ISPRS Journal of Photogrammetry and Remote Sensing 129 (2017) 131–150

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Gold – A novel deconvolution algorithm with optimization for waveform LiDAR processing

Tan Zhou [a,*], Sorin C. Popescu [a], Keith Krause [b], Ryan D. Sheridan [a], Eric Putman [a]

[a] LiDAR Applications for the Study of Ecosystems with Remote Sensing (LASERS) Laboratory, Department of Ecosystem Science and Management, Texas A&M University, 1500 Research Parkway Suite B 217, College Station, TX 77843, USA
[b] National Ecological Observatory Network, 1685 38th St., Suite 100, Boulder, CO 80301, USA

ARTICLE INFO

Article history:
Received 1 November 2016
Received in revised form 27 March 2017
Accepted 27 April 2017

Keywords:
Waveform LiDAR
Deconvolution
Gold
Richardson–Lucy (RL)
Decomposition
Parameter uncertainty

ABSTRACT

Waveform Light Detection and Ranging (LiDAR) data have advantages over discrete-return LiDAR data in accurately characterizing vegetation structure. However, we lack a comprehensive understanding of waveform data processing approaches under different topography and vegetation conditions. The objective of this paper is to highlight a novel deconvolution algorithm, the Gold algorithm, for processing waveform LiDAR data with optimal deconvolution parameters. Further, we present a comparative study of waveform processing methods to provide insight into selecting an approach for a given combination of vegetation and terrain characteristics. We employed two waveform processing methods: (1) direct decomposition, (2) deconvolution and decomposition. In method two, we utilized two deconvolution algorithms – the Richardson-Lucy (RL) algorithm and the Gold algorithm. The comprehensive and quantitative comparisons were conducted in terms of the number of detected echoes, position accuracy, the bias of the end products (such as digital terrain model (DTM) and canopy height model (CHM)) from the corresponding reference data, along with parameter uncertainty for these end products obtained from different methods. This study was conducted at three study sites that include diverse ecological regions, vegetation and elevation gradients. Results demonstrate that two deconvolution algorithms are sensitive to the pre-processing steps of input data. The deconvolution and decomposition method is more capable of detecting hidden echoes with a lower false echo detection rate, especially for the Gold algorithm. Compared to the reference data, all approaches generate satisfactory accuracy assessment results with small mean spatial difference (<1.22 m for DTMs, <0.77 m for CHMs) and root mean square error (RMSE) (<1.26 m for DTMs, <1.93 m for CHMs). More specifically, the Gold algorithm is superior to others with smaller root mean square error (RMSE) (<1.01 m), while the direct decomposition approach works better in terms of the percentage of spatial difference within 0.5 and 1 m. The parameter uncertainty analysis demonstrates that the Gold algorithm outperforms other approaches in dense vegetation areas, with the smallest RMSE, and the RL algorithm performs better in sparse vegetation areas in terms of RMSE. Additionally, the high level of uncertainty occurs more on areas with high slope and high vegetation. This study provides an alternative and innovative approach for waveform processing that will benefit high fidelity processing of waveform LiDAR data to characterize vegetation structures.

1. Introduction

Full waveform airborne laser scanners (ALS) are increasingly available to remote sensing data providers. The data acquired by such systems are widely applicable to vegetated ecosystem assessment and monitoring (Gwenzi and Lefsky, 2014; Hollaus et al., 2009; Lefsky, 2010; McGlinchy et al., 2014). Waveform airborne laser scanning is an active form of remote sensing technique that could provide additional geometric and physical information of the scattering substance along the path. It also gives users more control of data interpretation (Chauve et al., 2007) compared with the conventional discrete-return Light Detection and Ranging (LiDAR) technique. The physical principle of the waveform LiDAR system is similar to the conventional LiDAR system, but waveform LiDAR system can record the entire echo scattered from illuminated objects with different temporal resolutions (such as 1/2/4

* Corresponding author.
E-mail addresses: tankwin0@tamu.edu (T. Zhou), s-popescu@tamu.edu (S.C. Popescu).

http://dx.doi.org/10.1016/j.isprsjprs.2017.04.021
0924-2716/

nanosecond(s) (ns)) through digital sampling. Therefore, the waveforms not only include responses from ground level, but also comprise multiple-scattering responses of illuminated surfaces along the laser line which give more information of objects through waveform shapes, widths, intensities and skewness. These characteristics of waveform LiDAR make its processing more difficult than discrete-return LiDAR processing which only needs to combine the time between the emitted signal and received signal, the speed of light and geolocation information (like GPS, platform altitude of scanner). Through extracting and decoding these characteristics, the physical attributes of vegetation like canopy height (Gao et al., 2015; Gwenzi and Lefsky, 2014), target cross section (Roncat et al., 2011), stem volume (Reitberger et al., 2009), and above ground biomass of forest (Boudreau et al., 2008) can be modeled. Thus, gaining knowledge of the forest from waveform LiDAR data is a pivotal step toward efficient and comprehensive understanding forest roles in biomass change and carbon cycle under climate change.

Generally, the waveform processing method can be categorized into two types: one is the direct decomposition method and another is the deconvolution and decomposition method. Each method has been successfully applied for echo detection based on its own physical background. For direct decomposition, the emitted pulse is generally assumed to be Gaussian shape, as well as the scatterers' differential backscatter cross section (Wagner et al., 2006). The return waveform is obtained through the convolution of the Gaussian shape emitted pulse and the Gaussian substance scattering function (Wagner et al., 2006). Therefore, the Gaussian decomposition is the most frequently used approach to process the received signals (Mallet and Bretar, 2009). Many studies have been carried out on Gaussian decomposition to perform processing and analysis of different types of waveform LiDAR data such as LVIS (Laser Vegetation Imaging Sensor) data (Zhuang and Mountrakis, 2014), ICESat (Ice Cloud and land Elevation Satellite) data (Gwenzi and Lefsky, 2014; Harding, 2005; Keller, 2007; Lefsky et al., 2005), airborne data (Chauve et al., 2007; Hancock et al., 2017; Wagner et al., 2006; Wu et al., 2011). These researches have demonstrated that the Gaussian model is sufficient for processing waveform LiDAR data to characterize the vegetation structure, no matter whether its footprint is large or small. Fitting the waveform with the Gaussian function can provide peak amplitude, the width of each Gaussian component and peak location information. The peak amplitude can be used as a criterion to filter the points from below ground (Rowe, 2013) and it also provides us information about surface of objects along the pulse. The echo width has been employed as tool to characterize the crown depth, crown variability, and topographic relief (Harding, 2005; Keller, 2007). The range or elevation of a specific reflecting surface can be calculated using the peak location provided through Gaussian decomposition (Hofton et al., 2000). Furthermore, these pieces of information combined can be utilized to estimate woody cover and biomass, classify tree species (Reitberger et al., 2008) and map land-cover, etc. (Wang and Glennie, 2015).

The commonly used approaches to fit a sum of Gaussian functions are the Non-linear Least Square (NLS) method with Levenberg-Marquardt (LM) optimization algorithm (Hofton et al., 2000), the maximum likelihood methods (Persson et al., 2005) with the Expectation Maximization (EM) algorithm, and the Progressive Waveform Decomposition (PWD) method (Zhu et al., 2011). The limitations of the first two methods are the lack of prior knowledge about waveforms and the difficulties of identifying initialization of waveform parameters, such as the number of peaks, peak amplitude and width. Mallet et al. (2009) employed a stochastic method to reconstruct the waveform LiDAR by decomposing each echo with suitable functions like the generalized Gaussian function, Weibull function, Nakagami function and Burr function. These methods are robust and have shown good potential for applications in waveform processes and analysis (Chauve et al., 2007; Fieber et al., 2015; Reitberger et al., 2008). Whereas the pulse detection method such as the Average Square Difference Function (ASDF) (Zhu et al., 2011) may omit the important waveform parameters, the PWD method may lead to false echo detection due to the ringing effect.

Based on the standard LiDAR equations and real world constraints (Carlsson et al., 2001), the power of the received pulse can also be expressed as the sum of echoes from N targets with system and environment contributions (Mallet and Bretar, 2009). The direct decomposition method does not take into account the detector and system's contributions to the waveform, which results in the loss of illuminated surface information. To reduce unwanted system contribution and recover the true distribution of the illuminated surface, the second approach, the deconvolution and decomposition, is proposed. Several published studies have successfully applied different deconvolution algorithms such as B-spline, Richardson-Lucy (RL), Non-negative Least Squares (NNLS), Wiener Filter (WF) (Cawse-Nicholson et al., 2014; McGlinchy et al., 2014; Neuenschwander, 2008; Roncat et al., 2011; Wu et al., 2011), sparsity-constrained regularization approach (Azadbakht et al., 2016) and Bayesian inference method (Jalobeanu and Gonçalves, 2014) to recover the true cross-sectional profile of an illuminated object.

Though the widely used deconvolution algorithms in the waveform LiDAR are RL (Lucy, 1974), NNLS and WF (Jutzi and Stilla, 2006), the studies of Nordin (2006) and Wu et al. (2011) have demonstrated that the RL algorithm is superior to other algorithms for the estimation of tree biomass and detection of unobservable peaks. The detailed information of the above three algorithms can be found in Wu's study (2011). However, each algorithm has its own advantages and limitations when applied to the deconvolution. For example, the RL and NNLS can lead to more accurate results at the expense of taking a longer time to complete the iterative process; WF may require less implementation time but results in less accurate solution. Additionally, developing open source tools for the waveform processing is also a pressing need for the extensive applications of waveform LiDAR data with different format. Hancock et al. (2008) proposed the Gold's method to process the large-footprint waveform GLAS (Geoscience Laser Altimeter System), however, the information contained in the large-footprint and small-footprint data is different (Mallet and Bretar, 2009). Large footprint data, e.g., up to 65 m, have the waveform returned from multiple tree crowns and is affected by topography, especially on high slope terrain, therefore such data have a significantly different shape compared to small footprint waveform data that samples a small portion of tree crowns intersected by the laser beam, possibly not reaching the ground under dense vegetation. Thus, the Gold's method has not been proven in prior studies that it is suitable for small-footprint waveform LiDAR processing. Additionally, at the time of developing our study, we could not identify any publication that has conducted the parameter optimization of waveform deconvolution for vegetation applications. Results of waveform deconvolution depend significantly on the choice of parameters used with deconvolution functions.

To enrich the existing waveform processing methods and enhance the performance of the deconvolution, the optimized Gold algorithm described in Section 2.3.1 is proposed to reconstruct the differential backscatter cross section from the waveform LiDAR collected by the National Ecological Observatory Network (NEON). Meanwhile, there is a lack of quantitative and comprehensive comparisons of waveform LiDAR processing methods. The overall goal of this research is to propose a novel deconvolution approach to process waveform data and contribute to a better understanding of advantages and limitations of different small-footprint

waveform LiDAR processing approaches. Specific objectives are to: (1) introduce a novel deconvolution algorithm, the Gold algorithm, which is a non-negative iterative solution toward generating more accurate and representative ground elevation and canopy height; (2) develop an optimization methodology for finding appropriate deconvolution parameters; (3) explore advantages and limitations of various waveform processing techniques to derive topography and canopy height information; (4) perform comprehensive comparisons of results with different approaches and assess each approach's accuracy and parameter uncertainty.

Our hypothesis is that better results are expected with the new algorithm in terms of the echo detection, accuracy assessment and parameters uncertainty analysis. The innovative aspects of this study consist of: (1) extending the current small-footprint waveform processing methods by adapting the Gold algorithm to process the small-footprint waveform LiDAR data and then investigating their performance in different topography and vegetation conditions, (2) introducing an optimization process of determining the deconvolution parameters and (3) implementing processing steps within an open source software or tools such as R (2013) and LAStools (Isenburg, 2012).

2. Methodology

2.1. Study sites

This study was conducted using the data from three NEON terrestrial sites: (1) the Harvard Forest (HF), north-central Massachusetts; (2) the San Joaquin Experimental Range (SJER), north of Fresno, California; and (3) the Ordway-Swisher Biological Station (OSBS), near Melrose of Putnam County, Florida (Fig. 1). We hypothesize that the performance of approaches will be affected by factors such as topography and elevation gradients. These study sites were selected to test the robustness of different approaches for processing waveform LiDAR data. They were extended over diverse ecological regions, climate and elevation gradients with different number of flight lines.

The HF is a core wild-land site and statistically represents unmanaged wildlife conditions across the NEON's 30-year history (Kampe, 2010). One flight line of the cropped waveform sample area is chosen. The data covers about 60 m × 60 m with the center located at 731156.6 Easting, 4712671.4 Northing, and UTM Zone 18 N (Fig. 1A). This site primarily consists of dense mixed hardwood trees with dominant species being white pine (*Pinus strobus*) and red oak (*Quercus rubra*) in the center 20 m × 20 m area. The landscape is characterized by flat terrain with an elevation difference of approximately 5 m.

The second site, the SJER, is located in the foothills of Sierra Nevada Mountains, about 32 km north of Fresno, California. The cropped waveform sample is about 6.25 ha (250 m × 250 m) with the center at 256840.0 Easting, 4110820.0 Northing, and UTM Zone 11 N (Fig. 1B). The elevation ranges from 380 to 425 m dominated by sparse blue oak (*Quercus douglasii*), interior live oaks (*Quercus wislizeni*) and digger pine (*Pinus sabiniana*). The topography is complex with coarse, large hills and valleys.

The OSBS is located near Melrose of Putnam County, Florida with an elevation range from 21 to 48 m. The cropped area is covered by four flight lines and it is about 60 ha (1172 m × 534 m) with the center at 402507.6 Easting, 3282045.1 Northing, and UTM Zone 17 N (Fig. 1C). It is composed of homogenous forest dominated by Longleaf Pines (*Pinus palustris* Mill.) and Loblolly (*Pinus taeda*), areas of mixed patches of vegetation structure and heterogonous land cover types, including water body, wetland, open ground and road. These make the OSBS site well suited for comparing and testing performance of different processing waveform LiDAR processing over a range of simple to complex vegetation communities.

2.2. LiDAR data

2.2.1. Waveform LiDAR data

The three waveform LiDAR datasets were acquired with an Optech Gemini instrument at a nominal range of 1000 m (the aircraft flew at 1000 m above ground level). It achieved a nominal density of 3.82 points per square meter with a 0.8 m diameter spot and a spot spacing of about 0.524 m in the across-track direction and 0.5 m in the along-track direction. Both datasets were collected during the leaf-on condition on August 8, 2012 for the HF site, June 13, 2013 for the SJER site, and May 7 and May 19, 2014 for the OSBS site. All data were distributed by the NEON data center (http://www.neonscience.org/content/airborne-data). The detailed technical data specifications are shown in Table 1.

There was one flight line with 13,902 waveforms included in the HF sample area. For the SJER site, two flight lines were available with 167,019 waveforms for flight line 03 and 91,648 waveforms for flight line 12. For the OSBS site, four flight lines were available for our study region as shown in Fig. 1C. The number of waveforms for four flight lines is 660,995, 859,919, 597,455 and 1,371,186, respectively. Each waveform was segmented into 500 time bins with 1 ns temporal spacing.

