccurate energies, and thus reflectances, over vegetation. In addition the effect of system pulse length, sampling interval and noise on accuracy for different targets was assessed, which has implications for sensor design.

1. Introduction

Lidar has been shown to be a valuable tool for characterising vegetation, offering many advantages over other techniques due to its non-destructive measurement of structural and spectral information (Dubayah & Drake, 2005). Maps of global forest height have been determined from the spaceborne ICESat (Harding & Carabajal, 2005; Los et al., 2012), forest cover can be accurately derived from airborne lidar without the need for site-specific calibration (Armston et al., 2013), airborne lidar's structural information improves land cover classification accuracy (Mallet, Bretar, Roux, Soergel, & Heipke, 2011), airborne and terrestrial laser scanners (TLS) have also been used to characterise vegetation canopies and their effect on hydrological processes (Musselman, Margulis, & Molotch, 2013; Reid et al., 2014) and TLS can accurately measure woody volume (Raumonen et al., 2013), potentially allowing rapid biomass measurement. All of these studies rely on accurately extracting target properties from lidar signals.

* Corresponding author.
E-mail addresses: stevenhancock2@gmail.com (S. Hancock), rachel.gaulton@ncl.ac.uk (R. Gaulton), f.m.danson@salford.ac.uk (F. Mark Danson).

Lidar directly measures the 3D distribution of energy reflected from a target surface. The vast majority of research has focused on determining the range to a single feature (Stilla & Jutzi, 2008; Wagner, Ullrich, Melzer, Briese, & Kraus, 2004), with very few studies assessing the accuracy of the estimated return energy (backscattered radiation from the target). Recent research has tried to identify targets through clutter, such as the ground surface under vegetation (Jalobeanu & Gonçalves, 2014; Los et al., 2012), but they still assumed that there was a hard feature that lay at a single range. Commercial lidar systems have been optimised to measure the range to hard features.

The return energy contains useful information about complex targets, such as vegetation. The amount of energy returned can be used to calculate the size of illuminated objects (Hancock et al., 2014; Ramirez, Armitage, & Danson, 2013) and determine canopy cover (Armston et al., 2013). The return energy of certain wavelengths is related to leaf biochemistry and can be used to calculate leaf moisture (Gaulton, Danson, Ramirez, & Gunawan, 2013) and chlorophyll content (Nevalainen et al., 2014). It has been shown that lidar return energy is the most important lidar metric when classifying land cover type (Mallet et al., 2011; Neuenschwander, Magruder, & Tyler, 2009). Therefore an accurate method for retrieving return energy from lidar is vital for making physically based measurements of vegetation. In addition, vegetation canopies contain many small elements, making it likely that a lidar beam will hit multiple targets within a single return feature, potentially violating the assumptions of methods developed for single targets.

Many previous papers have assessed the accuracy of the range estimate from lidar data (Stilla & Jutzi, 2008; Wagner et al., 2004). Fewer have explored the accuracy of return energy and these previous energy extraction studies are described in Section 3.4. This study deals exclusively with methods for determining the energy of a return, with a view to improving the accuracy of physically based measurement methods of vegetation (which require no site specific calibration). This study assessed the methods in terms of their return energy accuracy, robustness and computational expense for both single, hard returns and for complex, vegetation returns. Return energy accuracy was assessed in terms of absolute accuracy and also in terms of consistency, as the value retrieved by a method can be calibrated to give true energy as long as there is a consistent, well defined relationship between the retrieved value and energy. Computational efficiency is an important consideration as Jalobeanu and Gonçalves (2014) state that "the best performing methods are too computationally intensive to be used on large datasets".

Empirical methods have been proposed to characterise vegetation without the need for target biophysical parameters to be directly derived from the lidar signal, e.g. Height Of Median Energy, HOME (Drake et al., 2002) and the leading edge extent (Lefsky, Keller, Yong, de Camargo, & Hunter, 2007). However these empirical methods require local calibration. Physically based methods, directly extracting target properties from the lidar signal, require no external calibration and so can be applied globally. The resulting parameters have a physical meaning that can be directly measured on the ground such as canopy cover, tree height or leaf area.

Fig. 1 shows maps generated using two different methods for determining return strength from a Leica ALS50-II airborne lidar. Fig. 1(a) shows the intensity reported by Leica's proprietary discrete return algorithm (sum of multiple returns per beam) whilst Fig. 1(b) shows the sum of full waveform intensity. For both, the return strength was calculated per beam and averaged into a 2 m raster. These lead to very different outputs, with forests providing a stronger return in Fig. 1(b) than (a) and so researchers might draw different conclusions about the biophysical nature of the vegetation depending on which method they used. This paper will explore the different methods to measure lidar return energy and which are more accurate over different surfaces. Several new lidar systems optimised for vegetation are in development covering terrestrial, airborne and satellite based systems (Danson et al., 2014; Douglas et al., 2015; Murooka et al., 2013; Wallace, Nichol, & Woodhouse, 2012) and so this type of data will become ever more common.

2. Lidar systems

There are two broad classes of lidar, time of flight (TOF) and phase shift systems. Phase shift systems have been shown to struggle at determining if and where a hit occurs in diffuse targets, such as vegetation due to their assumption that all reflected light comes from a single surface (Newnham, Goodwin, Armston, Muir, & Culvenor, 2012), and so these systems will not be covered here. TOF lidar systems emit a short pulse of light and measure the reflected energy, allowing the range to, and apparent reflectance of, the target surface to be determined. Within TOF lidars there are two further categories, discrete return and full waveform systems. Discrete return systems use proprietary algorithms to extract the range and energy of one or more targets along the laser beam's path (Disney et al., 2010; Jalobeanu & Gonçalves, 2012). Full waveform systems record all the reflected energy as a function of range, giving a more complete description of the scattering event (Harding, Blair, Garvin, & Lawrence, 1994) and allowing more accurate measurement of target properties over diffuse targets such as vegetation (Hancock, Disney, Muller, Lewis, & Foster, 2011). The data and processing requirements are much greater for full waveform than for discrete return lidar, although some doubts have been raised about the accuracy of range and energy derived from the proprietary discrete return algorithms over vegetation (Disney et al., 2010). This study focuses on methods for full waveform lidar, where the energy can be extracted from waveform lidar in a number of ways, ensuring that the optimal information is available for the task.
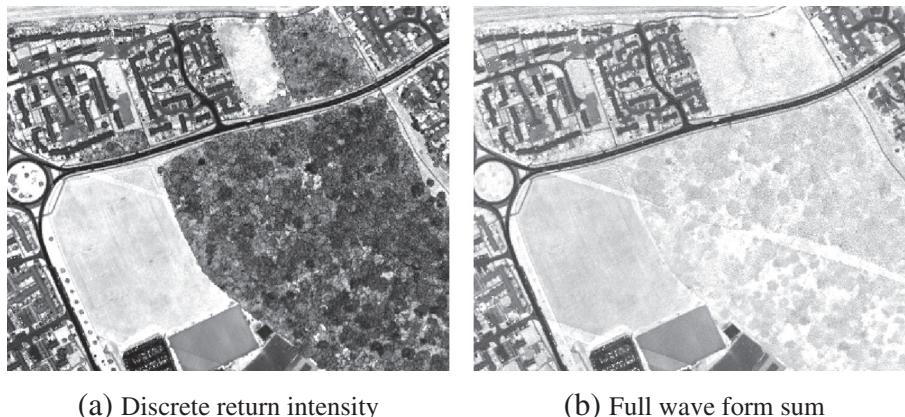

(a) Discrete return intensity   (b) Full wave form sum

Fig. 1. Two different ways to measure return strength from lidar. The site is an area in Luton, UK, with forest, grassland and buildings.
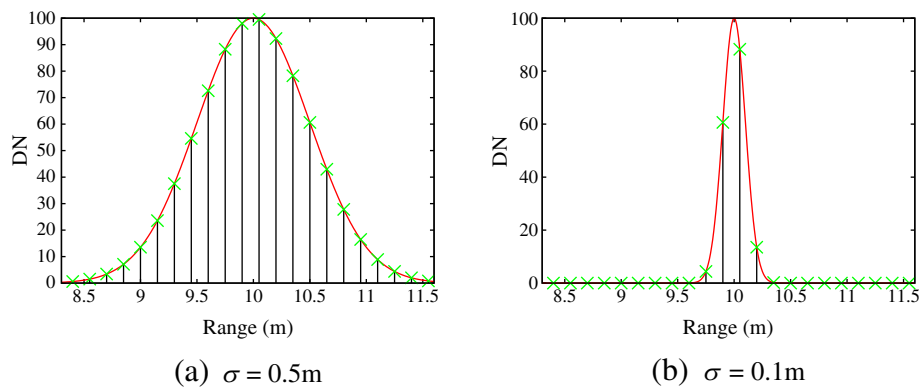
Fig. 2. Two different pulse lengths, σ, sampled at 1 ns intervals.

For full waveform systems, a digitiser records the intensity measured by a photodetector as a digital number (DN) at a preset sampling interval, referred to as a "bin". The outgoing laser pulse will have some distribution function whilst the detector will have a response function (typically assumed to be Gaussian). The convolution of these two components gives the system pulse (referred to as "the pulse" for the rest of this paper). The measured waveform is the convolution of this system pulse with the target profile, sampled at the sampling interval, plus noise. These three effects are explained in the following sections.

2.1. System pulse

Fig. 2 illustrates two Gaussian returns sampled at 1 ns (see online version for colour figures). This shows the interaction between pulse length and sampling interval to define the amount of information available (number of green crosses). The fewer samples there are across a feature the more uncertain the location and reflectance of that feature (Landi, 2002). Note that the amount of information available to the extraction algorithm depends on the ratio of the pulse length to the sampling interval, so that a 2 m pulse sampled every 1 ns contains the same amount of information as a 1 m pulse sampled every 0.5 ns and the two should be retrieved with identical accuracies, though the shorter pulse will be able to separate more closely spaced objects.

Table 1 shows the characteristics of the majority of waveform lidar systems that have been used for vegetation. Note that pulse widths are reported in terms of σ, defined as half the distance between the points at which the power drops below 61% of the maximum rather than the full-width-half-maximum (*FWHM*). Fig. 3 shows some examples of lidar system pulses. Fig. 3(a) shows the very short ($\sigma = 0.1$ m)

Gaussian pulse of SALCA (Danson et al., 2014). Fig. 3(b) shows the longer ($\sigma = 0.53$ m), asymmetric pulse of the Leica ALS50-II. Fig. 3(c) shows the short ($\sigma = 0.23$ m) but well sampled DWEL pulse (Douglas et al., 2015). This is near-Gaussian, but the detector filters out high-frequency returns, leading to an oscillation at the trailing edge. Fig. 3(d) shows the short ($\sigma = 0.35$ m) asymmetric pulse for a Riegl VZ-400. Of note is the Riegl VZ-400's long trailing edge, due to strong returns causing non-linearity in the detector electronics.

There have been recent developments to use single photon counting systems to measure the lidar waveform instead of traditional photodiode detectors. These are much more sensitive to low intensity signals, requiring lower powered lasers and so making the instrument lighter and cheaper. NASA's SIMPL (Dabney et al., 2010) and ICESat II (Harding et al., 2011) instruments currently in development are examples and Moussavi, Abdalati, Scambos, and Neuenschwander (2014) demonstrated that they can measure forest height. However they require repeated measurements to measure energy, either from a fixed position (Hernandez-Marin, Wallace, & Gibson, 2007; Wallace et al., 2012) or by averaging over an area (Moussavi et al., 2014), which introduces statistical errors on top of the error from inversion methods. Due to the different sampling methods described above these will not be investigated here.

