Radiometric Calibration of the Finnish Geospatial Research Institute Hyperspectral LiDAR

Julian Nicolae Ilinca

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.
Espoo 19.7.2018

Supervisor

Prof. Miina Rautiainen

Advisor

Prof. Sanna Kaasalainen

Aalto University
School of Electrical Engineering

Aalto University
School of Electrical
Engineering

Author Julian Nicolae Ilinca	

Title Radiometric Calibration of the Finnish Geospatial Research Institute
Hyperspectral LiDAR

Degree programme Nano and Radio Sciences

Major Space Science and Technology	**Code of major** ELEC3039

Supervisor Prof. Miina Rautiainen

Advisor Prof. Sanna Kaasalainen

Date 19.7.2018	**Number of pages** 58	**Language** English

Abstract

The algorithmic processing of multiwavelength sampled radiance data recorded by multispectral and hyperspectral LiDAR-instruments improves upon the accuracy of reflectance retrieval and target material characterization capabilities of the instrument. A simulation script was written to study different algorithmic waveform reconstruction procedures for intensity calibration of the Finnish Geodetic Research Institute hyperspectral LiDAR, considering environment of operation, processing speed, and digitization frequency. A Gaussian parametrization, a polynomial least squares, a cubic spline, and a Levenberg-Marquardt algorithm were analyzed in terms of acquiring waveform peak amplitude, spatio-temporal peak location, FWHM, and area parameters from the samples of an approximately 1 ns FWHM Gaussian pulse. The results show that the cubic spline algorithm is best suited for implementation with FGI-HSL, as it provides an error of $0.2575 \pm 0.191\%$ in waveform peak amplitude retrieval at a sampling frequency of 4 GHz, and real-time processing capabilities at a pulse repetition frequency of 2 MHz. Based on the insight of this study, suggestions are given for algorithm choice depending on the spatio-temporal shape of the full-waveform and the required accuracy of waveform parameter retrieval as function of sampling frequency.

Keywords hyperspectral , radiometric calibration , LiDAR, spectroscopy

Preface

I want to thank Professor Sanna Kaasalainen and my supervisor Professor Miina Rautiainen for their insightful suggestions and guidance through the process of writing this thesis.

Otaniemi, 28.4.2018

Julian N. Ilinca

Contents

Symbols and abbreviations

Symbols

A_I	area illuminated by laser beam [m^2]
A_R	effective system aperture [m^2]
A_t	effective target area [m^2]
a_m	peak amplitude
C_{Calib}	calibration constant
C	coefficient in polynomial equation
D	laser aperture diameter
D_S	target diameter [m]
f_{BRDF}	bidirectional reflectance ditribution function
f_s	sampling frequency [Hz]
$\mathbf{J}$	Jacobian matrix
k	damping term
L	radiance [W m^{-2} sr^{-1}]
$\mathbf{M}$	matrix
M	number of sampling grids
m	mean
N	number of generated pulses
P_a	backscatter echo amplitude
RSTD	relative standard deviation
$R = \frac{d\Phi_r}{d\Phi_{rid}}$	reflectance factor [dimensionless]
r	range [m]
$\hat{S}$	amplitude of Gaussian system waveform
S_I	power density illuminated by laser beam
s^2	standard deviation
t	time [s]
t_{peak}	location of waveform peak in time
u	fitted value
v	absolute difference between generated and true value
w	generated value
x	position on x-axis
y	position on y-axis
β	laser beamwidth
λ	wavelength [m]
$\rho = \frac{d\Phi_r}{d\Phi_i}$	reflectance [dimensionless]
Φ	radiant flux [W]
ϕ	beam azimuth angle
θ	beam elevation angle
ϑ	error parameter used in LM algortihm
σ_E	differential scattering cross section [m^{-2} sr^{-1}]
σ_S	backscatter cross section [m^{-2} sr^{-1}]
σ	standard deviation of a Gaussian pulse
ξ	backscatter coefficient

Sub- and superscripts

atm	atmospheric
e	error
i	incident
id	ideal
R	reflected
s	scattered
m	sampling grid number
l	generated pulse number
p, e	percentual error

Operators

$f()$	function
$f'()$	first order derivative of function $f()$
$f''()$	second order derivative of function $f()$
$\int$	integral
$\sum_i$	sum over index i
T	matrix transpose

Abbreviations

A/D	analog to digital
APD	avalanche photo-diode
BRDF	bidirectional reflectance distribution function
CCRF	conical-conical reflectance factor
CS	cubic spline
DN	digital number
DSM	digital surface model
FGI	Finnish Geospatial Research Institute
FWHM	full width at half maximum
GP	Gaussian parametrization
HCRF	hemisperical-conical reflectance factor
HS	hyperspectral
HSL	hyperspectral LiDAR
LM	Levenberg-Marquardt
LSM	least squares method
LiDAR	light detection and ranging
MS	multispectral
MSCL	multispectral Canopy LiDAR
MSL	multispectral LiDAR
MWCL	multi-wavelength canopy LiDAR
OBC	on-board computer
PG	point group
PLSQ	polynomial least squares
PRF	pulse repetition frequency
SC	supercontinuum
TOF	time-of-flight
TLS	terrestrial laser scanner

1 Introduction

Terrestrial laser scanning (TLS) has found numerous applications since its birth in the 1960s in fields such as land surveying, urban planning, land cover classification, and in various population, health and safety monitoring situations (Lim et al. 2003, Du et al. 2016, Gao et al. 2015, Park et al. 2007). Throughout the mid-1980s the development of TLS-technology continued and became industrially feasible (Park et al. 2007). Within the past decade multispectral (MS) and hyperspectral (HS) TLS instruments capable of material characterization, through measurement of multiple spectral components of the backscatter waveform (also commonly called the *echo*), have emerged (Kaasalainen et al. 2008, Jutzi et al. 2003, Wagner et al. 2006, Hakala et al. 2012, Woodhouse et al. 2001). The technology has since received attention from the mining industry and space research community, as it can be used to separate various minerals and map the surrounding area simultaneously in dangerous situations. To further enhance the characterization capabilities of HS-TLS instruments, further research concerning instrument calibration and data post-processing is called for (Wagner et al. 2006).

1.1 Multi- and hyperspectral terrestrial LiDAR

Light Detection and Ranging (LiDAR) is an *active remote sensing technique*, in which the same instrument emits a coherent laser pulse, and measures the backscatter radiance of the emitted pulse (Li et al. 2001). The technique involves recording the time-of-flight (TOF) of the emitted pulse from the instrument to the object causing the scattering and back to the instrument. The distance between the instrument and the target object is called the *range* and it is calculated from the TOF. The position of the target object is recorded based on the direction a pulse is emitted in. The final product of a terrestrial LiDAR instrument is a point cloud constructed from the range, position, and radiance data (x, y, z, R) (Hakala et al. 2012).

Multispectral LiDAR (MSL) and Hyperspectral LiDAR (HSL) instruments improve upon conventional LiDAR instrumentation by separating the incident light into spectral components. An array of photo-detectors in conjunction with a high speed digitizer are used to sample the response of each detector in the array (Hakala et al. 2012, Li et al. 2001). The sampled values partially represent the waveform of the scattered pulse at different wavelengths. Algorithms and fitting procedures are applied to the sampled data to reconstruct the *full-waveform* of the scattered pulse (Jutzi et al. 2003). Reconstruction of the full-waveform allows for parameters such as the peak amplitude and full-width at half maximum (FWHM) of the echo pulse to be estimated.

The calibration of *spectral radiance* (more commonly referred to as radiance) L [W m^{-2} sr^{-1}] is a requirement for any optical remote sensing instrument used for quantitative analysis of spectral properties of the target (Schaepman-Strub et al. 2006). The optical properties of the target are distinguished and characterized by relating the spectral radiance measured to the target object through a dimensionless quantity called the *reflectance factor R*. Schaepman-Strub et al. 2006 define the

reflectance factor as "*the radiant flux* Φ *reflected by a surface to that reflected into the same reflected-beam geometry and wavelength range by an ideal and diffuse standard surface, irradiated under the same conditions*". In practice the reflectance factor is calculated as the ratio of the peak amplitude of the fitted waveform (scattered from the target surface) and the peak amplitude of the pulse that is reflected from a calibration target, (which has close to Lambertian and diffuse scattering properties) (Wagner et al. 2006, Kaasalainen et al. 2008, Hakala et al. 2012). The final point cloud product of a hyper- or multispectral LiDAR instrument contains three-dimensional position information and the spectral response of the target $(x, y, z, R(\lambda))$ (Hakala et al. 2012).

Calibration is implemented in order to correct the data recorded by a TLS instrument. The fitted waveform values are usually normalized and the reduction of incident power relative to target area and beam divergence is corrected for as a function of range (Wagner et al. 2008, Jutzi et al. 2003). System induced noise and other electronic effects that alter the radiance measurement are also accounted for (Li et al. 2001). Finally, post processing is performed on computer workstations or on an on-board computer (OBC) of a mobile instrument. Post processing includes rendering of the map from the acquired data, discarding erroneous data and using algorithms for material and target classification (Puttonen et al. 2015).

1.2 Motivation

Research concerning TLS has moved strongly toward material characterization through the use of hyper- and multispectral data in the last decade (Gaulton et al. 2013, Kaasalainen et al. 2008, Du et al. 2016). Work has also been conducted to define rigorous calibration methods for performing full waveform analysis using TLS (Wagner et al. 2006, Kaasalainen et al. 2005, Kaasalainen et al. 2016). Nonetheless, robust, real time recognition between different materials in the environment using the backscatter radiance has not been demonstrated using the HSL, calling for systematic calibration methods needed to fulfill the requirements for robust target characterization when considering DSM and intensity data integration (Li et al. 2001, Kaasalainen et al. 2008, Hakala et al. 2012, Wagner 2010).

According to the knowledge of the author, analysis concerning the effect of *sampling frequency* on the accuracy of full-waveform retrieval using multi or hyper-spectral TLS instrumentation has not been conducted. While different waveform fitting methods and algorithms have been used and suggested (Hofton et al. 2000, Jutzi et al. 2003), the relative accuracy between fitting methods in determining various waveform characteristics has not been discussed. To aid in the creation of more rigorous calibration standards of HSL and MSL instrumentation and to find the most suitable waveform analysis procedure for the Finnish Geospatial Institute hyperspectral LiDAR instrument, this thesis answers the following research questions through the use of a simulation script assessing the performance of different waveform fitting methods and algorithms at sampling frequencies between 1 GHz and 5 GHz.

- Which of the tested algorithms is best suited for reconstructing a narrow FWHM

Gaussian laser pulse considering the parameters of interest and the speed of the process?

- How much does the accuracy of waveform retrieval improve for a single peak pulse due to increased sampling frequency when using different algorithms?

A simulation based approach was chosen, since it allows a direct comparison between a computer generated laser pulse and the algorithmically fitted pulse. Such a comparison cannot be made with measurement data since the original waveform of the emitted pulse and the echo cannot be known beforehand. The reconstructed waveform provided by the algorithms is used to calculate the peak amplitude, the spatio-temporal location of the peak, the FWHM and the area enveloped by the waveform. The processing time needed to complete the fit is also analyzed. The preciseness, efficiency, and suitability of the algorithms are compared at different sampling frequencies in order to find the best suited methodology for waveform parameter retrieval. The possibility of individual erroneous measurement results and their removal is discussed.

2 Literature review

2.1 Terrestrial LiDAR

Terrestrial LiDAR technology enables the extraction of both geographical and environmental information from a non-destructive interaction between light and matter. This information can be extracted through analysis of the waveform shapes of both the transmitted and the reflected laser pulses. TLS is a so-called *active-scanning* approach (Dassot et al. 2011), defined by the TLS-system reacting to part of the radiance it initially produced, allowing it to function independent of external illumination.

TLS instruments usually emit radiation in the form of laser pulses, while some instruments use a continuous source. Only pulsed laser TLS instruments are discussed in this thesis as the current method for retrieving target reflactance is based on measuring the peak amplitude of the pulse reflected from the target. Pulsed laser TLS instruments transmit a Gaussian pulse, while the shape of the reflected pulse depends on the three-dimensional structure of the target and the diameter of the *footprint* produced by the laser beam (Wagner et al. 2006). In the case of an *extended target*, (or in other words when the target area is larger than the beam footprint), the return waveform largely maintains the shape of the emitted pulse (Hofton et al. 2000). The return waveform shape does not resemble that of the emitted pulse when the footprint is larger than the target. For example, when a laser beam with a footprint diameter of 5 cm is incident upon a small leaf. In this case, part of the beam will continue to travel until it encounters another target, while part of the beam will be scattered by the leaf. The temporal offset between scattering events is observed as multiple peaks in the echo waveform as shown by Jutzi et al. 2006 and depicted in figure 1.

2.2 Discrete return TLS

Terrestrial LiDAR instrumentation can be divided into *discrete-return* and *full-waveform* instruments. A *discrete return* TLS instrument responds to a threshold defined increase in incident radiance upon a single photo-detector and records the TOF. Discrete-return TLS- instruments, which respond to multiple echoes within a single waveform are capable of distinguishing up to five separate scattering events (Dassot et al. 2011). Discrete-return TLS instruments enable the study of structural properties of the environment in the vicinity of the scanner based on the TOF measurement. For example Beinert et al. 2006 show that canopy structure parameters such as above ground biomass, tree diameter, and tree height can be retrieved using discrete-return angular orientation and range information of commercial TLS instruments combined with suitable object classification algorithms.