The waveforms stored the digital number (DN) of return pulses, which can be assumed to be the amplitude of the waveform (Fig. 2). Simultaneously, NEON provided geolocation information and corresponding outgoing pulses that consist of 100 time bins with a temporal resolution of 1 ns. The geolocation data comprised Easting, Northing, height, dx (m), dy (m), dz (m), and first return bin location. And dx, dy, and dz were the pulse direction vector that can be used to calculate the accurate geolocation of any other time bins in a given waveform without registration and rectification. All data were zero padded. NEON also provided us with a prototype system impulse which was a return pulse of single laser shot from a hard ground target with a mirror angle close to nadir and corresponding outgoing waveform (Fig. 2). This can help us remove the outgoing pulse and system response effect and perform a deconvolution on the waveform.

2.2.2. Reference data

To validate the performance of methods and end products of the waveform LiDAR, the discrete-return LiDAR data and the Digital Terrain Models (DTMs) and Canopy Height Models (CHMs) provided by NEON were used as the reference data. According to the NEON's discrete-return LiDAR Algorithm Theoretical Basis Document (ATBD), the maximum horizontal accuracy of discrete-return LiDAR is about 0.4 m, with maximum LiDAR vertical accuracy 0.36 m, respectively (Keith and Tristan, 2015). Discrete-return LiDAR data will be used as reference in the Number of echoes section test whether more points were extracted from the waveform LiDAR data. Additionally, we compared the waveform-based end products such as DTMs and CHMs with corresponding reference data provided by the NEON to conduct the accuracy assessment of our approaches.

2.3. Waveform processing

Waveform processing involves a series of steps, including noise detection, smoothing, radiometric calibration (Briese et al., 2008), deconvolution and decomposition, etc. Many studies have been conducted to interpret the waveform data and used it to estimate vegetation structure and function, such as canopy height and above-ground biomass (Chauve et al., 2007; Gwenzi and Lefsky, 2014; Roncat et al., 2011; Wagner et al., 2006). The major steps

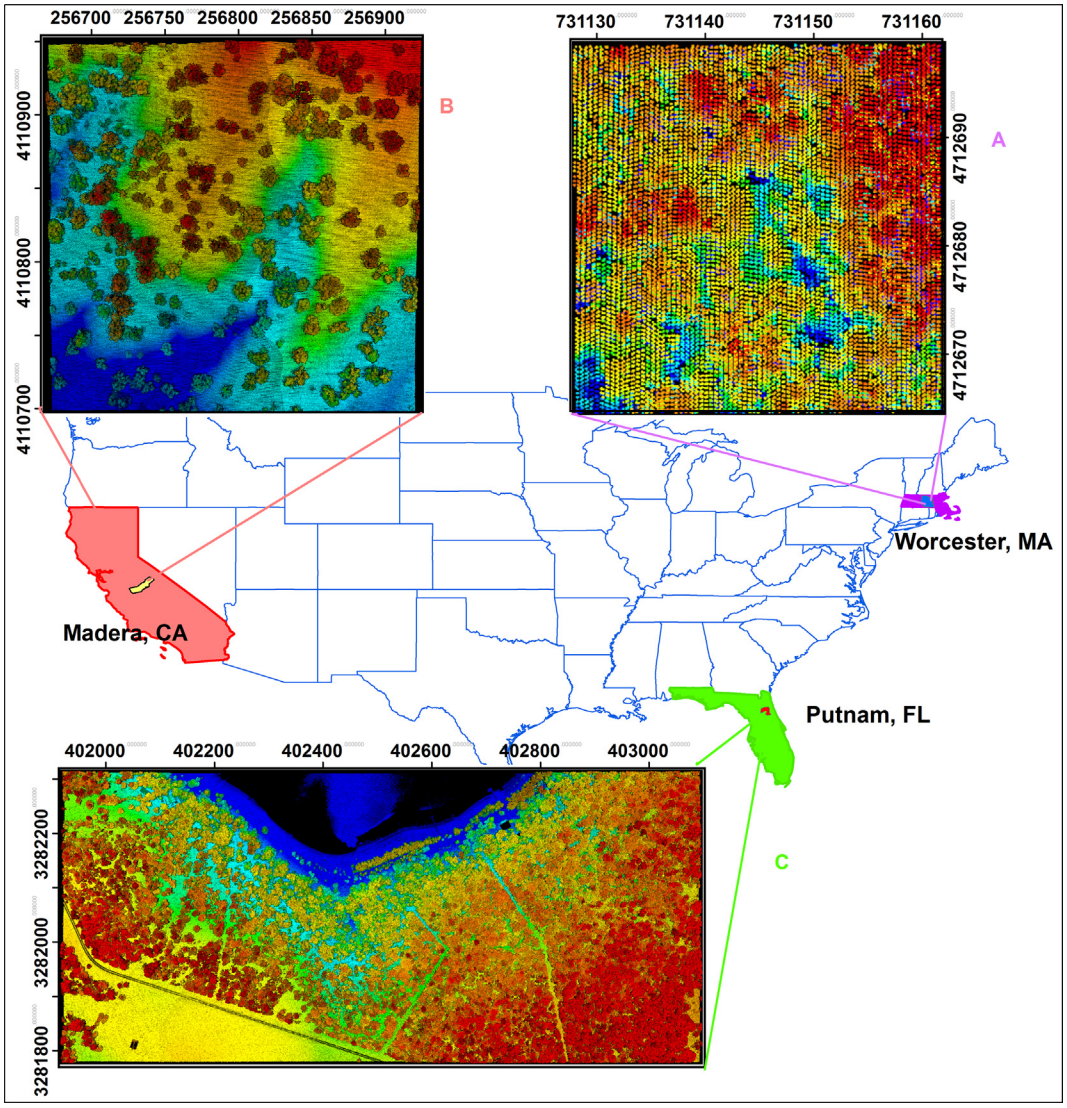

Fig. 1. Locations of three study sites in Massachusetts (one flight line), California (two flight lines) and Florida (four flight lines) with discrete-return LiDAR points.

Table 1
Main technical specifications of the NEON waveform data.

Study sites	Technical specifications
Operating altitude	~1000 m
Wavelength	1064 nm
Pulse repetition frequency	100 kHz
Scan frequency	50 Hz
Beam divergence	0.8 mrad
Scan angle range	±18.5 degree
Spot spacing	0.524 m (across-track), 0.5 m (along-track)
Footprint size	3.83 points/m^2 (0.8 m)
Digitizer	1 ns (12 bit A/D, baseline signal is 200)
Outgoing pulse width	~14 ns
Flying direction	HF: East, heading 90 degree SJER & OSBS: Northing to South or South to North, heading 180 or 0 degree

in the waveform processing used by these studies are signal deconvolution and decomposition. The deconvolution is an algorithm-based process that is used to reverse the effect of convolution on the recorded signals, and the decomposition is a process which can provide estimates of the location and properties of objects along the pulse (Wagner et al., 2006). In this study, we employed two distinct methods to process the waveform data. The first method was direct decomposition (I), which only applied the Gaussian decomposition to the waveform data. The second method was the deconvolution and decomposition method. For the second method, we utilized both new and classical deconvolution algorithms, the Gold and RL algorithms, to explore the advantages and limitations of deconvolution algorithms. The results from the above deconvolution were then subjected to Gaussian decomposition and we refer to them as the Gold approach (II) and RL approach (III) in this study. The RL algorithm was selected as a reference deconvolution algorithm because it is a superior widely used deconvolution algorithm (Harsdorf and Reuter, 2000; Nordin, 2006; Wu et al., 2011). Before performing analysis, we converted all delivered waveform data into ASCII format and then pre-processed them with steps, such as de-noising and mean filtering.

2.3.1. Parameter optimization for deconvolution

The returning pulses of the waveform were the product of interaction among outgoing pulses, atmospheric scattering, system noise and reflecting surfaces (Briese et al., 2008). For the NEON waveform LiDAR data, the backscatter responses can be expressed

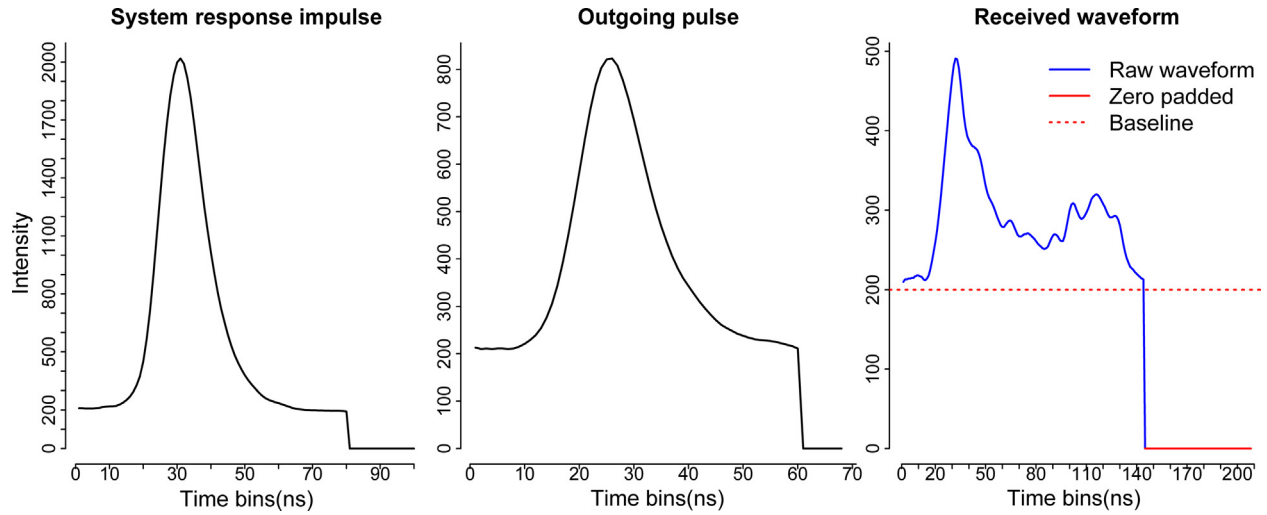

Fig. 2. A subset of the system response impulse, a sample of outgoing pulse and corresponding raw waveform recorded by NEON's full waveform LiDAR system.

as the convolution of the outgoing pulse, impulse response (atmospheric scattering, system noise, etc.) and effective target cross section (Eq. (1)). The deconvolution approach can remove the effect of the outgoing pulse and system impulse response and then improve the separability (Neuenschwander, 2008) of close peaks and reveal the true distribution of the scattering substances ($\delta_i(t)$) along the optical path (Cawse-Nicholson et al., 2014; Wagner et al., 2006; Wu et al., 2011).

$$P_r(t) = \sum_{i=1}^{n} \frac{D^2}{4\pi\gamma^2 R_i^4} P_t(t) * \tau(t) * \delta_i(t) \quad (1)$$

where $P_r(t)$ is the received laser power, $P_t(t)$ is the outgoing pulse, τ (t) is the receiver impulse function, D is the aperture diameter of the receiver optics, λ is the wavelength, R is the range from the LiDAR system to the target, $\delta_i(t)$ is effective target cross section and n is the number of targets detected along the pulse line.

The deconvolution is sensitive to the pre-processing of input data and the choice of parameters for deconvolution. For our optimization of processing parameters, we randomly selected 2000 waveforms of each dataset and processed this subset in two different ways with all other parameters as constant: one method is to keep all input data (the outgoing pulses, impulse response and return pulses) as raw values as supplied by the data provider; another method is to normalize the input data by subtracting the minimum non-zero value of each dataset. Through comparing the peaks' location of deconvolved waveforms with raw waveforms, we determined which method was employed in the subsequent analysis.

The parameters of the deconvolution function used in this study include boost, iterations and repetitions. The experiment demonstrated that boost was not as sensitive as the other two parameters and its recommended range was 1–2. Detailed information of these parameters can refer to the R package Peaks (Morhac, 2012). In our case, 1.5 was selected as a constant in the subsequent analysis. The number of iterations and repetitions in estimated impulse response and deconvolution algorithm are critical to the performance of the deconvolution and decomposition method. As such, our second step was to optimize these parameters for deconvolution. This step could be highly subjective and case-dependent for different datasets. To avoid subjectivity and a trial-and-error approach, we developed a general rule to find the reasonable range of these parameters. The total number of echoes generated from these 2000 waveforms and the percentage of matched waveforms were selected as criteria to narrow down the parameter ranges. The matched waveforms were defined as the deconvolved waveform with the similar peak locations (the difference within 3 ns) compared to the peak locations of raw waveforms after using the mean filter.

The preliminary experiment was conducted on about 2300 combinations of these parameters, and then we selected the potential parameter combinations based on the criteria above. In our case, we eventually narrowed down the estimate impulse response's iteration and repetition to 15–30 and 2–4, respectively. For the iteration and repetition in the deconvolution algorithms, the range for them was 30–55 and 3–5, respectively.

In this study, the Gold algorithm (Morhac et al., 1997) and RL algorithm were employed to deconvolve the raw waveform data, and then we compared these two algorithms in terms of detection of peaks, false detection rate, accuracy assessment, and parameter uncertainty analysis. The following two sections provide the principles of these two algorithms.

2.3.1.1. Richardson–Lucy (RL) algorithm. The RL algorithm was developed from the Bayesian's theorem which could reconstruct noisy images by taking into account statistical fluctuations in the signal (Fish et al., 1995). It was originally developed for recovering the image by searching iteratively for solutions to deconvolution problems. The basic idea is to calculate the most likely value $f_t(x)$ given the observed d(x) and the known point-spread function g (x). One waveform LiDAR profile can be seen as an image with the dimension $1 \times N$ and t^{th} iteration solution written as follows in terms of convolution (Fish et al., 1995):

$$f_{t+1}(x) = f_t(x) \cdot \left(\frac{d(x)}{f_t(x) * g(x)} * g(-x) \right) \quad (2)$$

where $*$ is the convolution operation, d(x) is the observed value at location x, $f_t(x)$ is the most likely value at location x and g(x) is the known point spread function, f(x) can be solved by iterating Eq. (2) until convergence.

2.3.1.2. Gold algorithm. The Gold algorithm is a non-oscillating and stable deconvolution method that can give us non-negative solutions (Morháč et al., 2003). This vital property is suitable for the waveform processing, since it is unreasonable and senseless if negative solutions appear. The Gold algorithm has been successfully applied to deconvolve the γ-ray spectra (Morhac et al., 1997) and

nuclear data (Morháč et al., 2003). For discrete values, it searched iteratively to solve the deconvolution problem using Eq. (3):

$$y(i) = \sum_{k=0}^{n-1} h(i-k)x(k) \quad i = 0, 1, 2, \ldots, m \tag{3}$$

where x, y are input and output vector, h(i) is the impulse response or outgoing impulse function, n is the number of samples of vector h, i is i^{th} sample point and x(k) represent the k^{th} waveform's differential backscatter cross-section. Here, the convolution system is one dimension. After a matrix transformation, the Gold algorithm can be expressed as Eq. (4) (Morhac et al., 1997):

$$x^{(k)}(i) = \frac{x^{(k-1)}(i)}{\sum_{j=1}^{n} h(i-j)x^{(k-1)}(j)} x^{(k-1)}(i) \tag{4}$$

For both algorithm applications, the impulse response and outgoing pulses must be known first. The NEON waveform LiDAR data provided each waveform with the corresponding outgoing pulse. In this study, three major steps were utilized to obtain the effective target cross section ($\delta_i(t)$): (1) the system response was used to deconvolve the corresponding outgoing waveform for each dataset to derive the estimated impulse response, (2) using the return waveform to deconvolve the outgoing pulse to get the immediate waveform, (3) the immediate waveform was employed to deconvolve the estimated impulse response to reveal the effective target cross section. After deconvolution, the results from the above were also decomposed using a mixture of Gaussian function. The major steps were the same as described in Section 2.3.2.