2.2. Target profile

The target profile is the product of reflectance, phase function and projected area along the path of the laser beam (Jupp et al., 2009) and the amount of energy returned is the integral of the target profile. This is the true waveform, without any effects from the lidar system or

Table 1
Full waveform lidar system characteristics. Pulse widths are in terms of σ. Footprint sizes are for typical flying altitudes for space and airborne systems and at 10 m for terrestrial systems.

| Instrument	Pulse width	Footprint	Digitiser	Wavelength	Reference
Spaceborne					
ICESat	0.76 m	52 m–90 m	1 ns	1064 nm	Harding and Carabajal (2005)
MOLI	TBD	25 m	TBD	1064 nm	Murooka et al. (2013)
GEDI	TBD	25 m	TBD	1064 nm	Krainak et al. (2012)
Airborne					
Lidar	1.8 m	75 cm	2.5 ns	337 nm	Nelson, Krabill, and Maclean (1984)
SLICER	1.27 m	10 m	0.7 ns	1064 nm	Harding (1994)
LVIS	1.27 m	10–30 m	2 ns	1064 nm	Hofton et al. (2000)
Optech ALTM 3100	1.02 m	30–80 cm	1 ns	1064 nm	Wagner et al. (2006)
Riegl LMS-Q560	0.51 m	50 cm	1 ns	1545 nm	Wagner et al. (2006)
Leica ALS50-II	0.53 m	33 cm	1 ns	1064 nm	Hill and Broughton (2009)
Terrestrial					
Echidna	1.90 m	≈50 mm	0.5 ns	1064 nm	Jupp et al. (2009)
Riegl VZ-400	0.35 m	≈2 mm	1 ns	1545 nm	Calders et al. (2014)
SALCA	0.10 m	≈5 mm	1 ns	1064 nm & 1545 nm	Danson et al. (2014)
DWEL	0.23 m	≈5 mm	0.5 ns	1064 nm & 1548 nm	Douglas et al. (2015)
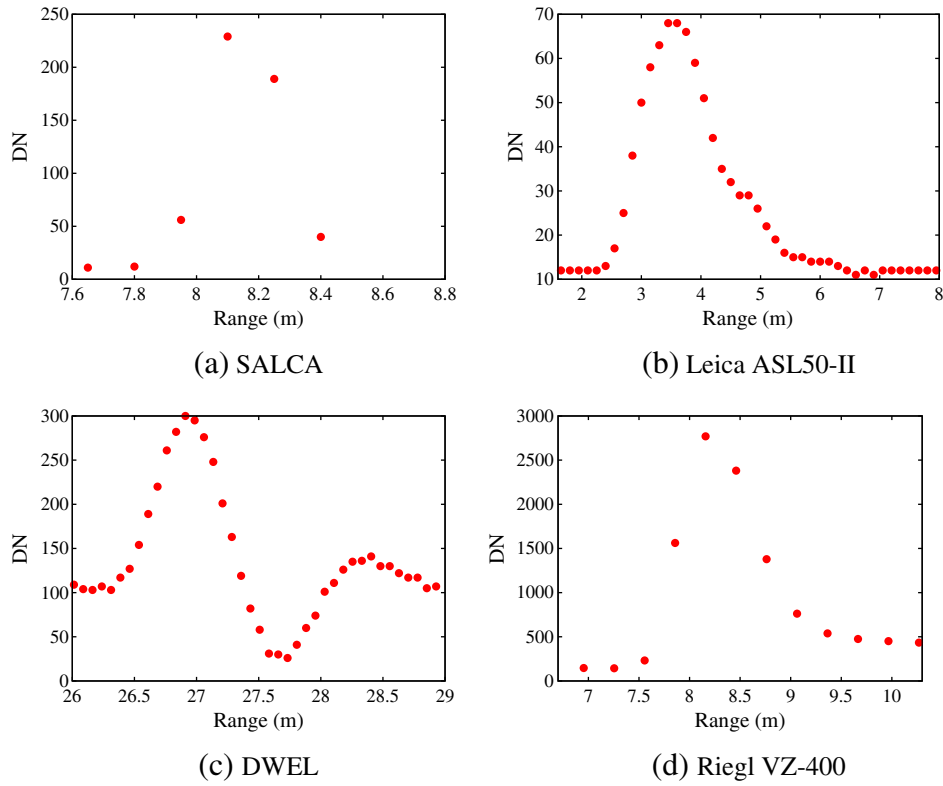
Fig. 3. System pulse examples.

multiple scattering, from which target properties can be directly derived. Fig. 4(a) shows a single surface that fills the field of view and at a shallow enough angle of incidence that the difference in range across the laser footprint is less than a range bin, such as bare Earth, a rock face, building or tree trunk (a single Dirac-delta function). In this case the peak intensity and energy are the same and this is what most commercial lidars have been designed to measure. Fig. 4(b) shows multiple short returns, such as a small footprint laser beam passing through branches or clipping the corner of a building. Vegetation canopies have many small elements that can lead to diffuse target profiles, such
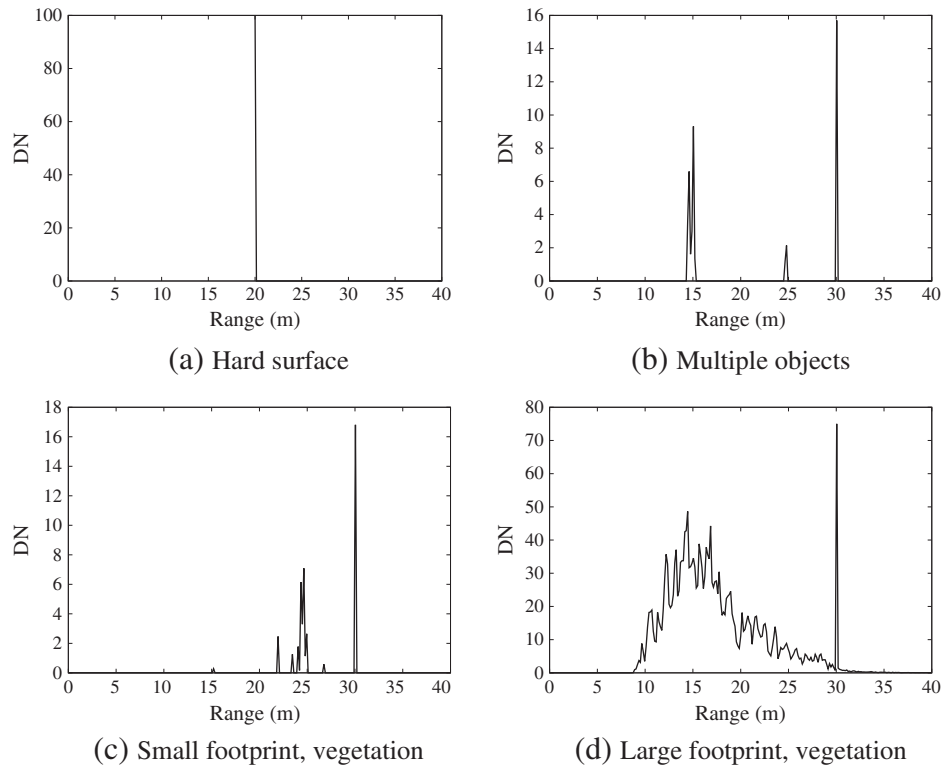

Fig. 4. Target profiles for a range of targets.

as Fig. 4(c) and (d), where the peak intensity and energy are not directly related. Fig. 4(c) shows a typical small footprint (33 cm diameter circle) airborne return from vegetation, with a number of overlapping elements giving a complex return. Fig. 4(d) shows a typical large footprint (10 m) return from vegetation, with a continuous smear of reflected energy.

In these cases the profile cannot be described by a single range and so the assessment of range accuracy is not trivial. Therefore this paper will deal exclusively with the extraction of energy. It is possible to use separate algorithms for determining range and energy and so the results of this study are still general. Section 5.2 will explore which of the cases in Fig. 4 occur for the different footprint sizes outlined in Table 1 and how that affects energy retrieval accuracy.

2.3. Noise

All lidar systems suffer from noise (Baltsavias, 1999), consisting of background, detector and photon noise (Hancock, 2010), Section 4.1.3. Fig. 5(a) shows the frequency distribution of all returns measured by SALCA in 37 separate scans between October 2010 and January 2014 and Fig. 5(b) shows all data collected during four different flights with the NERC ARSF Leica ALS50-II between July and October 2012. Leica waveforms were 37 m long and SALCA 15 m to 105 m long, sampled every 15 cm, giving 100–700 samples per waveform with energy from targets in only a few metres of bins and so the vast majority of bins contain only background and detector noise, dominating the distributions. These scans were collected in a range of environments (indoors, outdoors, in cloud and bright sunshine for SALCA and over different surfaces for the Leica ALS50-II). That all waveforms, with different ambient light levels, show a similar noise level suggests that this is not due to background light but must be due to detector noise. The implications for signal processing are discussed in Section 3.3.

3. Extraction methods

Table 2 lists the different methods used for extracting energy (and so reflectance) and a three letter acronym by which they will be referred to in this paper. This list is comprehensive as far as the authors are aware. There are other methods that only measure range, but only methods capable of estimating energy were tested here. The methods can be split into two main approaches, linear and iterative.

3.1. Linear methods

These are computationally efficient, requiring only a single pass through the data. They also tend to be numerically stable with the majority giving an output no matter what the input, although some of the more complex methods rely on matrix inversion and so can fail for certain inputs. Complete descriptions of all methods are given in Appendix A.

After the processing for this paper had been completed Jalobeanu and Gonçalves (2014) introduced a more robust version of the three point Gaussian using a very constrained Gaussian fitting to smooth and suppress ringing and Bayesian methods to constrain the three point fit. This increased the robustness and computational efficiency of the method over single, hard targets but the effect on return energy accuracy was not discussed and it is not clear that there would be any improvement. Because of this, this method was not tested here.

3.2. Iterative methods

These are more mathematically complex and computationally expensive than the linear methods (Neuenschwander et al., 2009). All studies so far have used the Levenberg–Marquardt iterative, nonlinear optimisation to fit assumed pulse shapes to the waveform. The MINPACK implementation of the Levenberg–Marquardt algorithm, with constraint, was used for the investigation (Garbow, Hillstrom, More, Markwardt, & Moshier, 1980). For overlapping returns (from diffuse targets) points of inflection were used to identify the number of functions within a feature and multiple shapes were fitted (Hofton, Minster, & Blair, 2000). Turning points were used to set initial estimates. Complete descriptions of all methods are given in Appendix B.

3.3. Denoising

Background noise must be removed before identifying returns. Fig. 5 suggests that, for SALCA and the Leica ALS50-II, the background noise is very stable and so can easily be removed by thresholding. Thresholding will remove some true signal and so the noise tracking method presented in (Hancock et al., 2011) was used to set the threshold (mean detector noise plus five standard deviations) and then tracking back to the mean noise level to get the start and end of the feature. This reliably removes the effect of background noise for instruments with such stable behaviour, though less stable instruments may need more complex methods in order to prevent mistaking noise for a real return.

In addition, there will be distortions in the received signal due to detector and photon noise (Hancock, 2010, Section 4.1.3). Smoothing can help make inversions more robust but has the disadvantage of blurring the signal, reducing the ability to separate nearby objects. It should also be noted that smoothing does not add to the information content available for the inversion and some suggest that it is better to use the raw data (Press, Tuekolsky, Vetterling, & Flannery, 1994).

3.4. Previous comparisons

A small number of previous studies have assessed the accuracy of energy extraction algorithms. However, the vast majority have focused
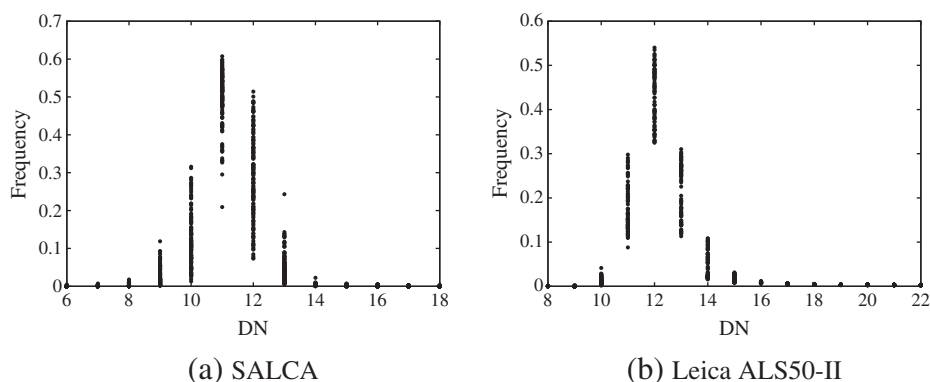
(a) SALCA

(b) Leica ALS50-II

Fig. 5. Frequency distributions of the DN of all bins showing the stability of the noise.

Table 2
Waveform energy interpretation methods along with a reference for a paper using each.