Discrete-return instruments are in some cases used to estimate the intensity of the scattered signal, and thus enable target classification if the materials within the scanning environment are known. For example, Eitel et al. 2011 employ a pulsed green laser to estimate foliar nitrogen concentration. However, the fact that only one intensity value is recorded based on a predefined threshold for each waveform pulse
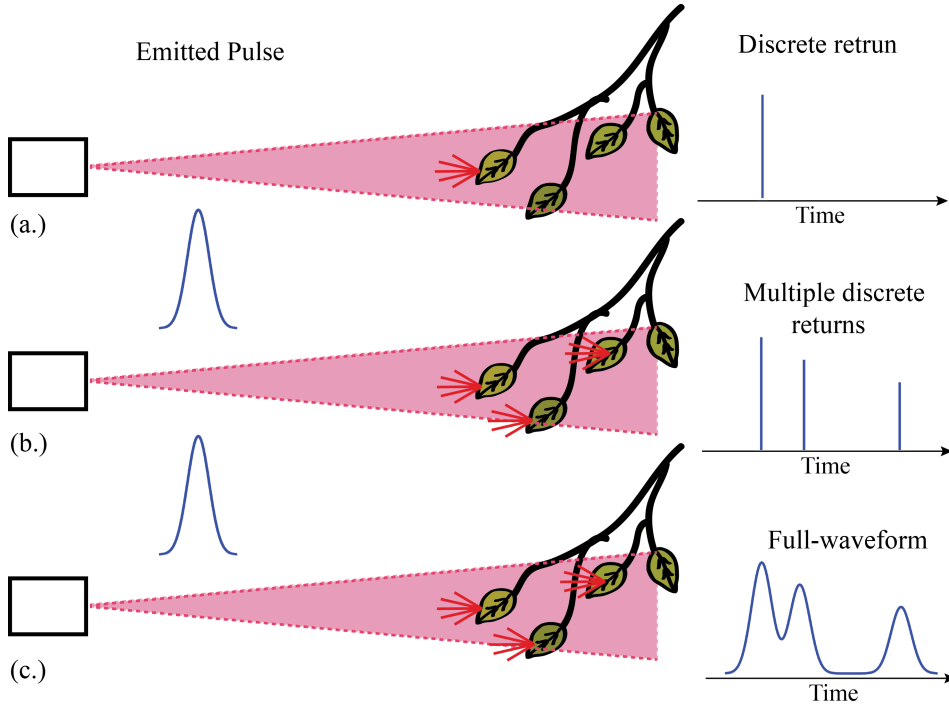
Figure 1: The different types of retrieval methods used in TLS.

makes the retrieval of the waveform peak inaccurate compared to a full-waveform detection approach.

2.3 Full-waveform TLS

TLS instruments capable of *full-waveform* reconstruction use large dynamic range high speed A/D converters to sample the echo (Gaulton et al. 2010). A suitable equation or algorithm is used to fit a line through the sampled points. Various fitting algorithms have been proposed and tested. Hofton et al. 2000 employ a fitting approach in which the recorded multi-peak waveform is decomposed into individual Gaussian components, each representing a single echo. A Gaussian equation is then used to reconstruct the waveform for each echo separately. However, the characteristics of individual echos depend on the complexity of the target, and may differ significantly from one another (Mallet et al. 2009), making simple Gaussian characterization ill-founded. To overcome this problem Chauve et al. 2009 propose the use of modified Gaussian equations. Least squares fitting method (LSM), such as the Levenberg-Marquard method, are often used to retrieve the full-waveform from the sample data (Parrish et al. 2011, Reitberger et al. 2006, Jutzi et al. 2006). LSM refine the parameters of the equation chosen to represent the waveform through multiple iterations, until a predefined number of iterations is completed or a predefined error margin is reached.

In order to uncover the full-waveform from the digitized samples, the reciprocal of the sampling frequency has to be small compared to the temporal width of the recorded waveform. For instance, digitizing a waveform with a full-width at

half-maximum (FWHM) of 1 ns with a digitization rate $f_s = 1$ GHz will provide approximately three samples from the waveform. While this might be enough to reconstruct a simple Gaussian waveform as shown by Zhu et al. 2017, more complex waveforms such as shown in figure 1 (c.) cannot be characterized properly if the digitization rate is too small.

Compared to discrete return retrieval, the full-waveform approach enables more accurate intensity and target reflectance measurement. The peak of the recorded waveform (commonly used to represent the peak intensity of the echo) can be detected when using a full-waveform approach, while the measured echo signal strength in discrete return devices is limited by the thresholding measurement approach. Peak intensity detection has enabled time series and spatial distribution analysis of chemical and pigment content in vegetation and man-made objects (Zhu et al. 2017). Full-waveform TLS instruments also enable the study of rigorous intensity calibration methods. For example, Krooks et al. 2013, and Kaasalainen et al. 2016 show through accurate intensity retrieval that the reflectance measured by the TLS instrument decreases as a function of decreasing incidence angle between the beam and the target, resulting in better means to describe the quality of the point cloud.

2.4 Multispectral LiDAR

Multispectral LiDAR instruments utilize either multiple lasers tuned to specific wavelengths (Gaulton et al. 2013), lasers configured for synchronous emission at different wavelengths (Gaulton et al. 2010), or supercontinuum generation of broadband (white light) laser pulses (Hakala et al. 2012). Supercontinuum generation is achieved through frequency mixing of high peak power narrow-band laser pulses within a non-linear photonic-crystal (optical) fiber (PCF) (Dudley et al. 2006). Broadband light is produced as a result of multiple nonlinear dispersion effects occurring in the PCF given that the input power of the laser is high enough.

The spectral bandwidth of a single return provided by a supercontinuum source is much larger than that achieved by multiple lasers or synchronous emission. Simultaneous measurement of the target spectral is limited only by the number of available sensors. A spectrograph is used to separate the different wavelength (λ) components from the scattered broadband pulse (Hakala et al. 2012), from which they are coupled via optical fibers to an array of photosensitive detectors, each responding to a different λ-component. Other methods for wavelength separation such as the use of a grating or spectral filters have been implemented by Wallace et al. 2014 and Wei et al. 2012 respectively.

Multi- and hyperspectral TLS instruments employ full-waveform processing to ensure accurate radiance retrieval for each wavelength component. Sensitive and mechanically stable optics are required as the recorded radiance changes as a function of the *phase angle*, (which is the angle between the incident beam upon the target and the backscatter direction toward the instrument) (Kaasalainen et al. 2016).

A list of hyper- and multispectral instruments is presented in table 1, along with the main technical specifications.

The non-commercial laboratory based multi- and hyperspectral TLS instrument
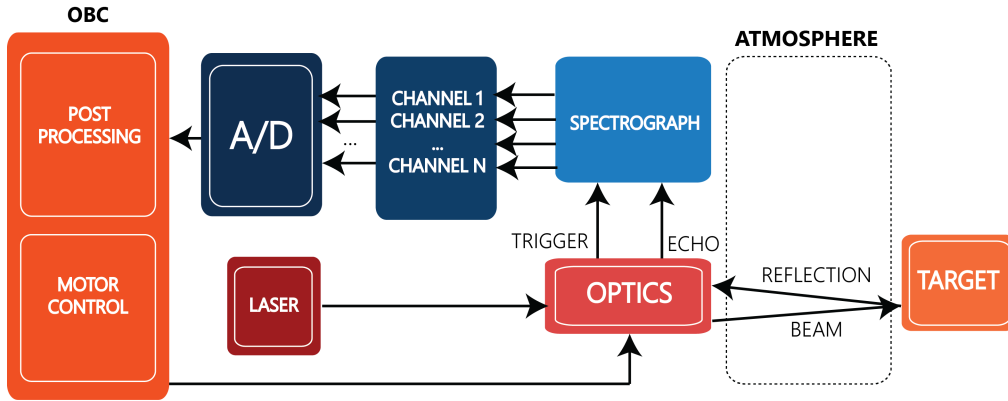
Figure 2: A narrow broadband laser pulse propagates through the PCF to a rotatable mirror configuration, from which it is directed toward the target. The backscatter radiance is focused using a concave mirror and directed to the spectrograph input. The spectrograph separates the broadband pulse into wavelength components. Individual avalanche photodiodes, electron multipliers, or CCDs respond to the radiance of different wavelength components of the scattered pulse. The response of each detector is sampled using high speed digitizers and stored for processing, or passed on to an OBC. Finally, effects due to noise or the atmosphere are removed and the peak amplitude of backscatter waveform is retrieved using a full-waveform approach.

Table 1: Terrestrial Hyper- and Multi-Spectral Instruments developed to date

Instrument	Pulse Width (ns)	f_s (GHz)	N-channels	Wavelength (nm)	PRF f_p (kHz)
SALCA	1	1	Dual	1040, 1550	5
DWEL	5.1	2	Dual	1064, 1548	2 (20)
FGI-HSL	1	1	8-16	400-2500	5
MSCL (1)	4.75	5	1	531, 550, 690, 780	20×10^{-3}
MSCL (2)	0.05	0.001	4	531, 570, 670,780	200
MWCL	N/A	N/A	4	556, 670, 700,780	0.8
Wuhan-Uni	1-2	N/A	32	538-910	20-40

prototypes presented in table 1 have been developed by individual research groups. All of the instruments differ from each other in hardware design. Field measurements showing differences in acquired intensity at different wavelengths have been completed with each instrument listed in 1. The operating wavelengths of all instruments (excluding the FGI-HSL), presented in table 1 are chosen specifically for retrieval of biochemical vegetation- or forest structural properties (Wei et al. 2012, Wallace et al. 2014, Woodhouse et al. 2001, Gaulton et al. 2010, Douglas et al. 2012). Artificial

material recognition in a forest environment is demonstrated by Hakala et al. 2012, (using the FGI-HSL instrument), through simultaneously incorporating intensity and point cloud descriptions of numerous targets over multiple measurements.

2.5 Reflectance measurement and intensity calibration

Reflectance measurement can be described as the process of converting photo-sensitive sensor data into physical parameters. Reflectance itself is a dimensionless quantity defined as the ratio of incident flux density (*irradiance*) to the portion of the exitent flux density (*exitance*) incident upon the sensor aperture (Schaepman-Strub et al. 2006). Contrary to traditional remote sensing calibration, in which sensor functions are internally monitored, current TLS instrumentation relies on calibration using external reference targets to account for system response as a whole (Wagner 2010).

Multi- and hyperspectral instruments require calibration of individual wavelength responses at each measurement channel to determine arbitrary properties from the backscatter waveform (Wallace et al. 2014). Calibration targets with different *reflectances* (defined as the ratio of the scattered and incident power spectral densities at the target (Nicodemus 1965)) are used to calibrate the system spectral response. A calibration target is used to ensure that all the channels produce reflectance values corresponding to those given by the manufacturer of the target. Employing calibration targets allows for a straightforward relation to be established between the system response and the known spectral response of the calibration target (Pfeifer et al. 2007, Krooks et al. 2013, Kaasalainen et al. 2008).

Intensity calibration can be described as an iterative process in which both mechanical system parameters and post processing software are calibrated for the system output to describe the backscatter properties of the target. The system response is composed of the optical response, the spectral power response of detectors, and system noise caused by electronics (Li et al. 2001). Wavelength-specific correction factors are implemented in the post processing software for each channel depending on the detector response. A monochromator can be used to extract the fundamental wavelength component of interest from a white laser pulse to calibrate for the detector response (Hakala et al. 2012). The amplification of the small voltage response supplied by a photo-detector is also to be accounted for. An empirical function for the amplifier response can be constructed as shown by Kaasalainen et al. 2008, to correct for nonlinearities.

In order to determine the reflectance of an arbitrary target the detector response of both the transmitted pulse (often called the *trigger*) and backscatter waveform are required by definition. However to make the measured reflectance meaningful it has to be related to some known reflectance. This is accomplished through the use of a dimensionless quantity called the reflectance factor. In practice the reflectance factor for an arbitrary target is computed as the ratio of peak amplitude of the echo and the peak amplitude of an echo measured from a calibration target (with close to Lambertian and diffuse scattering properties) (Kaasalainen et al. 2008).

The accuracy of the radiance measurement is not only influenced by the spectral response of the sensors, but also by the optical response of each individual component in the optical configuration (Kaasalainen et al. 2008). The optical configuration of a typical TLS instrument consists of a beam splitter or beam sampler and mirrors used to direct both the trigger and the backscatter beam to the spectrograph through separate paths. Optical components pertaining a constant large bandwidth spectral

response would be optimal for MS- and HS-TLS instruments. However, in reality the spectral response of optics is a function of wavelength. The wavelength dependencies are taken into account through the introduction of correction factors in the post processing software. High reflectivity mirrors with special coatings are used to minimize propagation losses within the instrument. In order to mitigate systematic error,w the mirrors are arranged to deliver maximum available radiance to the spectrograph.

The spectral response of the beam splitter is also wavelength dependent. Wavelength-specific correction factors can be implemented in the post-processing software. However, a beam splitter, as the name implies, reflects a fraction of the incident radiance and transmits the remaining fraction of the radiance toward the target. Hence, the wavelength-dependent power between the transmitted and the reflected beams has to be taken into account before performing the reflectance calculation.

Supercontinuum sources, like all laser sources, exhibit small variations in transmit power. These variations can be mitigated by normalizing the waveforms with the intensity of the transmitted pulse (Hakala et al. 2012). To characterize the target reflectance, Hakala et al. 2012 normalize the backscatter radiance with waveform peak radiance values collected at various distances using a 99% reflectivity calibration target. Individual backscatter radiance values are normalized using calibration data collected from different target ranges. The measured radiance data are normalized with the radiance values collected from echoes that correspond to the same measured range.

