Remote Sensing of Environment 200 (2017) 43–62

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Bayesian decomposition of full waveform LiDAR data with uncertainty analysis

Tan Zhou*, Sorin C. Popescu

LiDAR Applications for the Study of Ecosystems with Remote Sensing (LASERS) Laboratory, Department of Ecosystem Science and Management, Texas A & M University, College Station, TX 77450, United States

ARTICLE INFO

Keywords:
Waveform LiDAR
Bayesian inference
Decomposition
Uncertainty analysis
Tree canopy height
Model reasonableness

ABSTRACT

A thorough understanding of full waveform (FW) LiDAR data processing and associated uncertainty is critical to vegetation applications such as retrieving forest structure variables and estimating forest biomass. This paper applies the Bayesian non-linear modeling concept to process small-footprint FW LiDAR data (the Bayesian decomposition) collected at a study site of the National Ecological Observatory Network (NEON) to investigate its potential for waveform decomposition and uncertainty estimation. Specifically, several possible models suitable for fitting waveforms were assessed within the Bayesian framework, and the Gaussian model was selected to perform the Bayesian decomposition. Subsequently, we conducted performance evaluation and uncertainty analysis at the parameter, derived point cloud and surface model levels. Results of the model reasonableness show that the Gaussian model is superior to alternative models with respect to uncertainty, physical meaning and processing efficiency. After converting waveforms to discrete points, the model comparisons demonstrate that the Bayesian decomposition can be utilized for FW LiDAR data processing, and its results are comparable to the direct decomposition (DD), Gold and RL (Richardson–Lucy) approaches in terms of the root mean squared error (RMSE $<$ 0.93 m) of the point distances between the waveform-based point cloud and the reference point cloud. Additionally, more points can be extracted from FW LiDAR data with these methods than discrete-return LiDAR data, especially at the mid-story of vegetation based on the results of height bins, percentile heights and canopy LiDAR density at the individual tree level. Moreover, uncertainty estimates from the Bayesian method enhance the credibility of decomposition results in a probabilistic sense to capture the true error of estimates and trace the uncertainty propagation along the processing steps. For example, results of the surface model yield larger RMSE values (1.38 m vs. 0.65 m) with a wider credible interval than quantile point clouds with a more compact distribution. In contrast to commonly used deterministic approaches, the Bayesian decomposition method can produce an ensemble of reasonable parameter estimates with probability through Markov Chain Monte Carlo (MCMC) sampling from the posterior distribution of model parameters. These parameter estimates and corresponding derived products can be queried to provide meaningful interpretation of results and associated uncertainty. Both the flat priors and empirical priors can achieve good performance of the decomposition while the empirical priors tend to significantly speed up the model convergence. The Bayesian approach also renders an important insight into the uncertainty of the model performance evaluation using field data by generating reasonable prediction intervals to reduce inherent errors of field measurements.

1. Introduction

Light Detection and Ranging (LiDAR) has been adopted as a valuable survey tool reducing the need for field measurement to accurately characterize vegetation structure for the past decades (Hancock et al., 2015; Popescu et al., 2003; Wulder et al., 2012; Zhao et al., 2009). Especially, the advent of the full waveform (FW) LiDAR system, which is capable of recording entire reflected energy along the pulse line, has enabled this advantage to become more conspicuous (Cawse-Nicholson et al., 2014; Wagner et al., 2006; Wulder et al., 2012). The FW LiDAR data primarily consists of two parts: a pulse part that keeps geo-reference locations derived from range measurement between the laser sensor and the reference location, and a wave part which fully stores digitized return energy starting from the reference location till the end of digitized samples. Within a forest environment, FW LiDAR energy could penetrate dense canopies through small gaps in the canopy and achieve a full time-versus-intensity profile. Consequently, more detailed vertical information of vegetation structure can be revealed through

* Corresponding author.
E-mail address: tankwin0@tamu.edu (T. Zhou).

http://dx.doi.org/10.1016/j.rse.2017.08.012
Received 28 December 2016; Received in revised form 18 July 2017; Accepted 4 August 2017
Available online 09 August 2017
0034-4257/ Published by Elsevier Inc.

these data.

Generally, FW LiDAR data can be classified as large- (~50 m or larger), medium- (~10–30 m) or small-footprint waveform (< 1 m) based on their transmitted laser size (Wang and Weng, 2013). Some of the first FW sensors known as large footprint profilers included SLICER (Scanning LiDAR Imager of Canopies by Echo Recovery, 10 m footprint), LVIS (Laser Vegetation Imaging Sensor, 25 m footprint) and GLAS (the Geoscience Laser Altimeter System, 70 m footprint). All of them have been successfully applied to estimate various forest parameters and vegetation studies worldwide (Blair et al., 1999; Drake et al., 2002; Harding and Carabajal, 2005).

Recent advances of commercial LiDAR systems have promoted the availability of small-footprint FW LiDAR data from remote sensing industry providers. However, extensive applications of such systems to characterize forest structure and biomass are limited (Wulder et al., 2012). There are three main reasons behind this: (1) there is no standard format of FW LiDAR data, (2) large data volume required for storing the information, which leads to difficulties of data distribution and processing, (3) high cost of data acquisition, with added cost compared to discrete-return (DR) LiDAR data that hinders their adoption for many potential applications (Pirotti, 2011). In addition, while there are many software packages and applications available for processing DR LiDAR data, only few software developments are currently available for processing FW LiDAR. Therefore, the development of robust and dedicated methods and non-proprietary software for processing small-footprint FW LiDAR data are urgently needed.

Existing methods for FW LiDAR processing can be mainly categorized into two types: the decomposition method and the combined deconvolution and decomposition method. The most commonly used approach for the decomposition method is the direct decomposition (DD) which models the waveform with a mixture of Gaussian functions. Typical approaches such as Non-linear least-squares (NLS) (Hofton et al., 2000) or maximum likelihood estimation using the Expectation-Maximization (EM) algorithm (Persson et al., 2005) have been developed for fitting the waveform to extract 3D points and related parameters. However, these approaches are sensitive to the initialization of unknown parameters. Another popular method for recovering the true cross section of objects along the pulse line is the combined deconvolution and decomposition method. Multiple deconvolution algorithms, such as the Gold, Richardson-Lucy (RL), Non-negative least squares (NNLS), and Wiener Filter (WF) (Cawse-Nicholson et al., 2014; McGlinchy et al., 2014; Neuenschwander, 2008; Roncat et al., 2010; Rowe, 2013; Wu et al., 2011; Zhou et al., 2017) have been successfully introduced for reconstructing the differential backscatter cross section. One practical issue of these methods is that they are pertinent to the choice of proper parameter combinations for the deconvolution, which typically requires parameter optimization before data processing (Zhou et al., 2017). Although these methods have been proven to be able to generate sufficient fitting models, we cannot characterize the uncertainty with these models. They are calculated based on the deterministic models, which only seek a point value for the parameter of interest (Edwards et al., 2003). Without uncertainty analysis, the models are less informative or even useless when they are applied to the real-world problems (Zhao et al., 2011).

In the domain of LiDAR applications, the observations or data are inherently subject to various errors such as system setting, system calibration, and range measurement errors (Griewank and Walther, 2008). Additionally, the LiDAR vendors often do not clearly state what errors are considered when the data are provided. Thus, uncertainty of "truth" is ubiquitous and inherently present in the realities of LiDAR data modeling. Some studies associated with uncertainty mainly focus on the DR LiDAR data applications (Chauve et al., 2009; Chen et al., 2015; Frazer et al., 2011), while few published studies have explored FW LiDAR data's uncertainty for vegetation characterization. Furthermore, the models used here are based on the non-linear functions that generally suffer from problem of non-uniqueness (Sen and Stoffa, 1996), which can generate different parameter combinations given the same observational data and model, or several models can fit observational data at the expense of violating the physical meaning and theoretical assumption of the "real" model. These problems are more evident for the sophisticated models with multiple peak components in the waveform decomposition. Thus, estimating model uncertainty is imperative for an in-depth understanding of information derived from data and the estimation accuracy. This kind of uncertainty analysis has been inadequately addressed in many preceding studies due to the great diversity of remote sensing-based information retrieval procedures (Zhao et al., 2011) and the absence of efficient and universal methods to capture the uncertainty of data modeling.

One strategy to overcome these problems is to adopt a Bayesian paradigm of statistical inference by considering the model parameters to be realizations of random variables. Within a Bayesian framework, we combined prior information about unknown parameters with observed data using the Bayes' rule. Using a Markov Chain Monte Carlo (MCMC) (Gelfand and Smith, 1990) algorithm, we were able to sample from the posterior distribution of the unknown model parameters of interest. Through the posterior distribution, uncertainty bounds on the resulting model parameters and model reasonableness can be measured. This approach generally takes more time to reach a solution, but the non-uniqueness of the model parameter estimates can be avoided by describing these parameters in terms of probability density functions (PDFs) in the model space (Hong and Sen, 2009; Sen and Stoffa, 1996). Additionally, computational advances and the introduction of the more efficient Hamiltonian Monte Carlo (HMC) algorithm (Neal, 2011) have contributed enormously to the growing interest in applying Bayesian approaches to remote sensing data processing.

Recently, Bayesian analytical approaches have been applied to diverse domains related to waveform data as an alternative to traditional deterministic techniques. For example, Qin et al. (2016) analyzed ground-penetrating radar (GPR) data to detect the defect of underground structure using a Bayesian inversion method. Roonizi and Sassi (2016) elaborated how the electrocardiogram (ECG) waveform separation was conducted in a Bayesian framework to evaluate cardiac health status. In the geophysical field, the Bayesian approach was used for marine seismic waveform data to characterize subsurface reflectivity (Ray et al., 2013). For LiDAR applications, Bayesian methods are mainly used in a spatial modeling context to predict and map forest variables (Finley et al., 2013) or image construction (Hernandez-Marin et al., 2008). However, employing Bayesian approaches to decompose FW LiDAR data for vegetation studies is rarely reported in the current literature.

Therefore, the overall goal of this paper is to explore a Bayesian statistical method with the HMC algorithm to process small-footprint FW LiDAR data and quantify the uncertainty from data and models. More specifically, we attempt to (1) evaluate the reasonableness of models suitable for FW LiDAR data within a Bayesian framework; (2) develop a robust and dedicated Bayesian decomposition method to process FW LiDAR data for vegetation, and implement thorough comparisons with other FW LiDAR data processing methods; and (3) obtain reliable estimates of error and uncertainty in different steps (the parameter estimates, point cloud and surface models) using the Bayesian decomposition method. The motivation for the first objective is to check the validity of previous studies' underlying assumptions that the Gaussian model is sufficient for FW LiDAR data decomposition (Mallet and Bretar, 2009; Pirotti, 2011; Wagner et al., 2006), and to further reduce the uncertainty caused by the theoretical model error. The innovative aspects of this study consist of (1) integrating the nonlinear Bayesian concept with waveform data to provide a novel decomposition method for small-footprint FW LiDAR data; and (2) generating a consistent, transparent knowledge learning framework to quantify the uncertainty emerging from data and trace the uncertainty propagation along the processing steps. In this study, we did not intend to use the Bayesian decomposition as a proof of concept, but instead we applied

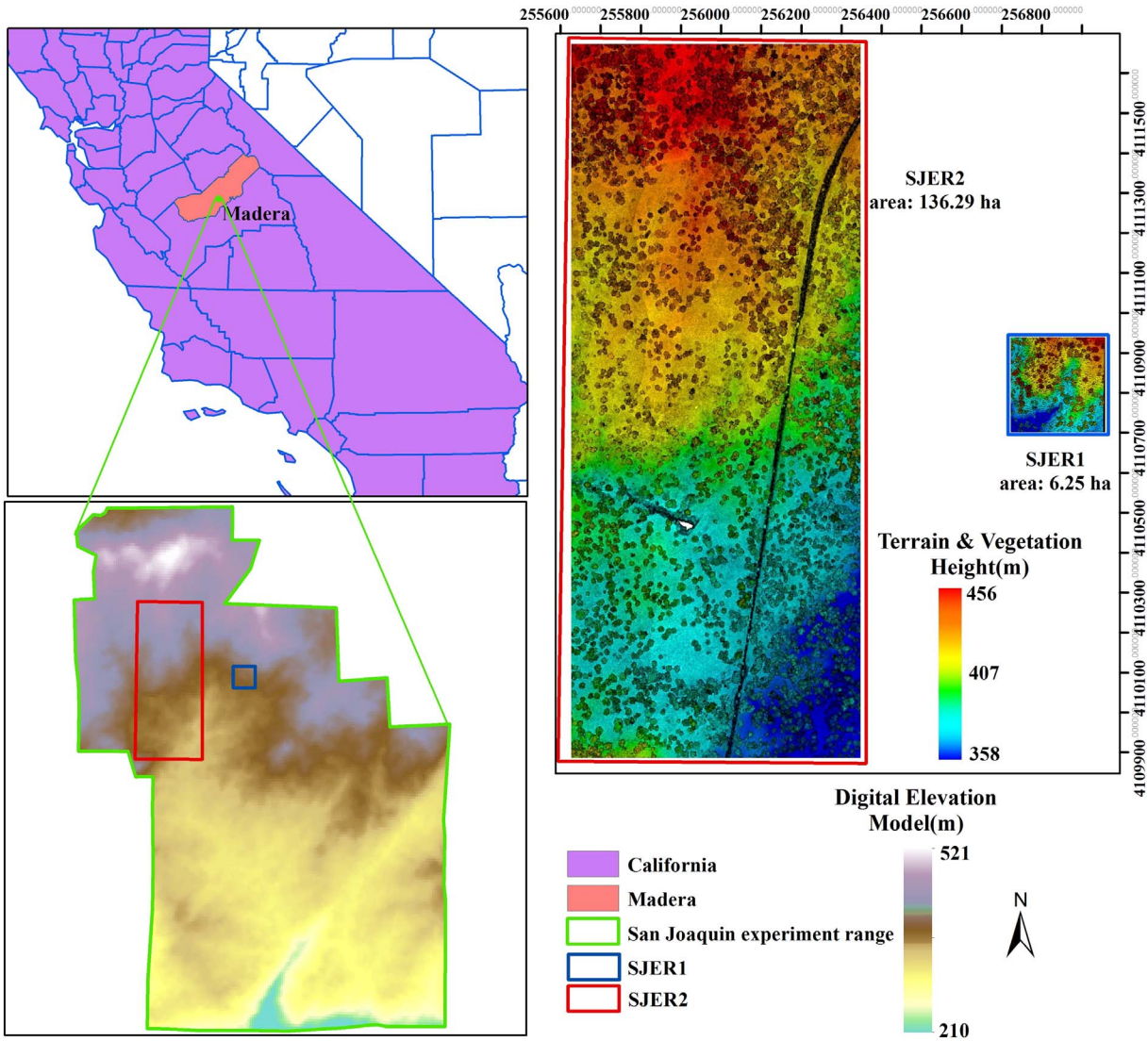

Fig. 1. Map of the San Joaquin Experimental Range (SJER) with location of study regions in California (left panel) and discrete-return LiDAR point image with terrain and vegetation height (right panel).

this approach to processing millions of waveforms in our study sites to provide insights into model justification and derive a benchmark for the uncertainty quantification of FW LiDAR data.