Method	Acronym	Description	Reference
Linear			
Peak	PEA	Maximum DN bin	Gaulton et al. (2013)
Sum	SUR, SUT, SUS	Sum of DN over all bins	Jacobus and Chien (1981)
Windowed sum	WS3, WS5, WS7	Sum of DN within window around peak	Fisher and Naidu (1996)
Quadratic	QUA, QUP	Analytical fit of quadratic to three points	Blais and Rioux (1986)
Three point Gaussian	TPG	Analytical fit of Gaussian to three points	Li et al. (1999)
Spline	SPL	Uniform B-spline curve fit to all points	Roncat et al. (2011)
Caruana's	CAR	Natural logarithm transform of range and intensity to linearise Gaussian fit	Guo (2012)
Iterative			
Gaussian	GAU	Non-linear Gaussian fit	Hofton et al. (2000)
Lognormal	LOG	Non-linear lognormal fit	Chauve et al. (2007)
Generalised Gaussian	GGA	Non-linear generalised Gaussian fit	Chauve et al. (2007)

on more mathematically complex methods (Chauve et al., 2007; Roncat, Bergauer, & Pfeifer, 2011; Wagner, Ullrich, Ducic, Melzer, & Studnicka, 2006; Wu, van Aardt, & Asner, 2011). Only a handful of papers have investigated the simple methods and none have comprehensively compared their performance to the mathematically complex methods.

Jutzi and Stilla (2005) compared the peak, rectangular sum and Gaussian (PEA, SUR and GAU) and concluded that, for their simple targets, the Gaussian performed slightly better than the peak and both performed better than the sum, though their focus was on range accuracy. Lin, Shih, and Teo (2010) explored the three methods used in Riegl's RiANALZE software (PEA, SUR and GAU), however no validation was attempted and so they could not conclude which method was more accurate. Armston et al. (2013) looked at the difference between Gaussian fitting and the sum for calculating return energy in the estimation of canopy cover, concluding that the latter gave more accurate energies for vegetation but that the Gaussian gave as good a result in the majority of cases and so, as one of their datasets (Riegl LMS-Q680i) only reported Gaussian parameters, the Gaussians were used to calculate canopy cover. Jalobeanu and Gonçalves (2014) used the three-point Gaussian but did not compare to other methods.

As far as the authors are aware no other studies have tested the suitability of the simple methods to extract return energy from waveform lidar. This study expands on previous work by including all waveform lidar algorithms capable of calculating energy and using realistic simulations to compare their performance across a range of lidar systems, targets and noise levels in order to see how accurately each method can extract return energy for different situations, paying particular attention to whether a computationally efficient method can achieve the same results as the more widely used, computationally expensive and mathematically complex methods.

4. Experiments

The performance of the fifteen algorithms (listed in Table 2) over a wide range of conditions was tested on both single and multiple, complex returns over the full range of pulse widths, pulse shapes, amplitudes and noise levels observed in current instruments. Simulators were used, for which the truth is known precisely, allowing accurate assessment of errors. In all cases the methods were tested on the raw waveforms and after smoothing by Gaussians with a range of widths.

The accuracy was assessed in terms of the bias (difference between true and retrieved energy), root mean square error (RMSE), failure rate (fraction of inversions that failed due to instability or gave a return energy greater than 1000, a nonsensical value that would skew the statistics) and the mean standard deviation of the energy estimates for each true energy. The last of these indicates whether a method gave a consistent estimate of energy for a given return shape at different ranges and in the presence of noise. If a method has a low standard deviation, bias can be removed by calibration, though it is preferable to

have a low RMSE. Note that due to the order of averaging, the standard deviation and RMSE were not directly comparable.

Throughout the paper a sampling interval of 1 ns was used. As noted in Section 2.1, this can represent any sampling interval after scaling the pulse lengths quoted by the ratio of the sampling intervals. Obviously the higher the sampling rate the more information there is available to a shorter pulse and so the smaller the minimum separation at which objects can be identified (Danson et al., 2014).

4.1. Single returns

A simulator, written in C, was used to create example waveforms for single returns, as shown in Fig. 3. Two different pulse shapes were used: (i) Gaussian (such as SALCA and ICESat); and (ii), the system pulse of the Leica ASL50-II to represent an asymmetric pulse (Fig. 3(b)). These will be refereed to as "Gaussian" and "asymmetric" pulses for the rest of this paper. Within the simulator the pulses could be centred at any point to test the method's ability to deal with sub-bin shifts, with any amplitude (measured in digital numbers, DN) and with any pulse width (σ). As the returns are from a single object, the target profile was a single spike, so that the energy was directly proportional to the intensity. The noise added to the simulator was based on that in Hancock, Lewis, Disney, Foster, and Muller (2008) except that, as Fig. 5 shows that the background signal level was independent of background illumination, the equations for solar energy were not needed, only detector noise. The noise was set by n, the width (σ) of the Gaussian distribution in terms of DN from which random numbers were drawn. Fitting a Gaussian to Fig. 3 suggests that n = 1 is an appropriate noise level for SALCA and the Leica ALS50-II.

For each pulse shape sixteen separate amplitudes (DN 10–255 above mean detector noise), forty separate pulse widths (0.1 m to 2.15 m) and fifteen separate ranges (between 10 m and 10.15 m to have the peak at a range of positions within the 15 cm bins) were simulated. Noise levels between 0 (no noise) and 20 (a very high noise case) were simulated. For each noise level fifty separate sets of noise (using different random number seeds) were added and a separate inversion performed. This gave 522,750 individual energy estimates at each noise level. Fig. 6 shows some examples of the simulated waveforms. The inversions were performed on the raw signal and after smoothing by a Gaussian of between 0.1 m and 2.15 m long (σ). As all returns were from a single target only a single function was fitted to each.

4.2. Diffuse targets

Some lidar algorithms assume that a return comes from a single interaction (such as the peak method). This is valid for hard targets, such as rocks and buildings, but is often not valid for vegetation, where the pulse can hit multiple targets through the canopy. It has been suggested that vegetation elements tend to be arranged as a Gaussian so that the
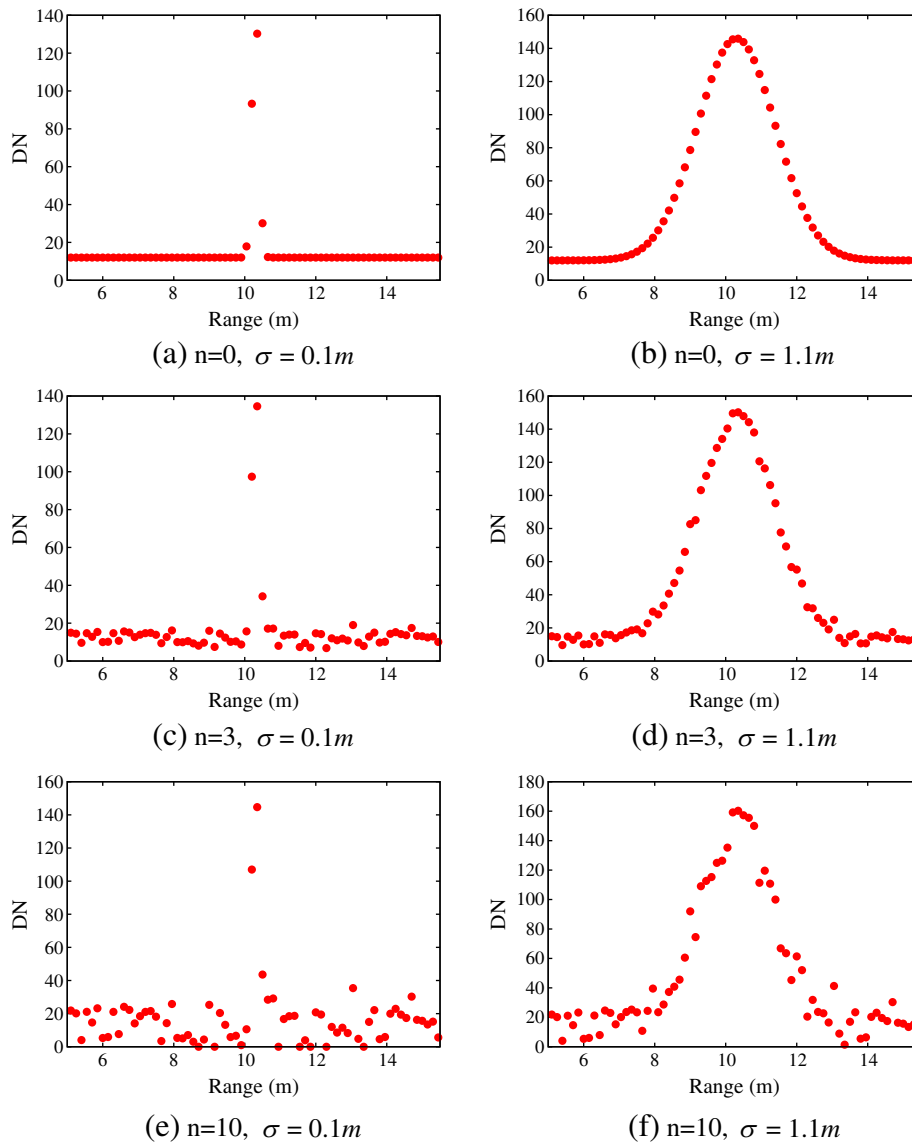
Fig. 6. Examples of generated waveforms for different noise levels, *n*, and pulse widths, σ.

returned waveform (convolution of the target and system pulse) can also be represented by a Gaussian (Hofton et al., 2000). If there are deviations from these assumptions the methods that rely on them will struggle.

To test these assumptions the ray tracer of Lewis (1999) and the mixed age sitka spruce forest model described in Hancock et al. (2011) were used to generate waveforms. This tool has been widely used for simulating remote sensing measurements (Disney, Lewis, & Saich, 2006; Hancock, Lewis, Foster, Disney, & Muller, 2012), has been validated (Widlowski et al., 2007) and forms part of the RAMI surrogate truth (Widlowski et al., 2008). libRAT allows realistic lidar signals to be produced where the truth is known and so forms a virtual laboratory (Lewis, 1999). Three different footprint sizes were simulated, a very small footprint terrestrial system (1 cm at 10 m), a small footprint airborne system (33 cm) and a larger footprint air or spaceborne system (15 m). 15 m was chosen to lie between the spaceborne (GEDI at 25 m) and airborne (SLICER at 10 m) system footprints. For each footprint size three different pulse lengths were simulated, σ = 0.1 m, σ = 1.1 m and σ = 2.15 m. Only a Gaussian system pulse was used to reduce computational expense.

To ensure that a range of structures were covered at each pulse length, 25 waveforms were created for the 15 m footprints, 100 for the 33 cm footprints and 283 for the 1 cm footprints. Different numbers were used because as footprint size increases more orders of multiple scattering and rays need to be simulated in order to achieve convergence in the Monte Carlo solution (Hancock, 2010, chapter 4), increasing the computational expense. In addition the target becomes more heterogeneous at smaller footprint sizes, so that more waveforms are needed to ensure that all possibilities are represented.

4.2.1. Return complexity

The measured waveforms and true target profiles visible to the instrument were examined to see if the above assumptions about target arrangement were valid at the different footprint sizes. It was expected that the largest footprint's waveforms would most closely resemble a Gaussian (Hofton et al., 2000), whilst the smallest footprint would be more heterogeneous and less likely to contain returns from multiple targets.

4.2.2. Fitting to diffuse targets

The energy retrieval algorithms were then applied to these simulations using the noise removal method described in Section 3.3. Extracting the biophysical parameters or performing a classification is beyond the scope of this paper and will be covered in a future study.

Here the fit was evaluated by the total energy returned and failure rate. The extractions were performed over the same range of noise levels ($0 \leq n \leq 20$ with fifty repetitions with different random number seeds for each waveform) as the single return experiments. The signals were smoothed by a range of Gaussians with widths between 0 (no smoothing) and 8 m.

Some of the methods rely on the identification of turning points in order to pick initial estimates (GAU, LOG and GGA) or to determine the number of individual features within a single return (GAU, LOG, GGA, PEA, WS3, WS5, WS7, QUA, QUP, TPG and CAR) and so could benefit from smoothing the data in order to identify these points, but not using the extra smoothing for fitting (Hofton et al., 2000). Therefore an additional "pre-smoothing", separate to the smoothing mentioned above, was included. This was by a Gaussian between 0 (no smoothing) and 8 m long. To make sure that the accuracy was optimal for each method, only the accuracies for the smoothing combination with the lowest RMSE and a failure rate less than 2% were presented for each method. Smoothing was only used if it gave more than a 1% RMSE improvement above the unsmoothed accuracy. For the iterative fitting methods (GAU, LOG and GGA) the parameters were constrained within MINPACK. The position (μ) was forced to lie between the start and end of the feature, with the initial estimate at the centre of gravity. The amplitude (A) was forced to lie between twice and a quarter of the observed peak. The width (σ) was forced to lie between 0.00001 m and the total feature width. The additional parameters in the lognormal and generalised Gaussian were not fixed.