2.5.1 Backscatter cross section

The radiance measurement performed with a TLS instrument depends on numerous physical factors of the target, which can be characterized using a term called the *backscatter cross section*. The backscatter cross section characterizes the electromagnetic interaction that occurs at the target when it is illuminated by the incident laser beam. A part of the radiation incident on the target will be absorbed, and the rest will be scattered as dictated by the theory of radiative transfer (Wagner et al. 2006). The scattered radiant flux density in the direction of the sensor aperture is not only dependent on range, but also on the beam spread angle and the morphology, composition and angular orientation of the target relative to the incident beam direction. The beam geometry of the incident and scattered radiance is described through the solid angles of the transmitted beam Ω_i and the scattered beam Ω_s observed at the instrument aperture. Assuming an ideal Lambertian surface diffuse with scattering properties, the scattering situation is depicted in figure 3

The total scattered radiance can be viewed as the integral over each individual infinitesimal scattering target area elements dA_t. The ratio of incident to scattered power from an effective target area A_t is mathematically described through a dimensionless quantity called the *biconical reflectance* $\rho(\Omega_i, \Omega_s)$.

$$\rho(\Omega_i, \Omega_s) = \frac{P_s(\theta_s, \phi_s; \Omega_s)}{P_i(\theta_i, \phi_i; \Omega_i)} \tag{1}$$
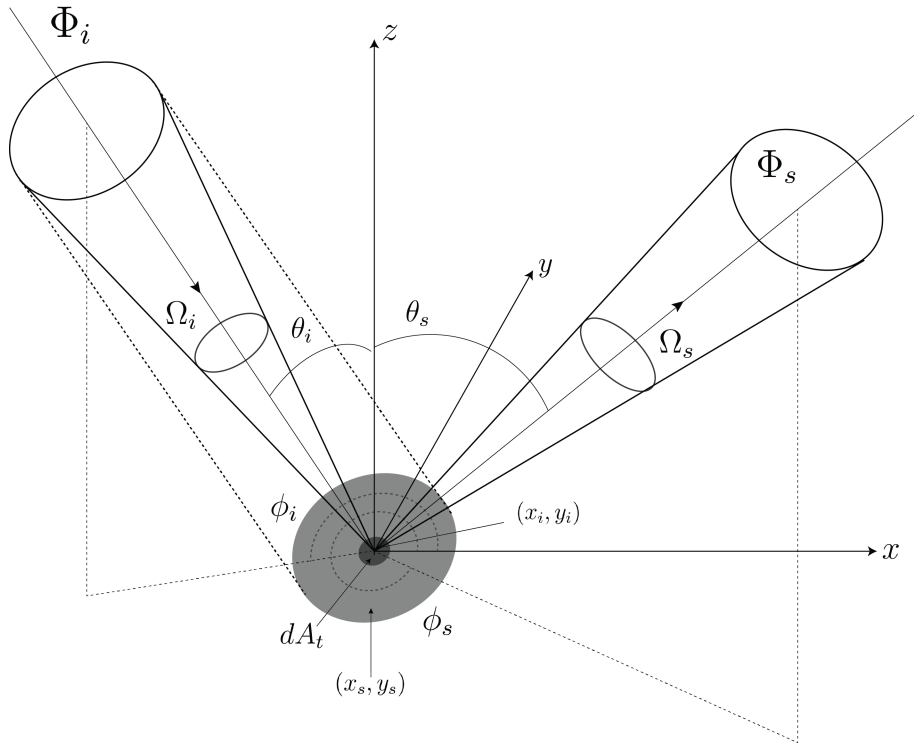
Figure 3: The incident light is reflected from differential area dA_t, and its flux density within a cone Ω_i is characterized as function of the incidence angles θ_i and ϕ_i. The scattered radiant flux density is characterized using the same geometrical approach as for the incident radiation.

To characterize the electromagnetic (EM) interaction occurring between the beam and the target, EM scattering theory begins by defining a quantity called the *differential scattering cross-section* σ_E (Jackson 1998). The EM interaction is described through the electric field vectors $\mathbf{E}_i$ and $\mathbf{E}_s$ of the incident and scattered EM-waves in the directions (θ_i, ϕ_i) and (θ_s, ϕ_i) respectively. The resulting σ_E will always be different for different targets, provided that the incident electric field maintains constant polarization and magnitude. Hence, σ_E can be used to characterize the target.

$$\sigma_E(\theta_i, \phi_i; \theta_s, \phi_s) = \lim_{r \to \infty} 4\pi r^2 \frac{|\mathbf{E}_s(\theta_s, \phi_s)|^2}{|\mathbf{E}_i(\theta_i, \phi_i)|^2} \tag{2}$$

However, the differential scattering cross section cannot be measured directly, as it is defined using infinitesimal quantities. The measurement depends on the sensor aperture area A_R. In practice an average scattering cross section is detected, which is the average of all the infinitesimal quantities over the solid angle Ω_s. The average scattering cross section is commonly referred to as the backscatter cross section σ_S with dimensions area per solid angle (m² sr⁻¹) (Wagner 2010).

A relation can be established between the incident and scattered power and the

backscatter cross section through the relation (Leader 1978)

$$\frac{P_r}{P_i} = \frac{A_R \, |\mathbf{E}_s|^2}{A_I \, |\mathbf{E}_i|^2} \tag{3}$$

where A_I is the area illuminated by incident radiation and is usually referred to as the *footprint*. By combining equations 1 and 3 and expressing the receiver aperture area in terms of the scattering solid angle ($\Omega_s = A_R/r^2$) the backscatter cross section can be written in terms of the reflectance (Wagner 2010).

$$\sigma_S = \frac{4\pi P_R}{\Omega_s P_I} A_I = \frac{4\pi}{\Omega_s} \rho(\Omega_i, \Omega_s) A_I \tag{4}$$

2.5.2 LiDAR equation

The LiDAR equation relates the measured *radiant flux density* (often called the optical intensity or just intensity), defined as the optical power per unit area, to the physical environment by taking into account the geometry of the measurement situation, backscatter cross section (Wagner et al. 2006) and the properties of the laser beam. The received power can be expressed in terms of σ_s by rearranging equation 4.

$$P_R = \frac{P_I \Omega_s}{4\pi A_I} \sigma_S \tag{5}$$

The size of the footprint area element depends on the beam spread angle β and the range r (Wagner 2010).

$$A_I = \frac{\pi r^2 \beta^2}{4} \tag{6}$$

The footprint area is proportional to the square of the range and thus targets far from the instruments may have a smaller surface area than the footprint. Assuming an extended target, the footprint area and the target area are equal ($A_t = A_I$). Furthermore the incident power at the target will be equal to the transmitted power from the TLS instrument ($P_T = P_I$). The scattering solid can also be expressed in terms of the aperture of the detector.

$$\Omega_s = \frac{\pi D_r^2}{4r^2} \tag{7}$$

By inserting equations 6 and 7 into equation 5 the received power can be expressed only in terms of the transmitted power and the measurement geometry.

$$P_R = \frac{P_T D_r^2}{4\pi r^4 \beta^2} \sigma \tag{8}$$

2.5.3   Calibration equation

Wagner et al. 2008 suggest the use of a *backscatter coefficient* ξ, which is the cross-section σ normalized relative to the footprint area. Due to the footprint area A_I being dependent on the direction of the incident beam compared to the surface normal, it is subject to change as function of $cos(\theta_i)$. The definition presented in equation 9 allows for comparison between backscatter measurements of extended targets irrespective of sensor type or instrument.

$$\xi = \frac{\sigma}{A_I cos(\theta_i)} \tag{9}$$

The backscatter coefficient is used in radar to relate the target to the system response due to the target usually being larger in size than the footprint (Woodhouse et al. 2001). However, the backscatter coefficient cannot be used for straightforward comparison if the footprint falls on multiple scatteres. This is the case in forest environments where multiple echoes are detected for one transmitted pulse (Mallet et al. 2009, Woodhouse et al. 2001). In the case the footprint falls on an extended target, the reflected waveform will exhibit only one peak as discussed by Persson et al. 2005, and ξ can be a more convenient parameter to use (Wagner et al. 2006).

Inserting equations 6 and 9 into equation 8 allows for the LiDAR equation to be expressed in terms of backscatter coefficient.

$$P_R = \frac{P_T D_r^2}{16r^2} \eta_{sys} \eta_{atm} \xi \tag{10}$$

Parameters η_{sys} and η_{atm} account for power loss due to system sensitivity and atmospheric propagation respectively (Wagner et al. 2008). Assuming that the scattered radiance (scattered from infinitesimal scattering area elements dA_t) can be described through a Gaussian function with a standard deviation s_i and that the echo waveform is also Gaussian, with an amplitude $\hat{S}$ and standard deviation s_s, the system waveform amplitude P_a can be described with a function resulting from the convolution of the two Gaussian functions.

$$P_a = \frac{D_r^2 \hat{S} s_s}{16r^2 \sqrt{s_s^2 s_i^2}} \eta_{sys} \eta_{atm} \xi \tag{11}$$

The full derivation for equation 11 is given by Wagner et al. 2006. To relate the backscatter coefficient ξ to the pulse amplitude and to construct a calibration equation, equation 11 is solved for ξ and the constant terms are separated from the variables.

$$\xi = C_{calib} \frac{r_i^2 P_a \sqrt{s_s^2 + s_i^2}}{\hat{S} \eta_{atm}} \tag{12}$$

2.5.4   Reflctance factor

The reflectance factor is a dimensionless quantity (closely related to the backscatter cross section), through which the measured reflectance is related to the radiance

reflected from ideal Lambertian surface under isotropic and diffuse illumination conditions. The reflectance factor hence provides a means to describe and normalize every radiance measurement relative to a known radiance. An abridged derivation of the reflectance factor is given in this section based on the original work of Nicodemus 1965.

Nicodemus 1965 constructed and defined a purely physical and unified approach to characterize the reaction between an incident beam of light and reflecting surface through a function termed as the *bidirectional reflectance distribution function* (BRDF). The definition has since been used to mathematically characterize various scattering geometries reflectance factors (Schaepman-Strub et al. 2006). The derivation of the reflectance factor begins from the definition of the BRDF in which the incident and scattered radiant fluxes are characterized through solid angles. The BRDF is then finally related to the *biconical reflectance factor* (or conical-conical reflectance factor CCRF), which characterizes most terrestrial remote sensing geometries (Schaepman-Strub et al. 2006).

The BRDF describes the target reflectance as a function of irradiance incidence angles θ_i and ϕ_i (between the plane of the target and the plane of the incident radiation) and the angles of scattered radiation θ_s and ϕ_s (between the plane of the target and that of the radiance). The beam geometries relative to an ideal target are depicted in figure 3 and the function itself is mathematically described by equation 16.

For isotropic, uniform, and monochromatic illumination the incident radiant flux $d\Phi_i$ is expressed in terms of the incident radiance L_i, the differential area element within the footprint dA_t, and the solid angle $d\Omega_s$ containing the observed radiant flux density.

$$d\Phi_i = L_i cos(\theta_i) d\Omega dA_t = dE_i dA_t \qquad (13)$$

dE_i in equation 13 expresses an infinitesimal part of the incident radiance upon the differential area element dA_t. The reflected radiance L_s can thus be expressed as the collection of all the dE_i elements and a proportionality factor $F(\theta_i, \phi_i, x_i, y_i; \theta_s, \phi_s, x_s, y_s)$ which describes the target response to incident radiance. The points (x_i, y_i) and (x_s, y_s) denote the locations of the differential area elements on which the incident radiance falls and from which it is reflected respectively. Nicodemus et al. 1977 make two assumptions about the proportionality factor; (1) that the radiant flux incident at (x_i, y_i) affects the exitance at (x_s, y_s) through some interaction between matter and radiation, and (2) that this interaction is a function of the distance between points (x_i, y_i) and (x_s, y_s) . The exitance at the location (x_s, y_s) as a contribution of all the incident radiance from the direction (θ_i, ϕ_i) is expressed with the integral equation:

$$dL_r(\theta_i, \phi_i; \theta_s, \phi_s, x_s, y_s) = dE_i \int_{A_t} F(\theta_i, \phi_i, x_i, y_i; \theta_s, \phi_s, x_s, y_s) dA_t \qquad (14)$$

Assuming isotropic and uniform target properties along with uniform and diffuse radiance upon the target, both assumption (1) and (2), made about the EM-interaction within the target and the distance between the (x_i, y_i) and (x_s, y_s) can

be ignored due to symmetry when integrating over the proportionality factor F. Equation 14 simplifies to

$$dL_r = dE_i \int_{A_t} F(\theta_i, \phi_i; \theta_s, \phi_s) dA_t \tag{15}$$

The integral term on the RHS of equation 15 describes the response of an isotropic and uniform target to the incident radiation as function of incidence and exitance angles and is called the *bidirectional reflectance distribution function.*

$$f_{BRDF} = \frac{dL_r(\theta_i, \phi_i; \theta_s, \phi_s)}{dE_i(\theta_i, \phi_i)} \tag{16}$$

The BRDF can not be measured due to it being derived using infinitesimal quantities. However, the reflected radiant flux density Φ_s is a measurable quantity, which can be expressed in terms of the BRDF and the integral over all the irradiance components as explained above.

$$dΦ_s = dA_t \int_{\Omega_s} \int_{\Omega_i} f_{BRDF}(\theta_i, \phi_i; \theta_s, \phi_s) L_i(\theta_i, \phi_i) cos(\theta_i) d\Omega_i \Omega_s$$
$$= \int_{\Omega_s} f_{BRDF} \Phi_i \Omega_s \tag{17}$$

The ratio of the incident and reflected flux densities $(d\Phi_s/d\Phi_i)$ define the *biconbical reflectance* $\rho_{BCR}(\Omega_i, \Omega_s)$ given in equation 18. By inserting equation 18 into equation 4 the backscatter coefficient can be expressed in terms of the BRDF (Wagner 2010).

$$\rho_{BCR}(\Omega_i, \Omega_s) = \frac{1}{\Omega_i} \int_{\Omega_s} \int_{\Omega_i} f_{BRDF}(\theta_i, \phi_i; \theta_s, \phi_s) d\Omega_i d\Omega_s \tag{18}$$

The *biconical reflectance factor* (CCRF) denoted by $R(\Omega_s, \Omega_{s,id})$ is defined as the ratio of the radiant flux reflected from the surface described by a biconical refelctance distribution function f and the radiant flux reflected from an ideal surface with $f_{BRDF} = 1/\pi$ within the same beam geometry under isotropic and diffuse illumination (Schaepman-Strub et al. 2006).

$$R_{CCRF} = \frac{\Phi_s}{\Phi_{r,id}} = \frac{\int_{\Omega_s} \int_{\Omega_i} f(\theta_i, \phi_i; \theta_s, \phi_s) L_i(\theta_i, \phi_i) d\Omega_i d\Omega_s}{(1/\pi) \int_{\Omega_s} \int_{\Omega_i} L_i(\theta_i, \phi_i) d\Omega_i \Omega_s}$$
$$= \frac{\pi}{\Omega_i \Omega_s} \int_{\Omega_s} \int_{\Omega_i} f_{BRDF}(\theta_i, \phi_i; \theta_s, \phi_s) d\Omega_i d\Omega_s \tag{19}$$

Hence, the CCRF gives a good approximation of target surface characteristics when the measured radiance is normalized with the post-processed sensor response data collected from a smooth, high reflectance calibration target, with close to ideal reflectance properties (Hakala et al. 2012).