2. Materials and methods

2.1. Study site and data

2.1.1. Study site

The study site is located at the San Joaquin Experimental Range (SJER), which is in the foothills of Sierra Nevada Mountains, about 32 km north of Fresno, California. Two study regions were investigated as shown in Fig. 1. One waveform sample region (SJER1) is about 6.25 ha (250 m × 250 m) with the center at 256,840.0 Easting, 4,110,820.0 Northing, and UTM Zone 11N. The SJER1 is composed of vegetation dominated by blue oak (*Quercus douglasii*), interior live oaks (*Quercus wislizeni*) and digger pine (*Pinus sabiniana*) with scattered shrubs and a nearly continuous cover of herbaceous plants. This study region was used mainly to develop the proposed model of the present paper and to compare its performance with existing approaches such as the DD, Gold and RL. Another study region (SJER2) covers approximately 136 ha with the center at 255,977.6 Easting, 4,110,780.2 Northing, and UTM Zone 11N, which primarily aims to demonstrate that the proposed model can be applied to a relatively large area, instead of a small concept-area. This region consists of mixed patches of vegetation structure and heterogeneous land cover types, including grassland, forest, water body, open ground and road.

2.1.2. LiDAR data

The LiDAR data were collected through the National Ecological Observatory Network (NEON) Airborne Observation Platform (Kampe, 2010) which carried sensors such as a hyperspectral imaging spectrometer, a FW LiDAR sensor and a DR LiDAR sensor flying at about 1000 m above ground level. This design can achieve sub-meter to meter scale ground resolution of study sites. Detailed technical specifications of data can be found in the study of Zhou et al. (2017).

In this study, two regions were chosen as displayed in Fig.1. The SJER1 region had 258,667 waveforms with two flight lines, while the SJER2 region had 20,040,883 waveforms with four flight lines. The original waveform is composed of 500 time bins with 1 ns temporal resolution. Each time bin stores the digital number (DN) or intensity of corresponding backscattered pulse. The time bins with non-recorded values are assigned as zero (zero-padded) to keep the length of the waveforms constant. For geolocation of corresponding waveform, 16 basic geolocation information attributes associated with waveforms are provided. Among them, the dx, dy, and dz. are the pulse direction

vector that can measure position change per nanosecond. In the subsequent analysis, eight items were used for calculating the geolocation of desired time bin after decomposition. The eight items are the Easting of first return x_0 (m), the Northing of first return y_0 (m), the height of first return z_0(m), dx (m), dy (m), dz. (m), the outgoing pulse reference bin location (leading edge 50% point of the outgoing pulse), and first return reference bin location (leading edge 50% point of the first return).

The DR LiDAR data for corresponding study regions were also collected with the maximum horizontal accuracy of 0.4 m and maximum vertical accuracy of 0.36 m based on the NEON'S LiDAR Algorithm Theoretical Basis Document (ATBD) (Keith and Tristan, 2015).

2.1.3. Field data

As part of the NEON's data collection efforts, extensive annual ground measurements of vegetation structure were conducted by the NEON Airborne Observation Platform and the Terrestrial Instrument System (TIS) programs. The plot design followed the protocol of the NEON Terrestrial Observation System, and each plot is restricted to a 20 × 20 m region. The plot locations have been established by NEON's Field Sentinel Unit for long-term plant, insect and soil measurements. There were six field plots with 151 individual trees collected during June 2013, which were available for these two study regions. One field plot including 16 trees was located in the SJER1 study region and the other five field plots with a total of 135 individual trees were located in the SJER2 study region. The key vegetation structure variables for each tree were measured such as the location (Easting, Northing), the maximum height, and the tree species.

2.2. Bayesian decomposition

2.2.1. Theoretical background

In a Bayesian statistical framework, deterministic models are specified via mathematical equations, e.g., linear or nonlinear functions, and unknown model parameters are treated stochastically with various probability distributions.

Based on Bayes' rule, the unknown parameters of a statistical model can be written as:

$$p(\theta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \theta)p(\theta) \tag{1}$$

where $\mathbf{y}$ is a vector of observed data which has a probability distribution depending on an unknown vector of parameters denoted as $p(\mathbf{y}|\boldsymbol{\theta})$, which is also known as the likelihood function. The prior distribution of model parameter vector $\boldsymbol{\theta}$ is a probability distribution that represents the experimenter's beliefs about unknown parameters prior to observing the data and was denoted as $p(\boldsymbol{\theta})$ (Hoff, 2009). $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution of the unknown vector of parameters.

Eq. (1) is fundamental to understanding that the posterior distribution of unknown parameters is proportional to the prior belief about unknown model parameters, $p(\boldsymbol{\theta})$, and the probability distribution of observed data given $\boldsymbol{\theta}$ ($p(\mathbf{y}|\boldsymbol{\theta})$). In this way, the posterior distribution expresses the experimenter's updated beliefs about $\boldsymbol{\theta}$ in light of the observed data $\mathbf{y}$.

The main controversy in the Bayesian approach lies in the preparation of prior information which is subjective (Ulrych et al., 2001). Three legitimate arguments for this subjectivity are: (1) it is natural that one's conclusion is affected by one's prior opinions, (2) the priors have little effect on the posterior when a large amount of data are available, and (3) the non-informative priors can be used to express ignorance about the unknown parameters which can be assumed to be objective (Hoff, 2009). Generally, there are two kinds of prior distributions frequently used. One is the non-informative priors that are commonly used when nothing is known about the value of parameters. Another is the informative priors or the empirical priors, which can be obtained from the previous evidence or empirical data.

Based on the previous study (Wagner et al., 2006), an individual waveform can be modeled with a mixture of Gaussian distributions. Therefore, it is natural to assign the distribution of the waveform as $p(\mathbf{y}|\boldsymbol{\theta})$ with a mixture of Gaussian distributions, which can be also interpreted as the likelihood of model. The shape of the waveform is determined by the parameters of the Gaussian distribution that can be easily obtained as the prior distribution of parameters through the peak identification algorithm (Zhou et al., 2017). The core of this algorithm is to identify the peaks by comparing the three adjacent intensities of the waveform and then selecting peak(s) when the corresponding peak intensity is higher than one-fifth of maximum intensity of the given waveform. Meanwhile, the number of peaks for the waveform is also obtained through this process.

Once we formulated the prior distribution of the parameters and the likelihood function, the concept of Eq. (2) can be used to derive the posterior distribution of model parameters through MCMC simulation (Gelfand and Smith, 1990). MCMC is a crucial technique for the rapid expansion of the Bayesian inference in science. There are cases that some parameters' posterior distributions are difficult or impossible to sample when the non-conjugate priors are used or the integration of parameters is conducted over a high dimensional parameter space (Hoff, 2009). In such conditions, the MCMC method can be helpful by approximating the true posterior distribution using the joint distribution $p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ instead of directly sampling from the integration of posterior distribution for parameters of interest $p(\boldsymbol{\theta}|\mathbf{y})$.

2.2.2. Waveform decomposition application

In a Bayesian context, the nonlinear model can be formulated in the following form

$$\mathbf{y}_i = f(\mathbf{x}_i, \theta)^* \epsilon_i \tag{2}$$

where y_i is the observed data, $f(\mathbf{x}_i, \boldsymbol{\theta})$ is a nonlinear function with parameters $\boldsymbol{\theta}$ and predictor $\mathbf{x}_i$, ϵ_i is an independent error with $\log\epsilon_i \sim N(0, \tau^2)$, τ is the standard deviation of $\log\epsilon_i$, and $\mathbf{x}_i$ is the ith time bin of the waveform. Each waveform is reconstructed in terms of the main model part ($f(\mathbf{x}_i, \boldsymbol{\theta})$) with a set of parametric functions and the error part. For the main model part, the coherence between proposed configurations and the real waveforms is measured. The multiplicative error is used here mainly because FW LiDAR data are typically restricted to be nonnegative (Gelman et al., 2015), and the multiplicative format is also convenient to formulate distribution for the error part. These formulations are consistent with the real FW LiDAR data as demonstrated in Fig. 2(a). There is a difference between the ideal Gaussian waveform (IGW) $f(\mathbf{x}_i, \boldsymbol{\theta})$ (black dash line) and the raw waveform (RW) $f(\mathbf{x}_i, \boldsymbol{\theta})^* \epsilon_i$ (purple line) that corresponds to error part $(\epsilon_i - 1)^* f(\mathbf{x}_i, \boldsymbol{\theta})$. The ideal Gaussian waveform is derived from Eq. (3) by summing four Gaussian components (j = 1, 2, 3 and 4) as shown in Fig. 2 with dash lines with different colors.

$$f(\mathbf{x}, \theta) = \sum_{j=1}^{n} A_j \exp\left(-\frac{(\mathbf{x} - u_j)^2}{2\delta_j^2}\right) \tag{3}$$

where n is the number of the Gaussian components, A_j, δ_j and u_j are the amplitude of the peak, the standard deviation and the time location of the peak for j^{th} waveform component, respectively. Eq. (3) gives rise to total $3 * n$ parameters associated with the number of Gaussian components. A_j, u_j and δ_j are restricted to nonnegative values.

To reduce the impact of detected "false" peaks of the raw waveform resulted from noise, especially at the beginning and tail of the waveform, a mean filter was conducted prior to subsequent processing. We called the waveform after filtering as the smoothed waveform (SW) (red line) that would be employed for subsequent analysis. A visual inspection showed that the smoothed waveform was nearly overlapping with the ideal Gaussian distribution waveform (Fig. 2(a)) which may justify the use of the Gaussian model for fitting FW LiDAR data. We explored more details of model choices in the Section 2.4 to quantitatively test whether this assumption was valid. The number of Gaussian

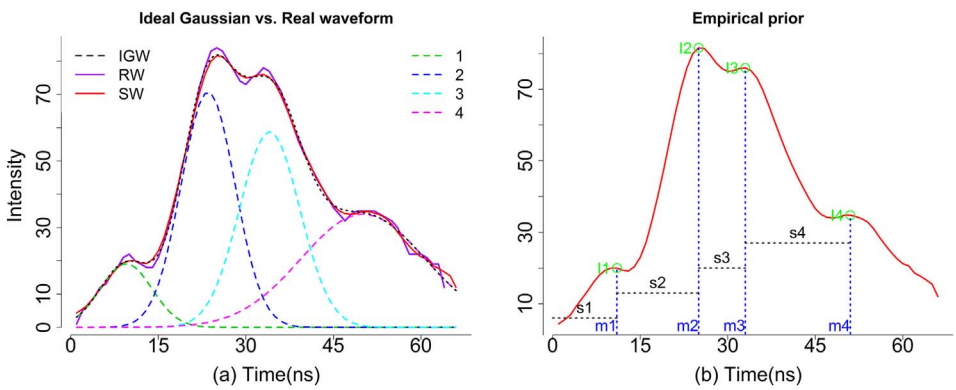

Fig. 2. (a). Illustration of ideal Gaussian distribution waveform (IGW, black dash) vs. real waveform (RW, purple) and smoothed waveform (SW, red). The number (1, 2, 3 and 4) represent the individual Gaussian components. (b). Empirical priors derived from the SW through peak identification algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

components, n, for each waveform varies depending on the number of peaks detected along the pulse line. The unknown parameters $\boldsymbol{\theta}$ include $A_1, u_1, \delta_1, \ldots, A_n, u_n, \delta_n$.

After log transformation of the Eq. (2), the log-likelihood of y_i can be written as:

$$\log \varepsilon_i = \log \mathbf{y}_i - \log f(\mathbf{x}_i, \theta) \tag{4}$$

$$p\,(\log \mathbf{y}_i \mid \theta, \tau) = N(\log f(\mathbf{x}_i, \theta), \tau^2) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\log \mathbf{y}_i - \log f(\mathbf{x}_i, \theta))^2}{2\tau^2}\right) \tag{5}$$

To compare the influence of prior information on the model performance, the non-informative priors and the empirical priors derived from the raw waveforms were assigned to the model. Here, the non-informative priors indicated that assigning equal probabilities to all possible values of parameter space, named it the flat priors in subsequent analysis.

Based on the statistical summary of raw FW LiDAR data, we narrowed down the reasonable range of these parameters to [10, 150], [15, 100] and [4, 15] as the parameter space of A_j, u_j and δ_j, respectively. Hence, a uniform distribution (U) was assigned to each parameter to express the ignorance of the effect of parameters' prior distribution on the outcomes, and the prior distribution for each parameter followed $A_j \sim U(10, 150), u_j \sim U(15, 100)$ and $\delta_j \sim U(4, 15)$. Another option for specifying the priors were to use the empirical priors that were derived from the corresponding SWs through peak identification algorithm (Zhou et al., 2017). This algorithm is mainly to estimate the number of Gaussian components n and approximated peak locations. According to Fig. 2(b), the time bin for the peak of one waveform component m_j (m_1, m_2, m_3 and m_4) was associated with the location of the peak that corresponds to u_j. The corresponding intensity I_j (I_1, I_2, I_3 and I_4) at the peak was related to A_j in Eq. (4). δ_j was much more difficult to interpret, therefore we used a third of the difference between consecutive peaks ($s_j/3$) to roughly represent prior information of δ_j. To sum up, we specified the prior distribution of A_j, u_j and δ_j to follow the normal distribution (N) with $A_j \sim N(I_j, 10^2)$, $u_j \sim N(m_j, 5^2)$, $\delta_j \sim N(s_j/3, 3^2)$ and $\tau \sim N(0, 0.5^2)$.

A log posterior distribution of the model was obtained through all prior information about parameters of interest and the data distribution $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}, \tau)$. For the flat priors, the posterior distribution was the likelihood of data (Eq. (6)). For the empirical priors, the posterior distribution of model was expressed in Eq. (7):

$$p(\theta, \tau \mid \mathbf{y}) \propto p\,(\mathbf{y} \mid \theta, \tau)\, p(\theta)\, p(\tau)$$

$$\propto \prod_{i=1}^{m} \frac{1}{\tau} \exp\left(-\frac{1}{2\tau^2}(\log \mathbf{y}_i - \log f(\mathbf{x}_i, \theta))^2\right)^{*} p(\theta) p(\tau)$$

$$\text{Flat priors:} \propto \frac{1}{\tau^m} \exp\left(-\frac{1}{2\tau^2} \sum_{i=1}^{m} (\log \mathbf{y}_i - \log f(\mathbf{x}_i, \theta))^2\right) \tag{6}$$

Empirical priors

$$:\propto \frac{1}{\tau^m} \exp\left(-\frac{1}{2\tau^2} \sum_{i=1}^{m} (\log \mathbf{y}_i - \log f(\mathbf{x}_i, \theta))^2\right) \prod_{j=1}^{n} p_1(A_j) p_2(u_j)\, p_3(\delta_j)\, p(\tau) \tag{7}$$

$$p_1(A_j) \sim N(I_j, 10^2);\ p_2(u_j) \sim N(m_j, 5^2);\ p_3(\delta_j) \sim N\left(\frac{s_j}{3}, 3^2\right);\ p(\tau) \sim N(0, 0.5^2)$$

It is reminded that m is the number of observations for each waveform and n is the number of Gaussian components of the corresponding waveform.

2.3. Model implementation

Our aim was to decompose the waveform data using the above models to perform inference about quantities of the unknown parameters of interest. The model was implemented in R using the brms packages (Buerkner, 2016) which can fit generalized non-linear mixed models using Stan by performing the Bayesian inference and the optimization for the user-specified model (Gelman et al., 2015). Stan is a C++ program to perform Bayesian inference which is composed of four main blocks: variable declarations, parameter statements, transformed parameters and model blocks. Detailed descriptions of model structure and procedures are given in Appendix A.