2.3.2. Gaussian decomposition

Since the outgoing laser pulse of the NEON data is nearly Gaussian in shape, as shown in Fig. 2, the returned waveform can be fitted by a mixture of Gaussian function (Wagner et al., 2006). The Gaussian components characterize the different targets when the laser beam interacts with objects along the path like vegetation and ground (Harding, 2005). The analytical expression of the Gaussian function (f(x)) can be written as:

$$f(x) = \sum_{i=1}^{n} A_i \exp\left(-\frac{(x-u_i)^2}{2\delta_i^2}\right) \tag{5}$$

where n is number of Gaussian components, A_i is the amplitude of peak at i^{th} waveform component, δ_i is the standard deviation of i^{th} waveform component, and u_i is the time location of peak at i^{th} waveform component.

In this study, a mean filter was first performed on each individual waveform to remove the noise and then we normalized these waveforms through subtracting the minimum value of each waveform. The return pulses were fitted with a mixture of Gaussian functions using a NLS method and optimized using the LM algorithm, which was implemented in the R package minpack.lm (Elzhov et al., 2013). Detailed steps can refer to the study of Chauve et al., 2007. The difficult part of using NLS to fit the Gaussian function was to determine the initial values of parameters. For the direct decomposition, we estimated the number of Gaussian components (n) by finding the number of peaks of raw waveform data and then used amplitude threshold to delete the "fake" peak (s). The initial amplitude (A_i) was assumed to be 2/3 of the corresponding peak value, initial echo width (δ_i) of each component was estimated as the half-widths of consecutive peaks and initial peak locations (u_i) were derived from corresponding raw waveforms' peak locations. For the deconvolution and decomposition, the initial n, u_i and A_i were derived from corresponding deconvolved waveform with the same step as the direct decomposition. However, the initial δ_i for each deconvolved waveform was estimated as half of the difference between peak location and the closest time bin with negative lagged difference.

2.3.3. Geolocation extraction

Through deconvolving the waveforms and fitting echoes to a mixture of Gaussian function with the LM algorithm, the time of peak positions, intensity and width were obtained. The 3D point cloud was generated based on the time of peak positions, the location of the first time bin of the return pulse and position change of pulse per nanosecond (dx, dy, dz).

For the direct decomposition method, the leading edge position of each waveform was used to calculate the geolocation of any time bin in a return waveform by incorporating the full width at half maximum (FWHM). The waveform is fitted with Gaussian function, so FWHM can be obtained through the standard deviation (σ). Therefore, the leading edge position can be calculated by Eqs. (6) and (7).

$$FWHM = 2\sqrt{2\ln 2}\sigma \tag{6}$$

$$t_l = t_p - 0.5 * \text{ FWHM} \tag{7}$$

where t_l is the leading edge position for each echo, t_p is the time of peak position for each echo and σ is the standard deviation of individually fitted function. t_p and σ can be obtained from Gaussian decomposition. Then the new geolocation of any time bin for given waveform can be calculated by Eq. (8).

$$\left.\begin{aligned} X &= (t_l - t_r) * dx + X_r \\ Y &= (t_l - t_r) * dy + Y_r \\ Z &= (t_l - t_r) * dz + Z_r \end{aligned}\right\} \tag{8}$$

where X, Y, Z is the new geolocation of peak, t_r is the first return reference bin location, dx, dy, dz are the position change for every ns, X_r, Y_r, Z_r are the Easting, Northing and height of the first return. t_r, dx, dy, dz, X_r, Y_r, Z_r are provided by the NEON geolocation dataset.

The new geolocation was determined by using the time of peak location (t_p) for the deconvolution and decomposition method, since the deconvolution can reveal the real geometry of objects. The NEON datasets provided us another geolocation dataset for deconvolution and decomposition method. The new geolocation is computed as:

$$\left.\begin{aligned} X &= [(t_p - t_r) - (t_{op} - t_{or})] * dx + X_r \\ Y &= [(t_p - t_r) - (t_{op} - t_{or})] * dy + Y_r \\ Z &= [(t_p - t_r) - (t_{op} - t_{or})] * dz + Z_r \end{aligned}\right\} \tag{9}$$

where t_{op} is the time of peak location for the each outgoing pulse and t_{or} is the time of corresponding reference bin location for the outgoing pulse. Both can be found in the NEON geolocation datasets.

2.3.4. Digital models extraction

The original point cloud derived from the decomposition step had some noisy points, since some raw waveforms were not exactly Gaussian shape. We filtered these points based on the intensity and height, which was achieved through the LAStools (Isenburg, 2012).

Digital terrain models (DTMs) and canopy height models (CHMs) were generated from these filtered 3D point cloud for each method using the LAStools (Isenburg, 2012). DTMs and CHMs were chosen mainly because sets of vegetation metrics were derived from CHMs, and any error of DTMs would propagate to affect the accuracy of CHMs. To derive a DTM from waveform LiDAR data, the point cloud has to be classified into ground and non-ground points. Generally, the intensity and width of the last echoes can be used as criteria to remove the non-ground laser points. The

intensity of the echoes can provide additional information about the reflectance properties of an object, such as judging whether the echoes came from below ground response (Chhatkuli et al., 2012). The echo width not only gives information on the range distribution of individual object that produce a single echo, but it also can assist to decide whether a pulse was reflected from solid ground or vegetation (Doneus et al., 2008; Ioannides et al., 2006). In this study, we employed the intensity and width of echoes to filter the non-ground points or noise. After exploring different thresholds of width and intensity of echoes, a width threshold of 20 was applied to remove noise or wrongly fitted echoes, and 1/10 of the corresponding maximum intensity of each waveform was selected as a threshold to remove the below ground response. These thresholds may not be universally valid, as they may vary by regions and types of data. Finally, the filtered points were imported into the LAStools to generate a refined DTM. To further evaluate the performance of the methods, the CHM was generated from the non-ground points based on the steps described by Khosravipour et al. (2014).

2.4. Algorithms' performance comparisons

2.4.1. Accuracy assessment

The discrete-return LiDAR data and its derived end products can be regarded as the ground truth data due to that they can potentially achieve better accuracy than direct field measurement (Chen, 2007). In this study, the discrete-return LiDAR data were adopted as reference data to verify whether more points can be extracted from the waveform LiDAR data. Additionally, we employed the reference DEMs and CHMs provided by the NEON to conduct accuracy assessment of our results generated from waveform LiDAR data from both visual and statistical perspectives. The SJER site was selected to show the visual comparison results for its complex topography. The results of the statistical comparisons were measured by mean difference, standard deviation (SD), root mean square error (RMSE) and percentage of spatial difference within 0.5, 1, 2 and >2 m of each study site.

2.4.2. Parameter uncertainty

The predictive parameter error estimates not only enable us to objectively quantify the expected quality of the results from available data but also allow us to estimate the rigorous error propagation through to the end products. Through the approaches we employed in this study, the standard error (se) of peak location for each estimated echo was obtained. Parameter uncertainty was represented by the 95% confidence interval (95% CI) of peak location. For each echo, the estimated peak location's confidence bounds were calculated using Eqs. (10) and (11):

$$t_{ll} = t_l - 1.96 * se \tag{10}$$

$$t_{lu} = t_l + 1.96 * se \tag{11}$$

The above equations were for the direct decomposition approach, while t_l became t_p for the Gold approach and RL approach. The above biased peak locations (t_{ll}, t_{lu}) would form two datasets for each approach: the 95% lower confidence level dataset (Lower dataset) and 95% upper confidence level dataset (Upper dataset). Once six DTMs and six CHMs for these three approaches were generated using these uncertainty datasets, we compared them with the DTMs and CHMs derived from the original decomposition to get the biases and quantitatively assess the approaches' robustness. The parameter uncertainty for each method was calculated as follows:

$$LU = EGE_L - EGE \tag{12}$$

$$UU = EGE_U - EGE \tag{13}$$

where EGE was the estimated ground elevation generated from dataset of peak locations, EGE_L was the estimated ground elevation generated from Lower dataset of peaks locations, EGE_U was the estimated ground elevation generated from Upper dataset of peak locations, LU was the lower uncertainty and UU was the upper uncertainty. The uncertainty of maximum CHM was calculated in the same way. The visual comparisons of parameter uncertainty were also conducted using the site which had the largest uncertainty based on the statistical results.

In order to quantify the effect of factors such as slope and vegetation height on the parameter uncertainty, the uncertainty was divided into three levels: high (>2.00 m and <−2.00 m), medium (−2.00 to −0.51 m and 0.51–2.00 m) and low (−0.5 to 0.5 m). The slope and vegetation height for each corresponding level were also extracted. The Upper dataset and Lower dataset of SJER site were combined for each approach. The Analysis of variance (ANOVA) was used to analyze the effect of factors like slope and vegetation on uncertainty levels, and identify which region may be more likely to have higher uncertainty. Box plots were chosen to visualize the uncertainty's statistical results. An overview of the whole waveform processing and comparisons procedure is given in Fig. 3.

3. Results and discussion

3.1. Deconvolution parameters optimization

Since deconvolution was sensitive to the input data, we first explored different pre-processing steps of the input data. The three sample results of deconvolution by the RL and Gold algorithms were plotted in Fig. 4, with different pre-processing steps against corresponding original waveforms after noise deletion.

Fig. 4 shows that the peaks of the waveforms with deconvolution were much easier identified and that the shape of waveform components was closer to the Gaussian distribution. This verified our assumption that the waveform can be simulated by a mixture of the Gaussian components after removing the system impulse's contribution. Also, the intensity of the waveforms was much higher and the width of the waveform was narrower than the raw waveform, which could be conducive to precisely revealing the geometry of objects along the pulse.

It was evident that input data with different preprocessing steps could impact the deconvolution results. Compared with the original waveform, the results using the raw data were more likely to detect a wrong peak (green[1] circle) as shown in Fig. 4(b). Therefore, we adopted adjusted data to do the subsequent steps in this study, which was consistent with the same pre-processing steps described by Wu et al. (2011). However, when using the adjusted data, we observed a downward shift in the time bin axis for both deconvolution algorithms. This was also found in the study of Cawse-Nicholson et al. (2014). The main reason may be that the system impulse was not acquired under the ideal condition, which led to the inaccurate deconvolved impulse response. However, this kind of inaccuracy of the deconvolution could be improved by obtaining more accurate system impulse.

In Fig. 4(a), the original waveform had noise in both edges that made the direct decomposition very difficult when we had no knowledge about the number of the reflecting objects along the pulse line. But, after removing the system response, the number and position of the peaks were evident. As shown in Fig. 4 (b) and (c), the local peaks could also be clearly distinguished after

[1] For interpretation of color in Figs. 4, 5, and 12, the reader is referred to the web version of this article.

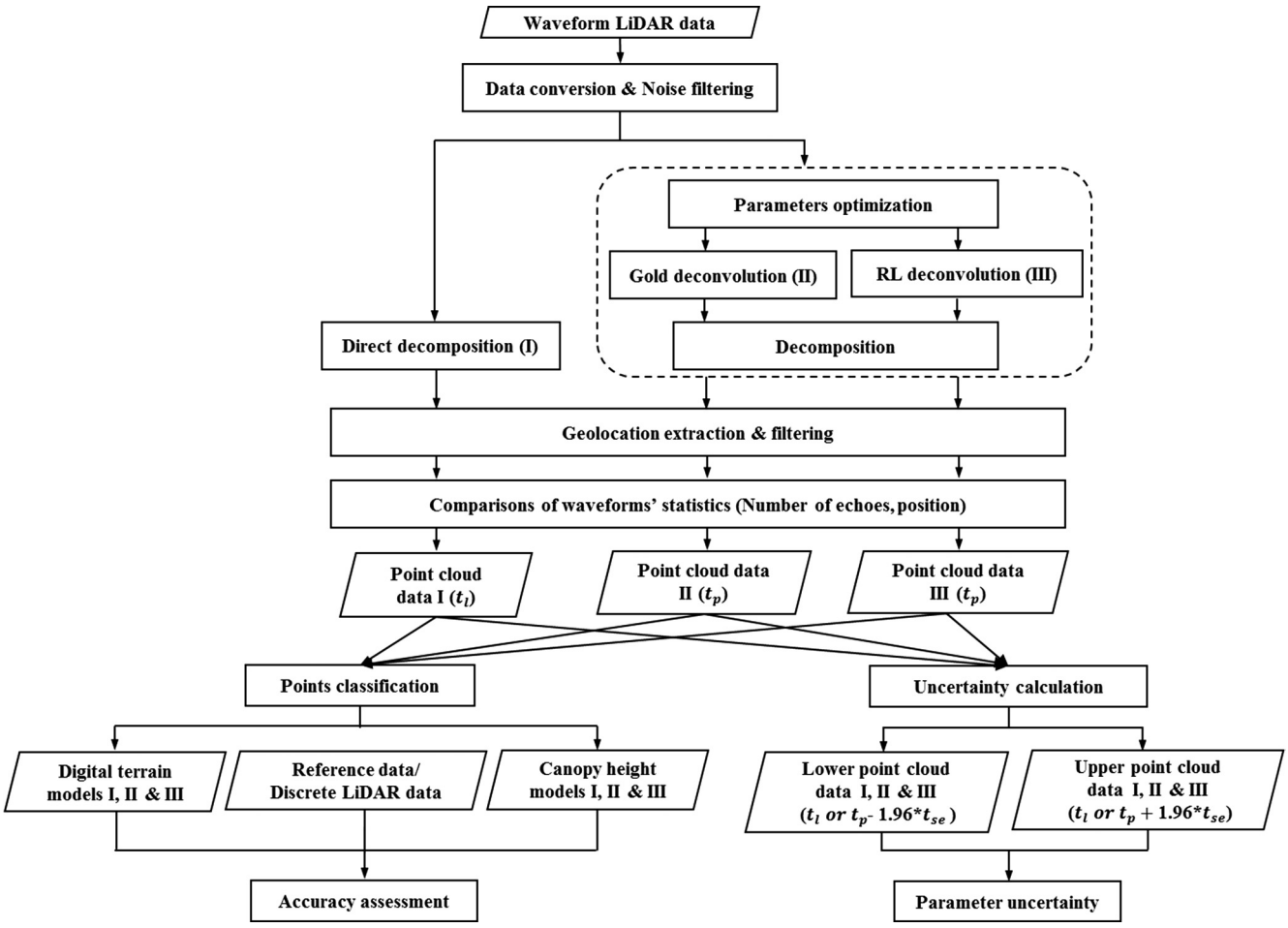

Fig. 3. Flowchart for waveform LiDAR data processing and comparisons.

deconvolution for more complicated waveforms. However, it should be noted that the deconvolution cannot break down pulses from surfaces that are too close which could result in one larger pulse as shown in Fig. 4(b). This kind of peaks overlaying could be solved by increasing the number of iterations and repetitions in the deconvolution algorithm, but this may cause additional minor peaks that are adjacent to the major local peak as described in the study of Wu et al. (2011). Therefore, to set up the optimal number of iteration was critical to the detection of peaks. In this study, exploratory analysis helped us narrow down the range of parameters and final optimal combinations of iteration and repetition were (30, 4), (35, 5), and (40, 5) for the HF, SJER and OSBS sites, respectively.