3   Materials

3.1   FGI-hyperspectral LiDAR instrument

The main components of the hyperspectral TLS (FGI-HSL) instrument developed at the Finnish Geospatial Research Institute are presented in the block diagram of figure 2. The instrument is calibrated using an external Spectralon reflectance target. A supercontinuum (SC) laser source and PCF manufactured by LEUKOS are used to excite a narrow, approximately FWHM=1 ns, broadband pulse with an average power of 200 mW and a spectral range between 500 nm-1900 nm. The pulse repetition frequency (PRF) of SC-source is tunable between 5 kHz and 30 kHz. A refracting collimator is attached to the transmitting end of the PCF from which the beam is transmitted to beam sampler, which reflects 10% of the incident spectral power to trigger the TOF measurement. The beam sampler is characterized by stable spectral response characteristics, a large bandwidth, and selective polarization properties. Furthermore, the beam sampler is manufactured with an optical coating canceling out internal reflections. The optical configuration is illustrated in figure 4.
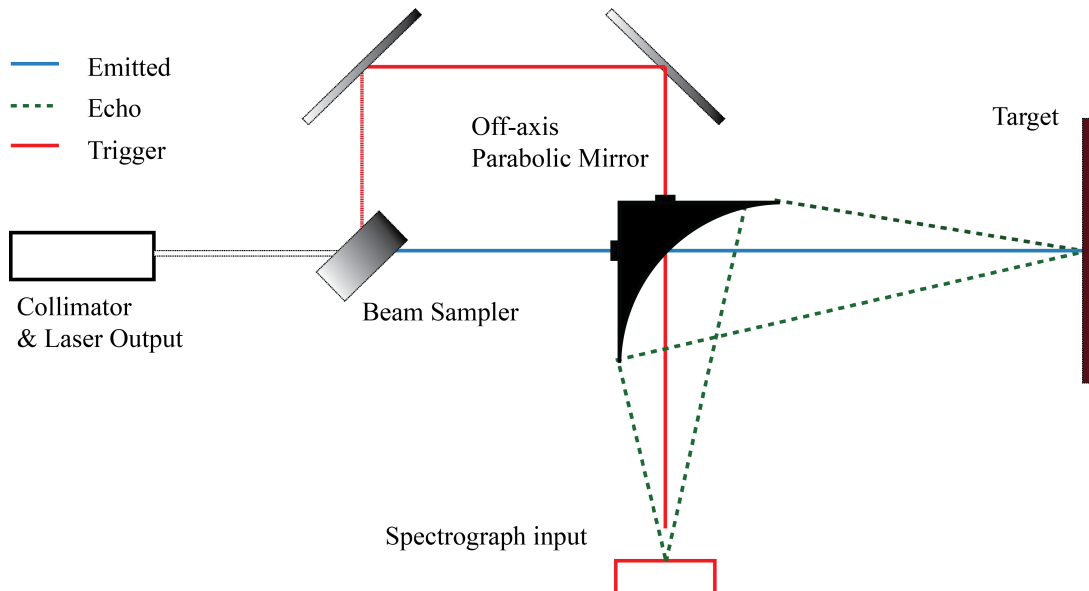
Figure 4: The beam sampler divides the laser beam into a trigger component and transmitted component. The trigger beam is directed to the spectrograph input by two mirrors and passes through a hole drilled in the off-axis parabolic mirror. The transmitted beam travels to the target. The echo is focused by an off-axis parabolic mirror to the spectrograph input, after which it is separated into its spectral components.

The beam exiting the instrument is directed through a hole drilled in the direction of the optical center line of an off-axis parabolic mirror. The same parabolic mirror is used as the primary focusing optic for the incident radiance at the collection aperture of the instrument. The focused beam is directed to a Specim ImSpectro V10 spectrogrph, which is equipped with a 50 μm slit, and has a spectral range between 400 nm and 1000 nm. The spectrally separated radiance components are

detected and converted to analog voltages with an avalanche photo-diode (APD) array composed of 16 individual sensors. The APDs (manufactured by First Sensor) have a nonlinear and wide spectral response bandwidth between 450 nm and 1050 nm. Calibration of the spectral response of the APDs is performed with the Oriel, Cornerstone 74125 monochromator. The analog voltage output of each sensor is separately digitized with a dynamic range of 14 bits and 1 GHz sampling frequency. To further enhance the full-waveform retrieval capability of the instrument, A/D converters with a digitization rate between 1 GHz and 5 GHz are being considered based on the trade-off between waveform accuracy and processing speed discussed in section 5.

Two motorized rotators with an accuracy of $\pm 0.0115°$ are used to control the three dimensional scanning geometry of the instrument. The commands dictating rotator movement, and thus the azimuth and elevation of the transmitted beam, are stored and associated with each time stamped return waveform.

3.2 Sampling problem

Only a few different algorithmic waveform fitting approaches concerning TLS full-waveform data processing and calibration for obtaining the full-waveform information from the digitized radiance input data have been utilized and proposed (Persson et al. 2005, Hofton et al. 2000, Wagner et al. 2006, Jutzi et al. 2003, Shen et al. 2017, Chauve et al. 2009). Furthermore, these approaches have not been discussed in the context of hyperspectral TLS instrumentation and calibration in literature. A collection of these methods is presented in the following sections along with some unanswered questions concerning HSL calibration and the accuracy of waveform retrieval.

The waveform of the scattered pulse is usually assumed to be Gaussian or to consist of multiple superimposed Gaussian components (Hofton et al. 2000, Wagner et al. 2006). A method utilizing a purely Gaussian waveform model in conjunction with the Levenberg-Marquardt algorithm is suggested by Hofton et al. 2000. Jutzi et al. 2003 also assume a Gaussian waveform but use a Gauss-Newton method to find the Gaussian parameters that best fit the sampled data, while Persson et al. 2005 use an expectation maximization algorithm to accomplish the same. However, in reality the temporal shape of the waveform is not always best represented by a Gaussian equation as shown by Zhu et al. 2017. To overcome the fitting problem presented by non-Gaussian waveforms, Chauve et al. 2009 suggest the use of modified Gaussian equations. Finally, Shen et al. 2017 suggest an algorithm based on B-Splines, making no preliminary assumptions about the waveform shape.

Information regarding the use of fitting methods for HSL- and MSL-instruments when considering accuracy of intensity retrieval has not been discussed in-depth. Only Puttonen et al. 2015, Woodhouse et al. 2001, Du et al. 2016, and Wallace et al. 2014 provide information about the fitting methods used for waveform parameter retrieval when considering the instruments presented in table 1. Furthermore, the accuracy of fitting methods employed by existing HSL- and MSL -instruments (presented in table 1) have only been discussed in the context of spectral characteristic separation

capabilities and not absolute accuracy. While the waveform reconstruction algorithms listed above have been shown to provide reflectance estimates that are accurate enough for the characterization of vegetation parameters at well defined wavelengths with a relatively large spectral separation between the measured components (Zhu et al. 2017, Du et al. 2016, Puttonen et al. 2015), the accuracy of individual measurements provided by the fitting methods have not been discussed. When moving toward hyperspectral remote sensing, where the spectral difference between measured wavelength components becomes smaller, and the final goal is to perform material recognition of multiple targets with less distinct absorption spectra than for example biological objects such leafs have, the accuracy of individual measurements becomes increasingly important.

The relation between the digitization rate and the accuracy of waveform parameter retrieval has received little attention in literature. This presents a problem concerning the need for increased physical data storage capabilities of HS-TLS instrumentation as more data needs to be stored as the sampling frequency is increased and as more detection channels are added. Most research groups have utilized digitization frequencies that provide ample sampled data to describe the waveform but have not reported the optimization between waveform accuracy, processing speed and data storage needs for HS-TLS instrumentation. The amount of radiant flux data gathered can be optimized, while maintaining high intensity detection accuracy, by studying the relation between the A/D conversion frequency and the accuracy of the fitting algorithms. As contemporary literature does not report absolute accuracy of parameter retrieval algorithms it is hard to make conclusions about which full-waveform algorithms are appropriate for analyzing specific waveforms. In the next section we present the methods used to find a suitable fitting algorithm and digitizing frequency for the FGI-HSL instrument.

Finally, the waveform peak amplitude representing the radiance has been used in all research as parameter to categorize between different material groups. Waveform parameters such as the pulse FWHM and the spatio-temporal area enveloped by the pulse have not been considered as distinguishing characteristics for material recognition application, based on the literature considered in this thesis.

4 Methods

Altogether four different fitting methods and algorithms are presented and their accuracy in retrieving four different parameters characterizing a Gaussian return waveform are analyzed as a function of sampling frequency. The parameters are the peak amplitude a_m, the FWHM of the Gaussian pulse, the area enveloped by the waveform A, and the spatio-temporal location of the waveform peak t_{peak}. These parameters are also presented in figure 6. The algorithms tested for consist of, a nth order polynomial least-squares (PSLQ) algorithm, a cubic spline (CS) fitting method, a Gaussian waveform parametrization (GP) method, and a Levenberg-Marquardt (LM) algorithm.

The algorithms analyzed were chosen based on previously used fitting methods for TLS full-waveform retrieval that have provided sub-centimeter range accuracy, or have shown to provide high enough accuracy in peak amplitude retrieval to enable material characterization through full-waveform reconstruction. When choosing the algorithms to be tested, consideration was given to the computational requirements and mathematical complexity of the algorithms. Complex algorithms involving support vector machines or expectation maximization (as presented by Du et al. 2016 and Persson et al. 2005 respectively) demand larger processing times, and thus are not discussed. This is due to the trade-off between accuracy of parameter retrieval and processing speed being of interest when considering the FGI-HSL instrument.

Both the MSCL (1) (Woodhouse et al. 2001) and the FGI-HSL-instrument (Puttonen et al. 2015) (presented in table 1) employ a Gaussian parameterization algorithm for retrieving the full-waveform and performing material characterization. The GP algorithm is considered due to its mathematical simplicity and straight forwad Gaussian parameter retrieval capability. The choice for testing the LM algorithm is based on its wide use in determining range with sub-centimeter accuracy by using the peak amplitude of the echo waveform (Hofton et al. 2000). The processing time required by LM algorithm depends on the accuracy of input parameters given as initial guesses for a waveform parametrization equation.

The choice for the CS algorithm is supported by the work of Shen et al. 2017. WhileShen et al. 2017 do not discuss waveform parameter retrieval capabilities of the B-spline fitting method (which is a cubic spline algorithm) in terms of material characterization, the group shows that the waveform shape can be retrieved by comparing the B-spline method to the well established Gaussian parametrization method, thus making the spline method fitting accuracy worth investigating. Furthermore, a spline method has not been used so far in conjunction with MSL or HSL instrumentation. The CS algorithm does not involve iteration for finding the parameters of interest (unlike the LM algorithm), making it appealing when considering real-time processing capabilities.

The PLSQ method was chosen on the basis that it is a widely used fitting approach and does not involve iteration in finding the parameters of interest. The PLSQ algorithm does not employ an iterative approach for finding the waveform parameters, making it appealing when considering real-time processing capabilities. The algorithms used are presented along with the script in the following sections.

Further information regarding the algorithms is given by Seber et al. 2003, Papoulis 1984, Levenberg 1944, and Marquardt 1963.

4.1 Algorithms

4.1.1 Cubic spline algorithm

A cubic spline (CS) is a piecewise polynomial function, consisting of $n-1$ cubic polynomial equations f_i where n is the number of data points fitted. Each polynomial f_i represents a piece of the spline between two consecutive data points at x_i and $x_i + 1$.

$$f_i(x) = C_3(x - x_i)^3 + C_2(x - x_i)^2 + C_1(x - x_i) + C_0, \tag{20}$$

where x_i denotes the ith sampled data point and x represents a value between x_i and $x_i + 1$. To produce the spline, the polynomials are linked to each other at the sampled data points. In order for the spline to be smooth, each polynomial piece f_i and their first and second order derivatives f_i' and f_i'' between x_{i-1} and x_i have to be continuous. Continuity must also occur at points where the piecewise polynomials are joined so that $f_i(x_i) = f_{i+1}(x_i)$, $f_i'(x_i) = f_{i+1}'(x_i)$ and $f_i''(x_i) = f_{i+1}''(x_i)$. Furthermore, the following relations between the piecewise polynomials and their derivatives have to hold. (Seber et al. 2003)

$$f_i(x_i) = C_{i0} \tag{21a}$$
$$f_i(x_{i+1}) = C_{i3}(x_{i+1} - x_i)^3 + C_{i2}(x_{i+1} - x_i)^2 + C_{i1}(x_{i+1} - x_i) + C_{i0} \tag{21b}$$

$$f_i'(x_i) = C_{i1} \tag{22a}$$
$$f_i'(x_{i+1}) = 3C_{i3}(x_{i+1} - x_i)^2 + 2C_{i2}(x_{i+1} - x_i) + C_{i1} \tag{22b}$$

$$f_i''(x_i) = 2C_{i2} \tag{23a}$$
$$f_i''(x_{i+1}) = 6C_{i3}(x_{i+1} - x_i) + C_{i2} \tag{23b}$$

The coefficients C can be solved so that they are represented only by the data points x_i and the first order derivatives using equation 21 and equation 22. The first order derivatives are evaluated in terms of the continuity constraints and the constraints given by equation 23. This in turn produces $m-2$ equations for all the m unknown first order derivatives of each piecewise polynomial. In order to solve all the unknowns, boundary conditions are applied for the first and last points of the spline by setting $f_0'' = f_{n-1}''$, where n is the number of data points used to form the spline. (Seber et al. 2003)

4.1.2 Polynomial least squares algorithm

The polynomial least squares algorithm solves the coefficients of a mth order polynomial equation that best describes the waveform to be fitted.

$$y = c_0 + c_1 x_1^1 + c_2 x_2^2 + ... + c_m x_n^m = \sum_{j=0}^{m} c_j x_i^j \tag{24}$$

The least-squares algorithm minimizes the sum of differences between the observed y values and the polynomial by finding the coefficients c_j in the following equation

$$F(c_0, c_1, ..., c_n) = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{m} c_j x_i^j \right)^2 \tag{25}$$

in which i represents the data point number from altogether n data points and j denotes the order. The derivative of F in equation 25 is computed with respect to each $j = 1, 2...m$ separately and set to zero as shown in equation 26.

$$\sum_{i=1}^{n} x_i^j \left(y_i - \sum_{k=0}^{m} c_k x_i^k \right) = 0 \tag{26}$$

The degree with respect to which the derivative is being calculated is represented by j and k ranges from 0 to m. This creates $n+1$ linear equations for each coefficient. The coefficients can then be solved from the linear equations through applying matrix algebra. (Young 2014)

4.1.3 Gaussian parametrization algorithm

The parametrization method solves for the parameters A, μ and σ of the Gaussian equation through using three sampled values and their spatio-temporal locations, presented by the points $(x_1, y_1; x_2, y_2; x_3, y_3)$ as shown in figure 5. The parametrization method presented here is a slightly modified version of the method used by Puttonen et al. 2015 for full-waveform retrieval.

$$g(t) = A e^{\frac{-(t-\mu)^2}{2\sigma^2}} \tag{27}$$

The points are chosen so that one is the sampled peak value and the two others are sampled values at equal temporal distances from both sides of peak value. To solve analytically for the parameters no more than three points are required. From now on, these three points will be referred to as a *point group* (PG). The equations used to solve for the parameters are constructed by solving equation 27 separately for each parameter.

$$\mu = \frac{1}{2} \frac{x_1^2(ln(y_3) - ln(y_2)) + x_2^2(ln(y_1) - ln(y_3)) + x_3^2(ln(y_2) - ln(y_1))}{x_1(ln(y_3) - ln(y_2)) + x_2(ln(y_1) - ln(y_3)) + x_3(ln(y_2) - ln(y_1))} \tag{28}$$

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 - (x_1 - \mu)^2}{2(ln(y_1) - ln(y_2))}} \tag{29}$$

$$A = y_2 e^{\frac{(x_2 - \mu)^2}{2\sigma^2}} \tag{30}$$

Once the parameters have been solved, they are inserted in equation (27) and $g(t)$ is calculated using a sufficiently small Δt to represent the Gaussian smoothly.