2.3.1. Model dialogistic

We measured the model's convergence using the potential scale reduction factor, named Rhat ($\hat{R}$), which is a statistical criterion to test how well the Markov Chains are mixing, or moving around the parameter space. $\hat{R}$ close to one indicates convergence, while high $\hat{R}$ value implies that we should run a longer chain to improve convergence to the stationary distribution. The effective sample size was also generated to represent the equivalent number of independent iterations of the chain. It is a criterion for the estimation efficiency. Generally, the higher the effective sample size, the more reliable estimates can be achieved (Gelman et al., 2014).

2.3.2. Model inference

The total samples may have divergent values before the chain reaches a stationary state, therefore the model inference was conducted on the posterior samples after dropping burn-in samples. Generally, through these posterior samples, the distribution of parameters ($\boldsymbol{\theta}$) and summary measures, such as mean, mode, standard deviation and percentiles for each parameter can be derived. The main advantage of simulating from posterior samples is that we can generate as many values as we wish and thereby minimize errors in approximating quantities of interest. For each parameter, these draws could be used to approximate credible interval (CI) as shown in Fig. 3(b).

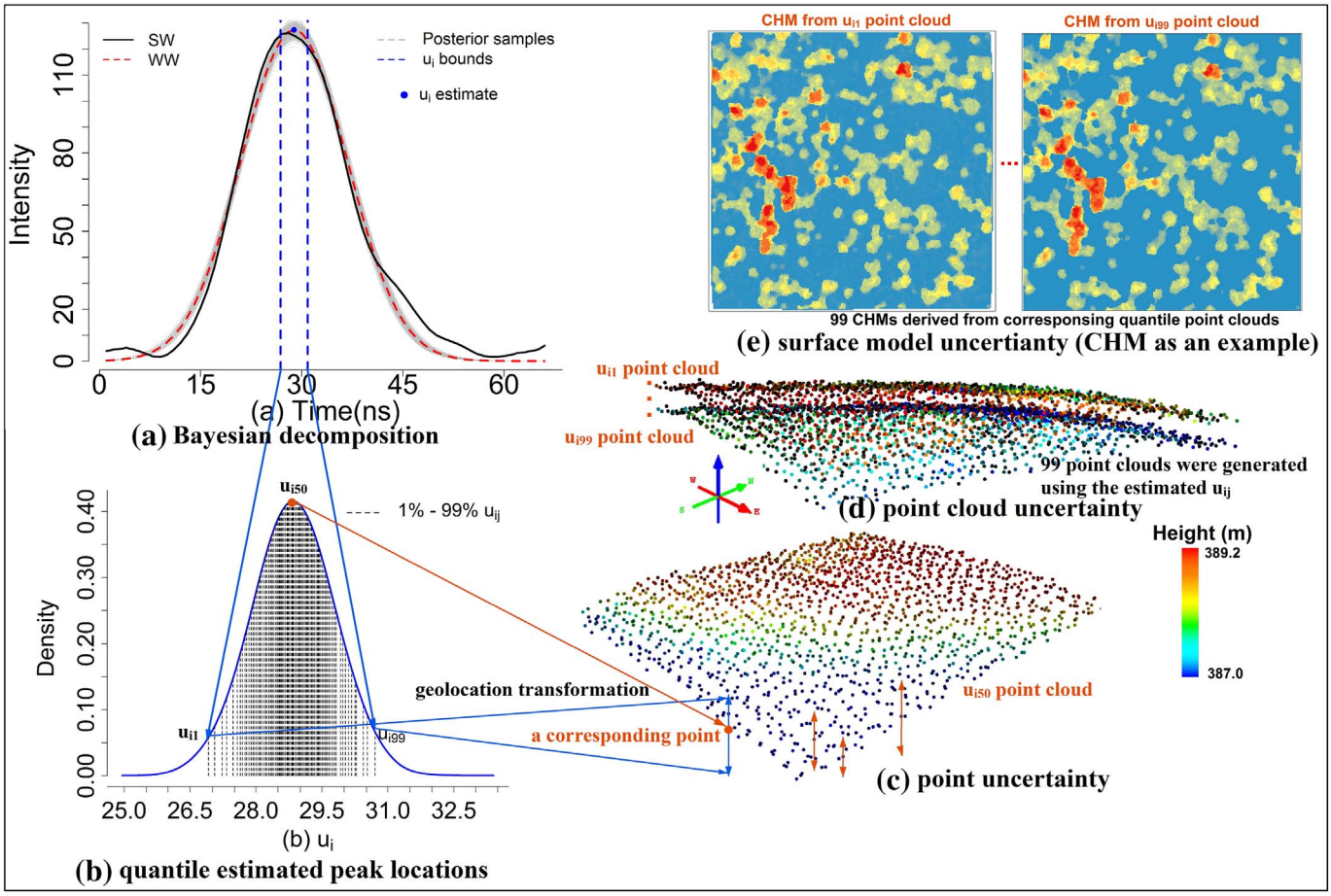

Fig. 3. Illustration of uncertainty propagation from data to the parameter estimates, point, point cloud and surface model such as CHM using Bayesian decomposition method. (a) The uncertainty of peak location (parameter uncertainty) using Bayesian method to fit the waveform. SW (black) represents the original waveform after smoothing, WW (red dash) represents the waveform using the mode estimates of Bayesian method, and the gray shadow represents the possible solutions for fitting the waveform. (b) The 99 quantile estimates of the possible peak locations from the u_i posterior distribution. (c) The point uncertainty propagated from the parameter uncertainty through geolocation transformation at a sample region with the background of u_{i50} point cloud. (d) Possible point clouds generated from 1% quantile estimate and 99% quantile estimated peak locations as examples. (e) Possible surface models such as Canopy Height Models (CHMs) generated from u_{i1} quantile point cloud and u_{i99} quantile point cloud as examples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.4. Model reasonableness

The Gaussian function has been mostly used to decompose the waveform LiDAR data, under the implicit assumption that the Gaussian model is capable of reconstructing detected objects based on waveform shapes. However, few studies have quantitatively or statistically justified the reasonableness of this assumption.

According to the previous study (Mallet et al., 2009), several models that could be used to fit waveform LiDAR data. In this study, we employed three representative models to explore the reasonableness of models in a Bayesian context: the Weibull model, the Adaptive Gaussian model and the Gaussian model. The Gaussian model is most frequently used model for waveform decomposition. The Adaptive Gaussian distribution has the form as Eq. (8) which can minimize the residual of the model by introducing another variable which is also known as rate parameter (λ).

$$f_{AG}(\mathbf{x}, \theta) = \sum_{i=1}^{n} A_i \exp\left(-\frac{(\mathbf{x} - u_i)^{\lambda}}{2\delta_i^2}\right) \tag{8}$$

In this study, the rate parameter λ, as a stochastic variable, was assigned to follow normal distribution N ~ (2, 0.25), because the rate parameter in the Gaussian model is 2 and the Adaptive Gaussian model's rate parameter should be close to this value.

The Weibull model was introduced since it enables us to simulate either symmetric or asymmetric peaks with four unknown parameters. This model has been successfully applied for Synthetic Aperture Radar (SAR) image processing (Tison et al., 2004). The Nakagami and Bur functions (Mallet et al., 2009) are also capable of simulating the waveform shape with four parameters as the Weibull function. These three functions share the same feature that all can simulate asymmetric and symmetric waveforms with the same number of parameters. Here, the Weibull function was selected to represent this class of potential models for the waveform decomposition. The Weibull distribution function used here can be written as:

$$f_W(\mathbf{x}, \theta) = \sum_{i=1}^{n} A_i \frac{k}{\delta_i}\left(\frac{\mathbf{x} - u_i}{\delta_i}\right)^{k-1} \exp\left(-\left(\frac{\mathbf{x} - u_i}{\delta_i}\right)^{k}\right) \tag{9}$$

where A_i is the amplitude, k (> 0) is the shape parameter that controls the behavior or the shape of the distribution, and δ_i (> 0) is the scale parameter that controls the spread of the distribution. The shape parameter can capture the asymmetry or skewness of the waveforms that overcomes the disadvantage of the Gaussian function, which is only suitable for symmetric distributions. u_i is a location parameter in the Weibull model. However, this parameter is not useful in waveform interpretation and subsequent geolocation transformation, since it does not have any physical meaning in our case.

The predictive accuracy of models is generally measured with the deviance information criterion (DIC) for the Bayesian model selections.

In the present study, a more relevant criterion, the Watanabe-Akaike information criterion (WAIC), was adopted to provide a basis for model assessment and model selection (Vehtari et al., 2016). WAIC is a fully Bayesian criterion that uses posterior distribution of existing simulation draws rather than a point estimate to approximate leave-one cross validation for estimating pointwise out-of-sample prediction accuracy.

One hundred forty waveforms with a different number of waveform components that represented different levels of model complexity were randomly selected as samples to check the model reasonableness of waveform decomposition. Each waveform was fitted with the above three models ($f(\mathbf{x},\boldsymbol{\theta})$, $f_{AG}(\mathbf{x},\boldsymbol{\theta})$, and $f_W(\mathbf{x},\boldsymbol{\theta})$) and WAIC, uncertainty bounds and residual standard error of model (SE) were reported. The model choice would be based on these criteria, physical meaning of corresponding parameters and processing efficiency.

2.5. Model efficiency

There are several commonly used MCMC methods such as Gibbs sampling, Metropolis algorithm, and Metropolis-Hastings (MH) algorithm. In this study, the MCMC simulation was achieved using the Hamilton Monte Carlo (HMC) algorithm to enhance the model efficiency.

The HMC is a relatively new MCMC algorithm that applies the concept of Hamiltonian dynamics to Metropolis update for simulating a Markov chain (Neal, 2011). The No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014) was used to implement the sampling procedures of HMC and more details can be found in Appendix B.

In addition, the effect of different priors on the efficiency of the model was also explored using the same sample waveforms as used in Section 2.4. Each waveform was decomposed twice using the flat priors and empirical priors with the Gaussian model. The computation time for each model was recorded and the average time for given number of waveform components was reported. In this step, the number of iterations and burn-in had been assumed to be adequate to derive the accurate approximation of parameters no matter which prior was used. According to Table B1, the empirical priors converge faster. As a result, we used the empirical priors to process all other waveforms.

2.6. Geolocation transformation

The Bayesian decomposition can provide estimated quantiles (u_{ij}, from 1% to 99% of posterior samples including mode estimate u_{i50}) for possible target time bin locations (u_i), corresponding standard error, effective sample size and $\widehat{R}$ of A_i, u_i, σ_i for each waveform. The 3D point clouds were generated by combing the original georeferenced data such as x_0, y_0, z_0, dx, dy and dz. provided by the NEON datasets with the estimated time bin (u_i, u_{i1}…u_{i99}). The leading edge position for each detected peak was used to compute the geolocation of desired time bin by incorporating the full width at half maximum (FWHM).The detailed calculation processes are given in (Zhou et al., 2017).

2.7. Performance evaluation

Both the DD and Bayesian decomposition methods can be classified as the decomposition method instead of the combined deconvolution and decomposition method in terms of processing steps. However, the DD method is different from the Bayesian decomposition method which belongs to the probabilistic approach, but the DD, Gold and RL methods are the deterministic approach which only has one estimate for a parameter or possible target position. Bayesian decomposition method here generates the distributions for the parameters, from which multiple possible estimates for the parameter values could be obtained e.g., using the quantiles. In our case, we generated 99 possible quantile estimates for the parameter u_i with probability as shown in Fig. 3(b), which resulted in 99 possible point clouds after geolocation transformation. To compare the performances of the Bayesian decomposition method with deterministic methods, the point cloud with highest probability using the mode estimate (u_{i50}) for geolocation transformation was selected to conduct the performance evaluation.

The Bayesian decomposition cannot converge for the extremely irregular or noisy waveforms, which results in some noisy points after geolocation transformation. With the aid of LAStools (Isenburg, 2012), the noisy points were deleted before conducting the method comparisons of the point cloud. In this study, the point cloud obtained using mode estimates (u_{i50}) with Bayesian decomposition method was compared with the DD, Gold and RL approaches from the previous study (Zhou et al., 2017) at two different levels: point cloud and individual tree's metrics such as the number of points at various height bins, percentile heights and canopy point density.

2.7.1. Point cloud comparisons

Point clouds are the primary result of waveform decomposition and their accuracy significantly influences the quality of their derived products such as percentile heights and Digital Terrain Models (DTMs). Thus, we computed the Hausdorff distances (Mémoli and Sapiro, 2004) between the waveform-based point cloud derived from different methods and the DR LiDAR point cloud. This comparison was named C2C in subsequent analysis. The thrust for the Bayesian decomposition method is to avoid the error brought by interpolation and variability of area based products such as DTM caused by various grid cell sizes. In addition, the point cloud comparison is a natural and direct way to evaluate the surface representation without adding any intermediate step (Mémoli and Sapiro, 2004).

The principle of the Hausdorff distance is that for each point of a compared cloud, the nearest neighbor method is used to search the nearest point in the reference cloud (the DR point cloud) and then compute their Euclidean distance. This process was implemented in the CloudCompare software (Girardeau-Montaut, 2015). Meanwhile, all points' X, Y, Z differences were also generated for the Bayesian decomposition, DD, Gold and RL approaches.

2.7.2. Individual tree metrics comparisons

In addition to the comparisons for the grand picture such as point clouds, the comparisons at the individual tree level were also explored to present a more comprehensive comparison of methods' performances. Individual tree dimension's LiDAR metrics such as percentile heights, median height and crown density are crucial for characterizing canopy structure and estimating biomass (Falkowski et al., 2009; Popescu, 2007; Zhao et al., 2011). We randomly selected 121 trees from the SJER to compare their total number of points, the number of non-ground points (the elevation larger than reference DTM), the non-ground canopy point density, the percentile heights and median height using the Bayesian decomposition method with corresponding DR LiDAR data results. Additionally, the results of individual tree level from the DD, Gold and RL approaches were also incorporated into comparisons that had been done in the previous study.

Due to fewer points at lower heights corresponding to the understory, the height bin width was 2 m when the tree height was below 4 m, and it became 1 m when the tree height was larger than 4 m. The total number of points and corresponding ratio of points in each height bin were summarized. To further compare the performance of different waveform processing methods, the normalized percentile heights (subtracting the minimum elevation for each tree boundary) were calculated. Given the point cloud of one tree, ten height metrics including 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th and 100th (maximum height) percentile heights were extracted to demonstrate the vertical structure of vegetation based on the height of LiDAR points. These metrics not only help to predict biomass, but also can quantitatively measure the waveform LiDAR data's penetration advantage.

The canopy point density for individual trees was also analyzed for different approaches since it was beneficial to map tree stem and crown (Lee and Lucas, 2007). We randomly selected 21 trees from 121 trees to

show detailed results of the comparisons. Here, the canopy point density was represented by the number of above ground points per square meter.