3.2. Number of echoes

To provide a more comprehensive comparison, the quantitative results for the three study sites are shown in Tables 2 and 3, respectively. For the deconvolution and decomposition method, the performance of the Gold algorithm was better than the RL algorithm from the perspective of the number of echoes detected. Additionally, the Gold algorithm mostly detected higher number of echoes with lower false detection rate. For instance, more waveforms were decomposed into three, four or five echoes for individual waveform using the Gold algorithm than the RL algorithm for all study sites. This indicated that the Gold algorithm had a higher potential to detect the hidden peaks than the RL algorithm for complex waveforms. The direct decomposition's performance was similar to the RL algorithm in terms of the number of echoes detected. The direct decomposition was more capable of detecting a higher number of echoes than the RL algorithm for those study sites when we compared the number of waveforms with different echoes. However, the likelihood of detecting the false echoes was increased with the direct decomposition method compared with the deconvolution and decomposition method (Tables 2 and 3).

Interestingly, the false echoes rate of the SJER site was lower than the HF and OSBS sites for all approaches (Tables 2 and 3). Comparing the three sites' decomposition results with their corresponding discrete-return LiDAR data demonstrated that the false echo detection rate of the SJER site was highest. The possible reason is that the SJER site was comprised of dense vegetation that could result in weak returns and overlapping echoes of the reflected waveforms, and potentially higher false echo detection rate was expected (Chauve et al., 2009). Another interesting finding was that the deconvolution and decomposition method can reduce the false echoes rate significantly as shown in Tables 2 and 3.

Hence, based on the visual inspection and quantitative comparisons of different methods, we concluded that the deconvolution and decomposition method outperformed the direct decomposition method. The Gold algorithm was superior to the RL algorithm based on the number of echoes and false echo detection rate. We also found that pre-processing of data significantly affected the echo detections; assigning zero-padded values of the return pulse to NA and utilizing the adjusted input data (the outgoing pulse, system response pulse and return pulse) could achieve better and more accurate results.

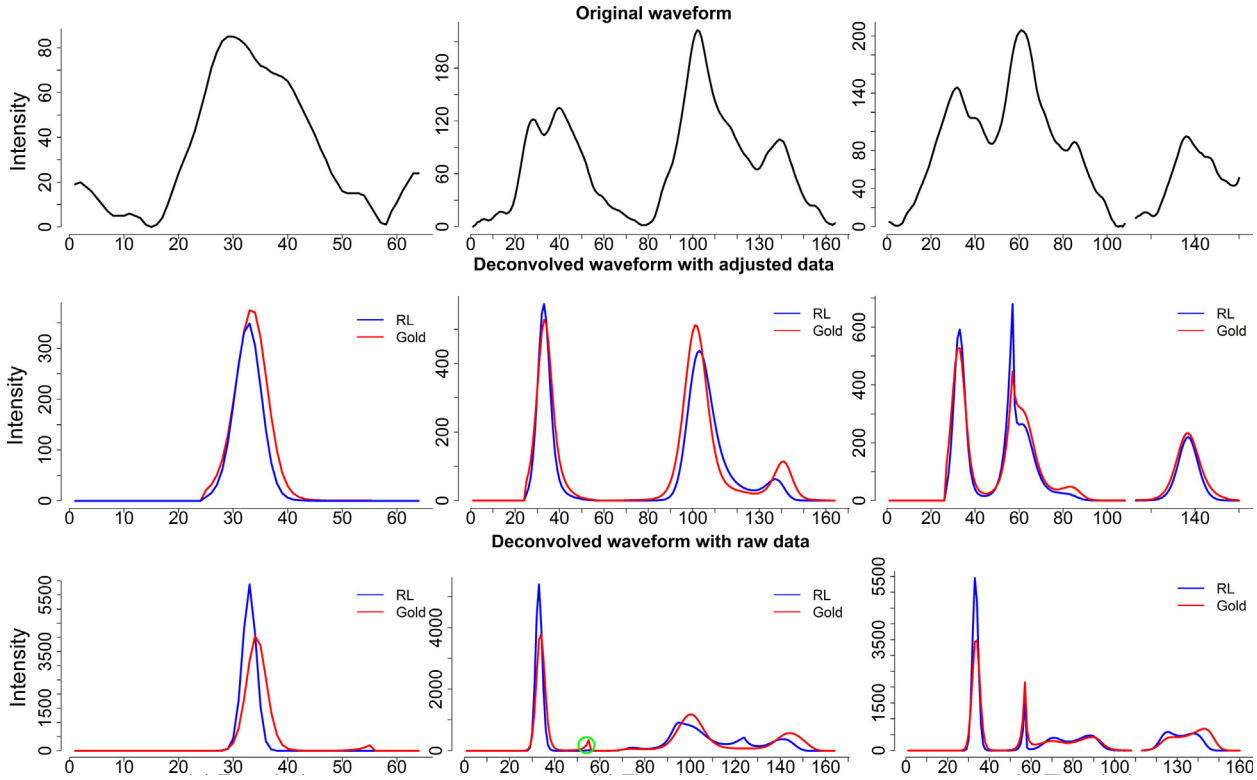

Fig. 4. Three samples of original waveforms (a, b and c) versus deconvolved waveforms by the RL algorithm and Gold algorithm with different pre-processing steps: *Deconvolved waveform with adjusted data* where the input data (the outgoing pulse, impulse response and return pulse) of deconvolution were normalized by subtracting minimum non-zero values; *Deconvolved waveform with raw data* where the input data (the outgoing pulse, impulse response and return pulse) of deconvolution were raw pulse values, excluding the zero padded values. The blank section in the original waveform of Fig. 4(c) resulted from unrecorded values in the raw data.

Table 2
Number of echoes estimated by the direct decomposition method and deconvolution and decomposition method with different input data for HF site.

Methods Number of echoes	Direct decomposition	Adjusted data			Raw data	
		Deconvolution + Decomposition			Deconvolution + Decomposition	
		RL algorithm (NA)[a]	Gold algorithm (NA)[a]	Gold algorithm (0)[b]	RL algorithm (NA)[a]	Gold algorithm (NA)[a]
	Number of waveforms					
1	5533	6057	4261	3160	8438	7454
2	4313	3088	2783	2990	3164	4024
3	2398	3493	5049	5626	1336	1438
4	685	322	765	1088	150	169
5	133	116	221	214	4	7
6	23	13	13	14	0	0
7	7	3	0	0	0	0
False echoes	2508 (10.05%)	370 (1.50%)	207 (0.71%)	296	56	102
Total echoes	24,945	24,679	29,217	31,524	19,394	20,527
Effective echoes	22,437	24,309	29,010	31,228	19,348	20,425

[a] The intensity with an original value of 0 assigned to NA after deconvolution.
[b] The intensity of each pulse maintained as 0 after deconvolution.

3.3. Position of echoes

To further demonstrate the performance of different approaches, we also explored the position of echoes. Fig. 5 shows that the three sample waveforms were decomposed by the three approaches: direct decomposition approach, RL approach and Gold approach. Here, the direct decomposition approach detected a higher number of echoes for each pulse than the other two deconvolution approaches, as shown in Fig. 5(a), (b) and (c) with 7, 6 and 3 Gaussian components, respectively. However, a closer examination revealed that the direct decomposition was likely to detect some unreasonable echoes. For example, the red Gaussian component in the direct decomposition of Fig. 5(b) is almost overlaid with the green component; the yellow component looks like a supplement to the other echoes and neither agrees well with the reality. Furthermore, the blue component in the direct decomposition of Fig. 5(c) is higher than the original waveform. These may explain why there was higher false echo detection rate for direct decomposition as shown in Tables 2 and 3.

It was evident that deconvolved waveforms performed better on decomposition with explicit Gaussian components in terms of visual comparisons. The performances of the RL and Gold approaches were similar in our example and almost had the same peak positions and shape, but the Gold approach worked better when the original waveform was composed of many peaks with noise. As shown in Fig. 5, the Gold approach could detect more

Table 3
Number of echoes estimated by the direct decomposition method and deconvolution and decomposition approach with different input data for SJER and OBOS sites.

Methods Number of echoes	SJER Adjusted data			OSBS Adjusted data		
	Direct decomposition	Deconvolution + Decomposition		Direct decomposition	Deconvolution + Decomposition	
		RL algorithm	Gold algorithm		RL algorithm	Gold algorithm
	Number of waveforms			Number of waveforms		
1	197,304	204,129	174,060	2,528,608	2,908,389	2,310,732
2	41,184	33,047	55,748	710,240	327,386	350,303
3	14,304	16,197	19,243	202,117	46,246	245,509
4	4392	4614	6862	41,085	4490	134,299
5	1166	620	1993	6537	306	67,185
6	267	59	551	866	18	19,544
7	42	1	171	81	2	4976
8	7	0	32	10	0	1861
9	1	0	6	0	0	632
10	0	0	0	0	0	160
11	0	0	1	0	0	28
12	0	0	0	0	0	3
False echoes	17,913 (5.15%)	378 (0.11%)	1223 (0.32%)	563,974 (11.85%)	14,709 (0.40%)	66,226 (1.38%)
Total echoes	34,7943	340,731	385,522	4,758,307	3,736,220	4,861,828
Effective echoes	330,030	340,731	384,299	4,194,333	3,721,511	4,795,602

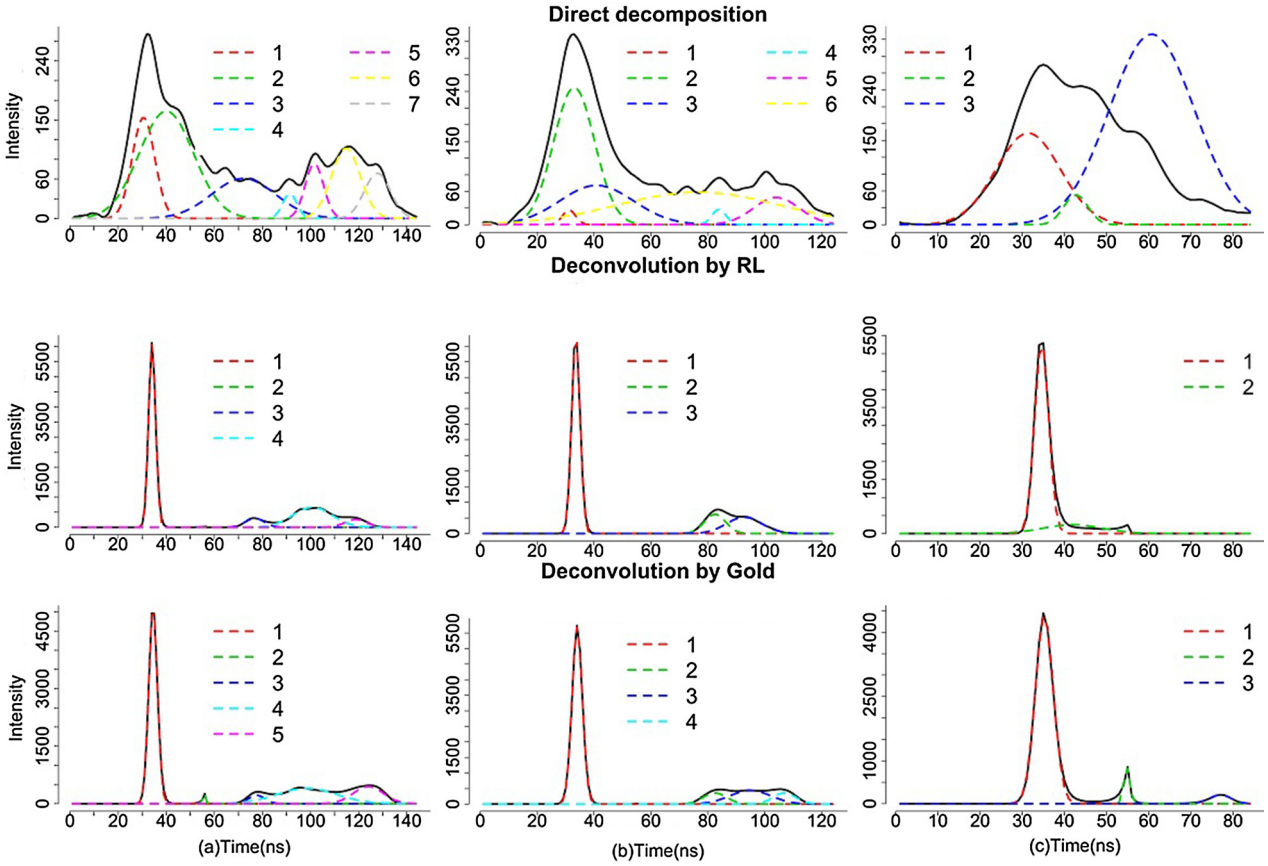

Fig. 5. Comparisons of the decomposition results with the direct decomposition approach, RL approach and Gold approach for three sample pulses (a, b, c). The solid black line is the original waveform. The colored dash lines are Gaussian components after decomposition.

echoes for the same pulse and may reconstruct more accurate cross sections of vegetation and terrain. The RL approach was less capable of detecting the last echoes that may represent the ground when we interpreted the waveform LiDAR data. However, it was worthy to note that the Gold approach may cause the ringing effect as shown in the second Gaussian component (green part) of the deconvolution by the Gold algorithm (Fig. 5(a)). This kind of minor peak around the major local peaks may be caused by a wavelike artifact that resulted from the sum of remaining low-frequency components after the loss of high-frequency components (Wu et al., 2011).

3.4. Point clouds

After geo-referencing, the point clouds generated from full waveform LiDAR data using different methods with number of points (n) are shown in Fig. 6. As the figure shows, the point clouds derived from waveform LiDAR data (Fig. 6(3), (4), (5)) are denser

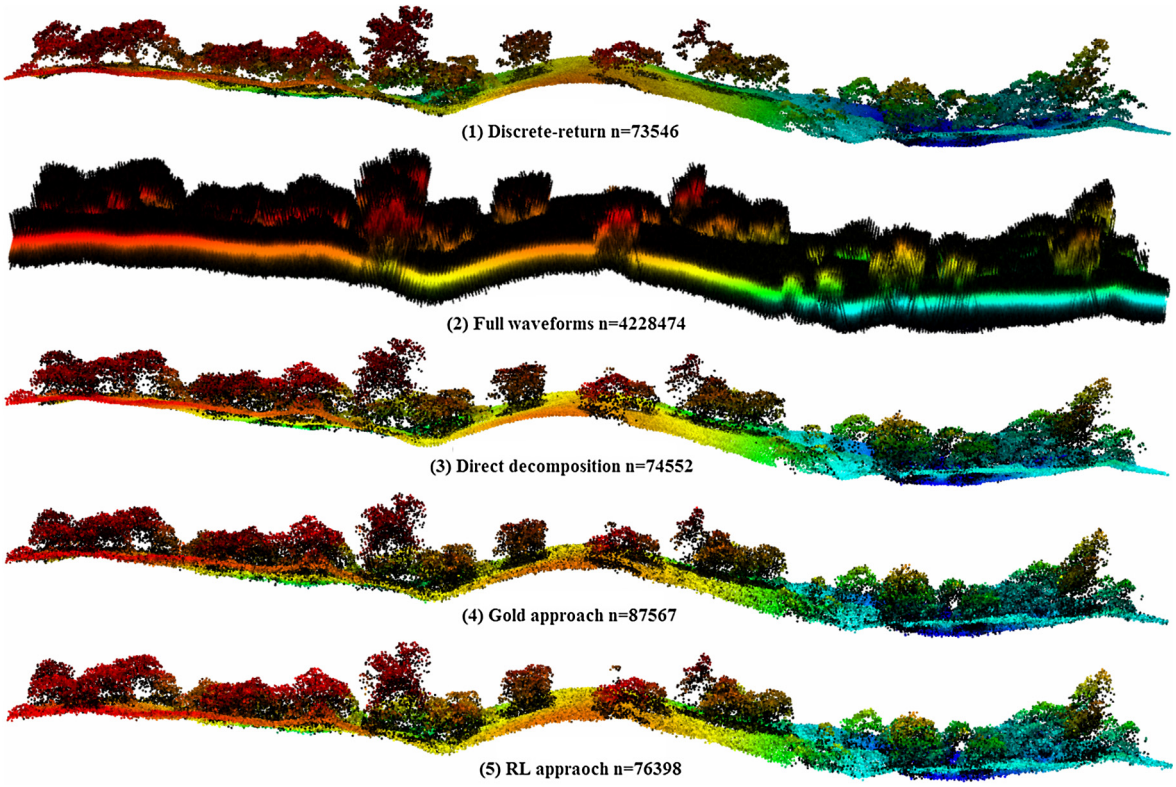

Fig. 6. Comparison of point clouds generated from discrete-return LiDAR data and waveform LiDAR data with different processing methods. (1) Point cloud derived from discrete-return LiDAR data. (2) Point cloud generated by geo-referencing every time bin of waveforms. (3) Point cloud derived from direct decomposition method. (4) Point cloud derived from Gold deconvolution and decomposition method (the Gold approach). (5) Point cloud derived from RL deconvolution and decomposition method (the RL approach).

than discrete-return LiDAR data (Fig. 6(1)), while they contained some noisy points in some vegetation parts. The point cloud generated by geo-referencing every time bin (Fig. 6(2)) could show a good shape of tree canopy with very high density points, but it was noisy with "false ground" and "false canopy height". The point cloud generated from the direct decomposition (Fig. 6(3)) was comparable to that used the deconvolution and decomposition (Fig. 6(4), (5)) results. However, the number of points generated from these methods of the same study extent revealed that the Gold approach is larger than the RL approach and the direct decomposition approach. Overall, these results provide important insights into the selection of waveform processing methods, as some of these approaches may provide more information on the three-dimensional vegetation structure.