The script performs altogether three parametrization procedures for each pulse using three different point groups when possible. Different point groups, all containing the sampled peak value, are formed in order to investigate the effect the location of the samples relative to the waveform peak have on the accuracy of the fit. The first point group includes two points which lie closest to the sampled peak value. The second point group contains two points approximately at half the maximum amplitude of the waveform. Finally, the third point group contains two points close to the base of the pulse as depicted in figure 5.
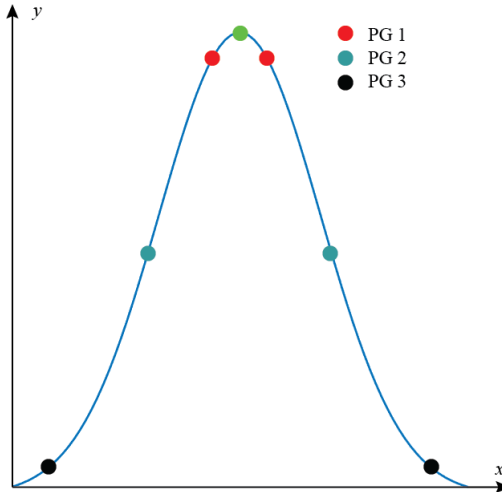
Figure 5: The three different point groups (PG) composed of sampled data points. The green data point is included in each point group.

4.1.4 Levenberg-Marquardt algorithm

The Levenberg-Marquardt (LM) algorithm is based on the work by Levenberg 1944 and Marquardt 1963. The algorithm takes an iterative approach to minimizing the differences of the squares between a given function $f(\mathbf{g})$ and a vector of measured data points $\mathbf{d}$. In $f(\mathbf{g})$, $\mathbf{g}$ is a vector holding initial guesses for the parameters of the given function. In the scope of this research the parameters used for the initial guesses are a_g, μ_g, and σ_g of the Gaussian equation. The initial guess for a_g is provided by the sample with the highest value, while the guess for μ_g is given by the position of a_g in time. Finally the guess for σ_g is generated through an iterative approach in which two samples on both sides of the maximum value as close as possible to $a_g/2$ are found. An approximate FWHM is calculated by subtracting the spatio-temporal position of these values, and σ_g is calculated using

$$\text{FWHM} = 2\sqrt{2ln2}\sigma_g \tag{31}$$

A Taylor series is used to linearly approximate $f(\mathbf{g})$.

$$f(\mathbf{g} + \delta_{\mathbf{g}}) \approx f(\mathbf{g}) + \mathbf{J}\delta_{\mathbf{g}} \tag{32}$$

In equation 32 $\mathbf{J}$ is the Jacobian matrix of $f(\mathbf{g})$ and $\delta_{\mathbf{g}}$ a vector holding factors that apply small increments or decrements to terms in $\mathbf{g}$ during the iteration. The condition for the minimization, where $\mathbf{d}$ represents a vector containing the sampled values and ϑ represents the error between sampled value and the guess function, is given by

$$|\mathbf{d} - f(\mathbf{g}) - \mathbf{J}\delta_{\mathbf{g}}| = |\vartheta - \mathbf{J}\delta_{\mathbf{g}}| \tag{33}$$

The minimum difference condition is met when $\mathbf{J}^T(\mathbf{J}\delta_{\mathbf{g}} - \vartheta) = 0$. The equation can then be rearranged to the form:

$$\mathbf{J}^T\mathbf{J}\delta_{\mathbf{g}} = \mathbf{J}^T\vartheta, \tag{34}$$

and solved for $\delta_{\mathbf{g}}$. Equation 34 is modified in the implementation of LM algorithm so that the term $\mathbf{J}^T\mathbf{J}$, which represents an approximation of the second order partial derivatives, is written as a single matrix $\mathbf{M}$ with the addition of a *damping term k* for the diagonal elements of $\mathbf{J}^T\mathbf{J}$.

$$\mathbf{M}\delta_{\mathbf{g}} = \mathbf{J}^T\vartheta \tag{35}$$

The algorithm begins by changing the initial guess and checking whether the error ϑ has decreased or increased. In the case an increase is observed, k is increased and the equation 35 is solved in an iterative manner until a value for $\delta_{\mathbf{g}}$ is found that decreases the error. If the error, on the other hand, is decreased, the term $\delta_{\mathbf{g}}$ is accepted, and the process of solving equation 35 is repeated until a predefined error threshold is reached or a predefined number of iterations is completed.
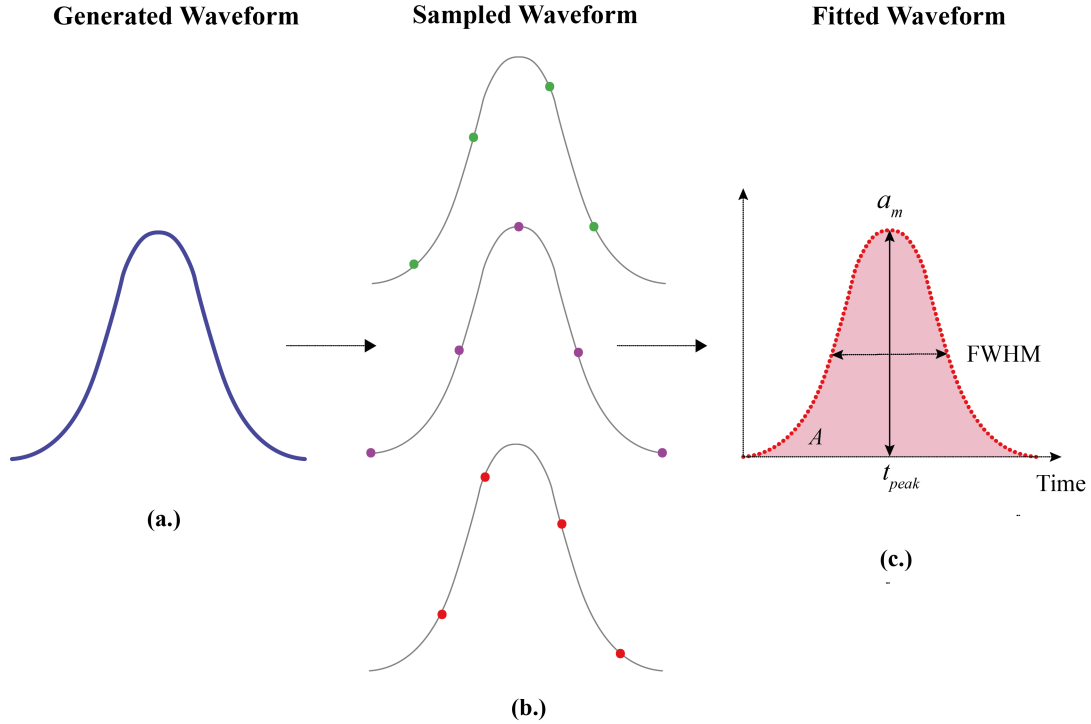
4.2 Script

The script evaluates the accuracy of the algorithmic fitting methods described above in finding the parameters of interest described. The script compares *generated (Gaussian) waveforms* to *fitted waveforms* produced by fitting data points sampled from the generated waveforms. Figure 6 shows a flow diagram of the three main stages of the fitting procedure and the nomenclature used for describing the process as a whole.

$$g(t) = A_g e^{\frac{-(t-\mu_g)^2}{2\sigma_g^2}} \tag{36}$$

In order to construct the *generated pulses*, a Gaussian function presented in equation 36 is used. The variables A_g, μ_g and σ_g are changed randomly within predefined intervals as each generated pulse is constructed. The intervals chosen are based on measured data observed during calibration test measurements. The Gaussian parameters are changed randomly in order to simulate sensor response at different wavelengths and to simulate fluctuations of the laser output power. Furthermore, Gaussian noise is added to each generated waveform to simulate system noise. The choice of Gaussian noise as the noise model is supported by noise measurements made during calibration tests for the FGI-HSL.

Generated Waveform   **Sampled Waveform**   **Fitted Waveform**

Figure 6: The three main stages of the script are described by a.) the generation of a Gaussian waveform, b.) the sampling of the generated waveform using different kernels to create *sampled waveforms*, and c.) the fitting of the sampled waveforms to produce the final *fitted waveforms*. The *parameters of interest* used for assessing the accuracy of the fit are shown in c.), where a_m is the maximum amplitude of the fitted pulse, A is the area enveloped by the pulse, FWHM is the full width at half maximum, and t_a is the spatio-temporal location of maximum amplitude.

A single generated pulse consists of 24000 points within a 300 ns interval. The pulse itself has a FWHM of approximately 1 ns. All the data points representing a generated pulse are saved in an array and used later to compute errors for the parameters of interest. Once a pulse is generated, it is sampled by a user defined number of different grids or kernels at a user defined sampling (or digitization) frequency of 1 GHz, 2 GHz, 3 GHz, 4 GHz, or 5 GHz. This allows the sampled points to be retrieved from different spatio-temporal locations of the pulse. This distinction is important, since in reality different returns are not sampled at identical locations of the temporal waveform due to the varying flight time and waveform characteristics.

After a generated pulse is sampled, the sample data is stored into a temporary matrix. The algorithms are then called one by one to produce the final *fitted pulses* from the sampled data. Each algorithm performs the fit on the same set of samples, so that a valid comparison can be made between the accuracy of different algorithms. Before the fitting begins, a threshold is set that dictates which sampled points will be taken into account in the fit. The thresholding operation works by choosing sampled points with an amplitude value higher than the noise floor level. This type of

procedure ensures, that only points within the actual pulse are used when performing the fit. Furthermore, the fitting procedure is sped up as the points outside the waveform of interest are not included in the fit.

After a fitted pulse is created (for each sampling grids at specified sampling frequency), its data is stored in an array. Before finding the parameters of interest from the fitted data, the contribution of the simulated system noise is removed. The system noise is estimated from the sampled points which do not represent the waveform itself. This fitted pulse is then used to find the peak amplitude, spatio-temporal location of the peak, area enveloped by the waveform, and FWHM. This process is repeated for all the fitted pulses.

To perform a statistical error assessment for the parameters of interest, altogether 100 pulses are generated. These pulses are then sampled using five sampling grids, constituting a set of 500 pulses for each different sampling frequency. Finally all the pulses are fitted. This process is then repeated 100 times. This constitutes a set of 50 000 fitted pulses for each algorithm at each sampling frequency. The error calculation is done via finding the difference between a value of interest (such as the peak amplitude) and the corresponding generated "true" value. The mean error and the standard deviation of the difference between the true and fitted values is then calculated separately for each algorithm. The whole script can be found via the link provided in the bibliography (Ilinca 2018).

4.3   Area and FWHM

The full width at half maximum (FWHM) of the fitted pulse for PLSQ and CS algorithms are calculated through an iterative approach. The values closest to half the maximum on both sides of a_m (depicted in figure 6) are searched for. Once the values are found, their locations in time are determined and subtracted from each other. In the case of the Gaussian Parametrization algorithm equation 37 is used to relate the σ to the FWHM.

$$\text{FWHM} = 2\sqrt{2ln2}\sigma \tag{37}$$

Two different methods are employed for calculating the area under the waveform. The first approach relies on the *trapezoidal rule* for computing the approximation of a definite integral. The trapezoidal integration method involves the summation of the areas of consecutive trapezoids as described by equation 38. The parameters y and x express the location of the points on the y- and x-axes respectively, and i denotes point number on the waveform. The individual regions are summed up to retrieve an approximation the total area under the pulse.

$$A_{pulse} = \sum_i \frac{1}{2}(y_{i+1} - y_i)(x_{i+1} - x_i) \tag{38}$$

The second method is presented in equation 39, where the maximum amplitude is multiplied by $\sqrt{2\pi}\sigma$ which is a solution for the Gaussian integral as given by Papoulis 1984.

$$A = \sqrt{2\pi}\sigma a_m \tag{39}$$

This method is justified as the generated pulse is Gaussian. Furthermore multi-peak waveforms can be analyzed as sum of Gaussian functions as discussed by Hofton et al. 2000 and Wagner et al. 2006.

4.4   Error calculation

The error analysis takes into account the mean error and standard deviation of the parameters of interest retrieved from the fitted pulse. To make the explanation of the error analysis clear, we refer to the parameters of interest retrieved from the fitted pulse as *fitted values* u_{lm} and the respective parameters of the generated pulse as *true values* w_l. The subscript $l = 1, 2, ..., N$ denotes the pulse number of the generated pulse which is sampled, and from which the fitted pulse is constructed, and N the total number of generated pulses. The subscript $m = 1, 2, ..., M$ denotes the sampling grid number out of $M = 5$ sampling grids.