2.7.3. Field data calibration

What the above comparisons share is that DR LiDAR data were adopted as reference data that had been successfully applied in many studies (Allouis et al., 2013; Chen, 2007; Zhou et al., 2017) for its high accuracy of measuring height. In this study, we not only compared the point cloud and individual trees' metrics derived from FW LiDAR data to corresponding DR LiDAR (reference) provided by the NEON, but also used the field-measured tree height to evaluate results. To facilitate comparisons of different methods' performances, the average bias, standard deviation and Root Mean Squared Error (RMSE) of their differences were computed. According to 151 field-measured trees' locations, the values from waveform-based Canopy Height Models (CHMs) using u_{i50} were extracted. To make our extracted values more representative, we generated a 1 m buffer for each location and then averaged the values fell in each buffer.

2.8. Uncertainty analysis

2.8.1. Uncertainty propagation

The essential feature of the Bayesian approach is the explicit quantification of uncertainty introduced by incorporating multiple levels of randomness or various sources of errors. Through estimating the predictive parameter uncertainty, rigorous error propagation along the processing steps can be quantified. In this study, there were different sources of uncertainty originating from data themselves and along the processing steps, which would accumulate and propagate to the final products. Most of the previous studies are based on the deterministic models or approaches that only seek a single value and ignore the noise or error inherent in the data and approaches. To this end, we conducted a comprehensive uncertainty analysis to quantify the uncertainty of results at different steps including parameter estimates, point cloud generation and surface model generation (Fig. 3).

Through the probability distribution of inference using the Bayesian approach, the summary measures of unknown parameters such as mode, percentiles and standard deviation are obtained instead of a single estimate. Unlike the deterministic method, the Bayesian credible region is characterized by the 95% highest posterior density (HPD) region rather than 95% confidence interval. The main difference is that the HPD region can be discontinuous when the parameter's posterior density is multimodal or asymmetric (Hoff, 2009), while the confidence interval region is always continuous. At this stage, the Gaussian model had been chosen to fit the waveforms that gave us suitable physical interpretation and accurate estimation of parameters. The posterior density of the individual parameter was generally used to obtain their CIs. The fitting functions used here followed the normal distribution which was symmetric. Thus, we directly used the empirical quantiles of the posterior samples to approximate the uncertainty of the peak locations (blue dash) as displayed in Fig. 3(a). The value of peak location was regarded as a realization of MCMC process. The distribution of possible peak locations with probability from the posterior samples after Bayesian decomposition was shown in Fig. 3(b). The estimated quantile peak locations u_{ij} starting from 1% to 99% of posterior samples were chosen to conduct geolocation transformation to derive possible points located along the blue arrow (Fig. 3(c)). We used the mode estimate u_{i50} of possible peak locations to generate the point cloud as the background in Fig. 3(c), and several points with different length of arrows were selected as examples to demonstrate the uncertainty of the corresponding point when we used their quantile estimates ($u_{i1} \ldots u_{i99}$). However, quantifying uncertainty of these individual points separately was not useful to some extent, since most of the subsequent studies were conducted on products derived from these points instead of individual points, such as the point clouds (Fig. 3(d)), DTM and CHM (Fig. 3(e)). The u_{ij} quantile point cloud was composed of the points derived from the geolocation transformation using u_{ij} for all waveforms located in the region. Fig. 3(d) demonstrates the u_{i1} and u_{i99} point clouds to demonstrate the possible maximum uncertainty of point cloud products in a sample region. Due to the relatively small differences among 99 point clouds, we showed only two point clouds. As mentioned earlier, the Hausdorff distance was employed to calculate the distance of these waveform-based u_{ij} quantile point clouds and the DR LiDAR point cloud. For each u_{ij} point cloud, the average bias (mean), standard deviation and RMSE of X, Y, Z and point distances were derived. The uncertainty of point cloud distances was characterized by the distribution of average bias and RMSEs derived from these quantile point clouds. A similar method was also conducted on the waveform-based surface models such as DTM and CHM. More specifically, the DTM for each quantile point cloud were generated through the *las-ground* and *las2dem* implemented in LAStools after deleting noisy points. Regarding the CHM, we followed the steps described by Khosravipour et al. (2014) and implemented these steps in LAStools to obtain the 99 quantile CHMs. To further quantify the performance of these methods, we compared the waveform-based DTMs and CHMs from quantile point clouds with corresponding reference data and computed the evaluation criteria such as the average bias and RMSE. The graphical and statistical methods were employed to analyze the comparison results and their corresponding uncertainties.

2.8.2. Uncertainty of accuracy assessment with field data

A limited number of field observations and imprecise field-measured data significantly affect the calibration results and accuracy assessment (Yao et al., 2012). Both require us to conduct uncertainty analysis of field data to enhance the credibility of calibration. However, the uncertainty of field data is difficult or impossible to quantify since data providers do not clearly state which error sources were considered. In this study, we assumed the field data were “true”, and the uncertainty of estimated tree height using the Bayesian decomposition method was analyzed. Specifically, 1 m buffers generated from field-measured individual tree locations (X and Y) were first used to extract possible tree points from the u_{i50} point cloud. For each buffer, the Z values of the points above 95th percentile height were selected as the possible tree height and these points were employed to identify the waveform(s) that fell in the tree region. The uncertainty of these individual trees' height was quantified through the 95% CI of peak locations from selected waveform(s) after the Bayesian decomposition. Moreover, we calculated the uncertainty of RMSE for the estimated tree height based on the nearest possible value and farthest possible value from each tree's 95% CI to the field-measured data. An overview of the proposed methodology was generated to summarize the major steps implemented in this paper as shown in Fig. 4.

3. Results

3.1. Model reasonableness

We modeled 140 sample waveforms with different complexities (the number of components) using the Weibull, Adaptive Gaussian and Gaussian models within the Bayesian framework, and three representative examples of decomposition results including waveform components (colored dash lines), uncertainty (gray shadow), the corresponding WAIC and SE of model are demonstrated in Fig. 5. As expected, all models worked well with a relatively small difference when we visually inspected the SW and MW. There was no noticeable loss of the fitting accuracy using these three models when the waveform was relatively simple.

However, the Weibull function model became less fitted and yielded a wider gray shadow with larger uncertainty when the waveform had higher number of Gaussian components. The statistics summary of the models further confirmed this inspection, as the Weibull model always

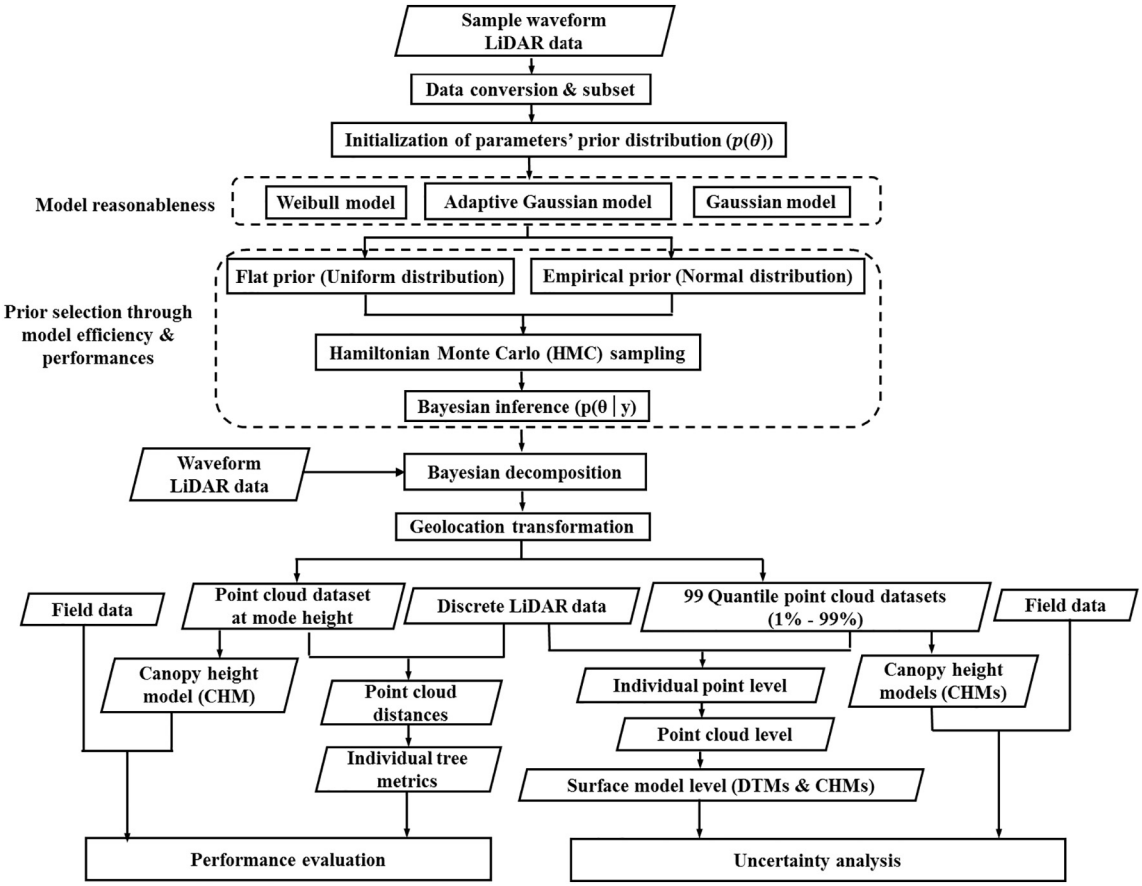

Fig. 4. Flowchart for the Bayesian decomposition of waveform LiDAR data with uncertainty analysis.

achieved the largest WAIC and SE given the same waveform. Moreover, most of the waveforms in the study site were close to symmetric distributions, which diminishes the advantage of the Weibull model that is capable of modeling the asymmetric waveforms. Among these three models, the smallest WAIC was achieved using the Adaptive Gaussian model to fit complex waveforms with more than one peak. However, this model might face the problem of overfitting, since the rate parameter (λ) of the Adaptive Gaussian model for each waveform was adjusted mainly for minimizing residuals of the fitted model. Consequently, the noise contained in the data might be considered as the main model part ($f(\mathbf{x}_i, \boldsymbol{\theta})$) instead of the error part. The experiment of sample waveforms showed that 68 out of 95 waveforms with one component could generate the smallest WAIC using the Gaussian model. Additionally, over 75% of all waveforms in the SJER study site were considered to be one component according to the inspection of the waveform components (n). These gave us more confidence to utilize the Gaussian model instead of the Adaptive Gaussian model to perform waveform decomposition in terms of model accuracy and uncertainty.

In addition, the physical interpretation of the parameters was also limited when we applied the Weibull model or Adaptive Gaussian model. For the four-parameter Weibull model, the estimates can't be employed as a location like Gaussian model's peak location to calculate the geolocation of the extracted points. It was meaningless or difficult to interpret the rate parameter of the Adaptive Gaussian model from the decomposition perspective and geolocation transformation. Furthermore, the experiment of these waveforms showed that the Weibull and Adaptive Gaussian models took more time to find optimized parameters and reach convergence for an additional parameter. Therefore, we concluded that the Gaussian distribution model was the most suitable model for FW LiDAR data decomposition based on the accuracy, uncertainty, physical meaning and processing efficiency.

3.2. Performance evaluation

3.2.1. Point cloud comparisons

The C2C validation of the SJER1 study region was executed as displayed in Fig. 6(b). To associate the C2C distances across the study site with the vegetation distribution, the CHM derived from the Bayesian decomposition method (Fig. 6(a)) was plotted against the C2C distances' spatial distribution pattern. It was worthy to note that larger C2C absolute distances (> 0.5 m) were more likely to occur at the vegetation part with higher CHM values. The ground region was the most accurate portion with relative small distance when we compared waveform-based point cloud to the corresponding DR LiDAR point cloud.

Fig. 7 depicts the distribution of point cloud distances, horizontal (X, Y) and vertical (Z) distances between the waveform-based point cloud using the mode dataset with the Bayesian decomposition method, and the DR point cloud. When examining these three 1D distances (gray), it was evident that the horizontal (X and Y) and vertical (Z) distances' distributions were symmetric around 0 m with almost the same distribution. However, a closer examination revealed that compared with the horizontal distances, the vertical distances were greater, but not markedly greater, with larger SD and RMSE. These three 1D distance's distribution together contributed to the point distances' distribution (3D) with the mean point distances and their corresponding RMSE were 0.51 m and 0.67 m, respectively. As expected, the absolute MDs and RMSEs of X, Y, Z coordinates' C2C distances were all smaller than point distances.

To further illustrate, more detailed C2C point cloud comparisons between waveform-based point clouds using four methods (the Bayesian decomposition, DD, Gold and RL) and reference point cloud were summarized in Table 1. All methods generated satisfactory point

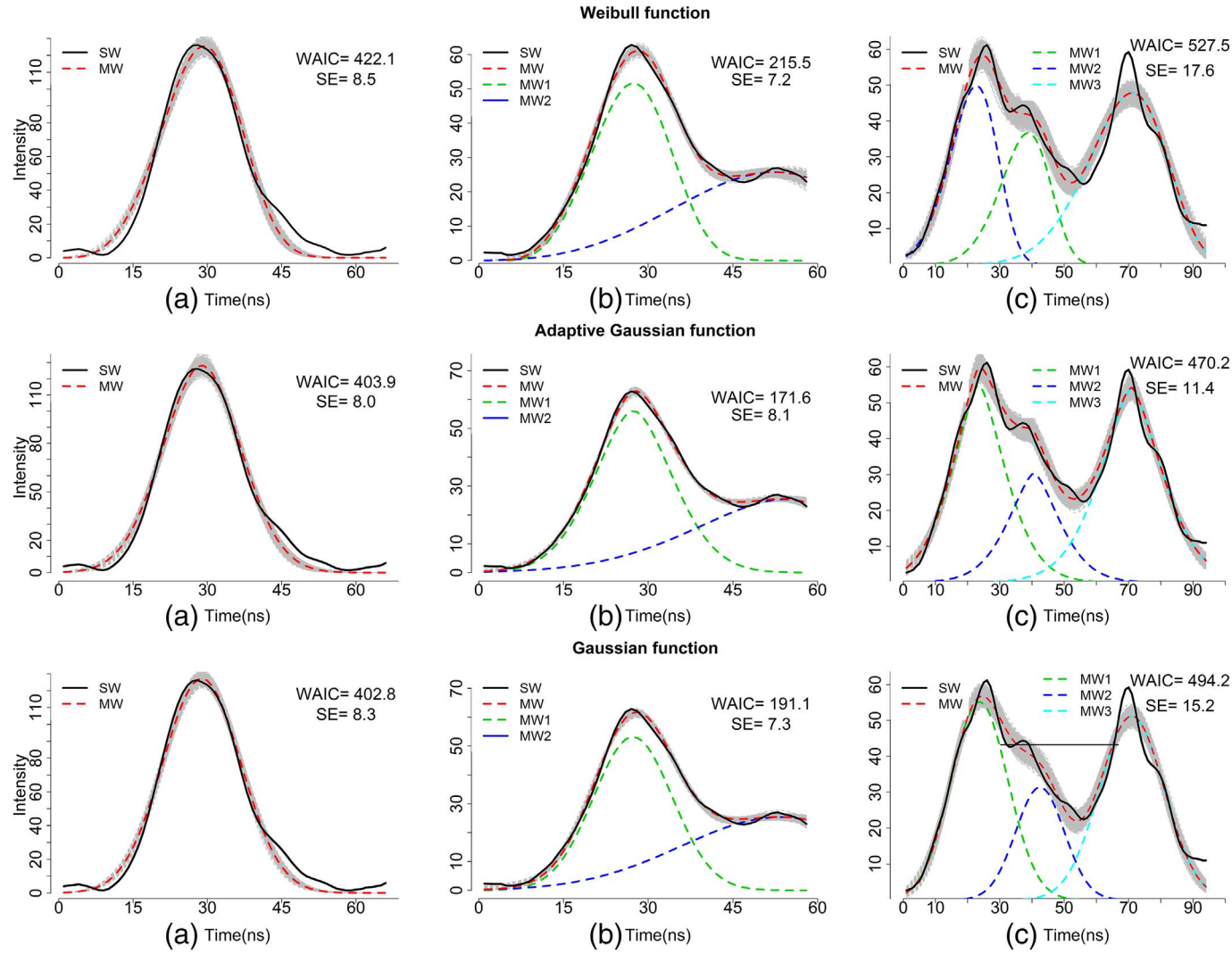

Fig. 5. Examples of Bayesian decomposition of FW LiDAR data using three models (the Weibull, Adaptive Gaussian and Gaussian models) and their corresponding uncertainties (gray shadow). Black solid line represents the smoothed waveform (SW) using a mean filter, red dash line represents the modeled waveform (MW), and other color dash lines represent modeled waveform components. WAIC is the Watanabe-Akaike information criterion and SE is the residual standard error of the model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

clouds with acceptable mean point distances (< 0.84 m) and RMSEs of point distances (< 0.93 m) compared to the DR point cloud. Upon closer examination of results with different methods, the DD method outperformed the other three methods with the smallest average point distances, X, Y, Z distances, and corresponding RMSEs. Especially for the horizontal distances (X and Y), there was a tiny difference of the four methods in terms of the MD and RMSE. However, a relatively larger difference of the methods' MD and RMSE occurred at the vertical direction (Z) that consequently leads to the same pattern of four methods' point distances.