4.2.3. Processing time

In order to assess computational efficiency for realistic waveforms, the processing time for each method to estimate return energy from the diffuse target simulations was assessed. This was done within the C program by calling the *clock*( ) function before and after each method was performed to give the number of central processing unit, CPU, clock cycles. The CPU time was measured for each separate return along a waveform and the mean reported. Some of the related methods were calculated within the same function and so the measured time was for all sub-methods (the sums, windowed sums and quadratic). This may have slightly increased the reported time, but due to their similarity and common loops, this increase is negligible.

5. Results and discussion

5.1. Single returns

Results are presented separately for the unsmoothed raw data (Section 5.1.1) and smoothed (Section 5.1.2) cases.

5.1.1. Unsmoothed raw data

Table 3 shows the results for all methods at noise level 1 (the closest to the SALCA and Leica ALS50-II behaviour). Examining the standard deviation of return energies for a given pulse energy, the peak (PEA), windowed sums (WS3, WS5 and WS7), quadratic (QUA and QUP), three point Gaussian (TPG) and Caruana's (CAR) all gave unacceptably large errors. These occurred at all noise levels suggesting that they are not robust enough to be used on raw data. The sum using Simpson's rule (SUS) performed worse than both the rectangular sum and the trapezium sum and so was also rejected.

Fig. 7 shows the four error metrics (mean across all pulse amplitudes, widths and ranges) against noise level for all feasible methods for Gaussian and asymmetric pulses. The lognormal method had higher errors at all noise levels, even for the asymmetric pulses it was intended for. This agrees with the findings of Chauve et al. (2007) and was most likely due to the instability of fitting a more complicated function.

For Gaussian pulses (Fig. 7), the five remaining methods gave near identical RMSEs and standard deviations for noise levels up to 8 (a high level), though the Gaussian gave a lower bias due to its ability to reconstruct the missing data. Above noise level 8 the generalised Gaussian started to give larger errors (due to numerical instability) whilst the Gaussian gave slightly larger standard deviations than the spline or two sum methods. The Gaussian gave slightly lower RMSEs (of the order of 0.5%) because of the lower bias. The generalised Gaussian and Gaussian started to fail to fit to some returns at noise levels of 3 and 5 respectively. The sum and spline showed negligible failures up to the highest noise level tested. These three linear methods (SUR, SUT and SPL) are superior for Gaussian pulses due to their computational efficiency and, at high noise levels, low standard deviations and failure rates. The differences between the performance of these three was negligible.

For asymmetric pulses (Fig. 7) again the lognormal showed higher errors due to instability. The other five methods gave near identical standard deviations until a noise level of 5 after which the spline, rectangular sum and trapezium rule gave lower errors than the Gaussian and generalised Gaussian methods. The iterative fitting methods started to fail at lower noise levels than for Gaussian pulses (at noise level 2). Again the spline and sum methods showed negligible failures at all noise levels. Interestingly the spline and sum methods showed lower bias and RMSE than the iterative methods at all noise levels. This must be due to the deviation of the pulse shape (Fig. 3(b)) from a simple analytical form making function fitting less robust. When successful the generalised Gaussian gave lower bias and RMSEs than the Gaussian due to its greater number of degrees of freedom, though this also caused more failures due to instability. It is concluded that the spline, rectangular sum and trapezium rule gave more accurate results for asymmetric pulses in all conditions.

Table 3
Energy errors for simulations of single returns at noise level 1. All values are in %.

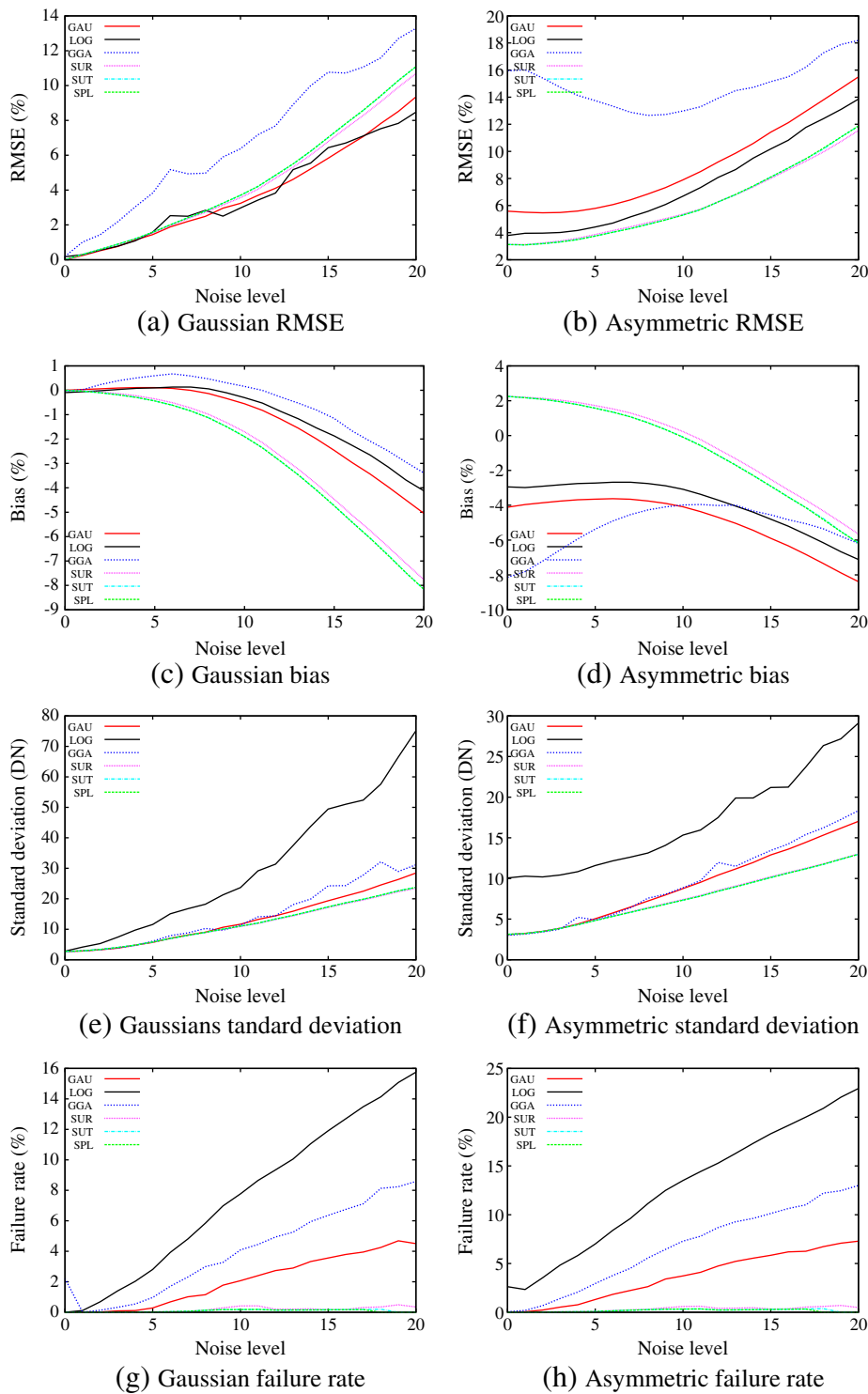| Method	Gaussian				Asymmetric			
	RMSE	Std. Dev.	Fails	Bias	RMSE	Std. Dev.	Fails	Bias
PEA	95.7	49.9	0.00	−63.79	57.8	44.0	0.00	−5.72
SUR	0.30	2.92	0.00	−0.02	3.12	3.23	0.00	2.20
SUT	0.30	2.92	0.00	−0.04	3.10	3.22	0.00	2.17
SUS	0.33	2.96	0.00	−0.06	3.22	3.46	0.01	2.17
WS3	114	20.3	0.00	−84.39	90.1	13.5	0.00	−62.94
WS5	104	29.2	0.00	−75.09	68.8	15.1	0.00	−45.92
WS7	94.8	35.1	0.00	−66.65	51.9	14.3	0.00	−33.62
QAP	95.6	50.4	0.00	−63.67	58.2	44.8	0.00	−4.86
QUA	64.5	61.7	0.02	−44.34	39.5	16.2	0.00	−27.96
TPG	47.4	82.5	0.09	−25.73	21.8	22.5	3.15	−9.09
SPL	0.30	2.92	0.00	−0.04	3.10	3.22	0.00	2.17
CAR	20.0	1774	7.39	2.46	86.4	3539	21.3	35.9
GAU	0.24	2.86	0.00	0.03	5.51	3.24	0.05	−3.96
LOG	0.29	2.89	0.12	−0.06	3.95	3.16	2.34	−2.98
GGA	1.01	4.16	0.01	0.03	16.0	10.3	0.19	−7.81

Fig. 7. Mean extracted energy errors against noise level for Gaussian and asymmetric pulses across the full range of ranges, amplitudes and pulse widths.

Examining the dependence of errors on amplitude and pulse width revealed that all the feasible methods gave higher errors for the very shortest ($\sigma < 0.2$ m) and lowest amplitude ($A < 50$) returns. They all performed equally badly for these hardest cases and gave low errors for the shortest pulses if the amplitude was greater than 50 and for low amplitude returns if width was greater than 0.2 m. Therefore SALCA's very short pulse width was only an issue for low amplitude returns (RMSEs of 17% for $DN < 10$ above noise). The RMSE could be reduced to 10% by doubling the sampling rate. This suggests that using a system pulse

longer than 0.2 m sampled at 1 ns would ensure that energy accuracy is not limited for single returns.

It is concluded that the spline, rectangular sum and trapezium rule were the most accurate methods for extracting return energy from raw lidar waveforms. They are far less computationally expensive than the iterative fitting methods, more robust (lower failure rate) and gave more accurate results for asymmetric pulses and for Gaussian pulses at higher noise levels (n greater than 6). At the low noise levels shown by SALCA and the Leica ALS50-II, the Gaussian method had a

Table 4
Energy errors for noise level 1 after smoothing waveforms by a 1.5 m Gaussian. All values are in %.

Method	Gaussian				Asymmetric			
	RMSE	Std. Dev.	Fails	Bias	RMSE	Std. Dev.	Fails	Bias
PEA	107	6.85	0.00	−80.55	98.6	1.43	0.00	−75.13
SUR	0.52	3.23	0.00	0.20	3.81	3.53	0.00	2.90
SUT	0.52	3.23	0.00	0.20	3.81	3.53	0.00	2.90
SUS	0.52	3.23	0.00	0.20	3.81	3.53	0.00	2.90
WS3	120	3.71	0.00	−91.54	117	0.81	0.00	−90.13
WS5	113	6.16	0.00	−85.96	107	1.39	0.00	−83.66
WS7	106	8.55	0.00	−80.46	98.5	1.97	0.00	−77.34
QAP	107	6.85	0.00	−80.55	98.5	1.43	0.00	−75.12
QUA	32.2	2.18	0.00	−24.67	30.8	2.48	0.00	−23.44
TPG	3.81	5.18	0.00	2.60	5.54	3.90	0.00	4.15
SPL	0.52	3.23	0.00	0.20	3.81	3.53	0.00	2.90
CAR	6.30	22.9	0.31	1.50	23.4	30.0	1.67	8.61
GAU	0.33	2.97	0.00	0.06	3.04	3.31	0.00	2.20
LOG	0.34	2.96	0.01	−0.05	3.07	3.33	0.08	2.20
GGA	0.71	3.62	0.19	−0.12	2.97	3.91	0.66	1.87

very slightly lower RMSE and bias (0.06% difference) for Gaussian pulses, but this came at the cost of extra computational expense and larger errors at higher noise levels.

5.1.2. Smoothed data

The analysis was repeated after smoothing by a 1.5 m long Gaussian and these are shown in Table 4 for noise level 1. All methods, except Caruana's, gave acceptable standard deviations and negligible failure rates. The RMSEs for the methods that performed well without smoothing were slightly larger after smoothing due to an increased bias. This may have been due to signal being smoothed beneath the noise threshold and this supports the argument for using raw data (Press et al., 1994).

Smoothing caused a drop in the standard deviation of the peak, windowed sums, quadratic and three point Gaussian methods, all of which gave acceptable results. The quadratic method (QUA) now gave the lowest standard deviation for a Gaussian pulse, though the RMSE and

bias were much higher than for the other methods. For an asymmetric pulse, the windowed sum over three bins (WS3) gave the lowest standard deviation (0.81%) followed by the windowed sum over five (WS5, 1.39%) then the peak and quadratic (PEA, QUA 1.48%) methods. It is not clear why the standard deviation for the asymmetric pulse was lower than for the Gaussian pulses, other than possibly due to a difference in definitions of pulse width (σ for an asymmetric pulse is not directly comparable to σ for a Gaussian).