The error (for each parameter of interest produced by each different algorithm) is calculated as the absolute difference between a fitted value and a corresponding true value $v_{lm} = |w_l - u_{lm}|$. The mean of all individual differences is then computed (corresponding to a single algorithm) to express the mean error m_e. A percentual mean error $m_{p,err}$ is also calculated which expresses the mean percentual difference between the parameters of interest (retrieved from the fitted waveforms) and the true values (retrieved from the generated waveforms). The computation for the percentual mean error is expressed by equation 40. The percentual mean error (provided by each different algorithm at each different sampling frequency) is calculated separately for each parameter of interest.

$$m_{p,err} = \frac{\sum\limits_{l=1}^{N} \sum\limits_{m=1}^{M} \left(\frac{|w_l - u_{lm}| + w_l}{w_l} - 1 \right)}{N \times M} \tag{40}$$

The standard deviation is calculated by subtracting the differences v_{lm} from the mean error m_e for all the pulses analyzed

$$s^2 = \sqrt{\frac{\sum\limits_{l=1}^{N} \sum\limits_{m=1}^{M} (m_e - v_{lm})^2}{N \times M - 1}} \tag{41}$$

Finally the relative standard deviation (RSTD) is given by equation 42.

$$RSTD = \frac{m_e}{s^2} \tag{42}$$

Both the PLSQ and CS algorithms use the same number of discrete data points for reconstructing a smooth waveform from the sampled values. This allows a fair comparison to be made for the time it takes for both algorithms to complete the waveform reconstruction and parameter retrieval processes. In the case of the LM and PG-algorithms reconstruction of the waveform is not necessary for retrieval of the parameters of interest, as they can be directly solved from the Gaussian parameters.

5 Results

5.1 Fitted waveform peak amplitude retrieval

The accuracy of peak amplitude retrieval directly from the sampled waveform depends strongly on whether the sampled peak values happen to fall close to the peak of the generated waveform. Figures 7a-7d contrast the accuracy of fitted waveform peak amplitudes a_m retrieved at 1 GHz and 5 GHz sampling frequencies between the CS algorithm and those retrieved directly from the sampled data. The method of retrieving the peak amplitude directly from the sampled waveform exhibits the largest mean error as can be seen from table 2. However, the mean error of the sampled waveform peak value remains within a range of approximately 1% of the generated waveform peak amplitude value at $f_s =$5 GHz. Furthermore, the sampled waveform mean error at 5 GHz is smaller than the mean error produced at $f_s \leq 2$ GHz by all other fitting methods. However, the RSTD for the sampled waveform mean error is approximately 130% at all analyzed sampling frequencies indicating a large mean error distribution.

In the case of the CS algorithm most of the retrieved peak values are underestimated relative to the generated peak amplitude as shown in figure 7b at a 1 GHz sampling frequency. This also holds true for the PLSQ algorithm. However, the retrieved values become evenly distributed on both sides of the generated values as the sampling frequency increases indicating that the accuracy of peak amplitude retrieval does not depend strongly on the spatio-temporal position of the samples on the pulse. This is seen from figure 7d in which the red line represents the generated true values and the colored dots represent the retrieved peak values using the CS algorithm. The same behavior is also exhibited by all the retrieved maxima when looking at the error distributions of the other algorithms. The Levenberg-Marquard algorithm produces the smallest mean error at $f_s = 1$ GHz , due to multiple completed iterations and good initial guesses for the model waveform parameters.

The LM algorithm produces the smallest percentual mean error for the peak amplitude at a 5 GHz sampling frequency. Compared to the CS algorithm, (which provides the second best accuracy out of the fitting methods tested), the mean error from the LM algorithm differs by only 0.09%. However, all the fitting methods provide a mean error of less than 0.5% from the true amplitude at $f_s \geq 3$ GHz. The PLSQ and GP algorithms show mean errors with only a difference of 0.002% to the CS algorithm at $f_s = 5$ GHz. This indicates that the effect the chosen algorithm has on the accuracy of the peak amplitude diminishes as the sampling frequency is increased.

Figure 8 depicts the results presented in tables 2 and 3. As the sampling frequency increases, both the mean error and the standard deviation of the mean error decrease for all fitting methods presented. Considering the CS and PLSQ algorithms, the mean error exhibits a sharp rise of approximately 15% as sampling frequency decreases from 2 GHz to 1 GHz. This jump can be directly attributed to the decreasing number of sampled points available for solving the polynomial equations used by the CS and PLSQ algorithms. The rise in mean error between the aforementioned
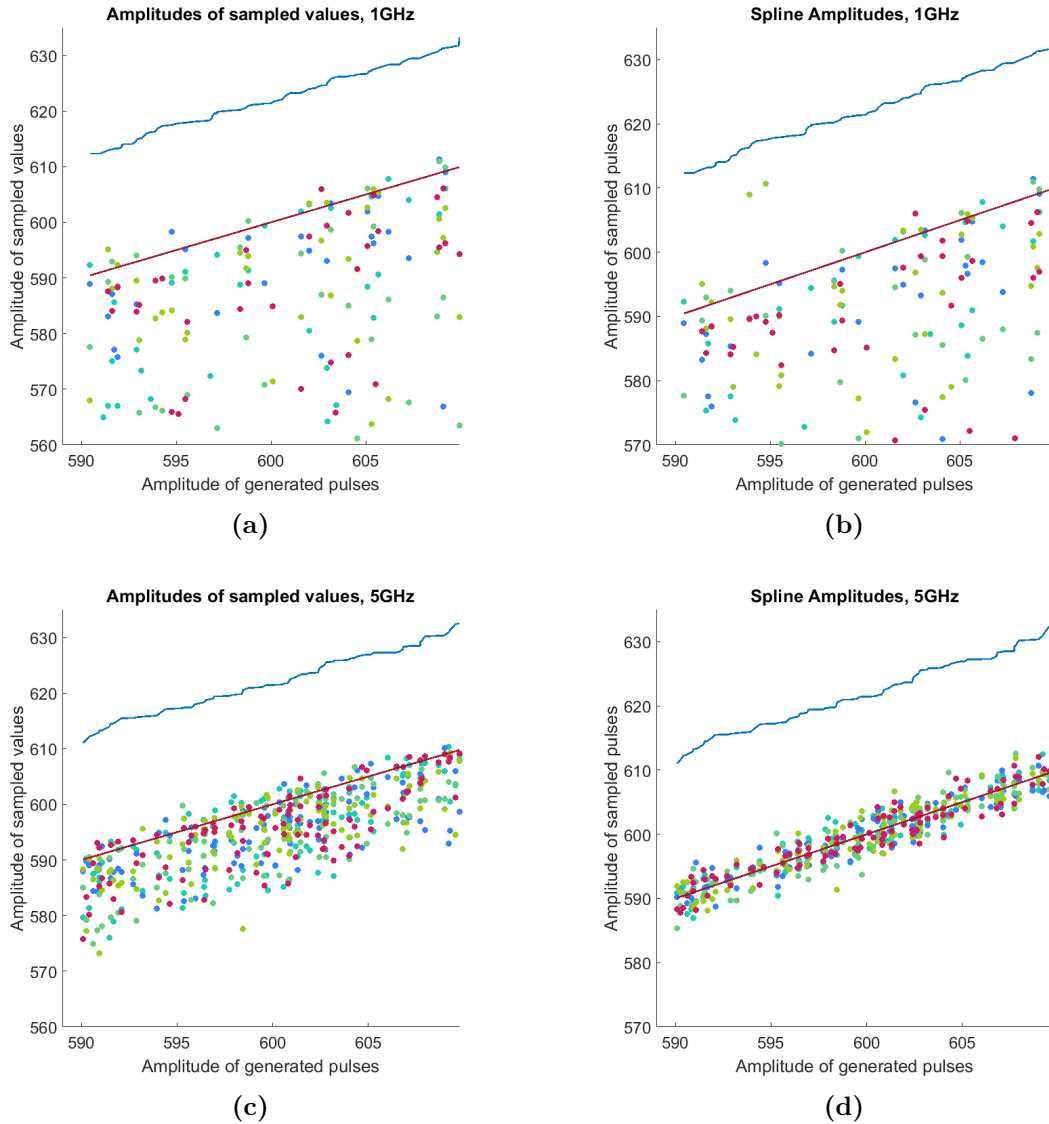
Figure 7: The Maximum amplitude a_m values retrieved using the CS algorithm from 100 individual wave forms are presented in figures (b.) and (d.). Maximum amplitude values retrieved from the sampled waveform are presented in figures (a.) and (c.). The different color dots represent peak values acquired using different sampling grids (or kernels). The red line represents the maximum amplitudes of the generated waveform without additional noise. The blue line represents maximum amplitude values of the generated waveform with additional Gaussian noise. The noise is removed from the sampled values before the algorithm produces the fit.

sampling frequencies is not as abrupt for the LM and GP methods, since they rely on parametrized equations to reconstruct the waveform. The mean errors resulting from the LM algorithm rise sharply as the sampling frequency decreases below 3 GHz, and become larger than the mean errors retrieved using the CS algorithm at $f_s = 2$ GHz. However, the LM algorithm still provides the smallest mean error and

Table 2: The mean errors and standard deviations of the peak amplitude parameter computed from the fitted waveforms, through application of the different fitting algorithms and methods.

f_s	5 GHz	4 GHz	3 GHz	2 GHz	1 GHz
Sampled Waveform					
m_{error} (%)	0.8907	1.3394	2.3200	5.0236	17.3624
s^2	6.9023	10.5555	18.4418	39.9748	135.3423
RSTD (%)	129.1607	131.3376	132.5118	132.8266	130.123
Cubic Spline (CS)					
m_{error} (%)	0.2504	0.2573	0.3563	1.7315	16.2436
s^2	1.1221	1.1486	1.5601	8.1157	81.6132
RSTD (%)	75.618	74.4210	73.0154	78.3128	84.1292
Least Squares Polynomial (PLSQ)					
m_{error} (%)	0.2526	0.2510	0.3248	2.2642	17.1658
s^2	1.1486	1.1254	14.4642	12.9023	86.8956
RSTD (%)	75.8148	74.7513	75.1597	95.2685	84.7178
Levenberg-Marquardt (LM)					
m_{error} (%)	0.1659	0.1850	0.2470	2.2518	5.3718
s^2	0.6275	0.9541	2.8406	20.7822	23.2638
RSTD (%)	57.7449	85.3791	188.1104	154.5274	72.4776

Table 3: Peak amplitude mean errors calculated by comparing the generated waveform and fitted waveform for the Gaussian Parametrization method.

s_f	PG	m_{error} (%)	s^2	RSTD (%)
5GHz	1	0.2545	1.1561	75.7551
	2	0.2705	1.1226	75.5506
	3	0.3122	1.4898	79.5589
4GHz	1	0.2551	1.1439	74.7472
	2	0.2784	1.2588	75.3960
	3	1.1405	18.6544	269.5448
3GHz	1	0.2600	1.1768	75.4660
	2	0.2812	1.2820	76.0332
2GHz	1	0.2801	1.2581	75.4759
	2	0.6401	5.5443	124.0723
1GHz	1	3.4600	21.2871	102.7154

RSTD at higher sampling frequencies. Considering the GP method, the smallest mean error occurs when the point-grid holding sampled values closest to the peak (PG 1) is used.

Considering all the fitting methods, only a minor decrease of approximately 0.1%
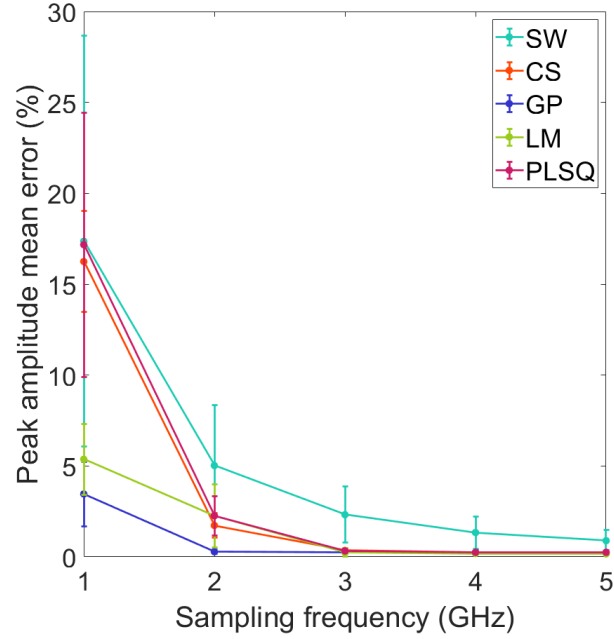
Figure 8: Peak amplitude mean errors and their maximum deviation indicated by the error bars. The symbol SW references the sampled waveform

is observed in the mean error when the sampling frequency is increased from 3 GHz to 5 GHz as can be seen from figure 8. The standard deviation of the PLSQ algorithm increases sharply at the $f_s = 3$ GHz indicating a decrease in the peak amplitude retrieval accuracy compared to the retrieval accuracy at $f_s = 4$ GHz and $f_s = 5$ GHz. Finally, it is important to note that the mean error values presented in the tables 2 and 3 are subject to slight variation between ±0.02. This is observed from running the simulation for 50k pulses multiple times. However, results only from one simulation are shown due to the large number of pulses analyzed.

5.2 Peak position retrieval

The retrieval of peak amplitude positions, (defined as the spatio-temporal location of the fitted-waveform peak amplitude value), is conducted through finding the location of the maximum amplitude value in time. The GP method shows the smallest mean error and relative standard deviation in finding the peak position out of all the tested algorithms.