3.2.2. Individual trees' metrics

This section provided a comparison and evaluation of the FW LiDAR processing methods at the individual tree level. To demonstrate the robustness of approaches and reduce the selection bias, 121 randomly selected trees' height bins, percentile heights and canopy point densities were generated (see examples in Fig. 8 and Table 2). There were two patterns observed with respect to the point density of individual trees. Thus, two representative trees derived from waveform-based point clouds using the Bayesian decomposition, DD, Gold and RL approaches were chosen to demonstrate these patterns: one tree with higher point density (Tree 1), and another tree with lower point density (Tree 2) were compared to the DR LiDAR data. Overall, about 91% of all selected waveform-based tree point clouds had more dense point clouds than their corresponding DR LiDAR data. As shown in Table 2, 16 out 21 trees' waveform-based point densities were higher than the corresponding region's DR LiDAR data, which followed the pattern of the Tree 1. Fig. 8 shows the absolute and normalized point frequency in each height bin of FW LiDAR data using four approaches and DR LiDAR data. As expected, more points were extracted from FW LiDAR data with these four approaches when examining the Tree 1's middle height bins from 6 to 22 m. The Gold and Bayesian decomposition methods outperformed other methods from the perspective of the number of points extracted from the mid-story of the tree.

This trend was not so obvious for the Tree 2, where higher point frequency in most of the height bins was observed using DR LiDAR data rather than FW LiDAR data, especially the absolute point frequency. While the normalized point frequency demonstrated advantages of FW LiDAR data for characterizing the mid-story of the tree, but the evidence was not as strong as for the Tree 1. The common feature of all trees shares was that DR LiDAR can detect more points in the lower part of their height or on the ground than FW LiDAR. This might be attributed to the fact that the tree had a dense canopy and the transmitted energy rarely reached the ground.

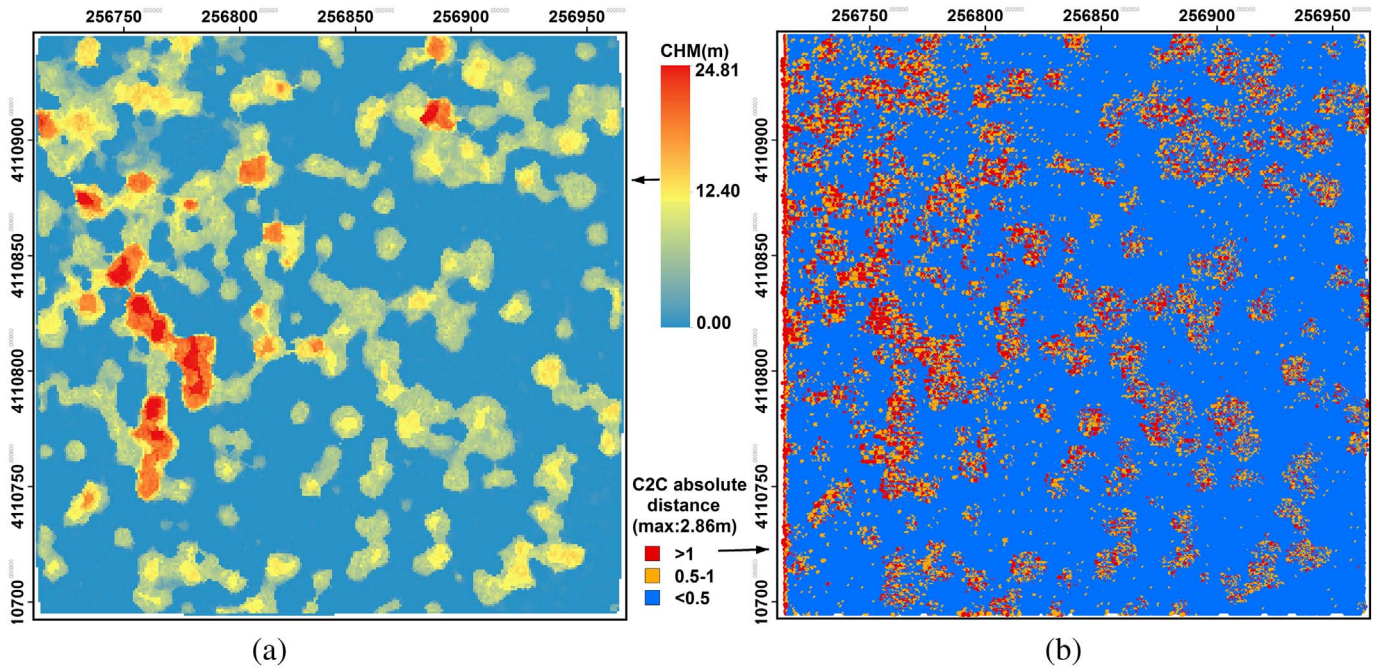

Fig. 6. (a). The canopy height model (CHM) generated from Bayesian approach (Left). (b). The spatial distribution of the distance between waveform-based point clouds using Bayesian decomposition method and the DR point clouds (C2C) at SJER1 region (Right)

The percentile heights' results for these two examples demonstrated similar trend as height bin results (Fig. 9). For Tree 1, the significant differences among different processing methods occurred around the median percentile, especially for the Bayesian decomposition and DD methods. Surprisingly, the four approaches and DR LiDAR data's percentile heights reached an agreement after 80th percentile height. This may indirectly imply that more points were extracted from the waveforms at the mid-story of tree height.

To reduce the ground points' effect on the comparisons, the trees' non-ground part was used to compare different methods' performances. We used 21 representative trees from the SJER1 to demonstrate comparison results. According to Table 2, only 5 out of 21 trees' DR LiDAR data results (with bold) yielded higher canopy point density than waveform LiDAR data using these four methods. Additionally, the DR LiDAR results detected more ground points than waveform-based results as shown in Table 2. There was no evident trend for the other three methods, however, the canopy point density results of all methods indicated that waveform LiDAR data can provide more non-ground points than DR LiDAR data. A closer examination of five individual trees with fewer points than corresponding DR LiDAR data shows that three of them (the Tree 2, Tree 12 and Tree 20) are located in the regions with fewer flight lines overlaid which may mainly contribute to the reduction of points detected.

3.2.3. Field data calibration

The accuracy of the maximum height derived from DR-based and waveform-based CHMs were assessed by comparing the field-measured data at the SJER1 study region. Overall, there was no significant difference among four waveform processing methods with regards to the average bias, standard deviation and RMSE. All of these methods

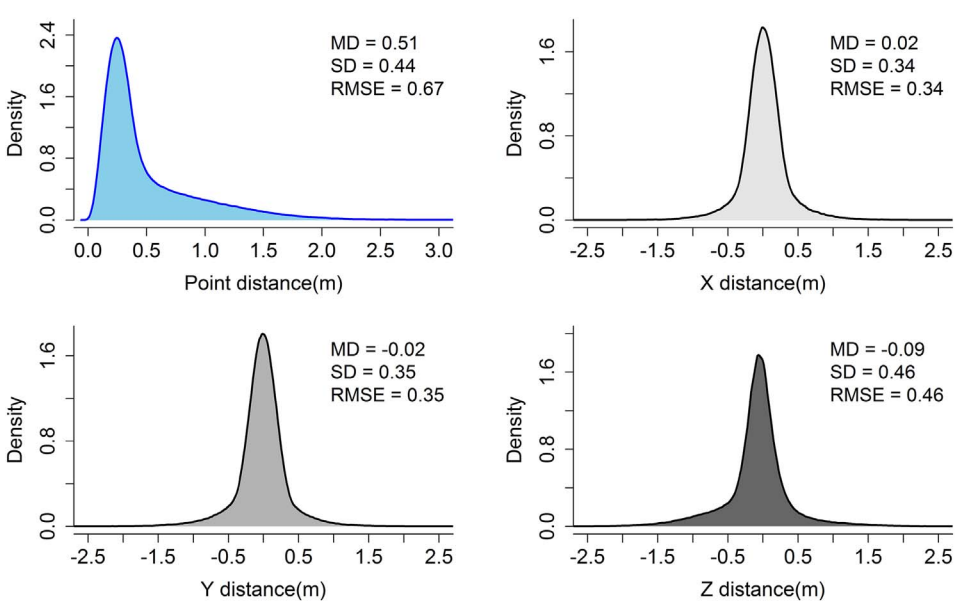

Fig. 7. Probability distribution of the C2C point distances (sky blue), horizontal (X, Y) and vertical (Z) distances (gray) using mode dataset with the Bayesian decomposition method at the SJER site. MD: Mean distances between waveform-based point clouds and DR point clouds. SD: Standard deviation. RMSE: Root mean square error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary statistics of the distances between waveform-based point clouds with the four methods (Bayesian decomposition (Bayesian), DD, Gold and RL) and the reference point cloud at the SJER site.

Items	Methods	MD	SD	RMSE	MAX	MIN
Point distances(m)	Bayesian	0.51	0.44	0.67	5.07	0.00
	DD	0.48	0.37	0.61	3.98	0.01
	Gold	0.78	0.36	0.86	4.74	0.00
	RL	0.84	0.40	0.93	4.57	0.01
X distances(m)	Bayesian	0.02	0.34	0.34	3.74	− 3.73
	DD	− 0.01	0.30	0.30	2.45	− 3.60
	Gold	0.00	0.36	0.36	3.64	− 3.68
	RL	0.00	0.39	0.39	3.47	− 3.44
Y distances(m)	Bayesian	− 0.02	0.35	0.35	3.69	− 3.81
	DD	0.02	0.29	0.29	2.59	− 3.00
	Gold	0.05	0.34	0.35	3.23	− 3.59
	RL	0.05	0.38	0.38	3.71	− 3.46
Z distances(m)	Bayesian	− 0.09	0.46	0.46	4.37	− 4.62
	DD	− 0.17	0.42	0.45	3.24	− 3.07
	Gold	− 0.48	0.51	0.70	3.98	− 4.61
	RL	− 0.47	0.59	0.75	3.56	− 4.23

*MD: the mean point distances between waveform-based point clouds and DR point clouds. SD: the standard deviation of distances. RMSE: the root mean square error of distances. MAX: the maximum of distances. MIN: the minimum of distances.

generated comparable and acceptable results compared with the field-measured data. Specifically, the comparison between DR LiDAR data and the measured data produced the smallest RMSE with 1.11 m, and the DD method consistently yielded the least accurate results with the largest average bias and RMSE. The superior waveform processing method varied with different statistics criteria used: the RL method was superior with the smallest RMSE (1.35 m), and the Gold method outperformed others with the smallest average bias.

3.3. Uncertainty analysis

3.3.1. Individual parameter uncertainty

Fig. 10 shows an example of the trace plot for three parameters' sampling processes and the marginal posterior distribution of the parameters (A, u, δ) using the flat priors and empirical priors. From the trace plot (Left), we can see an obvious difference at the beginning of the parameter sampling and the flat priors (sky blue) takes more steps to reach the stable status, especially for A and δ. Given the same number of iterations, the distributions of the parameters were nearly symmetric following the normal distribution with the empirical priors (gray), while the results of the flat priors (sky blue) were not symmetric which implied the posterior sampling of the parameters did not reach the stationary stage. The Model efficiency experiment using different parameters' prior distribution showed that the performance of the flat priors can reach the same level as the empirical priors with more burn-in steps and total iterations.

From Table 4, a general trend has emerged that more time is taken to process the waveform with a larger number of components no matter which priors were used. Interestingly, there was an abrupt rise of time cost when the waveform's number of components became four. The comparison between the flat priors and the empirical priors demonstrated that there was little difference of time cost when the waveform was relatively simple with one or two waveform component(s). However, the computation time using the empirical priors was much shorter than when using the flat priors for complex waveforms. This reduction of time could make a substantial difference when millions of waveforms need to process and the efficiency of using the empirical priors will become more evident. Without the consideration of the computation cost and time, the impact of different priors on the parameter estimates and the performance of the Bayesian models were negligible.

3.3.2. Point cloud uncertainty

The PDFs of descriptive statistics (Mean and RMSE) for point distances (sky blue), X, Y and Z distances (gray) through comparing the waveform-based LiDAR quantile point cloud datasets with the DR dataset are displayed in Fig. 11.