Fig. 8 shows the standard deviations and RMSEs of extracted DNs against noise level using a smoothed waveform. After smoothing, the quadratic method gave the lowest errors at low noise levels for Gaussian pulses and the three bin windowed sum gave the lowest errors at high noise levels and for asymmetric pulses, closely followed by the peak method. This must be due to smoothing's ability to fill in the gaps between samples by taking local averages and the suppression of noise. The three point Gaussian method showed a similar noise dependence to the quadratic method (they are mathematically very similar) but
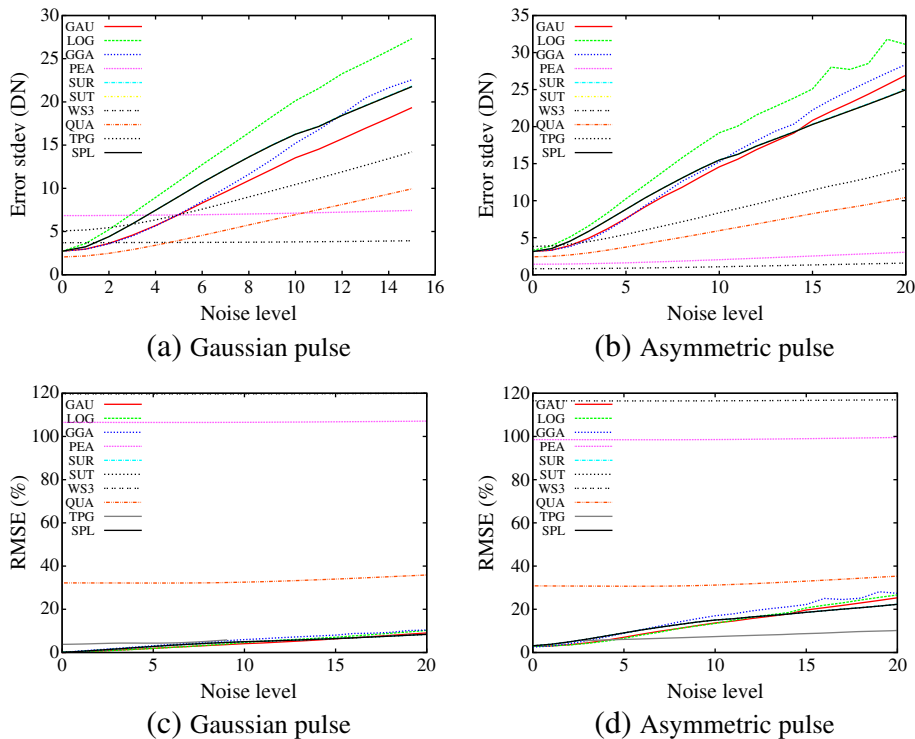

(a) Gaussian pulse

(b) Asymmetric pulse

(c) Gaussian pulse

(d) Asymmetric pulse

Fig. 8. Extracted energy errors against noise level after smoothing by a 1.5 m Gaussian across the full range of ranges, amplitudes and pulse widths. Note that the SUT and SUR lines are hidden beneath the SPL line.

with a consistently higher error. The Gaussian and general Gaussian fittings performed similarly with and without smoothing (though smoothing reduces failures) whilst the spline and sum methods showed higher errors than without smoothing (due to bias).

5.1.3. Single return conclusions

The results show that for Gaussian pulses the most accurate methods (lowest RMSE with the lowest failure rate) were the Gaussian (GAU), sum (SUR) and spline (SPL) without smoothing whilst for asymmetric pulses the most accurate were the spline (SPL) and sum (SUR and SUT) without smoothing. As the sum and spline gave the joint lowest errors no matter the pulse shape and did not fail at higher noise levels, it is concluded that these are the most appropriate methods for determining lidar return intensity.

After smoothing, the windowed sum (WS3), peak (PEA) and quadratic (QUA and QUP) methods gave the most consistent results (lowest standard deviation) and so can be calibrated to get accurate reflectance, closely followed by the sum (SUR, SUT and SUS) and three point Gaussian (TPG). However, these assume that the peak is related to the total energy in the pulse. This is the case as long as the return is from a single object but not if from multiple objects. This is tested in Section 5.2.

Smoothing reduced the failure rate of iterative fittings but did not increase accuracy.

5.2. Diffuse targets

Results for processing waveforms from diffuse targets are presented separately for target complexity (Section 5.2.1), fitting accuracy (Section 5.2.2) and computational processing time (Section 5.2.3).

5.2.1. Return complexity

Fig. 9 shows some typical simulated waveforms of sitka spruce forests at the three footprint sizes and three pulse lengths along with the target profiles (the waveform without any pulse shape or sampling effects). The 15 m and 33 cm footprints were looking down from above whilst the 1 cm footprints were looking up at the canopy. It can be seen that the waveforms became more complex as the footprint size increased due to more canopy elements being illuminated.

In the 15 m footprint (Fig. 9(a)) many elements blurred together to form a continuous waveform. For the 1.1 m and 2.15 m pulses this resembled a Gaussian, but the target profile (Fig. 9(b)) shows that this was a result of the system pulse. Therefore whilst a Gaussian (or lognormal or generalised Gaussian) could be fitted, the assumption that this
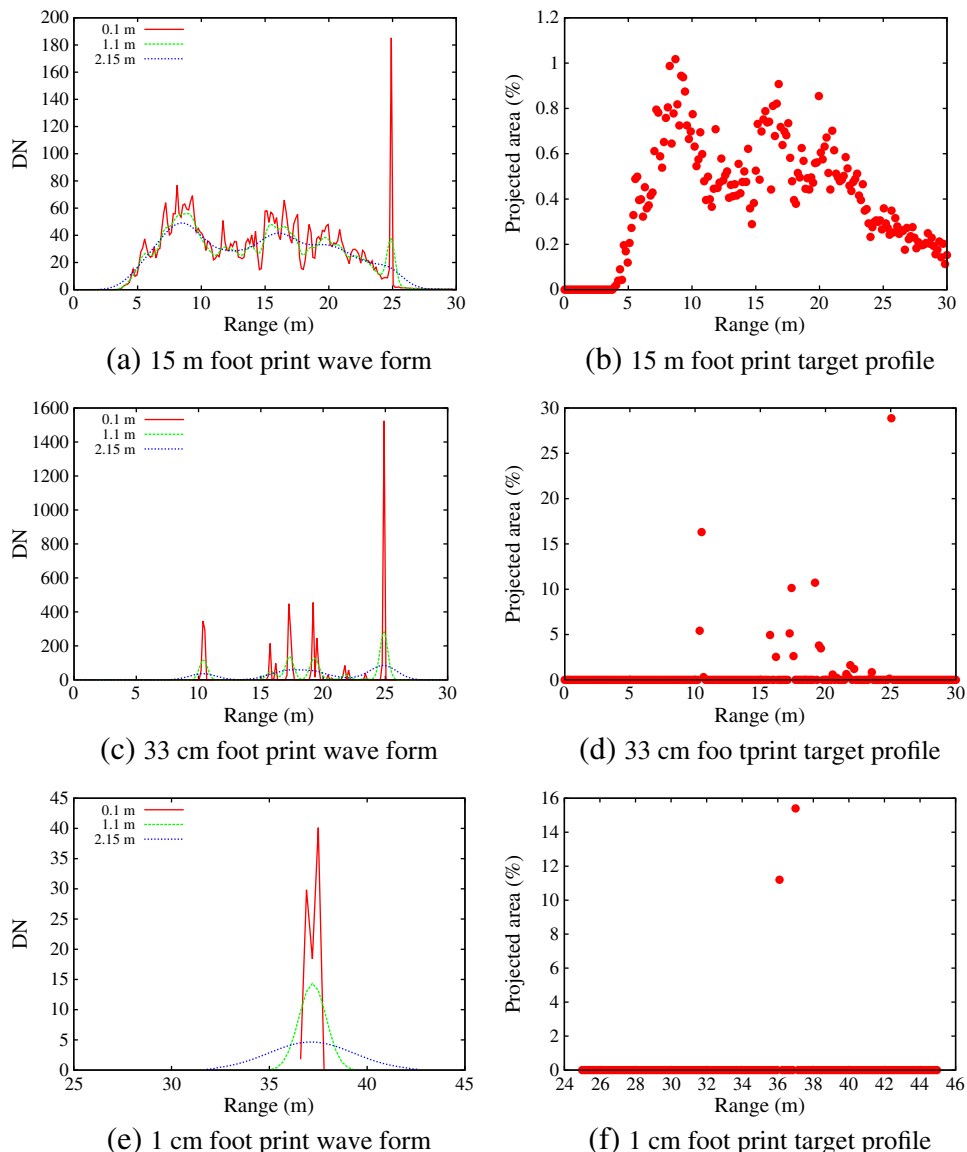

(a) 15 m foot print wave form

(b) 15 m foot print target profile

(c) 33 cm foot print wave form

(d) 33 cm foo tprint target profile

(e) 1 cm foot print wave form

(f) 1 cm foot print target profile

Fig. 9. Waveforms for different footprint sizes and pulse lengths from ray tracing simulations over a sitka spruce forest along with the corresponding target profiles.

will retrieve the true target profile may not be valid. Note that this includes no noise so all fluctuations are due to heterogeneity. The waveform was less complex for a 33 cm footprint but there were still areas with multiple returns combined into a single feature and so deviating from a simple analytical form. For the 1 cm footprint there were far fewer features with returns from multiple elements and so the assumption of single hits was more appropriate than for the larger footprints, but was not true in all cases.

If a return feature was more than one bin wide (1 ns) there were multiple elements contributing and so the assumption that the peak intensity is related to energy is not appropriate. For the 15 m footprint simulations there were no returns from single bins and for 33 cm footprints only 22% of returns came from single bins, the rest coming from more complex shapes. For 1 cm footprints all returns were from a single bin. To see whether these returns within a bin came from a single or multiple interaction the ray tracing simulations were repeated at 2 cm resolution and the number of separate returns within each 15 cm bin were added up. This showed that for 15 m footprints, only 1.3% of bins contained returns from single objects, for 33 cm footprints 20.6% and for 1 cm footprints 40.8%. Therefore at all footprint sizes the majority of returns over vegetation were complex and so we argue that the methods that assume single returns are not appropriate (the peak, windowed sum and quadratic).

5.2.2. Fitting to diffuse targets

The results for fitting to 15 m footprints are shown in Table 5. Only the results for noise levels 1 (as for SALCA and the Leica ALS50-II) and 10 (a high noise level) are shown as these show the overall behaviour at other noise levels. The standard deviation for the methods that measure peak intensity rather than an integral of energy (PEA and QUP) was examined and found to have larger errors than the standard deviations for other methods (such as GAU and SUR) and so these could not be reliably calibrated back to true intensity. The methods behaved similarly to the fits to single returns, with several giving unusable results in all cases no matter what smoothing was applied (CAR, QUP, QUA, WS3, PEA and TPG). The windowed sum (WS7 and WS5) seemed to perform well, but this was found to only be the case for low smoothing widths, where small fluctuations in signal would cause many turning points, each of which was picked up as a separate feature. The low RMSE is a result of a particular smoothing width leading to just enough features to add up all the energy, making them equivalent to SUR. The accuracy was unstable with changing smoothing widths and noise levels and so this is not a practical way to overcome bias. The LOG and GGA methods had low accuracies due to instability. The Gaussian, spline and sum methods all behaved similarly, with no more than 2% difference in accuracy. At low noise levels (n = 1) or long pulse lengths ($\sigma = 2.15$ m)

Gaussian fitting gave slightly (0.5% difference) lower errors whilst the sum and spline performed better (1% difference) in other situations.

Gaussian fitting required the maximum smoothing (8 m smooth and 8 m pre-smoothing for parameter selection) whilst the sum and the spline required none. Smoothing by a 5 m Gaussian was found to reduce the sum and spline RMSEs by 0.1% due to a change in the bias, as more signal was put beneath the noise threshold, but this small increase was not thought to be worth the additional computational expense. The errors reduced with shorter pulses, which is not what would be expected from the reduced information content (Landi, 2002). This was because these diffuse target profiles (as in Fig. 9(a)) meant that enough information was available for even the shortest pulse. This has implications for sensor design.