(a)　　　　　　　　　　　　　　　**(b)**
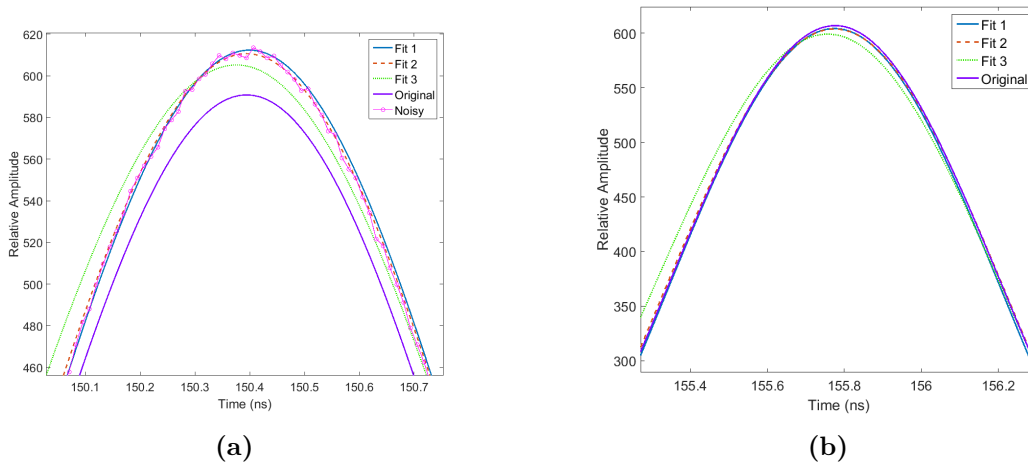
Figure 9: The fitted peak values and their locations as a result of the GP method. (a) No noise removal has been implemented and both the noise-induced and the original generated pulses are shown. In (b) the noise is removed. While fits 1 and 2 (produced using PG 1 and PG 2) provide an accurate reconstruction of the waveform, fit 3 (PG3) provides an erroneous fit due to the magnitude and location of the sampled points being distorted by noise.

The smallest mean errors are achieved when two out the three samples used in the GP method have a value approximately at half of the maximum amplitude (as presented in figure 5 by PG 2). Figure 9 further depicts the difference between the peak amplitude locations retrieved using different point groups of the same pulse by comparing them with the generated waveform peak amplitude position- and the noise-induced peak amplitude positions. The mean position error produced by the GP method remains relatively small and comparable to the mean errors produced by the PLSQ and CS algorithms when using PG 1 at $f_s = 5$ GHz and $f_s = 4$ GHz. All in all, the GP method gives the smallest m_{error} of 0.0015% for the position retrieved from the fitted waveform. However, it is very sensiteve to small changes in the spatio-temporal location and the amplitude of the samples.

The PLSQ algorithm provides a smaller mean error than the CS algorithm for sampling frequencies between 2 GHz and 4 GHz, but maintains a larger standard deviation. All the fitting methods produce a mean error in the order of 10^{-3}ns for $f_s \geq 3$ GHz, except the LM algorithm. The relatively large mean error produced by the LM algorithm is due to the possibility of the algorithm finding multiple values which satisfy the Gaussian objective function. The mean error of the LM algorithm

Table 4: Mean error and standard deviation of the spatio-temporal waveform peak position. The mean errors are calculated from true peak positions of the generated pulse and the peak amplitudes retrieved from the fitted waveforms.

| f_s	5 GHz	4 GHz	3 GHz	2 GHz	1 GHz
Sampled Waveform					
m_{error} (ns)	0.0498	0.0625	0.0837	0.1254	0.2496
s^2	0.0576	0.0722	0.0963	0.1442	0.2872
RSTD (%)	115.7411	115.4504	115.1733	115.0413	115.1435
Cubic Spline					
m_{error} (ns)	0.0044	0.0038	0.0069	0.0196	0.1852
s^2	0.0034	0.0029	0.0051	0.0108	0.1234
RSTD (%)	78.2008	76.1462	73.4775	55.3481	66.7396
Least Squares Polynomial					
m_{error} (ns)	0.0044	0.0033	0.0046	0.0167	0.2188
s^2	0.0078	0.0077	0.0032	0.0128	0.1517
RSTD (%)	142.1744	77.1895	70.0444	76.2752	69.5291
Levenberg-Marquardt					
m_{error} (ns)	0.0236	0.0353	0.0691	0.1429	0.5740
s^2	0.0236	0.0349	0.0683	0.1419	0.8782
RSTD (%)	99.7984	88.7433	86.6601	90.0019	135.3785

(at $f_s = 5$ GHz) is approximately 10 times larger compared to the mean errors achieved using the other algorithms at the same sampling frequency. However, the mean error retrieved using the LM algorithm (at $f_s = 5$ GHz) is approximately 2% of the pulse FWHM, and still provides sub-centimeter accuracy for range retrieval. Figure 10 presenting peak position accuracy of the PLSQ algorithm, shows that the spatio-temporal location of the samples on the waveform has a a minute effect on the accuracy of peak position (t_{peak}) retrieval. This same statement holds true for all of the tested algorithms in the scope of this analysis.

Noise is the largest contributing factor to the accurate acquisition of the fitted waveform peak position, especially at higher sampling frequencies, where multiple samples close to peak are collected. This is due to the generated waveform being deformed as a whole due to minor oscillations superimposed by noise. Due to the nature of the CS and PLSQ algorithms, this can lead to slightly erroneous polynomial equations. Hence, the spatio temporal location of the peak value becomes altered. This effect is a probable reason for the smaller mean error observed at $f_s = 4$ GHz, than that observed at $f_s = 5$ GHz when comparing the results of the PLSQ algorithm.

Estimating the peak position at $f_s = 5$ GHz shows the smallest mean error, with an order of 10^{-3}ns. This constitutes an error in the range measurement of 8.3 mm which is 41 times smaller than the possible range attainable from the sampled waveform.

Table 5: Gaussian parametrization peak position mean errors and standard deviation. The results are organized by point group (PG) and sampling frequency. An illustration of the sampled points organized in the point groups is given in figure 5.

s_f	PG	m_{error} (%)	s^2	RSTD (%)
5GHz	1	0.0028	0.0023	81.158
	2	0.0016	0.0012	75.4140
	3	0.0047	0.0038	81.205
4GHz	1	0.0023	0.0018	80.1265
	2	0.0021	0.0016	76.0079
	3	0.0208	0.0258	123.5744
3GHz	1	0.0018	0.0014	79.5694
	2	0.0017	0.0013	75.8165
2GHz	1	0.0015	0.0012	77.0572
	2	0.0057	0.0051	89.498
1GHz	1	0.0068	0.0033	48.0279
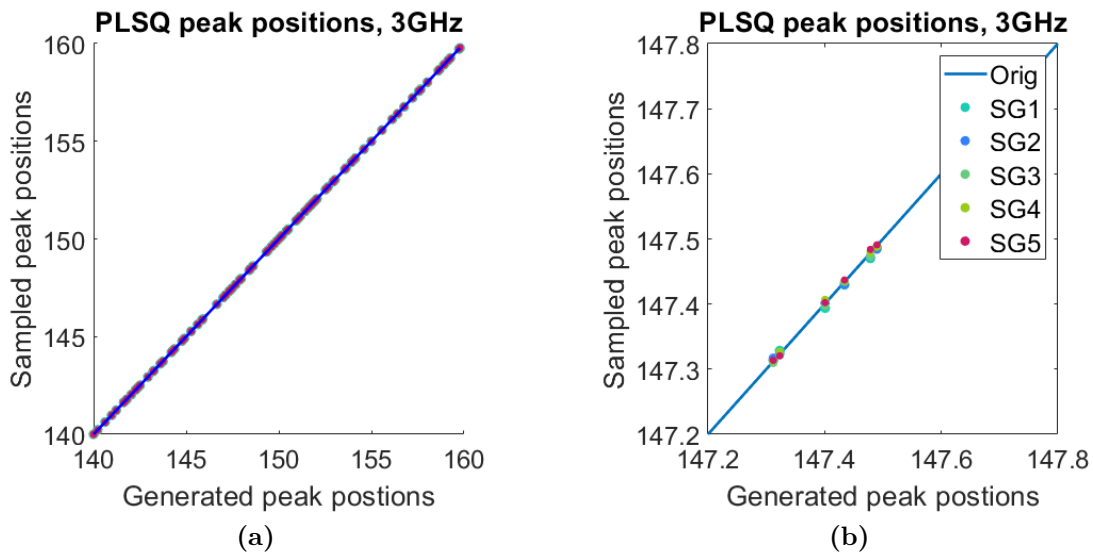
Figure 10: (a) Spatio-temporal peak value locations retrieved from all 5 sampling-groups (SG) with respect to the generated values (represented by the blue line) using the PLSQ algorithm at 3 GHz . A close-up of figure (a) is presented in figure (b). The points depict peak position location found for each fitted pulse, while the blue line depicts the peak positions of the generated pulse. The peak values attained from different sampling grids are depicted by different color points. All of the retrieved peak position values corresponding to same waveform fall on the same location, indicating that the sampling grid has no meaningful effect on the accuracy of the position retrieval when the sample spacing is one third of the pulse FWHM. The same pattern is repeated for all the algorithms analyzed.

5.3 FWHM retrieval

All the algorithms produce a mean error of less than 0.01 ns at tested sampling frequencies above 3 GHz as shown in tables 6 and 7. The RSTD values for CS and PLSQ methods are large compared to the mean error. Above $f_s = 3$ GHz the maximum error is approximately 0.1 ns, but at $f_s \leq 3$ GHz the error becomes comparable to the 1 ns FWHM of the pulse.

The LM algorithm provides the smallest mean error and RSTD values, with an error that is 4 times smaller than that for the PLSQ algorithm and two thirds of that produced by CS algorithm at $f_s = 5$ GHz. The RSTD values in table 6 show that the LM algorithm maintains an accuracy of $0.0024 \pm 6.8 \times 10^{-4}$ ns, at $f_s = 5$ GHz, for an approximately 1 ns FWHM pulse. The relative standard deviation changes by less than 20%, while the mean error remains at less than one hundredth of a nanosecond until $f_s = 3$ GHz .

In the case of the PLSQ algorithm the $f_s = 1$ GHz results are disregarded. This is due to the order of the polynomial equation decreasing to one due to too few data points. This means a linear fit is produced for a nonlinear pulse, yielding erroneous results.

Table 6: Mean errors, standard deviation and RSTD between the FWHM of the generated waveform and the FWHM of the fitted waveform.

f_s	5 GHz	4 GHz	3 GHz	2 GHz	1 GHz
Cubic Spline					
m_{error} (ns)	0.0032	0.0035	0.0052	0.0319	0.4937
s^2	0.0040	0.0044	0.0063	0.0387	0.5543
RSTD (%)	125.371	124.646	121.3410	121.5133	112.2830
Least Squares Polynomial					
m_{error} (ns)	0.0096	0.0037	0.0051	0.0508	N/A
s^2	0.0690	0.0034	0.0064	0.0704	N/A
RSTD (%)	718.486	125.970	125.553	138.8271	N/A
Levenberg-Marquardt					
m_{error} (ns)	0.0024	0.0022	0.0030	0.0182	0.1474
s^2	0.0012	0.0014	0.0024	0.0143	0.0967
RSTD (%)	57.7449	59.1508	76.2929	79.5319	65.7239

For sampling frequencies below 3 GHz the mean error rises sharply compared to the mean errors achieved using $f_s \geq 3$ GHz for all fitting methods except for the GP method. The mean error of the FWHM produced by the CS method at $f_s = 1$ GHz is approximately as large as half the FWHM of the generated pulse, (with a RSTD that is larger than the mean error). The LM algorithm can also exhibit mean errors larger than 20% of the pulse width at $f_s = 1$ GHz, which is due to the initial guess for the σ_g parameter becoming worse as the sampling frequency becomes smaller.

The GP method shows small mean errors for all sampling frequencies as long as PG2 is utilized. The noise induced amplitude variation of sampled points is relatively more pronounced close to the peak value than at mid-amplitudes of the waveform. It is also worth noting that at $f_s = 1$ GHz the RSTD of the GP method doubles compared to $f_s = 2$ GHz. At a sampling frequency of 3 GHz the mean error of the FWHM for an individual pulse can be over one tenth of a nanosecond. Since the generated pulses under analysis have an FWHM of approximately 1 ns, the error for sampling frequencies below 3 GHz is considered to be large.

Table 7: Gaussian Parametrization FWHM Errors

s_f	PG	m_{error} (ns)	s^2	RSTD (%)
	1	0.0186	0.0147	79.041
5GHz	2	0.0033	0.0025	75.237
	3	0.0054	0.0043	78.812
	1	0.0121	0.0096	79.058
4GHz	2	0.0034	0.0026	76.100
	3	0.0186	0.0234	125.902
3GHz	1	0.0074	0.0058	78.1935
	2	0.0034	0.0026	76.7180
2GHz	1	0.0044	0.0034	87.1479
	2	0.0082	0.0082	100.1216
1 GHz	1	0.0620	0.0459	204.6408

5.4 Area retrieval

The results presented for the CS and PLSQ algorithms in table 8 have been calculated using the Gaussian integral equation, given in equation 39. Utilization of the Gaussian integral provides smaller mean errors for the area enveloped by the waveform compared to using the trapezoidal rule (when calculating over 200 discrete points). However the trapezoidal method improves significantly when using the CS algorithm at 5 GHz, when 1000 points were used to reconstruct the fitted waveform. In this case the mean error is only 0.2243%, with an RSTD of 124%. Comparing the results between the trapezoidal and Gaussian area retrieval methods for the CS algorithm shows that minor errors in recovering the Gaussian parameters lead to large errors when determining the area enveloped by the pulse. However, the trapezoidal method (when integrating over a reconstructed waveform consisting of 1000 points) does not produce a smaller mean error than the Gaussian integral method at $f_s = 5$ GHz in the case of the PLSQ algorithm. This indicates that PLSQ algorithm better represents the waveform shape as a whole.

Table 8: Mean errors and standard deviations between areas enveloped by the generated waveform and the ones retrieved using the different algorithms. The areas are computed with the Gaussian integral equation.

f_s	5 wGHz	4 GHz	3 GHz	2 GHz	1 GHz
Cubic Spline					
m_{error} (%)	3.3514	3.4102	3.5088	4.7158	25.0202
s^2	22.7560	22.9001	23.7929	32.0699	171.1779
RSTD (%)	100.3373	100.2701	100.3727	101.2122	102.3882
Least Squares Polynomial					
m_{error} (%)	0.8273	0.3942	0.4491	2.5700	N/A
s^2	44.4134	3.4328	3.6725	23.2592	N/A
RSTD (%)	732.4650	125.6240	121.4668	137.2880	N/A
Levenberg-Marquardt					
m_{error} (%)	0.1659	0.1873	0.2243	0.8882	13.1700
s^2	0.8046	0.7139	0.9392	4.3298	69.7804
RSTD (%)	80.7553	56.3911	61.9640	72.9676	77.6586

The area mean error resulting from the application of the PLSQ algorithm (and the trapezoidal rule) is smaller for the 4 GHz than the 5 GHz sampling frequency as shown in table 9. This indicates that over-fitting may occur due to the large order of the polynomial equation at 5 GHz. The difference in mean error produced by the CS algorithm, between $f_s = 5$ GHz and $f_s = 4$ GHz is 0.0236% and the RSTDs differ only be 0.0267 % , showing that increasing the sampling frequency above 4 GHz (when using the CS algorithm), only makes a minute difference in accuracy.