The uncertainty of average bias and corresponding RMSE for point distances were larger than the other three individual coordinates' descriptive statistics when examining the mode and range of

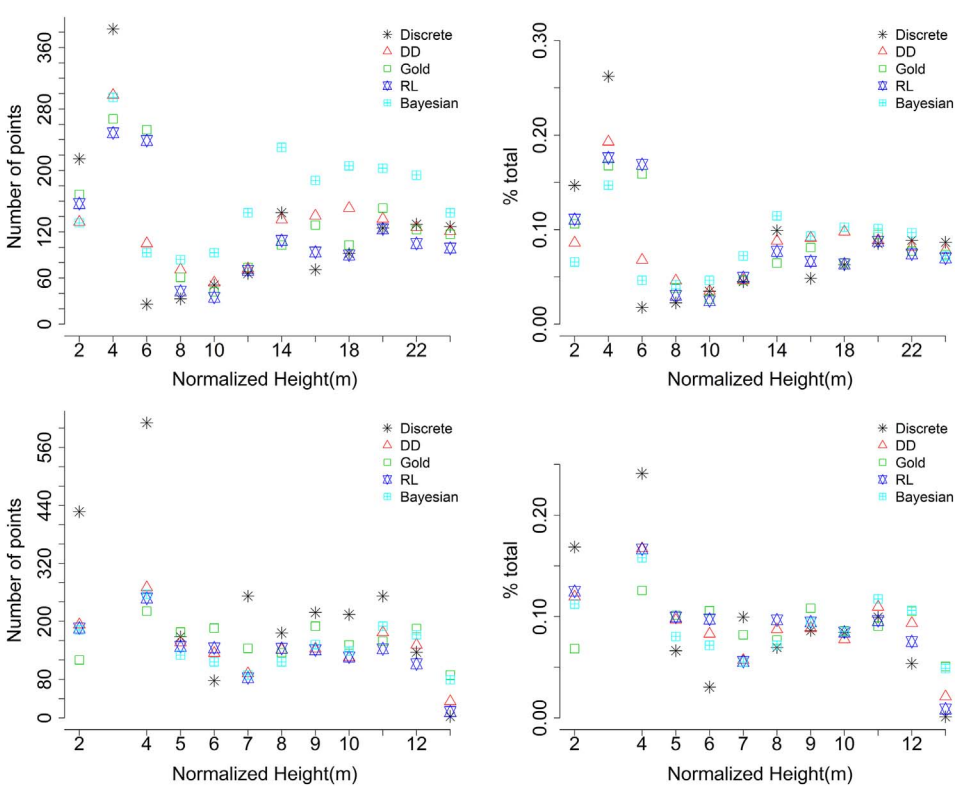

Fig. 8. Two representative examples (Top: Tree 1, Bottom: Tree 2) of individual trees' height bin vs. absolute point frequency and normalized frequency using DR LiDAR data and FW LiDAR data with the four waveform processing methods including the Bayesian decomposition, DD, Gold and RL approaches. Discrete represents results from DR LiDAR data. DD, Gold, RL and Bayesian represent results from the DD, Gold approach, RL approach and Bayesian decomposition method, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Summary of the total number of points/number of non-ground points/non-ground canopy point density for 21 individual trees using DR LiDAR data (Discrete), and waveform LiDAR data with the DD, Gold, RL and Bayesian decomposition (Bayesian) methods.

Tree index	Discrete	DD	Gold	RL	Bayesian
Tree 1	1465/867/7.3	1554/1114/15.0	1607/980/13.2	1429/831/12.7	2008/1593/21.3
Tree 2	**2534/1337/10.8**	**1616/947/7.4**	**1761/1014/8.0**	**1490/838/6.6**	**1982/1367/10.7**
Tree 3	827/353/7.0	858/448/18.1	879/425/19.1	833/419/15.9	1035/569/28.7
Tree 4	887/327/7.1	898/432/15.7	1053/399/18.0	903/439/16.5	102/499/19.3
Tree 5	**3029/1216/16.7**	**1128/610/13.2**	**1190/569/13.1**	**1085/546/10.7**	**1326/744/15.0**
Tree 6	**3873/1211/11.2**	**2306/1059/9.8**	**2315/1079/10.0**	**2257/1046/9.7**	**2312/1202/11.2**
Tree 7	2186/678/6.6	1880/880/8.6	1877/737/7.2	1859/862/8.5	1816/872/8.6
Tree 8	1698/999/7.2	2256/1335/9.7	2385/1243/9.0	2422/1448/10.5	2559/1681/12.2
Tree 9	543/295/7.5	659/445/11.5	669/432/11.0	652/430/10.6	671/448/11.3
Tree 10	552/316/7.2	405/283/6.4	558/322/7.3	569/358/8.1	563/387/8.8
Tree 11	499/193/8.1	444/184/7.8	521/239/10.1	509/268/11.3	500/243/10.2
Tree 12	**1945/649/17.7**	**893/445/12.1**	**946/386/10.5**	**885/398/10.9**	**897/412/11.2**
Tree 13	773/331/7.9	834/454/10.8	883/433/10.3	876/500/11.9	832/465/11.1
Tree 14	1278/718/7.2	1484/951/9.5	1563/964/9.7	1553/953/9.5	1609/1101/11.0
Tree 15	1275/647/7.4	1748/988/11.2	1910/1161/13.2	1853/1090/12.4	1988/1467/16.7
Tree 16	1788/1071/6.5	2098/1407/9.0	2439/1596/10.2	2428/1512/9.7	2624/2086/13.4
Tree 17	708/329/7.4	625/315/7.1	813/327/7.4	765/393/8.8	749/377/8.5
Tree 18	770/303/7.1	691/309/7.3	747/297/7.0	753/356/8.4	746/356/8.4
Tree 19	1847/902/7.2	2437/1696/13.6	2687/1678/13.5	2737/1700/13.6	3027/2423/19.5
Tree 20	**1116/575/11.4**	**758/531/10.5**	**829/525/10.4**	**824/531/10.5**	**811/548/10.9**
Tree 21	1660/805/6.8	2004/1182/9.9	2041/1104/9.3	2042/1100/9.2	2267/1589/13.4

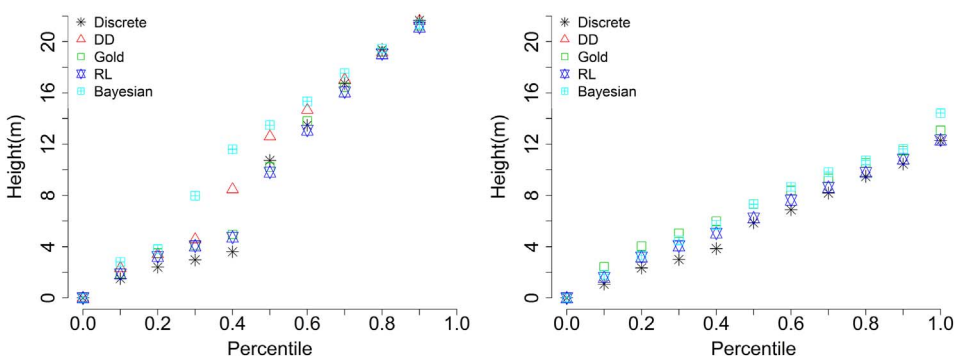

Fig. 9. The individual trees' percentile heights (Left: Tree 1; Right: Tree 2) using DR LiDAR data (Discrete), and waveform LiDAR data with the DD, Gold, RL and Bayesian decomposition (Bayesian) methods. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

corresponding distribution. For example, the mode of average distances of the X, Y and Z centered around −0.02, 0.02 and −0.08 m, respectively. The counterpart of point distances was much larger which centered around 0.51 m with more flat distribution. The RMSE demonstrated a similar pattern that the distribution of point distances was less compact than other three coordinates' distribution with a larger mode. Interestingly, there was no significant difference between the distribution of X distances and Y distances, while the distributions of Z distances varied with the X and Y distances with larger absolute mode and uncertainty.

3.3.3. Surface model uncertainty

The uncertainty of surface models derived from these quantile point clouds using the Bayesian decomposition method is demonstrated in Fig. 12. The distribution of the CHM's mean bias and RMSE (dark gray) tended to be wider than DTM counterparts. For instance, the estimated RMSE for the CHM was 1.38 m and 95% CI ranged from 1.31 to 1.61 m, both of which were larger than the RMSE of the DTM. The distribution of the average difference for DTM was likewise smaller than CHM's average difference from the perspective of 95% CI. Therefore, it was evident that the uncertainty of DTM was smaller than the uncertainty of CHM in terms of their corresponding distribution of the average bias and RMSE.

Compared to the point clouds' uncertainty (Fig. 11), the surface models yielded larger uncertainty with wider CIs, which could represent the propagation of true estimation error along the processing steps. Using the RMSE of CHM as an example, it demonstrated a larger median and more flat pattern distribution than point clouds' RMSE (Fig. 11).

3.3.4. Uncertainty of accuracy assessment with field data

Estimated tree height's uncertainty yielded from the Bayesian decomposition method at individual tree level against the field-measured data is displayed in Fig. 13. It was observed that data points (blue circle) of estimated tree height (ETH) versus field-measured tree height mostly followed tightly around Y = X line that indicated the Bayesian decomposition method can accurately capture the height of these individual trees. In addition, almost all trees' 95% CIs of ETH (145 out 151 individual trees), as indicated by the horizontal error bar (sky blue), were intercepted by Y = X line. The uncertainty of ETH's RMSE confirmed good agreement with the field-measured height that resulted in the uncertainty of RMSE ranges from 0.69 to 2.05 m. To avoid the proliferation of figures, two individual trees (Tree 1621 and Tree 1647, orange) were used as examples to further demonstrate final results of tree height estimation using the Bayesian decomposition method as shown in the right panel of Fig. 13. This figure provided more than just one value of ETH that generally obtained from the deterministic methods, and yielded more informative estimates of tree height with probability for the individual tree.

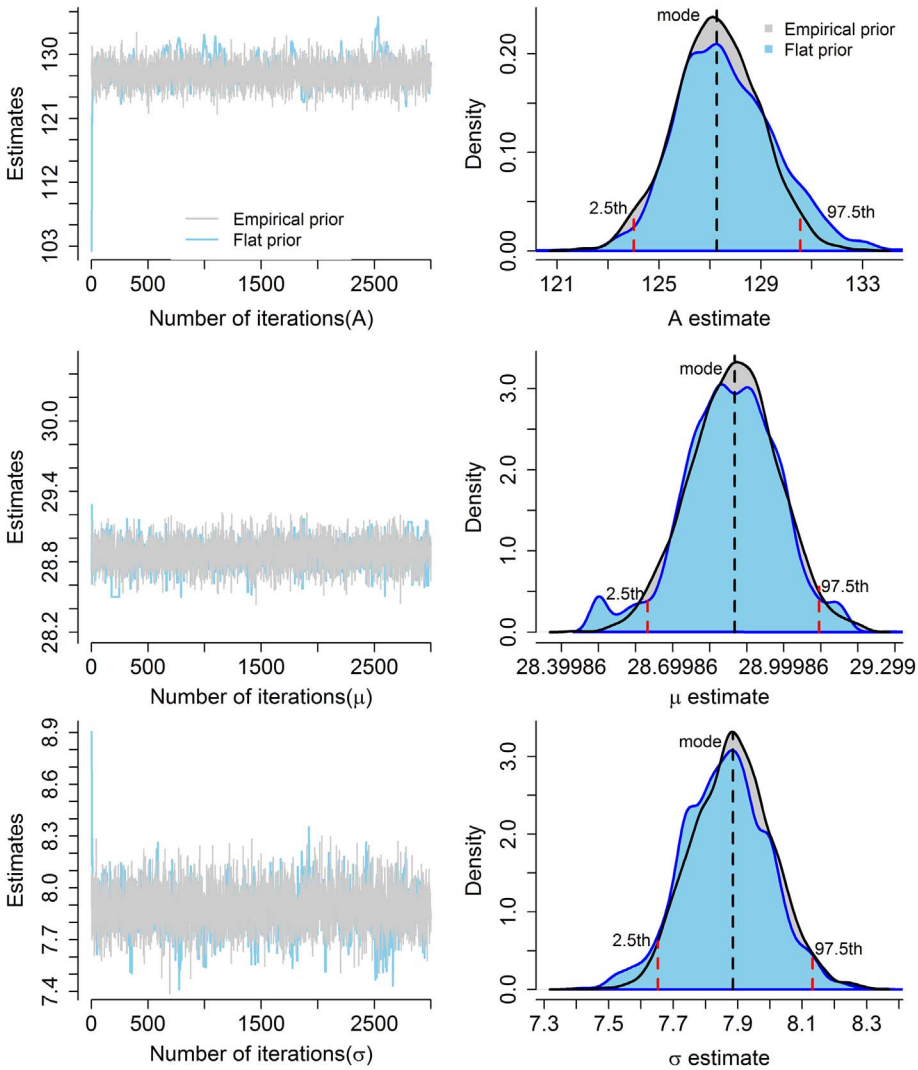

Fig. 10. An example of parameter estimates derived from the Gaussian model within the Bayesian framework using the flat priors and empirical priors with the same number of iterations. Left: The trace plot of model parameters A, u, δ using the flat priors and the empirical priors. Right: The probability density distribution of estimated parameters A, u, δ using the flat priors and the empirical priors.

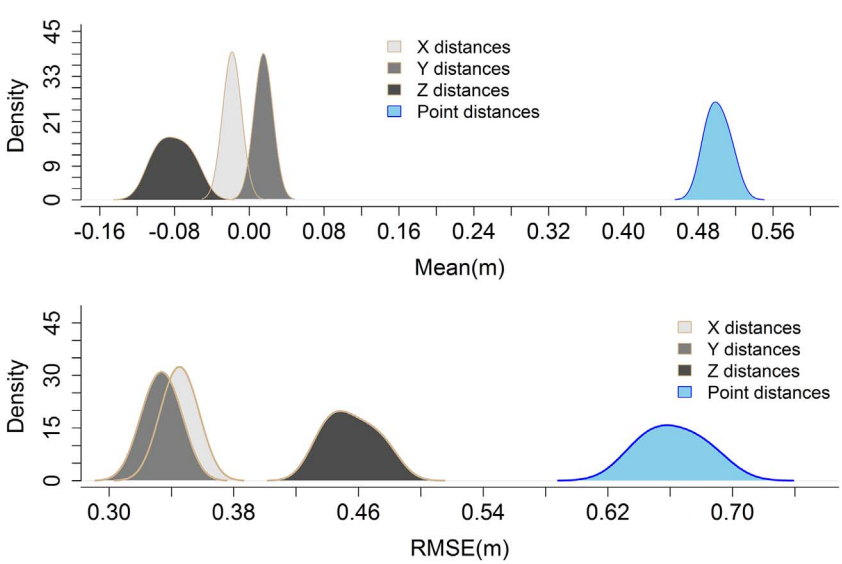

Fig. 11. Uncertainty of average distance (Mean) and corresponding RMSE between reference point cloud dataset and waveform-based LiDAR quantile point cloud datasets using the Bayesian decomposition method at the SJER site.

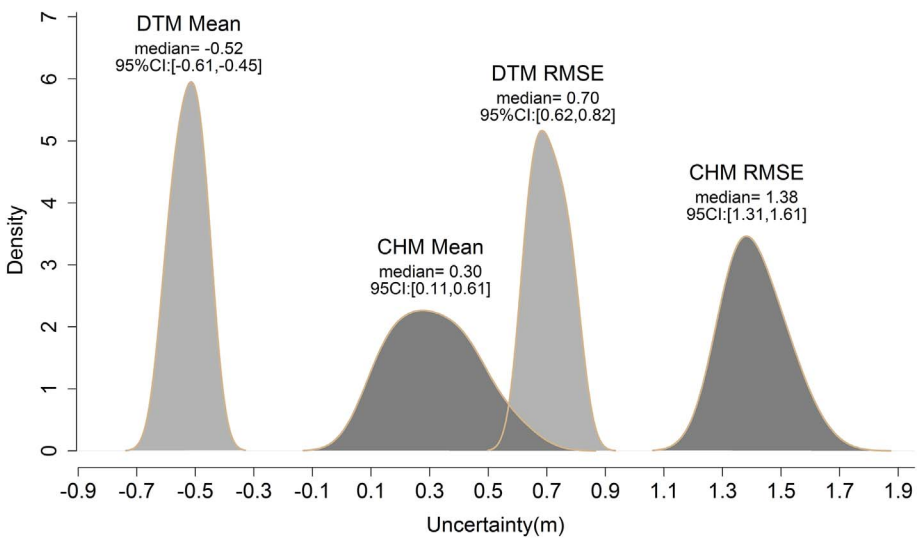

Fig. 12. Uncertainty of the average difference and RMSE between waveform-based surface models (the DTM and CHM) and reference surface models using Bayesian decomposition method at the SJER site. CI: Credible interval.