For 33 cm footprints (Table 6) again some methods did not give accurate results (PEA, QUP, QUA, TPG, CAR, LOG and GGA). The windowed sum (WS3) appeared to be the best for short pulses but again this was an artefact due to noise fluctuations in the signal. Gaussian fitting, the sum and the spline were the most accurate methods (SUS was more sensitive to noise than SUR and SUT). The sum (SUR and SUT) and the spline gave almost identical accuracies. The trapezium rule (SUT) performed better than the simple sum (SUR) for short pulse lengths, due to there being more missing data to fill in, whilst SUR performed better for longer pulses. The spline had near identical accuracies to the trapezium rule (SUT). Accuracies for all methods increased with longer pulses, as we would expect from the increased information content (Landi, 2002). It appears that at this footprint size the target profile is not extended enough to allow a short pulse to measure as accurately as a longer pulsed system.

Again the sum and spline required no smoothing. The smaller footprint gave rise to a less diffuse target profile (Fig. 9(c)) which did not show the same relationship between smoothing width and bias evident in the 15 m footprints. The Gaussian performed best after the longest smoothing widths (8 m).

The methods showed the same relative behaviour for the 1 cm footprints (Table 7) as for the 33 cm footprints (Table 6), though with higher errors and an increased sensitivity to noise at the shortest pulse length ($\sigma = 0.1$ m). The sum and spline performed better than the Gaussian in all situations and were less sensitive to noise (except for SUS which was unstable). Again the sum and spline required no smoothing whilst Gaussian fitting needed the maximum smoothing (8 m).

It is concluded that the spline (SPL) and rectangular sum (SUR as it gave near identical results but is simpler than SUT) were the most appropriate for extracting return energies from waveform lidar over vegetation, except for large footprint (15 m) systems, where Gaussian fitting gave a 1% lower RMSE. The results for smaller footprints (33 cm and 1 cm) back up the observation of Jacobus and Chien (1981) that

Table 5
Mean relative RMSEs for extractions from simulated 15 m footprint waveforms using optimal smoothing values for each method. All errors are in %.

Pulse method	0.1 m		1.1 m		2.15 m	
	n = 1	n = 10	n = 1	n = 10	n = 1	n = 10
PEA	36.4	30.2	85.0	84.6	89.8	89.4
SUR	6.51	10.3	12.7	12.8	12.8	12.9
SUT	6.54	10.3	12.7	12.9	12.8	12.9
SUS	6.04	10.3	12.7	12.9	12.8	12.9
WS3	32.6	39.8	71.6	39.1	80.2	35.0
WS5	15.8	34.0	57.4	24.6	73.3	21.7
WS7	9.22	32.9	46.1	18.2	68.0	16.8
QUP	80.7	78.6	NA	93.6	NA	NA
QUA	28.5	185.3	31.5	44.2	NA	36.0
TPG	99.6	100.1	NA	NA	NA	NA
SPL	6.53	10.3	12.7	12.9	12.8	12.9
CAR	NA	NA	NA	NA	NA	NA
GAU	5.71	12.1	12.2	11.8	12.3	12.0
LOG	96.3	96.4	72.4	89.2	69.7	NA
GGA	23.0	21.3	35.9	23.6	36.5	NA

Table 6
Mean relative RMSEs for extractions from simulated 33 cm footprint waveforms using optimal smoothing values for each method. All errors are in %.

Pulse method	0.1 m		1.1 m		2.15 m	
	n = 1	n = 10	n = 1	n = 10	n = 1	n = 10
PEA	93.3	81.2	42.0	41.6	62.5	60.1
SUR	9.65	11.7	0.57	6.93	1.31	3.90
SUT	7.90	11.6	0.56	7.28	1.32	4.01
SUS	11.6	16.1	0.57	7.35	1.32	4.04
WS3	6.05	15.2	55.2	49.7	82.3	44.4
WS5	7.08	13.8	33.8	30.4	72.8	30.9
WS7	7.53	13.6	20.1	18.7	64.1	22.8
QUP	95.3	86.7	65.8	55.2	83.1	49.5
QUA	85.4	54.5	30.4	216.2	26.2	84.9
TPG	94.0	88.8	87.4	80.8	100.2	100.2
SPL	7.90	11.6	0.56	7.29	1.32	4.01
CAR	NA	NA	NA	NA	NA	NA
GAU	11.2	41.2	1.37	13.0	1.40	4.71
LOG	104.1	58.6	94.2	92.7	89.9	92.6
GGA	37.6	41.1	37.8	30.5	22.9	28.8

Table 7
Mean relative RMSEs for extractions from simulated 1 cm footprint waveforms using optimal smoothing values for each method. All errors are in %.

Pulse method	0.1 m		1.1 m		2.15 m	
	n = 1	n = 10	n = 1	n = 10	n = 1	n = 10
PEA	105.7	68.9	87.4	87.5	53.3	51.0
SUR	10.8	19.9	0.48	4.00	0.22	2.04
SUT	9.13	22.1	0.43	4.01	0.22	2.03
SUS	15.8	42.4	0.45	4.25	0.25	2.14
WS3	8.32	33.0	55.5	53.9	95.5	57.3
WS5	8.58	33.6	28.1	26.9	83.2	41.7
WS7	8.61	33.8	12.3	11.4	71.5	30.9
QUP	108.6	82.3	31.3	34.6	75.2	50.2
QUA	23.8	31.6	28.4	33.1	28.3	29.2
TPG	24.9	71.0	3.55	31.3	2.66	5.18
SPL	13.9	24.2	0.44	4.16	0.23	2.06
CAR	NA	NA	NA	NA	NA	NA
GAU	14.4	40.2	2.16	7.88	1.22	5.52
LOG	18.5	41.5	5.80	9.96	6.86	4.35
GGA	16.4	57.7	2.45	24.7	1.07	6.94

assuming an expected waveform shape will hide any deviations from it, potentially reducing the accuracy. This is especially true over vegetation where complex shapes can arise that do not fit simple models, such as the shadowed ground return under heterogeneous canopies (Hancock et al., 2012).

5.2.3. Processing time

The processing times are given in Table 8. The peak method was the fastest (as we would expect), closely followed by the sum, windowed sums, then the quadratic, three point Gaussian, spline and finally the three iterative methods were several orders of magnitude slower; 1000 times slower for a 1.1 m pulse and 33 cm footprints and 20,000 times slower for 2.15 m pulses and 33 cm footprints. This supports the previous claims that iterative fitting methods are the most computationally expensive (Neuenschwander et al., 2009) which may limit their applicability to large datasets (Jalobeanu & Gonçalves, 2014). Processing time increased with large footprints and longer system pulses due to the larger volume of data, even for the methods that use a subset of the data, as they still need to search through the larger data volume to identify the points of interest. The different processing times for the grouped methods (the quadratic and sum methods) were due to different optimal smoothing widths.

Examination of the contribution of smoothing to the computational expense shows that TPG would have been the fastest method in the absence of smoothing and so may be the most computationally efficient method when using the filtering proposed by Jalobeanu and Gonçalves (2014). For the iterative fitting methods (GAU, LOG and

GGA) around 50% of the processing time was due to smoothing and so they would still be much slower with more efficient filtering methods. The windowed sums were faster than the sum but required more smoothing, slowing them down. The sum and spline required no smoothing.

5.2.4. Diffuse target conclusions

Due to broadening of returns by diffuse targets it cannot be assumed that the peak intensity is related to energy and so the peak, quadratic and windowed sum methods, which performed well for single returns after smoothing, were deemed unsuitable. The three point Gaussian gave reasonable estimates for small footprints, where returns are likely to be narrow, but gave inaccurate energies for large footprints as the brightest three points did not fully describe the feature. Caruana's method was as unstable for diffuse returns as for single returns and so was deemed unsuitable for analysing vegetation with waveform lidar. Vegetation presents complex targets that are not perfectly modelled by simple analytical forms. For large footprints (15 m) the target profile approaches a collection of Gaussians and, after smoothing, Gaussian fitting gave the most accurate results (around 1% better than SUR, SUR and SPL). In all other cases the sum (SUR and SUT) and spline (SPL) were slightly more accurate (around 1% RMSE better) and have the advantage of not needing any smoothing (maximising the resolution) and being far less computationally expensive.

Gaussian fitting is the most widely used method for interpreting waveform lidar, but these results show that equally accurate results can be achieved at a fraction of the computational expense by the sum and spline methods (around 1000 times faster for the sum and 15 times faster for the spline). Function fitting (GAU, GGA, LOG and SPL after deconvolution) gives easily understandable metrics (location, energy and width for each feature) that can be used for classifying vegetation (Mallet et al., 2011; Roncat, Briese, Jansa, & Pfeifer, 2014). However these can also be derived from turning points in the waveform when using the sum and this identification of turning points, peaks and widths is a necessary step in order to choose initial estimates required for function fitting. Alternatively deconvolution could be used to retrieve the target profile (Hancock et al., 2008; Roncat et al., 2014). The relative accuracies of the resulting metrics (location and width) need to be assessed and will form a future investigation, but for return energy (and so target reflectance) these conclusions will still hold and it is possible to use different algorithms for range and energy.

For large footprints (15 m) the accuracy was insensitive to noise and pulse length. For smaller footprints (33 cm) there was a strong relationship between pulse length, noise and retrieval accuracy. The longer the pulse the more accurate and less sensitive to noise the retrieval was. This has implications for sensor design. The results suggest that for a system with a noise level similar to SALCA or a Leica ALS50-II (n = 1),

Table 8
Mean processing time for a single return, in milliseconds, for the different methods at noise level 1, including any smoothing. NA indicates that fit errors were not acceptable.

Footprint pulse	15 m			33 cm			1 cm		
	0.1 m	1.1 m	2.15 m	0.1 m	1.1 m	2.15 m	0.1 m	1.1 m	2.15 m
PEA	0.008	0.005	0.006	0.001	0.001	0.003	0.001	0.001	0.001
SUR	0.007	0.008	0.010	0.001	0.003	0.005	0.001	0.002	0.003
SUT	0.007	0.008	0.010	0.001	0.003	0.005	0.001	0.002	0.003
SUS	2.58	0.008	0.010	0.001	0.003	0.005	0.001	0.002	0.003
WS3	0.025	0.016	0.017	0.002	0.004	0.008	0.002	0.002	0.002
WS5	0.025	0.016	0.017	0.002	0.004	0.008	0.002	0.002	0.002
WS7	0.025	0.016	0.017	0.002	0.004	0.008	0.002	0.002	0.002
QUP	0.037	NA	NA	0.003	0.008	0.012	0.002	0.003	0.004
QUA	2.56	2.54	NA	0.003	0.008	0.013	0.002	0.003	0.004
TPG	0.006	NA	NA	0.020	0.024	0.021	0.016	0.010	0.015
SPL	1.48	1.82	2.22	0.043	0.26	0.76	0.040	0.14	0.35
CAR	NA	NA	NA	0.004	NA	NA	0.003	0.004	0.005
GAU	5.27	5.01	5.31	0.59	3.65	54.8	0.33	0.80	1.51
LOG	126	112	111	125	166	542	4.72	8.93	15.9
GGA	6.60	5.77	5.76	0.46	1.68	4.11	0.35	0.73	0.87

a 2.15 m pulse sampled every 1 ns (or 1.1 m pulse sampled every 0.5 ns) would give a 2.7% error compared to a 3.4% error for a 1.1 m (0.6 m) pulse.

6. Conclusions

A range of methods for extracting return energy (and so reflectance) from waveform lidar have been tested on a range of targets and system parameters. As far as the authors are aware this included every method currently proposed for measuring waveform lidar energy and the full range of current and proposed full waveform lidar systems. The targets ranged from single returns to complex returns over vegetation and so this is the first time such an exhaustive cross-comparison has been carried out. Photon counting systems were excluded due to the repeat measurements needed to estimate energy.

For single returns the peak, quadratic and windowed sum methods gave the most consistent results after smoothing and so could easily be calibrated to give accurate energy from hard surfaces. However, this requires the peak intensity to be related to the total waveform energy, an assumption that has been shown not to hold over vegetation and so these methods (and the three point Gaussian, which gave reasonable results) are deemed unsuitable. Caruana's method was too unstable to provide accurate fits in all cases tested. This leaves the sum, spline and three iterative fitting methods (SUR, SUT, SPL, GAU, LOG and GGA).

It has been shown that vegetation causes complex returns. For large footprints (15 m) these approached a Gaussian after smoothing so that Gaussian fitting (GAU) gave the most accurate results (around 1% lower RMSE than the sum and spline). For smaller footprints (33 cm and 1 cm) the sum (SUR and SUT) and spline (SPL) gave slightly more accurate (around 1% lower RMSE than Gaussian fitting). The sum and spline have the advantage of being robust, computationally efficient (the sum several hundred times more so than the spline) and making no assumptions about the target shape. Some interpretation will be needed to identify the location and distribution of individual targets, especially when the system pulse causes their waveforms to overlap. This can be achieved by function fitting (Hofton et al., 2000), by examining turning points, which is a necessary step to choose initial estimates for function fitting or by deconvolution (Hancock et al., 2008; Roncat et al., 2014). An assessment of the relative accuracies of these will form a future paper.