3.5. Accuracy assessment

Qualitative and quantitative accuracy assessments for the derived end products (DEM and CHM) were conducted in terms of visual and statistical comparisons. The results from SJER were used as an example to demonstrate the visual comparison results.

3.5.1. Digital terrain model

Fig. 8 shows that the DTMs generated from waveform LiDAR data with different approaches in the SJER study site almost have identical elevation distribution in comparison to the reference DTM. The range of the DTM derived from waveform LiDAR was also consistent with the reference DTM. However, a lower elevation range was observed for all approaches with the direct decomposition approach ranging from 381.9 to 424.2 m, Gold approach from 381.3 to 423.2 m and RL approach from 381.3 to 423.0 m. It should be noted that there were some blank regions on the edges of DTMs, which may be due to no waveform or no information extracted from the sparse waveforms in those regions.

To further validate the performance of DTM results generated from different approaches, the statistical results for the DTMs are shown in Table 4. It was noticeable that for all study sites, DTMs derived from waveform LiDAR were lower than the reference DTM from the perspective of range, and the direct decomposition approach outperformed the other two approaches generally. More precisely, the direct decomposition and gold approaches had similar performances for the HF study area in terms of root mean square error (RMSE) and mean difference (MD). Unlike the results of the HF site, the SJER and OSBS sites' direct decomposition approach outperformed the Gold and RL approaches. Especially for the SJER site, larger variances and RMSEs (Table 4) were reported that probably resulted from the downward shift of detected peaks after we utilized the shifted estimated impulse response to deconvolve the waveforms. As shown in Fig. 7, the estimated impulse response of the SJER site (red) is shifted downward compared with the estimated impulse response of the HF site. Additionally, there is an extra peak at the end of the SJER site's impulse response which is an artifact, since the impulse response was reflected from flat ground that should have one peak. The deconvolution is sensitive to the input data such as impulse response and any shift of impulse response will propagate to the position accuracy of the detected peaks (Cawse-Nicholson et al., 2014). On the contrary, the impulse response for HF was more reasonable with only one significant peak. These may explain why the HF site's DTMs derived from the Gold approach and the RL approach were better than for the SJER site with the same approaches. However, through obtaining more accurate impulse response, the accuracy of the SJER DTM could be improved.

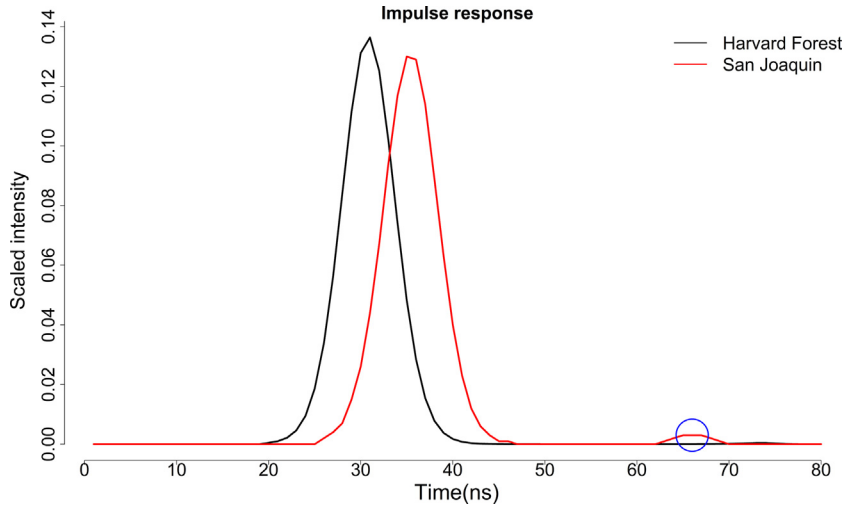

Fig. 7. The comparisons of the impulse response used for deconvolution provided by the NEON datasets.

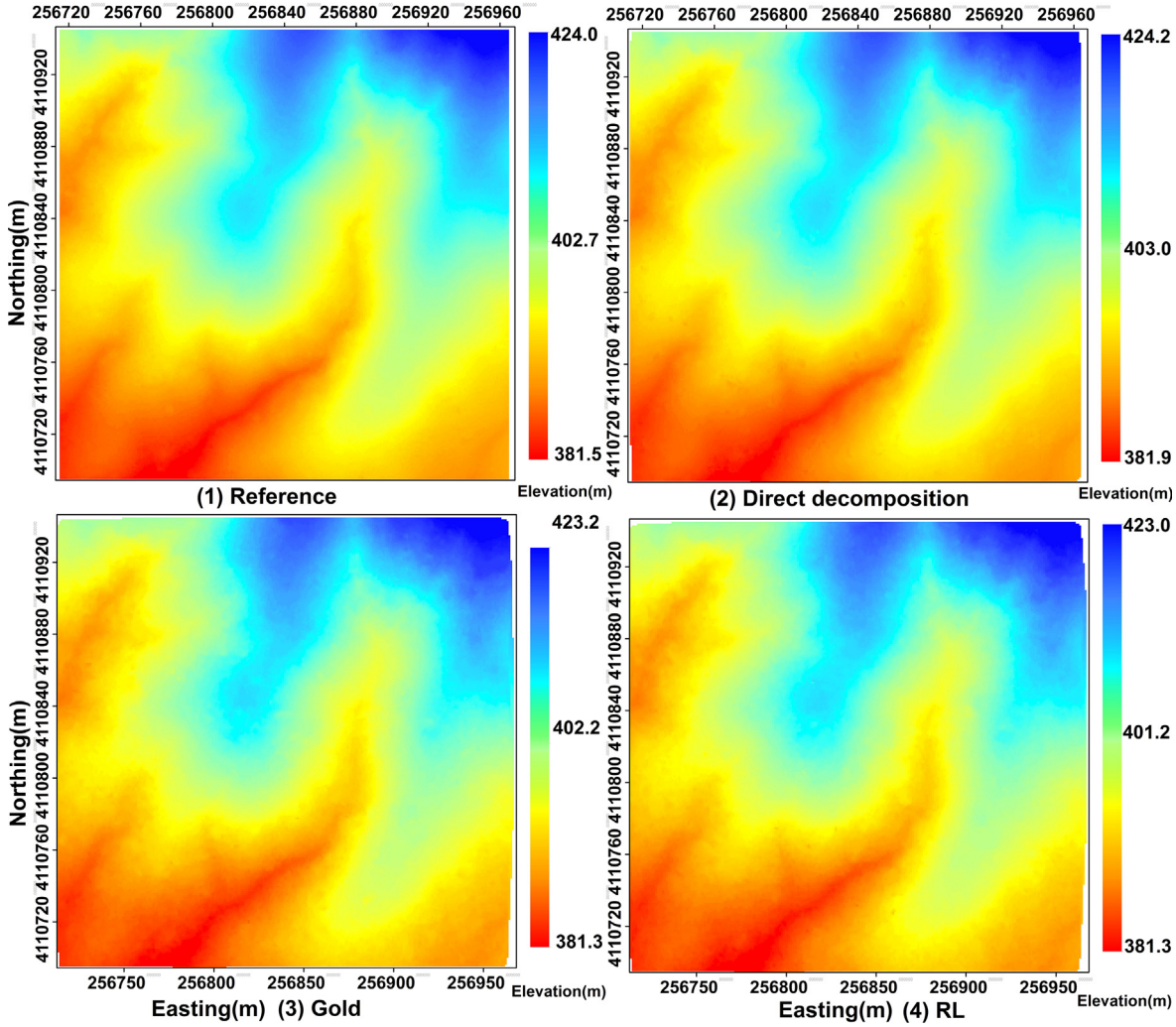

Fig. 8. Comparisons of (1) Reference DTM to waveform-based DTM generated from (2) the direct decomposition, (3) the Gold deconvolution and (4) the RL deconvolution approaches for the SJER site.

Assuredly, the topography of the study areas, resolution of the DTM, and ground points' classification method may also contribute to larger variance of the results.

The percentage of the spatial elevation difference between DTMs derived from waveform LiDAR and reference data for the three study sites was also analyzed as shown in Table 4. Almost

Table 4
Summary statistics of DTMs (1 m resolution) generated from the three approaches (Direct decomposition, Gold algorithm and RL algorithm) for the HF, SJER and OSBS sites.

Approaches	Range (m)	SD (m)	MD (m)	RMSE (m)	PW0.5 (%)	PW1 (%)	PW2 (%)	PW3 (%)
HF								
Reference	313.5–318.0							
Discrete	313.9–318.1	0.18	−0.28	0.24				
Direct	313.7–317.9	0.23	−0.35	0.42	75.36	99.91	100	100
Gold	313.6–317.7	0.25	−0.35	0.31	68.82	97.68	100	100
RL	313.9–318.3	0.70	−0.21	0.70	60.43	80.45	93.40	100
SJER								
Reference	381.5–424.0							
Discrete	382.1–424.5	0.15	−0.12	0.15				
Direct	381.9–424.2	0.25	−0.24	0.26	93.87	99.70	100	100
Gold	381.3–423.2	0.30	−1.04	1.15	3.53	25.65	99.20	100
RL	381.3–423.0	0.31	−1.16	1.26	2.23	18.80	99.50	100
OSBS								
Reference	21.0–46.9							
Discrete	21.7–47.6	0.12	−0.08	0.15				
Direct	21.0–47.4	−0.31	0.35	0.36	94.10	99.78	100	100
Gold	20.6–47.0	0.40	−0.14	0.42	60.56	94.56	99.88	100
RL	20.5–46.6	0.35	−0.54	0.62	49.68	92.47	99.94	100

Reference: Reference DTM; Discrete: Discrete-return LiDAR derived DTM; Direct: Direct decomposition approach; Gold: Gold approach; RL: RL approach. SD: standard deviation; MD: Mean elevation difference between DTM derived from waveform LiDAR and reference data; PW0.5: the percentage of difference within 0.5 m (−0.5 to 0.5); PW1: the percentage of difference within 1 m (−1 to 1); PW2: the percentage of difference within 1 m (−2 to 2); PW3: the percentage of difference beyond 2 m (>2.0 and <−2.0).

all elevation differences (>94%) were within 0.5 m when we used the direct decomposition approach for the SJER and OSBS study sites. The performance of direct decomposition, Gold and RL approaches significantly varied on the SJER and OSBS sites when only the elevation difference within 0.5 (PW0.5) or 1 m (PW1) was considered. However, the elevation difference within 2 m (PW2) for this site was almost the same for these three approaches. As we explained above, the estimated impulse response may be the main reason for the lower performance of this site. For the deconvolution and decomposition method, the Gold approach had a better performance than the RL approach with a higher percentage of spatial difference located in the ranges from −0.5 to 0.5 m and −1 to 1 m for all study sites.

3.5.2. Canopy height model

Fig. 9 displays the CHMs generated from the direct decomposition, Gold and RL approaches for the SJER site. All approaches yielded similar results to our reference data in terms of the range and spatial distribution. It was worthwhile to note that the values in some regions of the reference data were zero (dark red), while the canopy height of the same region derived from waveform data was slightly higher as marked with ellipse shape. As noted from the regions marked by the circle, the Gold and RL approaches' results give more detail information about low vegetation than the direct decomposition approach.

These differences may demonstrate that the waveform LiDAR data was more capable of detecting the low vegetation than the discrete-return LiDAR data, and the deconvolution and decomposition method had higher potential to detect low vegetation. The findings may provide insights into detecting understory layers below the forest canopy or grassland vegetation by using waveform LiDAR data. The range shift was not observed for CHMs using the Gold and RL approaches as noted for the DTMs, because the CHM was obtained by subtracting the DTM from the digital surface model (DSM) and the time shift was offset after subtraction.

In addition to the visual comparisons, the quantitative comparisons between waveform-based CHMs with different approaches and reference CHM data are shown in Table 5. The Gold approach had the best performance with smallest standard deviation and RMSE for all sites. The result of the RL approach at the OSBS site had the smallest height difference but larger RMSE than direct decomposition approach, which may be primarily caused by the wider range of the height difference. However, it was still comparable to the result of direct decomposition approach. These observations may further indicate that the three approaches were reliable to extract the vegetation structure from the waveform LiDAR data and the Gold approach outperformed the other two approaches.

Analysis of percentage of the region's difference between maximum CHM derived from waveform LiDAR and the reference data further confirmed this conclusion (Table 5). From the perspective of the percentage of height difference within 2 m, about 96% of regions in HF site and 90% of region in other two sites, the performance of these three approaches were satisfactory. The majority of height differences >2 m occurred at the boundary of trees and ground. The reason behind this was that the small change of peak (t) not only lead to the slight change of height (z), but also resulted in the synchronized change of XY locations, as implied in Eqs. (8) and (9). When compared with the reference data, the boundary of trees and ground most likely yielded larger height differences due to this kind of XY location shift.

However, the performance of the three approaches varied at different sites in terms of the height difference within 0.5 and 1 m. The Gold approach worked best at the HF and SJER sites, while it did not work as well as the direct decomposition in the OSBS site.

Globally, the CHMs generated from the waveform LiDAR data using the three approaches were satisfactory compared to the reference data for these three study sites. The Gold approach worked slightly better with smaller standard deviation and RMSE for all sites. However, the direct decomposition approach outperformed the Gold approach with higher percentage of area located in the given spatial difference range, especially when the spatial difference range was within 0.5 and 1 m at OSBS site. This may be because the HF study area was flatter than SJER study area and potentially proved that Gold approach could work well in regions with different topography.