4. Discussion

4.1. Model reasonableness

Several potential models such as the Generalized Gaussian, Burr and Weibull models have been proposed to fit the airborne LiDAR waveforms (Mallet et al., 2009), while most of them are unrealistic models which may violate theoretical considerations or misinterpret their real-world's physical meanings. Thus, the choice of model is crucial for the subsequent FW LiDAR data processing and outcome interpretation.

The results of model reasonableness in the present study justify that the Gaussian model is more suitable for reconstructing the waveform LiDAR signals in terms of efficiency, physical interpretation and uncertainty. Three representative models suitable for FW LiDAR data processing including the Weibull, Adaptive Gaussian, and Gaussian models were explored. In our case, the Adaptive Gaussian model better fitted waveforms compared to the Gaussian model. However, the Adaptive Gaussian model is prone to overfit the model and incorrectly considers data noise as real waveform signals by adjusting the rate parameter to pursue the smallest residual of model fitting. The Weibull model is more difficult to constrain than the other two models with larger CIs or uncertainty. Moreover, there is no explicit way to explain the physical meaning of the corresponding parameters. Hence, the Gaussian model is a suitable trade-off between accuracy and meaningful solution to extract useful information form FW LiDAR data.

Simultaneously, the quantification of uncertainty using these models demonstrates the attractive features of the Bayesian method and allows us to interpret results in a more natural way with posterior distribution of estimates. Moreover, we can explicitly trace the uncertainty inherent in any inference and monitor the uncertainty propagation through the use of quantiles for different parameters, quantile datasets and models (Fig. 3).

4.2. Performance evaluation

4.2.1. Point cloud

The calibration and comparison conducted on the point clouds eliminate the reliance on the underlying surface derived from local parametric estimates. Additionally, we compare every point of the waveform-based point cloud with the corresponding reference point cloud instead of mainly comparing a subset of the total points in the specified grid cells when the area based method (DTM or CHM) is used. This can avoid the error introduced by interpolation and poor choice of gridding size, and ultimately make the comparisons become more convincing and comprehensive. The C2C results show the Bayesian decomposition method is applicable for extracting information from FW LiDAR data to characterize the vegetation structures. The contribution made by the Z direction to the point distances is much larger than the X and Y directions. A reasonable explanation for the difference is that pulse direction vector of the Z direction (~ − 0.15 m/ns) is much larger than the X and Y directions (~ − 0.02 and − 0.01 m/ns). As a consequence, the relative change in the Z direction is much higher than in

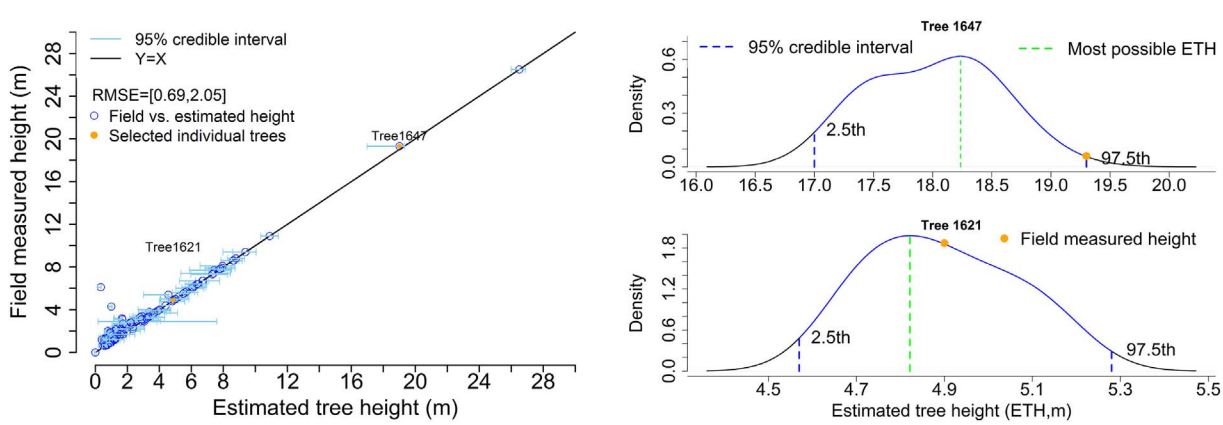

Fig. 13. Uncertainty of the individual trees' height as obtained from the Bayesian decomposition method vs. field-measured tree height at the SJER study site. The left panel refers to all individual trees' 95% credible interval (sky blue), and the right to the distribution of two possible estimated tree heights with 95% credible interval (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the X and Y directions given the same change of time. This may also suggest that FW LiDAR data's horizontal precision is higher than the vertical precision. The uncertainty of the point cloud (Fig. 11) further substantiates this finding with larger average distance and RMSE in the Z direction.

Overall, the C2C method's results support that the four waveform processing methods can generate relatively satisfactory results with all methods' RMSE values being < 0.93 m (Table 1). However, a closer examination of the point clouds comparison results reveal that the decomposition method (the DD and Bayesian decomposition) may generate more accurate point clouds than the combined deconvolution and decomposition method (the Gold and RL) in terms of the descriptive statistics such as the MD and RMSE. Specifically, the DD method is prone to outperform other three methods with slightly smaller average distance and corresponding RMSE. This conclusion backs up the results using the area based method of Zhou et al. (2017) that the DD method potentially yields more accurate DTMs and CHMs than the Gold and RL methods.

4.2.2. Individual tree metric

The C2C comparison yields a general view about the waveform processing methods on a large scale. Examining metrics derived from different methods at the individual tree level further reveal advantages and disadvantages of these methods. Most of the individual trees derived from waveform LiDAR data can obtain more points - by as many as 80% of trees (16/21) at the SJER1 study region - compared with the corresponding DR LiDAR data. Two illustrative examples of individual tree's point distribution at different height bins, percentile heights and canopy point density depict a similar trend that FW LiDAR data are capable of extracting more points at the mid-story of the vegetation. Surprisingly, by comparing the DR LiDAR data results, FW processing based methods are less likely to detect ground points that maybe the disadvantage of using FW LiDAR data.

4.2.3. Field data calibration

Various factors could affect the field calibration results, such as the number of sample plots, the accuracy of measured instruments and the mis-registration error between LiDAR data and field plots (Zhao et al., 2011). Although field calibration is necessary for LiDAR applications, caution should be exercised when using calibration results, given the survey error was inherently unavoidable and ubiquitous. Table 3 demonstrates that DR LiDAR data (RMSE ≤ 1.1 m) can be an alternative for field-measured data when they are not available. This conclusion is also consistent with the previous studies' claims that DR LiDAR data can accurately measure the tree height (Chen, 2007). In addition, the uncertainty of ETH (Fig. 13) is generated to indicate the true magnitude of estimation error and enhance the credibility of field calibration results by reducing the error introduced by the field measurements. For the Bayesian decomposition method, the calibration results are not just represented by one RMSE value like the DD, Gold and RL methods yielded (Table 3), but all possible RMSE values with probability are provided. This situation renders an intuitive and reasonable way to address the real-world model calibration problem.

4.3. Uncertainty analysis

Quantification of uncertainty is one of the notable advantages for adopting the Bayesian method. The uncertainty is retained throughout all processing steps including parameter estimates, geolocation transformation and surface model generation, a full disclosure of uncertainty of these processing steps is crucial for the result interpretation and the potential applications. In the present study, a thorough uncertainty quantification method is conducted on parameters, quantile point clouds and surface models (DTMs and CHMs) to ensure the quality of the results. The Bayesian approach gives us all possible estimates with probability for unknown parameters (Fig. 3) instead of just providing a single value within a frequentist framework. This enables the researchers to view where the most probable locations of the waveforms to represent the illuminated object along the pulse line. Furthermore, the probability distribution of the estimates also essentially precludes non-uniqueness problem, which is common in the waveform decomposition. Actually, this problem has been faced by the users who employed advanced statistical models to tackle inversion problems in various fields (Gouveia and Scales, 1998; Oh and Kwon, 2001).

The Bayesian method has been subjected to criticism for its subjectivity by introducing prior information. In the present study, the results of using different priors showed that this subjectivity can be overcome by using the flat priors (non-informative priors). The nearly identical performance as empirical priors can be achieved at the expense of more computation cost and time using the flat priors, which agrees well with Ellison's conclusion (Ellison, 2004). While this conclusion is not consistent with Denham's study (Denham et al., 2009), using the empirical priors tends to increase the precision of the parameters. There are factors that contribute to this inconsistency, such as data noise, setting of non-informative prior and the main model used in the simulation. However, all results demonstrate that the Bayesian approach is capable of data analysis such as parameters extraction from waveforms with reasonable accuracies. To reach the same performance of the empirical priors, the computation time using the flat priors is markedly longer (Table 4). This indirectly reflects that the empirical priors are capable of reducing the complexities of Bayesian decomposition method with accurate outputs.

The Bayesian decomposition method may appear to require more computation cost than the other three methods, however, it can overcome the parameter initialization problem of the DD method, and avoids the step of parameters optimization for deconvolution when the combined deconvolution and decomposition method is used. Results of this study demonstrate the potential and advantages of using the Bayesian approach to characterize uncertainty of parameter estimates. Moreover, it allows researchers to trace the propagated error and uncertainty explicitly through PDFs of corresponding evaluation criteria from parameters to points, point clouds and surface models (Fig. 3). Assuredly, the unsuccessful Bayesian decomposition may occur when the waveform is extremely noisy or irregular. Assigning the number of Gaussian components (n) as a random variable is expected to be one potential solution which could be one aspect of further research.

The uncertainty results for the quantile point cloud datasets are represented by a probability distribution, which can characterize the variability, extent of average distance and RMSE in a probabilistic sense to assess the performance of the Bayesian method. As expected, the

Table 3
Summary of the comparisons between the field-measured tree height and maximum tree height derived from the CHMs using DR LiDAR data, and FW LiDAR data with the DD, Gold, RL and Bayesian decomposition (Bayesian) methods.

Study site	Statistics criteria	Discrete (m)	DD (m)	Gold (m)	RL (m)	Bayesian (m)
SJER	Average	− 0.24	0.73	− 0.15	− 0.32	− 0.26
	Standard deviation	1.11	1.64	1.52	1.36	1.48
	RMSE	1.11	1.75	1.49	1.35	1.50

Table 4
Average processing time for a single waveform with different number of peaks.

Number of components per waveform	Number of waveforms	Average time (flat priors, seconds)	Average time (empirical priors, seconds)
1	95	2.43	1.91
2	21	12.40	10.49
3	10	67.38	45.36
4	6	235.60	114.70
5	4	364.65	162.00
6	2	462.60	267.00
7	2	511.20	276.90

horizontal precision of FW LiDAR data is better than their vertical precision and the main uncertainty of the point clouds come from the vertical direction.

The surface model's uncertainty results reveal that the DTM has smaller variance and uncertainty than the CHM, which yields a larger average bias and a wider range of the RMSE (Fig. 12). Various factors can degrade the accuracy of the DTM, such as sampling size, point density, terrain conditions and processing methods (Vincent et al., 2012). To some extent, the quality of the DTM significantly affected the CHM quality. Consequently, larger error or uncertainty of CHM accrued might mainly due to the fact that CHM was generated by subtracting DTM from Digital surface model (DSM) and an additional error was likely to be introduced with this step. This kind of uncertainty propagation was also observed through comparing the average bias and corresponding RMSE of the point clouds (Fig. 11) with the surface models (Fig. 12). More specifically, the point clouds' RMSE varied from 0.62 to 0.71 m, while the uncertainty of the RMSE for the waveform-based DTM and CHM were 0.62–0.80 m and 1.31–1.61 m, respectively. It is evident that the error and uncertainty are increasing along the processing steps and an additional processing step brings more error and uncertainty. The uncertainty analysis captures the fact the error introduced into the estimation steps and provides more informative information which consequently gives us more confidence to interpret and apply results to real-world problems such as tree species identification.

Actually, most of the remote sensing applications such as variable extraction and biomass estimation using various sources of remote sensing data are inadequately addressed or overlooked the uncertainty analysis of their results (McRoberts et al., 2010; Zhao et al., 2011). The Bayesian concept or approach can be transplanted to these applications to quantify the estimate of error or uncertainty by employing statistical inference of posterior distribution for the subjects of interest. At the current stage, the Bayesian approach appears complicated which requires users to have knowledge about the concepts and procedures, and needs extensive computation time (De Lannoy et al., 2014) for MCMC sampling. These aspects have hindered the broad applications of Bayesian approaches, however, the advances in computation and development of generalized tool such as BUGS, JAGS and Stan will likely contribute to the popularity of Bayesian models in the foreseeable further.

5. Conclusion

This paper has incorporated the Bayesian concept with waveform decomposition to develop an innovative method to extract information from FW LiDAR data and conduct a thorough uncertainty analysis along the processing steps. Built upon the deterministic method for waveform decomposition of Zhou et al. (2017), a more comprehensive exploration of the extant methods and the proposed Bayesian method are conducted at the point, point cloud and surface model levels.

The Bayesian method contributes to the waveform decomposition in several ways. As demonstrated, the solutions of decomposition using the Bayesian method are represented by a probability distribution over parameter estimates instead of producing a single value for each parameter using the deterministic approach. Additionally, it permits the users to interpret the results in a probabilistic sense that is stable enough to provide a feasible solution to the decomposition problem. Moreover, the adoption of Bayesian analytics generates a systematic and transparent knowledge learning framework to estimate uncertainty emerging from parameters, points, the point cloud and the surface model via the use of quantiles of the probability distribution. Meanwhile, the uncertainty propagation can be explicitly traced and observed from data to the parameter estimate, the point, the point cloud and the surface model (Fig. 3). We also explored the non-uniqueness problems of waveform decomposition within Bayesian framework to reduce the theoretical error from the model itself and justify the reasonableness of using the Gaussian model for FW LiDAR data decomposition in terms of uncertainty, physical meaning and processing efficiency. The superior method of FW LiDAR data processing varies from the perspective of different criteria and waveform-derived products. Results of point cloud comparisons demonstrate that the Bayesian decomposition method can achieve comparable accuracy as the DD, Gold and RL methods. Results from the individual tree level highlight that FW LiDAR data can characterize more detail of the mid-story of vegetation with more dense points. The combined deconvolution and decomposition method (the Gold and RL approaches) outperforms the decomposition method (DD and Bayesian decomposition) in terms of the surface model's results. In addition, field data for calibration results using the Bayesian method provide a reasonable way to reduce the effect of field measurement error on the model calibration. Further, the Bayesian method is expected to become more efficient and user-friendly with the aid of computational advances and convenient implementation tools. In addition, more research efforts are still needed to apply FW LiDAR data for vegetation studies, such as tree species identification and biomass estimation over extensive regions.

Acknowledgment

The authors gratefully acknowledge the support provided by NASA's ICESat-2 SDT (Grant # NNX15AD02G) and RAPID Response (Grant # NNX14AN99G), NSF Doctoral Dissertation Improvement Grant (DEB 1702008), and Texas A & M Institute for Genome Sciences and Society (TIGSS) for providing computational resources and systems administration support for the TIGSS HPC Cluster. We would like to thank Dr. Jeffrey D. Hart and Paul-Christian Bürkner for their helpful insights and discussions. We thank the National Ecological Observatory Network (NEON) for their help with providing waveform LiDAR and field data. Last but not the least, we greatly appreciate the constructive comments from the two anonymous reviewers.