For small footprints (33 cm) there was a strong dependence of accuracy on system pulse length, with long pulses giving lower errors and being less sensitive to noise. This agrees with the findings of Landi (2002). Therefore there is a trade-off between energy accuracy, and the minimum separation that targets can be distinguished (for longer system pulses) or data volume (for higher sampling rates). This dependence disappears for large footprints (15 m) due to decreased target heterogeneity at this scale. This is good news for the proposed spaceborne missions.

These results suggest that the spline and sum will give as good if not better results at a fraction of the computational expense, except for large footprint systems (15 m), where Gaussian fitting is marginally more accurate.

Datasets from some lidar instruments, such as the Riegl LMS-Q680i (footprint around 30 cm), were only available to some past studies in the form of Gaussian fit parameters due to proprietary file formats for waveform lidar data (Armston et al., 2013). Discrete return systems only report an energy derived from a proprietary algorithm, which also imposes additional restrictions on quantitative use of these data and may suffer from similar limitations. The results from this study show that these methods may limit the accuracy of retrieved lidar energy over vegetation. New, open source, file formats are making full waveform data more common (Bunting, Armston, Lucas, & Clewley, 2013). These findings have implications for all full waveform lidar vegetation missions, such as JAXA's proposed Multi-Footprint Observation LiDAR

and Imager (MOLI) (Murooka et al., 2013) and NASA's recently funded Global Ecosystem Dynamics Investigation (GEDI) missions.

Appendix A. Linear method descriptions

A.1. Peak (PEA)

This is the simplest method, with the brightest bin in a feature taken as the return energy. This is the value used by Gaulton et al. (2013) for single leaves and the most closely correlated to the discrete return Leica ALS50-II output, though the exact algorithm used by Leica is proprietary. The disadvantage is it makes no attempt to reconstruct the missing data and ignores the effect of pulse broadening by diffuse targets. For short pulses (such as Fig. 2(b)) this will lead to a change in measured energy as sub-bin range changes.

A.2. Sum (SUR, SUT, SUS)

This is the second simplest method, with the sum of the waveform values above a threshold taken as the energy (Jacobus & Chien, 1981; Landi, 2002). There are three variations of the sum method. The rectangular sum (SUR) adds up all bins above noise, making no attempt to reconstruct missing data. Using the trapezium rule (SUT) or Simpson's rule (SUS) goes some way to trying to fill in the missing data (Stroud & Booth, 2003). This is also referred to as the "area under the curve" method. Range can be calculated from the similar "centre of gravity" (Jutzi & Stilla, 2005).

A.3. Windowed sum (WS3, WS5, WS7)

This is the same as the rectangular sum but only uses points within a defined number of bins from the peak to help avoid noise and reduce computational expense. Fisher and Naidu (1996) tested three (WS3, one bin either-side), five (WS5, two bins either-side) and seven (WS7, three bins either-side) point windows. This does not attempt to fill in missing data and assumes that the windowed areas describe the full return. Jupp et al. (2009) used the mean of the brightest three bins, which is equivalent to WS3 but then applied a linear correction based on the sub-bin range. This fills in the gaps between datapoints but still assumes that the brightest three bins are representative of the whole return.

A.4. Quadratic (QUA, QUP)

A lidar pulse can be approximated as a quadratic function (Blais & Rioux, 1986; Jupp et al., 2009):

$$DN(r) = a + br + cr^2 \tag{A.1}$$

where DN is the waveform digital number, r is the range and a, b and c are the constants. This can be analytically fitted to the brightest three bins of a feature (Blais & Rioux, 1986). A feature will always have at least one point above noise and so if there are less than three above noise, the points in the noise can be used to fit the function. For the three points r_p, DN_p (the peak), r_{p-1}, DN_{p-1} and r_{p+1}, DN_{p+1} the quadratic coefficients can be calculated:

$$c = \frac{(d_1 - d_0)}{(r_{p+1} - r_{p-1})} \tag{A.2}$$

$$b = d_0 - c(r_p + r_{p-1}) \tag{A.3}$$

$$a = DN_p - b * r_p - c * r_p^2 \tag{A.4}$$

where d_0 and d_1 are the gradients between the three points:

$$d_1 = \frac{(DN_{p+1} - DN_p)}{(r_{p+1} - r_p)} \tag{A.5}$$

$$d_0 = \frac{(DN_p - DN_{p-1})}{(r_p - r_{p-1})}. \tag{A.6}$$

Taking the integral of the resulting quadratic between the two crossings of the noise threshold gives the energy (QUA). Alternatively the peak of the quadratic (DN value of maximum point) can be used as a measure of energy (QUP). Both versions attempt to fully reconstruct the missing data.

A.5. Three point Gaussian (TPG)

Li, Valentine, and Rana (1999) proposed analytically fitting a Gaussian to the brightest three bins of a feature (a similar approach to the quadratic fit). A Gaussian is given by:

$$DN(r) = Ae^{\frac{(r-\mu)^2}{2\sigma^2}} \tag{A.7}$$

where;

DN   is the waveform intensity
r is the range
μ is the range to the Gaussian centre
A is the amplitude
σ is the pulse width.

Using only the three brightest points, r_i, DN_i, this can be solved analytically:

$$\mu = \frac{1}{2} \frac{\ln \left(DN_1^{r_3^2 - r_2^2} DN_2^{r_1^2 - r_3^2} DN_3^{r_2^2 - r_1^2} \right)}{\ln \left(DN_1^{r_3 - r_2} DN_2^{r_1 - r_3} DN_3^{r_2 - r_1} \right)} \tag{A.8}$$

$$\sigma = \sqrt{\frac{1}{2} \frac{(r_1 - r_3)(r_2 - r_1)(r_3 - r_2)}{\ln \left(DN_1^{r_3 - r_2} DN_2^{r_1 - r_3} DN_3^{r_2 - r_1} \right)}} \tag{A.9}$$

$$A = (DN_1 DN_2 DN_3)^{\frac{1}{3}} e^{\left(\frac{(r_1 - \mu)^2 + (r_2 - \mu)^2 + (r_3 - \mu)^2}{6\sigma^2} \right)} \tag{A.10}$$

where r_2, DN_2 is the peak position and amplitude and r_1, DN_1 and r_3, DN_3 are the points on either-side. This assumes that the peak of the true return is at a known position within the bin (either the centre or the edge) and so it is not a true reconstruction of the missing signal. This approach was devised for interpreting nuclear spectrometer results (Li et al., 1999) which do not have the same sampling issues as waveform lidar.

Jalobeanu and Gonçalves (2014) introduced a more robust version using a very constrained Gaussian fitting to smooth and suppress ringing then used Bayesian methods to constrain the three point fit. This increased the robustness and computational efficiency of the method over single targets but the effect on return energy accuracy was not discussed and it is not clear that there would be any improvement. As Jalobeanu and Gonçalves (2014) was released after the processing for the current paper had been completed, this implementation was not tested here.

A.6. Spline (SPL)

Roncat et al. (2011) proposed fitting a uniform B-spline curve to the waveform. A spline is a piecewise continuous linear polynomial, allowing an analytical fit. Roncat et al. (2011) used the Cox–de Boor method (Farin, 2002) for a computationally efficient, well constrained fit. A B-spline can be fitted using standard computer libraries (Press et al., 1994), page 115 and the energy calculated as the integral above noise. Deconvolving with a uniform B-spline curve fit to the system pulse can give a direct measure of the target profile (Roncat et al., 2014).

A.7. Caruana's (CAR)

Caruana, Searle, Heller, and Shupack (1986) showed that a Gaussian can be linearised by taking the natural logarithm of the waveform intensity and range, allowing a direct linear fit. The equation for a Gaussian then becomes:

$$\ln DN(r) = a + br + cr^2 \tag{A.11}$$

where a, b and c are the parameters to fit (a is equal to the DC bias), r is the range and DN the waveform intensity. This method is sensitive to noise and low amplitude points have a larger effect on the final result (Guo, 2012). This could be avoided by removing low amplitude points by thresholding, however Guo (2012) proposed an elegant method to remove this effect without any pre-processing by weighting the linear fit by amplitude. This reduces the effect of low amplitude points without needing to set an arbitrary threshold (which can be problematic (Hancock et al., 2011)).

The matrix to be solved becomes:

$$\begin{pmatrix} N & \Sigma r & \Sigma r^2 \\ \Sigma r & \Sigma r^2 & \Sigma r^3 \\ \Sigma r^2 & \Sigma r^3 & \Sigma r^4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \Sigma \ln(DN) \\ \Sigma r \ln(DN) \\ \Sigma r^2 \ln(DN) \end{pmatrix} \tag{A.12}$$

where N is the total number of points, DN is the intensity values and r is the range. a, b and c can be found by solved using Gauss–Jordan elimination (Press et al., 1994) and the Gaussian parameters found using the following equations:

$$\sigma = \frac{1}{\sqrt{-2c}} \tag{A.13}$$

$$\mu = \frac{b}{-2c} \tag{A.14}$$

$$E = \sqrt{\frac{-\pi}{c}} e^{1 - \frac{4b^2}{c}} \tag{A.15}$$

where σ and μ are the Gaussian parameters (of Eq. A.7) and E is the energy (integral of the Gaussian).

Appendix B. Iterative method descriptions

B.1. Gaussian fitting (GAU)

This is the most widely applied method for analysing lidar over vegetation (Guo, Chehata, Mallet, & Boukir, 2011; Hofton et al., 2000; Murooka et al., 2013; Neuenschwander et al., 2009; Reitberger,

Krzystek, & Stilla, 2008; Wagner et al., 2006). Non-linear optimisation is used to fit a number of Gaussians (Eq. A.7) to the waveform.

This completely reconstructs missing data, if the system pulse convolved with the target profile is near Gaussian, and should give an accurate energy. Smoothing by convolution with a Gaussian can make the waveform more Gaussian like (Hofton et al., 2000). However, Gaussian fitting can be unstable (Jalobeanu & Gonçalves, 2012); Wagner et al. (2006) reported "successful" Gaussian fitting to 98% of the measured waveforms, but that some of these had nonsensical values. Here "successful" means that the algorithm gave an energy value, not that the value was correct as no ground truth was used.

B.2. Lognormal fitting (LOG)

Chauve et al. (2007) suggested fitting the more complex lognormal function in order to cope with asymmetric pulses and complex targets. The function to fit is given by:

$$DN(r) = Ae^{-\frac{(\ln(r-s)-\mu)^2}{2\sigma^2}}. \tag{B.1}$$

This is very similar to the Gaussian, except that an extra variable, s, is included and the natural logarithm of r is used. This allows more flexibility when fitting but is more computationally expensive and can be less robust than Gaussian fitting (Chauve et al., 2007).

B.3. Generalised Gaussian fitting (GGA)

Another more complex function proposed by (Chauve et al., 2007) is the generalised Gaussian, given by:

$$DN(r) = Ae^{-\frac{|r-\mu|^{\alpha^2}}{2\sigma^2}}. \tag{B.2}$$

Again this is very similar to a Gaussian except that the numerator of the exponent is raised to the power of α^2 rather than 2. Like the lognormal this allows more flexibility when fitting but is more computationally expensive and can be less robust than Gaussian fitting (Chauve et al., 2007).

B.4. Bayesian methods

Hernandez-Marin et al. (2007) and Wallace et al. (2012) proposed using Bayesian methods to fit a line through noisy data. This was developed for photon counting systems, with their much lower signal to noise levels and so was not investigated in this paper. Whilst it may be possible to apply it to standard waveform lidars, it is most likely equivalent to fitting a spline at these much lower noise levels and so was not tested.

References

Armston, J., Disney, M., Lewis, P., Scarth, P., Phinn, S., Lucas, R., et al. (2013). Direct retrieval of canopy gap probability using airborne waveform lidar. *Remote Sensing of Environment, 134*, 24–38.

Baltsavias, E.P. (1999). Airborne laser scanning: Basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing, 54*, 199–214.

Blais, F., & Rioux, M. (1986). Real-time numerical peak detector. *Signal Processing, 11*, 145–155.

Bunting, P., Armston, J., Lucas, R.M., & Clewley, D. (2013). Sorted pulse data (spd) library. Part i: A generic file format for lidar data from pulsed laser systems in terrestrial environments. *Computers & Geosciences, 56*, 197–206.

Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., Herold, M., et al. (2014). Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution.* http://dx.doi.org/10.1111/2041-210X.12301.

Caruana, R.A., Searle, R.B., Heller, T., & Shupack, S.I. (1986). Fast algorithm for the resolution of spectra. *Analytical Chemistry, 58*, 1162–1167.

Chauve, A., Mallet, C., Bretar, F., Durrieu, S., Deseilligny, M.P., Puech, W., et al. (2007). Processing full-waveform lidar data: Modelling raw signals. *International archives of photogrammetry, remote sensing and spatial information sciences 2007* (pp. 102–107).

Dabney, P., Harding, D., Abshire, J., Huss, T., Jodor, G., Machan, R., et al. (2010). The slope imaging multi-polarization photon-counting lidar: Development and performance results. *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE international* (pp. 653–656). IEEE.

Danson, F.M., Gaulton, R., Armitage, R.P., Disney, M., Gunawan, O., Lewis, P., et al. (2014). Developing a dual-wavelength full-waveform terrestrial laser scanner to characterize forest canopy structure. *Agricultural and Forest Meteorology, 198*, 7–14.

Disney, M.I., Kalogirou, V., Lewis, P., Prieto-Blanco, A., Hancock, S., & Pfeifer, M. (2010). Simulating the impact of discrete-return lidar system and survey characteristics over young conifer and broadleaf forests. *Remote Sensing of Environment, 114*, 1546–1560.

Disney, M.I., Lewis, P., & Saich, P. (2006). 3d modelling of forest canopy structure for remote sensing simulations in the optical and microwave domains. *Remote Sensing of Environment, 100*, 114–132.

Douglas, E.S., Martel, J., Li, Z., Howe, G., Hewawasam, K., Marshall, R.A., et al. (2015). Finding leaves in the forest: The dual-wavelength echidna lidar. *IEEE Geoscience and Remote Sensing Letters, 12*, 776–780. http://dx.doi.org/10.1109/LGRS.2014.2361812.

Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., et al. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment, 79*, 305–319.

Dubayah, R.O., & Drake, J.B. (2005). Lidar remote sensing for forestry applications. *Journal of Forestry, 98*, 44–46.

Farin, G.E. (2002). *Curves and surfaces for CAGD: A practical guide.* Morgan Kaufmann.

Fisher, R., & Naidu, D. (1996). A comparison of algorithms for subpixel peak detection. *Image technology* (pp. 385–404). Springer.

Garbow, B., Hillstrom, K., More, J., Markwardt, C., & Moshier, S. (1980). *MPFIT: A MINPACK-1 least squares fitting library in C.* Argonne National Laboratory (http://www.physics.wisc.edu/craigm/idl/cmpfit.html).

Gaulton, R., Danson, F.M., Ramirez, F.A., & Gunawan, O. (2013). The potential of dual-wavelength laser scanning for estimating vegetation moisture content. *Remote Sensing of Environment, 132*, 32–39.

Guo, H. (2012). *Streamlining Digital Signal Processing: A Tricks of the Trade Guidebook* (pp. 297–305) (Second Edition). A simple algorithm for fitting a Gaussian function. *Signal processing.*

Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multi-spectral image data for urban scene classification using random forests. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 56–66.

Hancock, S. (2010). *Understanding the measurement of forests with waveform lidar.* Ph.D. thesis UCL (University College London) (http://discovery.ucl.ac.uk/20221/).

Hancock, S., Disney, M., Muller, J. -P., Lewis, P., & Foster, M. (2011). A threshold insensitive method for locating the forest canopy top with waveform lidar. *Remote Sensing of Environment, 115*, 3286–3297.

Hancock, S., Essery, R., Reid, T., Carle, J., Baxter, R., Rutter, N., et al. (2014). Characterising forest gap fraction with terrestrial lidar and photography: An examination of relative limitations. *Agricultural and Forest Meteorology, 189*, 105–114.

Hancock, S., Lewis, P., Disney, M.I., Foster, M., & Muller, J. -P. (2008). Assessing the accuracy of forest height estimation with long pulse waveform lidar through Monte-Carlo ray tracing. *Edinburgh: Silvilaser. September, 17–18.*

Hancock, S., Lewis, P., Foster, M., Disney, M., & Muller, J. -P. (2012). Measuring forests with dual wavelength lidar: A simulation study over topography. *Agricultural and Forest Meteorology, 161*, 123–133.

Harding, D., Blair, J., Garvin, J., & Lawrence, W. (1994). Laser altimetry waveform measurement of vegetation canopy structure. *Geoscience and remote sensing symposium. IGARSS'94. Surface and atmospheric remote sensing: Technologies, data analysis and interpretation., International, Vol. 2.* (pp. 1251–1253).

Harding, D.J., & Carabajal, C.C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters, 32*, L21S10.

Harding, D., Dabney, P., Valett, S., Yu, A., Vasilyev, A., & Kelly, A. (2011). Airborne polarimetric, two-color laser altimeter measurements of lake ice cover: A pathfinder for NASA's ICESat-2 spaceflight mission. *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE international* (pp. 3598–3601). IEEE.

Hernandez-Marin, S., Wallace, A.M., & Gibson, G.J. (2007). Bayesian analysis of lidar signals with multiple returns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*, 2170–2180.

Hill, R., & Broughton, R.K. (2009). Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne lidar data: A case study in lowland Britain. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 223–233.

Hofton, M.A., Minster, J.B., & Blair, J.B. (2000). Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing, 38*, 1989–1996.

Jacobus, C.J., & Chien, R.T. (1981). Two new edge detectors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 581–592.

Jalobeanu, A., & Gonçalves, G. (2012). The full-waveform lidar Riegl LMS-Q680i: From reverse engineering to sensor modeling. *American Society of Photogrammetry and Remote Sensing Annual Conference, Sacramento, CA, USA.*

Jalobeanu, A., & Gonçalves, G. (2014). Robust ground peak extraction with range error estimation using full-waveform lidar. *IEEE Geoscience and Remote Sensing Letters, 11*, 1190–1194.

Jupp, D.L.B., Culvenor, D.S., Lovell, J.L., Newnham, G.J., Strahler, A.J., & Woodcock, C.E. (2009). Estimating forest LAI profiles and structural parameters using a ground-based laser called echidna. *Tree Physiology, 29*, 171–181.

Jutzi, B., & Stilla, U. (2005). *Measuring and processing the waveform of laser pulses.* Optical, 194–203.

Krainak, M.A., Abshire, J.B., Camp, J., Chen, J.R., Coyle, B., Li, S.X., et al. (2012). Laser transceivers for future NASA missions. *SPIE defense, security, and sensing.* International Society for Optics and Photonics (83810Y-83810Y).

Landi, G. (2002). Properties of the center of gravity as an algorithm for position measurements. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 485*, 698–719.

Lefsky, M.A., Keller, M., Yong, P., de Camargo, P.B., & Hunter, M.O. (2007). Revised method for forest canopy height estimation from geoscience laser altimeter system waveforms. *Journal of Applied Remote Sensing, 1*, 1–18.

Lewis, P. (1999). Three-dimensional plant modelling for remote sensing simulation studies using the botanical plant modelling system. *Agronomie, 19*, 185–210.

Li, J., Valentine, J.D., & Rana, A.E. (1999). The modified three point Gaussian method for determining Gaussian peak parameters. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 422*, 438–443.

Lin, W. -T., Shih, P.T. -Y., & Teo, T. -A. (2010). Full-waveform lidar processing with different computation and classification methods. *Asian conference on remote sensing proceedings*. Hanoi: Asian Association on Remote Sensing.

Los, S., Rosette, J., Kljun, N., North, P., Chasmer, L., Suárez, J., et al. (2012). Vegetation height and cover fraction between 60 S and 60 N from ICESat GLAS data. *Geoscientific Model Development, 5*, 413–432.

Mallet, C., Bretar, F., Roux, M., Soergel, U., & Heipke, C. (2011). Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, S71–S84.

Moussavi, M.S., Abdalati, W., Scambos, T., & Neuenschwander, A. (2014). Applicability of an automatic surface detection approach to micropulse photon-counting lidar altimetry data: Implications for canopy height retrieval from future ICESat-2 data. *International Journal of Remote Sensing, 35*, 5263–5279.

Murooka, J., Kobayashi, T., Imai, T., Suzuki, K., Sakaizawa, D., Yamakawa, S., et al. (2013). Overview of Japan's spaceborne vegetation lidar mission. *SPIE remote sensing*. International Society for Optics and Photonics (88940B-88940B).

Musselman, K.N., Margulis, S.A., & Molotch, N.P. (2013). Estimation of solar direct beam transmittance of conifer canopies from airborne lidar. *Remote Sensing of Environment, 136*, 402–415.

Nelson, R., Krabill, W., & Maclean, G. (1984). Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment, 15*, 201–212.

Neuenschwander, A.L., Magruder, L.A., & Tyler, M. (2009). Landcover classification of small-footprint, full-waveform lidar data. *Journal of Applied Remote Sensing, 3* (33544-33544).

Nevalainen, O., Hakala, T., Suomalainen, J., Mäkipää, R., Peltoniemi, M., Krooks, A., et al. (2014). Fast and nondestructive method for leaf level chlorophyll estimation using hyperspectral lidar. *Agricultural and Forest Meteorology, 198*, 250–258.

Newnham, G., Goodwin, N., Armston, J., Muir, J., & Culvenor, D. (2012). Comparing time-of-flight and phase-shift terrestrial laser scanners for characterising topography and vegetation density in a forest environment. *Proceedings of Silvilaser*.

Press, W.H., Tuekolsky, S.A., Vetterling, W.T., & Flannery, B.R. (1994). *Numerical recipes in C* (2nd ed.). Cambridge: Cambridge University Press.

Ramirez, F.A., Armitage, R.P., & Danson, F.M. (2013). Testing the application of terrestrial laser scanning to measure forest canopy gap fraction. *Remote Sensing, 5*, 3037–3056.

Raumonen, P., Kaasalainen, M., Åkerblom, M., Kaasalainen, S., Kaartinen, H., Vastaranta, M., et al. (2013). Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sensing, 5*, 491–520.

Reid, T., Spencer, M., Huntley, B., Hancock, S., Essery, R., Carle, J., et al. (2014). Spatial quantification of leafless canopy structure in a boreal birch forest. *Agricultural and Forest Meteorology, 188*, 1–12.

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*, 1407–1431.

Roncat, A., Bergauer, G., & Pfeifer, N. (2011). B-spline deconvolution for differential target cross-section determination in full-waveform laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 418–428.

Roncat, A., Briese, C., Jansa, J., & Pfeifer, N. (2014). Radiometrically calibrated features of full-waveform lidar point clouds based on statistical moments. *IEEE Geoscience and Remote Sensing Letters, 11*, 549–553.

Stilla, U., & Jutzi, B. (2008). Waveform analysis for small-footprint pulsed laser systems. *Topographic laser ranging and scanning: Principles and processing* (pp. 215–234).

Stroud, K.A., & Booth, D.J. (2003). *Advanced engineering mathematics, Vol. 175*, Palgrave Macmillan.

Wagner, W., Ullrich, A., Ducic, V., Melzer, T., & Studnicka, N. (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing, 60*, 100–112.

Wagner, W., Ullrich, A., Melzer, T., Briese, C., & Kraus, K. (2004). From single-pulse to full-waveform airborne laser scanners: Potential and practical challenges. *International Archives of Photogrammetry and Remote Sensing, 35*, 201–206.

Wallace, A., Nichol, C., & Woodhouse, I. (2012). Recovery of forest canopy parameters by inversion of multispectral lidar data. *Remote Sensing, 4*, 509–531.

Widlowski, J. -L., Robustelli, M., Disney, M., Gastellu-Etchegorry, J. -P., Lavergne, T., Lewis, P., et al. (2008). The rami on-line model checker (romc): A web-based benchmarking facility for canopy reflectance models. *Remote Sensing of Environment, 112*, 1144–1150.

Widlowski, J. -L., Taberner, M., Pinty, B., Bruniquel-Pinel, V., Disney, M., Fernandes, R., et al. (2007). Third radiation transfer model intercomparison (rami) exercise: Documenting progress in canopy reflectance models. *Journal of Geophysical Research, [Atmospheres], 112* (1984–2012).

Wu, J., van Aardt, J., & Asner, G.P. (2011). A comparison of signal deconvolution algorithms based on small-footprint lidar waveform simulation. *IEEE Transactions on Geoscience and Remote Sensing, 49*, 2402–2414.