3.6. Parameter uncertainty

3.6.1. Digital terrain model

Fig. 10 shows that the DTMs' spatial uncertainty of the SJER site for the three approaches mostly ranges from −1 m to 1 m, and the

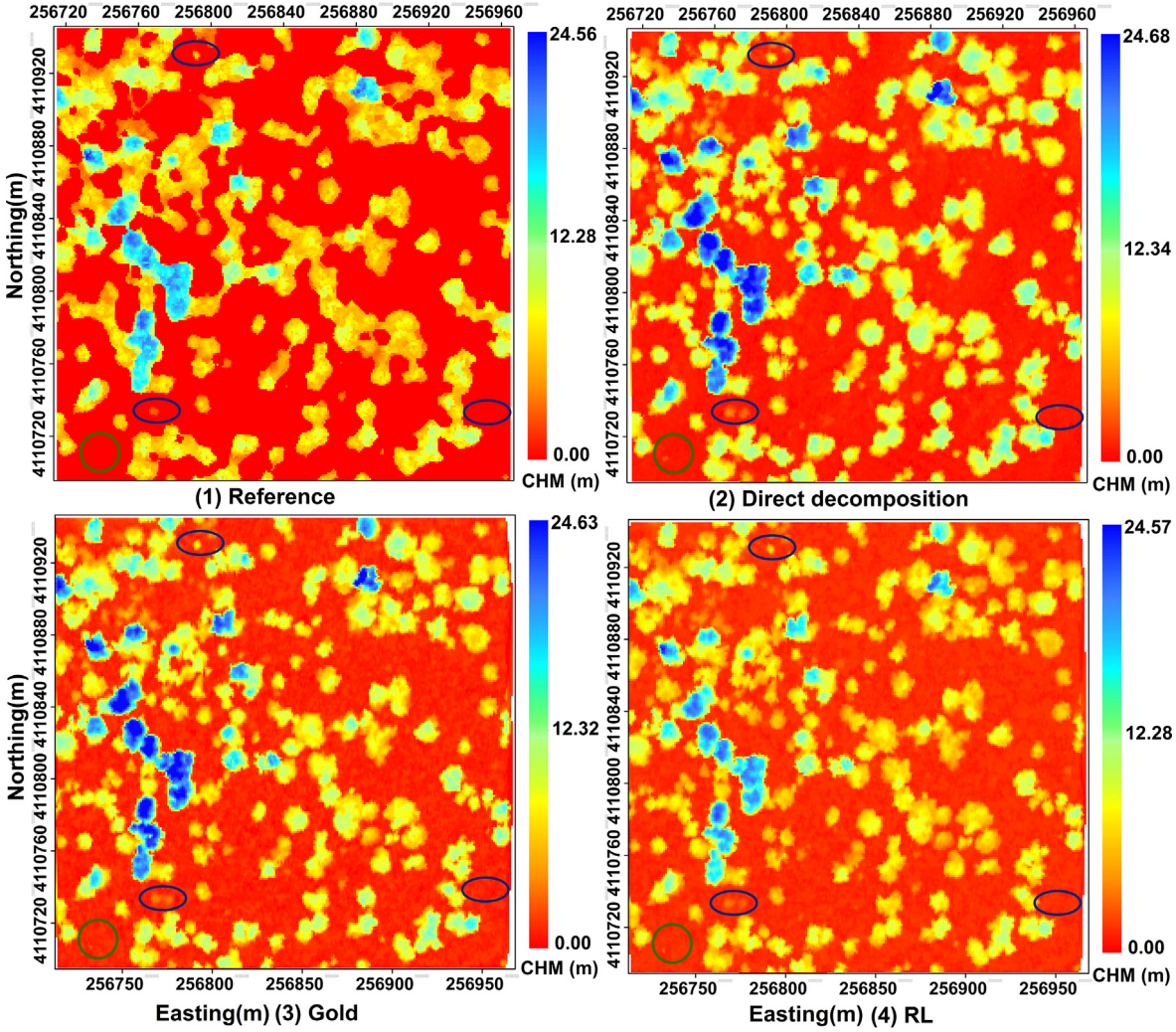

Fig. 9. Comparisons of (1) Reference CHM to waveform-based CHM generated from (2) the direct decomposition approach, (3) the Gold approach (4) and the RL approach for the SJER site.

Table 5
Summary of comparison of CHMs (resolution 1 m) generated from the three approaches (Direct decomposition, Gold approach and RL approach) for the HF, SJER and OSBS sites.

Approaches	Range (m)	SD (m)	MD (m)	RMSE (m)	PW0.5 (%)	PW1 (%)	PW2 (%)	PW3 (%)
HF								
Reference	9.10–23.06							
Discrete	9.19–23.06	0.42	0.25	0.51				
Direct	9.19–23.40	0.70	0.65	0.95	50.61	80.95	96.39	100
Gold	10.83–22.91	0.45	0.38	0.72	65.53	88.28	98.37	100
RL	9.95–23.67	0.98	0.28	1.02	42.45	69.31	95.28	100
SJER								
Reference	0.00–24.56							
Discrete	0.00–24.49	0.32	0.15	0.35				
Direct	0.00–24.68	1.40	0.50	1.65	60.87	75.08	88.03	100
Gold	0.00–25.12	1.28	−0.12	1.06	72.65	81.82	89.62	100
RL	0.00–24.57	1.39	0.35	1.51	34.56	80.34	88.89	100
OSBS								
Reference	0.00–27.35							
Discrete	0.00–27.77	0.28	0.11	0.31				
Direct	0.00–27.58	1.67	0.75	1.40	68.25	83.48	92.56	100
Gold	0.00–27.81	1.54	0.30	1.62	48.59	79.68	88.25	100
RL	0.00–28.50	1.93	0.13	1.87	38.25	68.37	86.87	100

Reference: Reference CHM; Discrete: Discrete-return LiDAR derived CHM; Direct: Direct decomposition approach; Gold: Gold approach; RL: RL approach. SD: standard deviation; MD: Mean height difference between CHM derived from waveform LiDAR and reference data; PW0.5: the percentage of difference within 0.5 m (−0.5 to 0.5); PW1: the percentage of difference within 1 m (−1.0 to 1.0); PW2: the percentage of difference within 1 m (−2.0 to 2.0); PW3: the percentage of difference beyond 2 m (>2.0 and <−2.0).

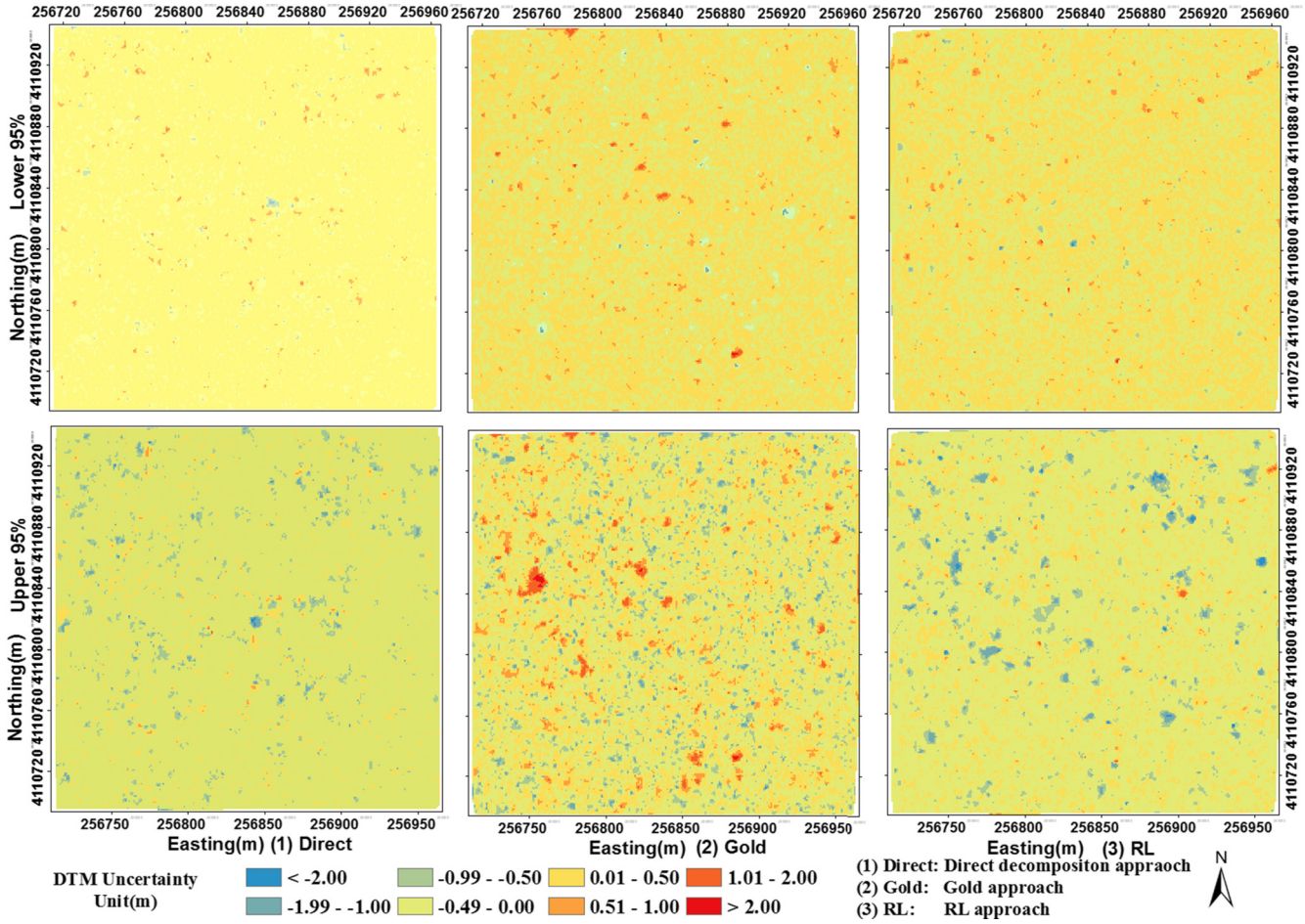

Fig. 10. The spatial uncertainty of the DTM caused by the parameter uncertainty in the SJER site using the direct decomposition approach (left), Gold approach (middle) and RL (right) approach, respectively. The above was the result from the corresponding Lower dataset and the bottom was the result from the corresponding Upper dataset.

DTMs derived from the Lower datasets are smoother than the Upper datasets. More specially, the direct decomposition approach had the smallest variation of spatial uncertainty with >86% of the prediction error within 0.50 m for the Lower dataset, and 89% of the prediction error was located in the range from 0.00 to 0.49 m for the Upper dataset. The RL approach yielded similar result as the Gold approach when using the Lower dataset and most of the uncertainty ranged from 0.51 to 1.00 m. The spatial uncertainty using the Upper dataset with the RL approach had less variation compared to the Gold approach that was consistent with the standard deviation in Table 6. It was worth noting that the spatial distribution of uncertainty derived from the Lower dataset was not consistent with that of the Upper dataset. This may be mainly attributed to the fact that the NEON data could provide x, y, z change per nanosecond for each waveform which can result in the synchronized change of points' x, y, z when the peak location with 95% uncertainty level was considered. In addition, the interpolation and the smoothing process may also lead to this kind of inconsistency.

Global statistics for the DTMs' parameter uncertainty is presented in Table 6. For all study sites, all approaches' absolute mean spatial uncertainties were below 0.25 m, except for the Upper dataset using the RL approach, indicating that biases caused by parameter uncertainty were relative low. The RMSE ranged from 0.08 to 0.36 m, which was almost consistent with the standard deviation. From the statistics in Table 6, there is no obvious difference among these three approaches in terms of the RMSE and SD. However, the Gold approach outperformed the other two approaches when the MU was taken into account.

The RL approach had larger minimum error and maximum error in the HF site that resulted in larger RMSE and mean absolute errors at the HF site. The RL approach's performance enhanced substantially when it came to the SJER and OSBS sites with smaller RMSE. For instance, the RL approach's RMSE for the Lower and Upper dataset were 0.14 and 0.28 m at the SJER site, when compared to 0.42 and 0.94 m at the HF site. The only one flight line may contribute to higher SD and RMSE at the HF site since there was no overlap in the study region with less dense raw waveform data. It was surprising to find that the Upper datasets for all approaches have larger range and RMSE than the corresponding Lower datasets. It may be attributed to the fact that most of the Upper dataset's points were lower than reference DTM, which had a more weight on the effect of the DTM generation than using the Lower dataset.

The direct decomposition method worked well and consistently in these three sites, but the deconvolution and decomposition method (either the Gold or RL approach) was more likely to generate smaller RMSE than the direct decomposition approach.

To further identify areas where DTM surfaces of low quality with high uncertainty, and compare the performances of approaches under different conditions, the slope and the vegetation height were taken into account as important predictors for the categories of uncertainty. Here, the HF and SJER sites were selected as examples to demonstrate the effect of the slope and vegetation height.

The ANOVA analysis showed that vegetation height and slope had a significant effect on the uncertainty levels of DTMs for the three approaches at the HF site, with all p-values smaller than 0.05 (Fig. 11a and b). The green line (median) increased with

Table 6
Global statistics summarizing validation errors caused by parameter uncertainty for DTMs.

Approaches	Dataset	Range (m)	SD (m)	MU (m)	MinU (m)	MaxU (m)	RMSE (m)
HF							
Direct	Low	313.89–318.15	0.15	0.23	−0.06	0.76	0.32
	Up	313.69–317.50	0.18	−0.24	−0.79	0.60	0.32
Gold	Low	313.40–318.00	0.07	0.04	−0.21	0.28	0.08
	Up	313.31–317.97	0.09	−0.03	−0.49	0.21	0.09
RL	Low	313.74–317.48	0.39	−0.15	−2.12	0.88	0.42
	Up	313.35–316.2	0.52	−0.78	−3.05	0.22	0.94
SJER							
Direct	Low	382.09–424.05	0.12	0.05	−1.24	1.53	0.21
	Up	381.69–424.11	0.18	−0.10	−2.51	1.29	0.13
Gold	Low	381.39–423.11	0.17	0.03	−1.61	1.43	0.17
	Up	380.78–423.15	0.36	−0.04	−2.59	2.53	0.36
RL	Low	381.39–423.11	0.14	0.03	−2.21	2.59	0.14
	Up	381.35–422.92	0.26	−0.10	−3.77	1.89	0.28
OSBS							
Direct	Low	21.88–47.10	0.16	0.08	−1.57	2.48	0.18
	Up	20.96–47.02	0.26	−0.13	−2.27	1.42	0.29
Gold	Low	20.94–45.46	0.11	0.01	−1.59	1.38	0.11
	Up	20.99–45.61	0.18	0.02	−1.88	1.51	0.18
RL	Low	20.48–46.67	0.10	0.01	−1.84	1.78	0.10
	Up	20.99–46.63	0.11	0.00	−2.54	1.74	0.11

Direct: Direct decomposition approach; Gold: Gold approach; RL: RL approach. SD: standard deviation; MU: Mean uncertainty caused by parameters between DTM derived from uncertainty dataset and DTM derived from peak location dataset; MinU: Minimum change caused by parameter uncertainty; MaxU: Maximum change caused by parameter uncertainty; low: Lower dataset; up: Upper dataset.

uncertainty levels for the three approaches, which demonstrated that larger slope and higher vegetation height were more likely to cause high uncertainty of DTM. For the SJER site, the uncertainty levels vs. slope showed that all approaches' p-values were zero except the Gold approach's p-value was 0.917. This indicated that the slope had no effect on the uncertainty levels and the Gold approach may be robust when dealing with complex topography. The average slope and slope distribution of the three uncertainty levels for the other two approaches were similar to each other, even the ANOVA analysis showed that slope was a significant factor for determining the uncertainty levels. This may potentially imply that slope could influence the uncertainty levels, but its cause-effect relationship is not so strong. Unlike slope, the vegetation height's effect on the uncertainty levels was more significant with regards to mean, median and interquartile range (IQR) as shown in Fig. 11c and d. Additionally, the higher the vegetation height, the more likely for this area to have higher uncertainty level. It was worthy to note that the median was not consistent with the mean for the corresponding uncertainty level. Most of the low and medium uncertainty occurred on the ground with mean and median being zero for low uncertainty levels (Fig. 11d).