Appendix A. Model implementation

A.1. Model structures

The models were built on the Stan inference engine. It consists of four blocks: variables declarations, parameter statements, transformed parameters and model blocks The data block is functional as input data declaration; the parameters block is to introduce all unknown parameters of interest; the transformed parameters block is mainly for the conversion of data and parameters to the readable way, and the model block is to

compute the log posterior density. Stan first translated a model into C + + and then compiled the code for each waveform. Different waveforms may have a different number of Gaussian components, which generates a different posterior distribution and model for each waveform. The challenge is that the model always changes based on the number of Gaussian components which requires users to compile a new model for each waveform. Meanwhile, a new dynamic link library (DLL) was generated to store the model. This would exceed the maximum number of DLLs can be loaded on a computer after the model fitting a certain amount of waveforms. In order to solve this problem, the *update* function of brms was used to avoid recompilation issues. This function requires the structure of the input model or the posterior distribution should be exactly the same format. Therefore, we grouped the waveforms with the same number of peaks, and then employed the *update* function for fitting these waveforms. This strategy could save the compilation time and make the processing much more efficient. In addition, the parallel computing is also automatically implemented in brms package by specifying the number of clusters and chain. In this analysis, we assigned them both as 2. All of these components in brms package make HMC converge much faster to a target distribution.

A.2. Model convergence

The model convergence is measured with $\widehat{R}$ and is computed as follows:

$$\widehat{\mathrm{Var}}(\theta) = \left(1 - \frac{1}{l}\right)W + \frac{1}{l}B \tag{A.1}$$

$$\widehat{R} = \sqrt{\frac{\widehat{\mathrm{Var}}(\theta)}{W}} \tag{A.2}$$

where W is the mean variances of stationary distribution for each chain, B is the variance of stationary distribution at the between-chain level, l is the number of draws in each chain, and $\widehat{\mathrm{Var}}(\theta)$ is the variance of the stationary distribution as a weighted average of W and B. Here, the chain of $\widehat{R}$ was the split chain that discarded the burn-in iterations.

Appendix B. Model efficiency

Recently, the inference engine Stan (Gelman et al., 2015) has been introduced to implement HMC sampling. Stan employs a reverse mode automatic differentiation rather than a numerical differentiation to compute the gradient (Griewank and Walther, 2008). Furthermore, the No-U-Turn Sampler (NUTS), a variant of HMC, was used to automatically tune two parameters in the leapfrog method. Specifically, the step length L is achieved by means of a recursive algorithm with doubling procedure devised by Neal (2003) for slice sampling, and step size ϵ via an adaptation of dual averaging algorithm of Nesterov (2009).

These enable NUTS to run more efficiently than other MCMC algorithms and to become desirable for those who have little experience of tuning HMC without user intervention. Fig. B1 displays the relationship between the log posterior of the model (x-axis) and acceptance rate (y-axis) using NUTS samplers to demonstrate the efficiency of HMC or NUTS. Our experiment demonstrates that the sampling acceptance rate of NUTS is much higher (around 85–95%) than the desired acceptance rate of Metropolis algorithm which generally ranges from 23% to 50% (Roberts et al., 1997) (Fig. B1).

The exploration of these waveforms also demonstrated that assigning the proper number of and total iterations of MCMC was critical to the model efficiency. The above sample waveforms were also used to explore optimized combinations of the number of burn-in samples and total iterations. To balance the processing time and accuracy, the optimization of these parameters was conducted. $\widehat{R} < 1.1$ was employed to judge the number of total iterations and samples were enough to obtain the acceptable results. The summary of these parameter combinations after optimization was shown in Table B1. The number of chains used in this analysis was 2. To save the computation time and reduce the autocorrelation of draws in MCMC simulations, we saved every third iteration of each chain by thinning posterior samples.

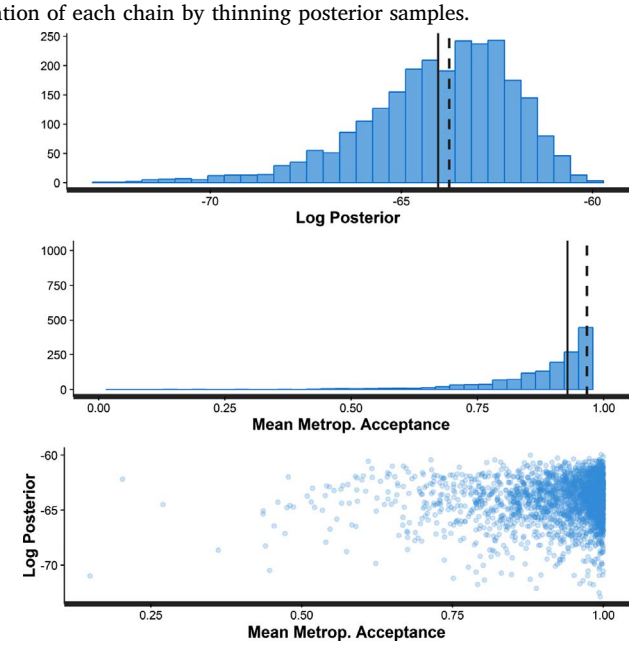

Fig. B1. An example of histogram for log posterior and final Metropolis acceptance rate, and distribution of log posterior vs. Metropolis acceptance rate using HMC algorithm.

Table B1
The final parameters of MCMC simulation for different waveform components.

Number of component per waveform	Number of total iterations	Number of burn-in
1	9000	2000
2	10,000	2500
3	12,000	3000
4	15,000	4000
5	18,000	5000
6	20,000	6000
7	22,000	7000
> 8	25,000	8000

References

Allouis, T., Durrieu, S., Véga, C., Couteron, P., 2013. Stem volume and above-ground biomass estimation of individual pine trees from LiDAR data: contribution of full-waveform signals. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6, 924–934.

Blair, J.B., Rabine, D.L., Hofton, M.A., 1999. The laser vegetation imaging sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. ISPRS J. Photogramm. Remote Sens. 54, 115–122.

Buerkner, P.-C., 2016. Brms: Bayesian Regression Models Using Stan. (In).

Cawse-Nicholson, K., van Aardt, J., Hagstrom, S., Romanczyk, P., Schaaf, C., Strahler, A., Li, Z., Krause, K., 2014. Improving waveform lidar processing toward robust deconvolution of signals for improved structural assessments. In: SPIE Defense + Security. International Society for Optics and Photonics pp. 90800I-90800I-90806.

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, 5211–5228.

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

Chen, Q., Vaglio Laurin, G., Valentini, R., 2015. Uncertainty of remotely sensed above-ground biomass over an African tropical forest: propagating errors from trees to plots to pixels. Remote Sens. Environ. 160, 134–143.

De Lannoy, G.J.M., Reichle, R.H., Vrugt, J.A., 2014. Uncertainty quantification of GEOS-5 L-band radiative transfer model parameters using Bayesian inference and SMOS observations. Remote Sens. Environ. 148, 146–157.

Denham, R., Mengersen, K., Witte, C., 2009. Bayesian analysis of thematic map accuracy data. Remote Sens. Environ. 113, 371–379.

Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F., Prince, S., 2002. Estimation of tropical forest structural characteristics using large-footprint lidar. Remote Sens. Environ. 79, 305–319.

Edwards, M., Huet, S., Goreaud, F., Deffuant, G., 2003. Comparing an individual-based model of behaviour diffusion with its mean field aggregate approximation. J. Artif. Soc. Soc. Simulat. 6.

Ellison, A.M., 2004. Bayesian inference in ecology. Ecol. Lett. 7, 509–520.

Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E., Hudak, A.T., 2009. Characterizing forest succession with lidar data: an evaluation for the Inland Northwest, USA. Remote Sens. Environ. 113, 946–956.

Finley, A.O., Banerjee, S., Cook, B.D., Bradford, J.B., 2013. Hierarchical Bayesian spatial models for predicting multiple forest variables using waveform LiDAR, hyperspectral imagery, and large inventory datasets. Int. J. Appl. Earth Obs. Geoinf. 22, 147–160.

Frazer, G.W., Magnussen, S., Wulder, M.A., Niemann, K.O., 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. Remote Sens. Environ. 115, 636–649.

Gelfand, A.E., Smith, A.F., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398–409.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2014. Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton, FL, USA.

Gelman, A., Lee, D., Guo, J., 2015. Stan: a probabilistic programming language for Bayesian inference and optimization. J. Educ. Behav. Stat. 40, 530–543.

Girardeau-Montaut, D., 2015. Cloud Compare—3D Point Cloud and Mesh Processing Software. Open Source Project.

Gouveia, W.P., Scales, J.A., 1998. Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis. J. Geophys. Res. Solid Earth 103, 2759–2779.

Griewank, A., Walther, A., 2008. Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. Siam.

Hancock, S., Armston, J., Li, Z., Gaulton, R., Lewis, P., Disney, M., Mark Danson, F., Strahler, A., Schaaf, C., Anderson, K., Gaston, K.J., 2015. Waveform lidar over vegetation: an evaluation of inversion methods for estimating return energy. Remote Sens. Environ. 164, 208–224.

Harding, D.J., Carabajal, C.C., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. Geophys. Res. Lett. 32.

Hernandez-Marin, S., Wallace, A.M., Gibson, G.J., 2008. Multilayered 3D LiDAR image construction using spatial models in a Bayesian framework. IEEE Trans. Pattern Anal. Mach. Intell. 30, 1028–1040.

Hoff, P.D., 2009. A First Course in Bayesian Statistical Methods. Springer Science & Business Media.

Hoffman, M.D., Gelman, A., 2014. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res. 15, 1593–1623.

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

Hong, T., Sen, M.K., 2009. A new MCMC algorithm for seismic waveform inversion and corresponding uncertainty analysis. Geophys. J. Int. 177, 14–32.

Isenburg, M., 2012. LAStools-efficient tools for LiDAR processing. Available at. http://www.cs.unc.edu/~isenburg/lastools/, Accessed date: 9 October 2012.

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, 043510.

Keith, K., Tristan, G., 2015. NEON L0-TO-L1 Discrete-return LiDAR Algorithm Theoretical Basis Document (ATBD). The National Ecological Observatory Network.

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, 863–872.

Lee, A.C., Lucas, R.M., 2007. A LiDAR-derived canopy density model for tree stem and crown mapping in Australian forests. Remote Sens. Environ. 111, 493–518.

Mallet, C., Bretar, F., 2009. Full-waveform topographic lidar: state-of-the-art. ISPRS J. Photogramm. Remote Sens. 64, 1–16.

Mallet, C., Lafarge, F., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2009. A stochastic approach for modelling airborne lidar waveforms. In: Laserscanning, pp. 201–206.

McGlinchy, J., van Aardt, J.A.N., Erasmus, B., Asner, G.P., Mathieu, R., Wessels, K., Knapp, D., Kennedy-Bowdoin, T., Rhody, H., Kerekes, J.P., Ientilucci, E.J., Wu, J., Sarrazin, D., Cawse-Nicholson, K., 2014. Extracting structural vegetation components from small-footprint waveform Lidar for biomass estimation in savanna ecosystems. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 480–490.

McRoberts, R.E., Cohen, W.B., Næsset, E., Stehman, S.V., Tomppo, E.O., 2010. Using remotely sensed data to construct and assess forest attribute maps and related spatial products. Scand. J. For. Res. 25, 340–367.

Mémoli, F., Sapiro, G., 2004. Comparing point clouds. In: Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing. ACM, pp. 32–40.

Neal, R.M., 2003. Slice sampling. Ann. Stat. 705–741.

Neal, R.M., 2011. MCMC using Hamiltonian dynamics. In: Handbook of Markov Chain Monte Carlo. 2. pp. 113–162.

Nesterov, Y., 2009. Primal-dual subgradient methods for convex problems. Math. Program. 120, 221–259.

Neuenschwander, A.L., 2008. Evaluation of waveform deconvolution and decomposition retrieval algorithms for ICESat/GLAS data. Can. J. Remote. Sens. 34, S240–S246.

Oh, S.-H., Kwon, B.-D., 2001. Geostatistical approach to Bayesian inversion of geophysical data: Markov chain Monte Carlo method. Earth Planets Space 53, 777–791.

Persson, Å., Söderman, U., Töpel, J., Ahlberg, S., 2005. Visualization and analysis of full-waveform airborne laser scanner data. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 36, W19.

Pirotti, F., 2011. Analysis of full-waveform LiDAR data for forestry applications: a review of investigations and methods. iForest – Biogeosci. For. 4, 100–106.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass Bioenergy 31, 646–655.

Popescu, S.C., Wynne, R.H., Nelson, R.F., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can. J. Remote. Sens. 29, 564–577.

Qin, H., Xie, X., Vrugt, J.A., Zeng, K., Hong, G., 2016. Underground structure defect detection and reconstruction using crosshole GPR and Bayesian waveform inversion. Autom. Constr. 68, 156–169.

Ray, A., Alumbaugh, D.L., Hoversten, G.M., Key, K., 2013. Robust and accelerated Bayesian inversion of marine controlled-source electromagnetic data using parallel tempering. Geophysics 78, E271–E280.

Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. Ann. Appl. Probab. 7, 110–120.

Roncat, A., Bergauer, G., Pfeifer, N., 2010. Retrieval of the backscatter cross-section in full-waveform LIDAR data using B-splines. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 137–142.

Roonizi, E.K., Sassi, R., 2016. A signal decomposition model-based bayesian framework for ECG components separation. IEEE Trans. Signal Process. 64, 665–674.

Rowe, J., 2013. Ground Classification and Below Ground Response Assessment of Forested Regions Using Full-waveform LiDAR.
Sen, M.K., Stoffa, P.L., 1996. Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. Geophys. Prospect. 44, 313–350.
Tison, C., Nicolas, J.-M., Tupin, F., Maître, H., 2004. A new statistical model for Markovian classification of urban areas in high-resolution SAR images. IEEE Trans. Geosci. Remote Sens. 42, 2046–2057.
Ulrych, T.J., Sacchi, M.D., Woodbury, A., 2001. A Bayes tour of inversion: a tutorial. Geophysics 66, 55–69.
Vehtari, A., Gelman, A., Gabry, J., 2016. Practical Bayesian Model Evaluation Using Leave-one-out Cross-validation and WAIC. (arXiv preprint arXiv:1507.04544).
Vincent, G., Sabatier, D., Blanc, L., Chave, J., Weissenbacher, E., Pélissier, R., Fonty, E., Molino, J.F., Couteron, P., 2012. Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure. Remote Sens. Environ. 125, 23–33.
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, 100–112.
Wang, G., Weng, Q., 2013. Remote Sensing of Natural Resources. CRC Press.
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, 2402–2414.
Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T., 2012. Lidar sampling for large-area forest characterization: a review. Remote Sens. Environ. 121, 196–209.
Yao, W., Krzystek, P., Heurich, M., 2012. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. Remote Sens. Environ. 123, 368–380.
Zhao, K., Popescu, S., Nelson, R., 2009. Lidar remote sensing of forest biomass: a scale-invariant estimation approach using airborne lasers. Remote Sens. Environ. 113, 182–196.
Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sens. Environ. 115, 1978–1996.
Zhou, T., Popescu, S.C., Krause, K., Sheridan, R.D., Putman, E., 2017. Gold – a novel deconvolution algorithm with optimization for waveform LiDAR processing. ISPRS J. Photogramm. Remote Sens. 129, 131–150.