In summary, the analysis identified that high prediction uncertainty of DTM was more likely to occur at larger slope and higher vegetation for all approaches in flat topography with dense vegetation. The vegetation height's effect on the DTM's uncertainty levels was more significant than slope when it came to the complex topography. The ground was more prone to lower spatial uncertainty level which may result from that waveforms in the ground region were simpler than in the slope and vegetation regions.

3.6.2. Canopy height model

Visual comparisons of the maximum CHM of the HF site for the three approaches are shown in Fig. 12. With regards to different datasets for the three approaches, the spatial distribution of uncertainty was similar, but not identical. The Gold approach had the smallest variance with similar spatial distribution using the two different datasets. The RL approach had the largest variance, which was coincident with the global statistic of Table 7. Somewhat surprisingly, the largest uncertainty level (dark red or dark blue) for the RL approach was more likely to occur at the edges of the HF site

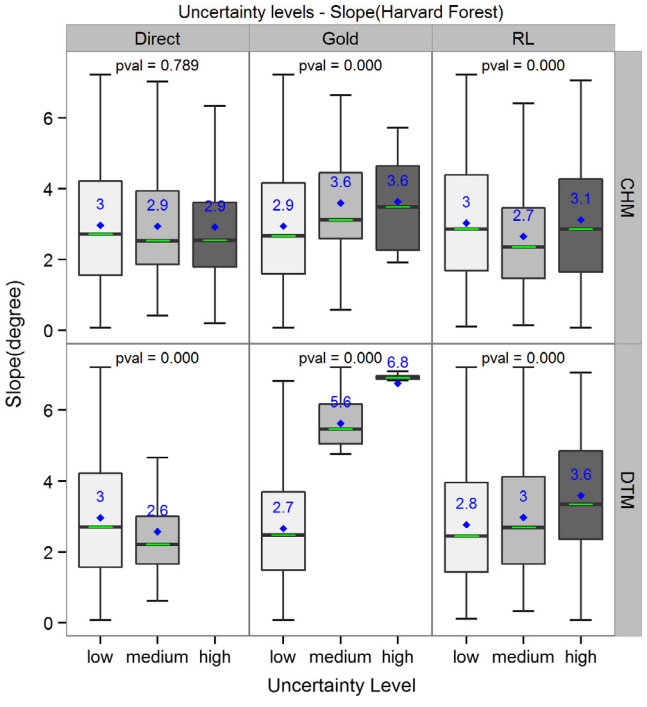

Fig. 11. a. Box-plot of DTMs and CHMs' Uncertainty levels vs. Slope for the HF site. b. Box-plot of DTMs and CHMs' Uncertainty levels vs. Vegetation Height for the HF site. c. Box-plot of DTMs and CHMs' Uncertainty levels vs. Slope for the SJER site. d. Box-plot of DTMs and CHMs' Uncertainty levels vs. Vegetation Height for the SJER site. *Green lines indicated median of dataset, the height of the box portion was given by the IQR of the dataset and the ends of the whiskers meant 1.5 IQR of lower quantile and 1.5 IQR of upper quantile. Blue points were the mean of the corresponding variable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

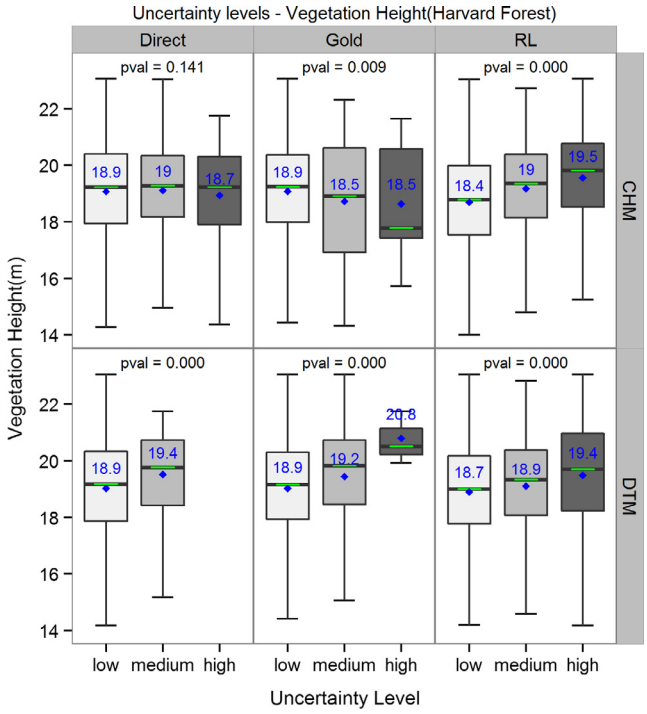

Fig. 11 (*continued*)

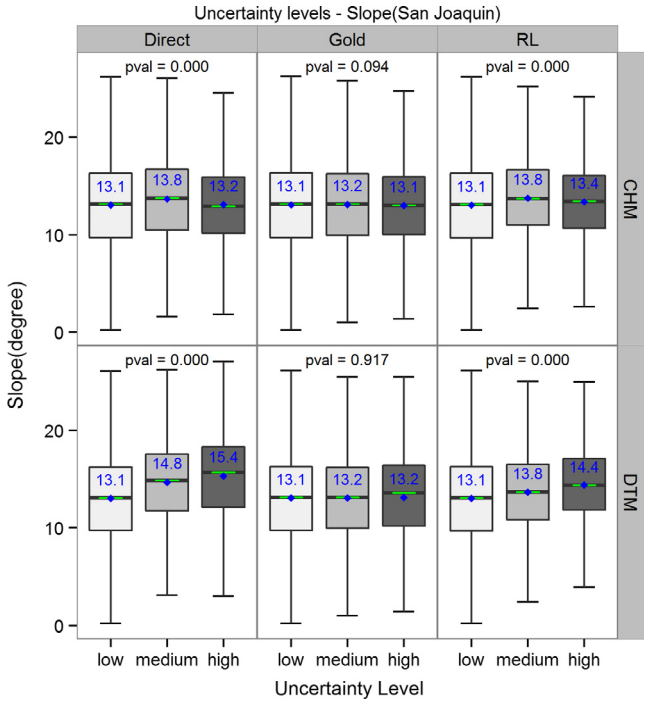

Fig. 11 (*continued*)

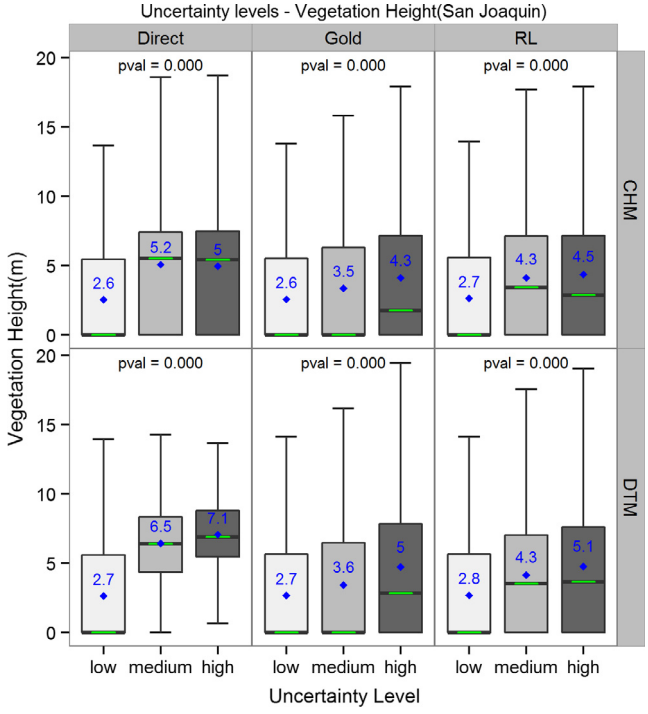

Fig. 11 (*continued*)

rather than at the boundary of trees and ground (Fig. 12). After a closer examination, we found that the large uncertainty level was located in dense vegetation areas, which may further imply that the Gold approach was less suitable for dense vegetation regions.

The quantitative assessment of the CHMs' parameter uncertainty (Table 7) yielded similar results to those of DTMs. The absolute mean spatial difference for all approaches ranged from 0.00 to 0.27 m, which was smaller than the DTMs' result derived from the corresponding datasets. It was not surprising to see that the minimum uncertainty and maximum uncertainty for each dataset of CHMs were larger than corresponding DTMs' results, since DTMs were much flatter than CHMs and relatively lower uncertainty was expected. In addition, the CHM was generated with additional steps compared with DTM that may bring more error into the CHM products. Most of these large uncertainties occurred at the boundary of trees and ground, and a small shift of XY location from parameter uncertainty would result in a large difference of canopy height.

The Gold approach had the smallest RMSE for CHMs in the HF and OSBS sites, but it had the largest RMSE for the SJER site. By contrast, the RL approach yielded the opposite results with the smallest RMSE at the SJER site, and the largest RMSE and mean error for the HF site. These trends were consistent with the DTMs' results. It confirmed the previous conclusion that the Gold approach may be more suitable in flat terrain areas and RL approach tended to perform better in complex topography conditions. The direct decomposition method performed well in both study areas, but not as good as the deconvolution and decomposition method (either the Gold or RL approach) that may further indicate the advantages of the deconvolution. Assuredly, the relationship between vegetation, and topographic conditions and deconvolution results is complex, the simulated waveform data with different topographic and vegetation conditions will be the ideal datasets to further test the approaches and provide insights into selecting approaches under different conditions of topography and vegetation.

The ANOVA analysis of CHMs at the HF site (Fig. 11a) demonstrated that the median, mean and IQR of slope were similar for different CHMs' uncertainty levels using the direct decomposition approach. This result indicated that the slope had no effect on the CHMs' uncertainty levels for the direct decomposition approach in terms of the statistical perspective. However, this factor's effect on the uncertainty levels was significant ($p < 0.050$) for the Gold and RL approaches. Likewise, the vegetation height also

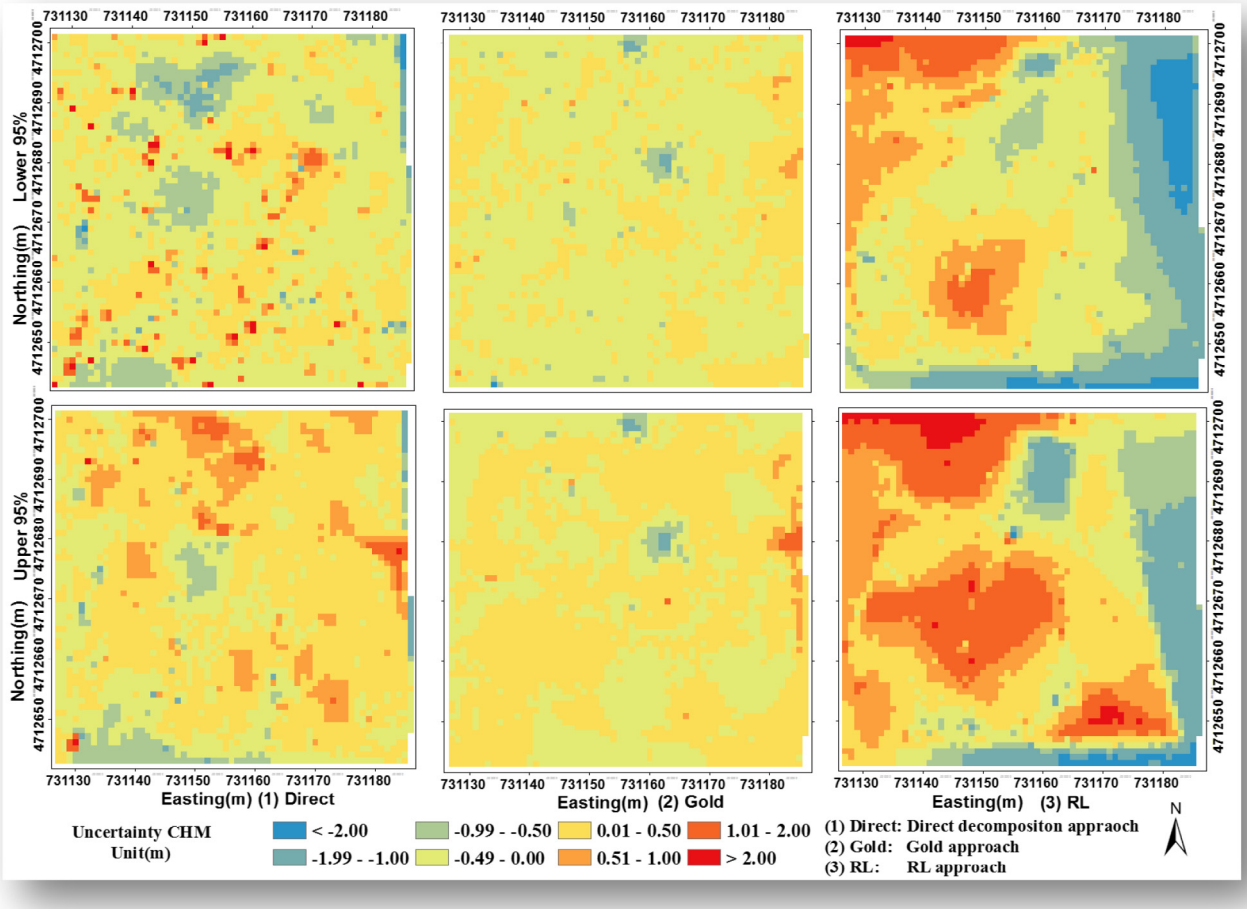

Fig. 12. The spatial uncertainty of CHM caused by the parameter uncertainty in HF region using the direct decomposition approach (left), Gold approach (middle) and RL (right) approach, respectively. The above was the result from the Lower dataset and the bottom was the result from the Upper dataset.

Table 7
Global statistics summarizing validation errors caused by parameter uncertainty for CHMs (unit: m).

Approaches	Dataset	Range	SD	MU	MinU	MaxU	RMSE
HF							
Direct	Low	11.44–24.13	0.46	−0.07	−2.85	4.82	0.47
	Up	11.76–23.38	0.38	0.13	−1.74	2.33	0.40
Gold	Low	10.71–22.75	0.18	−0.04	−2.34	0.92	0.19
	Up	10.84–23.16	0.19	0.01	−1.64	1.71	0.20
RL	Low	9.90–22.86	0.88	−0.27	−2.73	2.66	0.92
	Up	10.19–23.71	0.95	0.25	−3.77	2.56	0.98
SJER							
Direct	Low	0.00–24.85	0.43	0.06	−6.32	11.00	0.44
	Up	0.00–24.70	0.40	0.01	−6.42	8.87	0.39
Gold	Low	0.00–24.70	0.51	0.15	−7.06	10.22	0.53
	Up	0.00–24.24	0.49	0.04	−8.86	9.26	0.49
RL	Low	0.00–24.78	0.33	0.00	−9.44	8.70	0.33
	Up	0.00–24.63	0.35	0.09	−6.19	5.50	0.37
OSBS							
Direct	Low	0.00–27.36	0.61	0.00	−17.2	17.05	0.61
	Up	0.00–27.73	0.69	0.07	−14.56	17.70	0.69
Gold	Low	0.00–28.61	0.01	0.00	0.00	2.71	0.01
	Up	0.00–28.59	0.58	−0.03	−13.71	16.22	0.57
RL	Low	0.00–28.37	0.48	0.03	−18.18	19.20	0.49
	Up	0.00–28.12	0.46	−0.01	−17.20	17.12	0.46

Direct: Direct decomposition approach; Gold: Gold approach; RL: RL approach. SD: standard deviation; MU: Mean uncertainty caused by parameters between DTM derived from uncertainty dataset and DTM derived from peak location dataset; MinU: Minimum change caused by parameter uncertainty; MaxU: Maximum change caused by parameter uncertainty; low: Lower dataset; up: Upper dataset.

had an impact on the uncertainty levels using the Gold and RL approaches, but not for the direct decomposition approach at the HF site as shown in Fig. 11b. Generally, higher uncertainty level was more likely to occur at the higher slope and vegetation height for the Gold and RL approaches.

For the SJER site, the analysis showed that the uncertainty levels of CHMs were influenced by the vegetation height for all approaches ($p = 0.000$), and there was a large difference between mean and median of vegetation height for different uncertainty levels (Fig. 11d). Most of the uncertainty was more likely to occur at the lower vegetation and ground, especially for the Gold approach. The analysis of the slope (Fig. 11c) showed that the Gold approach's uncertainty levels were robust to slope changes ($p = 0.094$). The p-values of the other two approaches' were zero, which indicated that the uncertainty of CHM was likely to be affected by slope changes. However, these two approaches' median, mean and IQR were similar for different uncertainty levels which may imply that the slope did not play as an important role as vegetation height in determining uncertainty levels.

In summary, the analysis identified that higher uncertainty of CHM was prone to occur at higher vegetation for all approaches in the complex topography with less dense vegetation areas. For the flat topography, there was no obvious pattern for the Gold and RL approaches for different uncertainty levels. The direct decomposition approach outperformed other two approaches and was robust to the change of slope and vegetation height under such topography condition.

4. Conclusions

This study proposes a new waveform LiDAR deconvolution algorithm called the Gold algorithm as a preprocessing step, and comprehensively compares different methods of processing the waveform LiDAR data at three different ecological sites from both visual and quantitative perspectives.

Our work has demonstrated the advantage of the deconvolution and decomposition method with more echoes of waveforms detected and less false echoes generated, especially when the Gold approach is used. Furthermore, the accuracy assessment of the end products (DTMs and CHMs) shows that the three approaches can generate satisfactory results, while the best performances vary when different criteria are used: the Gold approach has better performance with smaller RMSE, and the direct decomposition approach outperforms others in terms of the percentage of spatial difference within 0.5 and 1 m. According to the parameter uncertainty of end products, the factors like the vegetation height and slope both have an effect on the robustness of approaches, while the slope becomes a less significant factor when it comes to the spatial uncertainty of CHMs. Specifically, the Gold approach tends to have better performance in the dense vegetation region and the RL approach works better in the sparse vegetation region. Therefore, the important contributions of this study lie in successfully introducing a novel deconvolution algorithm, the Gold algorithm, for waveform LiDAR processing, and providing a comprehensive comparison and a quantifiable basis selection of different waveform LiDAR processing methods for different topography and vegetation conditions. Potential future studies could use the proposed method to process waveform LiDAR data and extract semantical information, such as individual tree crown mapping, understory tree detection, and to estimate forest structure and biophysical parameters. In addition, future investigations could benefit from expanding the availability of new waveform LiDAR datasets to cover varied vegetation conditions in multiple ecosystem types and complex topography, urban areas, in rangelands and grasslands.

Acknowledgement

The authors gratefully acknowledge the support provided by NASA New Investigator Program (Grant #NNX08AR12G), NSF Doctoral Dissertation Improvement Grant (DEB 1702008), and the graduate student support provided by the Department of Ecosystem Science and Management and the College of Agriculture and Life Sciences at Texas A&M University. We also thank Dr. Thadeus Bowerman and Shruthi Srinivasan for patient proofreading of the first draft of the paper. We express our sincere gratitude to two anonymous reviewers for their insightful comments.

References

Azadbakht, M., Fraser, C., Khoshelham, K., 2016. A sparsity-based regularization approach for deconvolution of full-waveform airborne Lidar data. Remote Sensing 8 (8), 648. http://dx.doi.org/10.3390/rs8080648.

Boudreau, J., Nelson, R., Margolis, H., Beaudoin, A., Guindon, L., Kimes, D., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sens. Environ. 112 (10), 3876–3890. http://dx.doi.org/10.1016/j.rse.2008.06.003.

Briese, C., Höfle, B., Lehner, H., Wagner, W., Pfennigbauer, M., Ullrich, A., 2008. Calibration of full-waveform airborne laser scanning data for object classification. In: Paper Presented at the SPIE Defense and Security Symposium.

Carlsson, T., Steinvall, O., Letalick, D., 2001. Signature Simulation and Signal Analysis for 3-D Laser Radar. Month, 4, C4ISR.

Cawse-Nicholson, K., van Aardt, J., Hagstrom, S., Romanczyk, P., Schaaf, C., Strahler, A., Krause, K., 2014. Improving waveform lidar processing toward robust deconvolution of signals for improved structural assessments. In: Paper Presented at the SPIE Defense + Security.

Chauve, A., Mallet, C., Bretar, F., Durrieu, S., Deseilligny, M.P., Puech, W., 2007. Processing full-waveform lidar data: modelling raw signals. In: Paper Presented at the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 2007.

Chauve, A., Vega, C., Durrieu, S., Bretar, F., Allouis, T., Pierrot Deseilligny, M., Puech, W., 2009. Advanced full-waveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest. Int. J. Remote Sens. 30 (19), 5211–5228. http://dx.doi.org/10.1080/01431160903023009.

Chen, Q., 2007. Airborne lidar data processing and information extraction. Photogramm. Eng. Remote Sens. 73 (2), 109.

Chhatkuli, S., Mano, K., Kogure, T., Tachibana, K., Shimamura, H., 2012. Full Waveform Lidar Exploitation Technique and its Evaluation in the Mixed Forest Hilly Region. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 505–509.

Doneus, M., Briese, C., Fera, M., Janner, M., 2008. Archaeological prospection of forested areas using full-waveform airborne laser scanning. J. Archaeol. Sci. 35 (4), 882–893. http://dx.doi.org/10.1016/j.jas.2007.06.013.

Elzhov, T.V., Mullen, K.M., Spiess, A.-N., Bolker, B., 2013. minpack. lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm found in MINPACK, Plus Support for Bounds. R Package Version 1.1-8.

Fieber, K.D., Davenport, I.J., Tanase, M.A., Ferryman, J.M., Gurney, R.J., Becerra, V.M., Hacker, J.M., 2015. Validation of canopy height profile methodology for small-footprint full-waveform airborne LiDAR data in a discontinuous canopy environment. ISPRS J. Photogramm. Remote Sens. 104, 144–157.

Fish, D., Brinicombe, A., Pike, E., Walker, J., 1995. Blind deconvolution by means of the Richardson-Lucy algorithm. JOSA A 12 (1), 58–65.

Gao, S., Niu, Z., Sun, G., Zhao, D., Jia, K., Qin, Y., 2015. Height extraction of maize using airborne full-waveform LIDAR data and a deconvolution algorithm. IEEE Geosci. Remote Sens. Lett. 12 (9), 1978–1982.

Gwenzi, D., Lefsky, M.A., 2014. Modeling canopy height in a savanna ecosystem using spaceborne lidar waveforms. Remote Sens. Environ. 154, 338–344. http://dx.doi.org/10.1016/j.rse.2013.11.024.

Hancock, S., Anderson, K., Disney, M., Gaston, K.J., 2017. Measurement of fine-spatial-resolution 3D vegetation structure with airborne waveform lidar: Calibration and validation with voxelised terrestrial lidar. Remote Sens. Environ. 188, 37–50. http://dx.doi.org/10.1016/j.rse.2016.10.041.

Hancock, S., Lewis, P., Disney, M., Foster, M., Muller, J., 2008. Assessing the Accuracy Of Forest Height Estimation with Long Pulse Waveform Lidar through Monte-Carlo Ray Tracing. Silvilaser, Edinburgh (September, 17, 18).

Harding, D.J., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. Geophys. Res. Lett. 32 (21). http://dx.doi.org/10.1029/2005gl023471.

Harsdorf, S., Reuter, R., 2000. Stable Deconvolution of Noisy Lidar Signals. tc, 10, 1.

Hofton, M.A., Minster, J.B., Blair, J.B., 2000. Decomposition of laser altimeter waveforms. IEEE Trans. Geosci. Remote Sens. 38 (4), 1989–1996.

Hollaus, M., Mücke, W., Höfle, B., Dorigo, W., Pfeifer, N., Wagner, W., Regner, B., 2009. Tree species classification based on full-waveform airborne laser scanning data. Proc. Silvilaser 2009, 54–62.

Ioannides, M., Arnold, D., Niccolucci, F., Mania, K., 2006. Digital Terrain Modelling for Archaeological Interpretation within Forested Areas using Full-Waveform Laserscanning.

Isenburg, M., 2012. LAStools-Efficient Tools for LiDAR Processing. Available at: <http://www.cs.unc.edu/~isenburg/lastools/> (accessed October 9, 2012).
Jalobeanu, A., Gonçalves, G., 2014. Robust ground peak extraction with range error estimation using full-waveform LiDAR. IEEE Geosci. Remote Sens. Lett. 11 (7), 1190–1194.
Jutzi, B., Stilla, U., 2006. Range determination with waveform recording laser systems using a Wiener filter. ISPRS J. Photogramm. Remote Sens. 61 (2), 95–107. http://dx.doi.org/10.1016/j.isprsjprs.2006.09.001.
Kampe, T.U., 2010. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. J. Appl. Remote Sens. 4 (1), 043510. http://dx.doi.org/10.1117/1.3361375.
Keith, K., Tristan, G., 2015. NEON L0-TO-L1 Discrete-Return LiDAR Algorithm Theoretical Basis Document (ATBD). Retrieved from <http://data.neonscience.org/documents>.
Keller, M., 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. J. Appl. Remote Sens. 1 (1), 013537. http://dx.doi.org/10.1117/1.2795724.
Khosravipour, A., Skidmore, A.K., Isenburg, M., Wang, T., Hussin, Y.A., 2014. Generating pit-free canopy height models from airborne Lidar. Photogramm. Eng. Remote Sens. 80 (9), 863–872. http://dx.doi.org/10.14358/pers.80.9.863.
Lefsky, M.A., 2010. A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. Geophys. Res. Lett. 37 (15). http://dx.doi.org/10.1029/2010gl043622. n/a–n/a.
Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., de Oliveira, R., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. Geophys. Res. Lett. 32 (22). http://dx.doi.org/10.1029/2005gl023971.
Lucy, L.B., 1974. An iterative technique for the rectification of observed distributions. Astronom. J. 79, 745.
Mallet, C., Bretar, F., 2009. Full-waveform topographic lidar: state-of-the-art. ISPRS J. Photogramm. Remote Sens. 64 (1), 1–16. http://dx.doi.org/10.1016/j.isprsjprs.2008.09.007.
Mallet, C., Lafarge, F., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2009. A stochastic approach for modelling airborne lidar waveforms. Laserscanning, 201–206.
McGlinchy, J., van Aardt, J.A.N., Erasmus, B., Asner, G.P., Mathieu, R., Wessels, K., Cawse-Nicholson, K., 2014. Extracting structural vegetation components from small-footprint waveform lidar for biomass estimation in savanna ecosystems. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 7 (2), 480–490. http://dx.doi.org/10.1109/jstars.2013.2274761.
Morhac, M., 2012. Peaks: Peaks. R Package Version 0.2.
Morhac, M., Kliman, J., Matousek, V., Veselsky, M., Turzo, I., 1997. Efficient one- and two-dimensional gold deconvolution and its application to gamma-ray spectra decomposition. Nucl. Instrum. Methods Phys. Res. Sect. A-Accel. Spectr. Detectors Assoc. Equip. 401 (2–3), 385–408. http://dx.doi.org/10.1016/S0168-9002(97)01058-9.
Morháč, M., Matoušek, V., Kliman, J., 2003. Efficient algorithm of multidimensional deconvolution and its application to nuclear data processing. Digital Signal Process. 13 (1), 144–171.
Neuenschwander, A.L., 2008. Evaluation of waveform deconvolution and decomposition retrieval algorithms for ICESat/GLAS data. Can. J. Remote. Sens. 34 (Suppl. 2), S240–S246.
Nordin, L., 2006. Analysis of Waveform Data from Airborne Laser Scanner Systems (MS Thesis). Lulea Univ. Technol., Lulea, Sweden.
Persson, Å., Söderman, U., Töpel, J., Ahlberg, S., 2005. Visualization and analysis of full-waveform airborne laser scanner data. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 36 (Part 3), W19.
RC Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>.
Reitberger, J., Krzystek, P., Stilla, U., 2008. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. Int. J. Remote Sens. 29 (5), 1407–1431. http://dx.doi.org/10.1080/01431160701736448.
Reitberger, J., Schnörr, C., Krzystek, P., Stilla, U., 2009. 3D segmentation of single trees exploiting full waveform LIDAR data. ISPRS J. Photogramm. Remote Sens. 64 (6), 561–574. http://dx.doi.org/10.1016/j.isprsjprs.2009.04.002.
Roncat, A., Bergauer, G., Pfeifer, N., 2011. B-spline deconvolution for differential target cross-section determination in full-waveform laser scanning data. ISPRS J. Photogramm. Remote Sens. 66 (4), 418–428. http://dx.doi.org/10.1016/j.isprsjprs.2011.02.002.
Rowe, J., 2013. Ground Classification and Below Ground Response Assessment of Forested Regions using Full-Waveform LiDAR.
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 J. Photogramm. Remote Sens. 60 (2), 100–112. http://dx.doi.org/10.1016/j.isprsjprs.2005.12.001.
Wang, H., Glennie, C., 2015. Fusion of waveform LiDAR data and hyperspectral imagery for land cover classification. ISPRS J. Photogramm. Remote Sens. 108, 1–11.
Wu, J., van Aardt, J., Asner, G.P., 2011. A comparison of signal deconvolution algorithms based on small-footprint LiDAR waveform simulation. IEEE Trans. Geosci. Remote Sens. 49 (6), 2402–2414.
Zhu, J., Zhang, Z., Hu, X., Li, Z., 2011. Analysis and application of LiDAR waveform data using a progressive waveform decomposition method. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 38. 5/W12-31-36.
Zhuang, W., Mountrakis, G., 2014. An accurate and computationally efficient algorithm for ground peak identification in large footprint waveform LiDAR data. ISPRS J. Photogramm. Remote Sens. 95, 81–92. http://dx.doi.org/10.1016/j.isprsjprs.2014.06.004.