Using the Gaussian integral equation in conjunction with the GP method proves to be the most accurate method for waveform area retrieval at sampling frequencies

Table 9: Mean errors and standard deviations between generated waveform areas and the areas retrieved using the different algorithms computed with trapezoidal rule.

f_s	5 GHz	4 GHz	3 GHz	2 GHz	1 GHz
Cubic Spline					
m_{error} (%)	0.2243	0.2478	0.2848	0.6586	25.0202
s^2	1.8896	2.0921	23.7929	5.2072	171.1779
RSTD (%)	124.7271	124.9943	124.7971	119.7393	102.3882
Least Squares Polynomial					
m_{error} (%)	1.3051	1.2259	0.9697	5.4197	N/A
s^2	12.1596	9.1337	3.6725	37.3826	N/A
RSTD (%)	135.8520	111.1196	119.9568	102.1060	N/A

Table 10: Mean errors and standard deviations between generated waveform areas and the areas retrieved using the GP method and computed using the integral of the Gaussian equation.

s_f	PG	m_{error} (%)	s^2	RSTD (%)
	1	1.5545	12.928	80.0271
5GHz	2	0.2862	3.512	76.0720
	3	0.5107	4.811	76.3355
	1	0.9740	5.3031	79.7834
4GHz	2	0.3243	1.6429	75.1620
	3	1.3339	9.9429	112.8711
3GHz	1	0.5433	2.9388	79.3186
	2	0.2933	1.5316	75.8142
2GHz	1	0.3135	1.6544	77.8695
	2	0.4732	2.5377	79.7617
1 GHz	1	6.3242	6.1557	14.4099

below 3 GHz. This is most probably due to the generated waveform being a perfect Gaussian, and the integral used to evaluate the area being the Gaussian integral. At $f_s = 4$ GHz (for the GP method) the decrease in mean error compared to $f_s = 5$ GHz can be explained as a result of the points in point group 1 being located farther in time from the peak value. The overall smallest mean error for the area covered by the waveform is given by the LM algorithm at $f_s = 5$ GHz. However, this due to the parametrization equation used in conjunction with the LM algorithm being a Gaussian equation.

5.5   Computation efficiency

Table 11 presents the time taken by each algorithm to retrieve the parameters of interest from one pulse. The values presented do not take into account the time used to calculate the mean noise per channel, since the calculation can be completed using the first few 1000 recorded pulses per channel. The same channel-wise mean noise value can be subtracted from the acquired peak values without further processing.

The GP method is computationally the most efficient compared to the other algorithms and methods due to its simplicity and no need for iterative processes, (as it can retrieve the peak value using only three sampled data 'points). The GP method performs roughly 1000 times faster than the next fastest CS algorithm at $f_s = 4GHz$. The speed of parameter acquisition by the GP method is further enhanced by the fact that removal of noise has to be completed for a considerably smaller amount of data points. The CS algorithm preforms roughly 2.5 times faster than the PLSQ algorithm, when not taking the noise removal time into account.

Table 11: Mean elapsed time for algorithms to retrieve the parameters of interest. 1000 intermediate data points are utilized to reconstruct the waveform using the CS and PLSQ algorithms. The GP and LM algorithms are both able to compute the Gaussian parameters from which the peak value can directly be accessed, and thus the time for reconstructing the waveform is not calculated.

f_s	Cubic Spline	Polynomial-LSM	Gaussian Parametrization	Levenberg-Marquardt
5 GHz	7.6268×10^{-4} s	2.006×10^{-3} s	5.896×10^{-7} s	0.0223 s
4 GHz	$5.6270 \times 10^{-4} s$	1.523×10^{-3} s	6.206×10^{-7} s	0.0170 s
3 GHz	6.9343×10^{-4} s	1.814×10^{-3} s	6.637×10^{-7} s	0.0206 s
2 GHz	7.0926×10^{-4} s	1.712×10^{-3} s	6.502×10^{-7} s	0.0233 s
1 GHz	6.2065×10^{-4} s	1.248×10^{-3} s	1.142×10^{-6} s	0.0187 s

If the noise mean value is subtracted from the pulses directly after the CS fit is complete, one instance is timed to take 7.74×10^{-4} s. However a more computationally efficient way to retrieve the parameters of interest is to store multiple parameters in a temporary array and subtract the mean error due to noise from all values at once using matrix algebra. When this method was timed for 500 fitted waveforms, the elapsed time was 1.1×10^{-3}. For the CS and PLSQ algorithms the peak value acquisition is faster than retrieving the FWHM due to the absence of iterations required to determine FWHM from the reconstructed waveform. Even though the process of computing the FWHM is more cumbersome for the PLSQ and CS algorithms, it holds value in the calibration process since it allows the measurement of pulse widening as a function of range. (Li et al. 2001)

The computation of the FWHM when using the PLSQ and CS algorithms requires more processing power than the computation of the FWHM using the GP and LM algorithms. This is due to the iterative approach required to find the points at the

half-maximum location of the fitted waveform reconstructed by the CS and PLSQ algorithms, as explained in section 4.3.

The retrieval of peak values is more efficient if multiple peaks are retrieved at once from a large set of reconstructed waveforms stored in arrays, compared to retrieving the peak values directly after the fit has been completed. When considering the data of 500 fitted waveforms stored in a single array the retrieval of all maximum values took 1.4×10^{-3}s. By comparing the elapsed times between finding the peak amplitudes from an array with waveform data stored from 500 waveforms, to finding one peak at a time as each pulse was fitted, the procedure of finding the peak values from the array was 6.87 times faster. Thus, if the parameters of interest are retrieved separately from each measured waveform a large buffer for data storage is of importance when post processing is performed on an OBC as the sampling frequency is increased.

6 Discussion

The difference in the magnitude of the mean errors decreases for all the tested algorithms with increasing sampling frequency. Furthermore, all of the four fitting procedures tested maintain the shape of the generated waveform accurately when $f_s \geq 3$ GHz. Thus all the fitting methods presented can be utilized for parameter retrieval, given that the scattered waveform shape is Gaussian. The CS and PLSQ algorithms can be used more readily to fit non-Gaussian waveforms, compared to the GP and LM algorithms. However, the accuracy of the CS and PLSQ methods in such situations has not been tested, and reliable conclusions cannot be drawn without further investigation. To answer the question on which of the fitting methods is best suited for characterizing the reflectance, the question on which parameter best describes the return waveform is considered first.

The FWHM parameter exhibits the smallest mean errors and standard deviations, considering all sampling frequencies and fitting methods analyzed. The standard deviation of the FWHM results can be large even at sampling frequencies above 3 GHz when using the CS and PLSQ algorithms. The LM algorithm on the other hand, along with the GP method show RSTD values that are at least 45% smaller than those produced by the CS and PLSQ algorithms. Due to this, the FWHM has the potential of being the best parameter to characterize the laser pulse, before the reflectance is calculated. However, in order to relate the FWHM to the measured radiance, it has to be related to the peak amplitude. The formulation of such a relation is only possible if prior information of the echo waveform is available. Furthermore, computing the peak amplitude for the FWHM would have to be performed separately for each return, increasing the demand for processing power and thus making the FWHM a less appealing parameter for real-time reflectance retrieval.

While FWHM retrieval provides the smallest mean errors in terms of pulse characterization, the mean errors provided by the algorithms through area retrieval and peak amplitude retrieval are approximately the same. By comparing tables 2 and 9 for the CS algorithm at $f_s = 5$ GHz the mean error of the peak amplitude is smaller than that of the waveform area. However, at $f_s = 4$ GHz and $f_s = 3$ GHz the waveform area provides a smaller mean error. The lower mean error of the waveform area retrieval compared to the mean error of peak amplitude retrieval, (in this particular case), can be explained through noise induced distortions of the sampled values experienced by the fitted waveform as a function of sampling frequency. Such distortion effects will be more distinct when inspecting individual points such as the peak amplitude, compared to inspecting the waveform area (calculated as an integral of multiple sampled points). This indicates that the accuracy of waveform parameter retrieval is influenced more by the increasing amount of minor alterations at individual sampled points as a consequence of noise and increasing sampling frequency. This same distortion effect can be seen in the results provided in table 8 as decreasing standard deviation and RSTD of the mean errors provided by the PLSQ and CS algorithms as the sampling frequency is decreased from 5 GHz to 3 GHz. The RSTD and standard deviation become larger again at $f_s = 1$ GHz, as the sampling frequency becomes large compared to the pulse-width.

The RSTD of the peak amplitude parameter is generally smaller than the RSTD of the waveform area. Even though the mean errors are in some cases smaller for the waveform area retrieval, the peak amplitude parameter characterizes the measured flux density more accurately. Furthermore, the relatively small mean errors for the waveform area given by the Gaussian integral equation, are most probably due to the fact that the generated pulse is Gaussian.

Based on the analysis above, the FWHM is the best parameter to characterize a return waveform, given that the return waveform maintains the shape of the emitted waveform. However, in terms of reflectance retrieval for TLS instruments the peak amplitude parameter a_m is the most straightforward to retrieve and has mean errors comparable to those produced by FWHM and area retrieval. Furthermore, retrieval of the peak amplitude is applicable in situations in which the waveform shape is distorted or exhibits multiple peaks, without having to separate the different waveform components composing the so-called full-waveform. Next, the most suitable algorithms for retrieving the peak amplitude parameter are discussed, while keeping real-time processing in mind.

The LM algorithm provides the smallest mean errors and standard deviations for all the parameters of interest except the position of the peak amplitude t_{peak} at $f_s = 5$ GHz. However, the magnitude for the mean error in retrieval of the t_{peak} parameter is only 2% of the pulse width, providing sub-centimeter accuracy for range. Thus, the use of the LM algorithm is suggested when performing reflectance measurements during the hardware calibration. These results give further grounds for the use of the LM algorithm for full-waveform retrieval reported for instance by Hofton et al. 2000 and Jutzi et al. 2003.

As stated in section 4.1.3, the GP algorithm is based on the work done by Puttonen et al. 2015. The research group shows that distinct characterization capabilities between man-made compounds and organic materials is possible with the GP method at a sampling frequency of 1 GHz. From this we can draw the conclusion that both the CS and LM algorithms are suitably accurate for similar material characterization tasks as those performed by Puttonen et al. 2015, since they provide a mean error that is at least 13.62 times smaller at $f_s = 5$ GHz and 9.7109 smaller at 3 GHz, than that of the GP method at $f_s = 1$ GHz.

The GP method provides the smallest mean errors at $f_s \leq 2$ GHz for the retrieval of the a_m parameter, with the caveat that the generated pulse is Gaussian. The GP method in itself is not suitable for multi-peak waveform fitting, due to its strict Gaussian formulation, unless it is combined with a non-linear least squares algorithm, or some other iterative algorithm with the capability of refining the accuracy of parameter retrieval. In situations, in which the return waveform can be characterized as Gaussian, the GP method is suitable for waveform reconstruction, due to its simplicity and relative accuracy compared to other fitting methods discussed.

Furthermore, at a sampling frequency of 2 GHz the GP algorithm achieves a mean error that is only 0.03% larger than the mean error of the CS algorithm and 0.11% larger than the mean error of the LM algorithm at $f_s = 5$ GHz. In addition to the minute percentual mean error difference between the algorithms, the GP algorithm is approximately 3.8×10^4s than the LM algorithm and 1300 times faster than the

CS method in performing parameter acquisition. At a 5 GHz sampling frequency the GP method could work in real-time in conjunction with a TLS instrument having a pulse repetition frequency up to 5 MHz, considering the hardware used in the timing computation and taking into account retrieval of the peak amplitude parameter of both the trigger and the echo waveform.

This implies that, out of the fitting methods tested, the GP method would be best suited for both accurate and fast processing. However, the effective use of the GP method requires that the reflected waveform has a Gaussian temporal shape, which approximately occurs if the emitted pulse is Gaussian and the target is extended compared to the beam footprint. This situation can be taken advantage of in locations such as mines where the whole scanning area is covered by a relatively flat surface compared to the footprint area. However, to ensure the suitability of the GP algorithm, several on-site tests would have to be conducted, and the performance of the GP algorithm would have to be compared with results from fitting algorithms that take the possible non-Gaussian nature of the return waveform into account.

The mean errors of the peak amplitude parameter provided by the CS algorithm at $f_s \geq 3$ GHz, when compared to the LM algorithm, differ at most by 0.11%. The smallest difference between these mean errors is exhibited at $f_s = 5$ GHz, and it is 0.085%. This shows that the result provided by the LM algorithm is only marginally more accurate compared to the results provided by the CS algorithm. However, the RSTD calculated by using the LM algorithm is considerably larger than that retrieved through the CS algorithm at a 3 GHz sampling frequency. This leads to the LM algorithm being less accurate for maximum amplitude retrieval at 3 GHz, as it can produce an error as large as 0.7116% of the generated true maximum value, while the CS maximum error is 0.5688%. For sampling frequencies of 4 GHz and 5 GHz the LM algorithm outperforms the CS algorithm, but both algorithms produce a maximum error of less than 0.5%.

The smallest mean error produced by the CS algorithm is observed at a 5 GHz sampling frequency. However, the difference in mean error (for a_m) between $f_s = 5$ GHz and $f_s = 4$ GHz is 0.007%. Thus, only a minute improvement in accuracy is observed when transitioning to a higher sampling frequency. The difference between the processing time required for peak amplitude retrieval is decreased by approximately 20 ms when lowering the sampling frequency from 5 GHz to 4 GHz.

Considering that the mean error and the maximum error differences between the CS and LM algorithms are small, and that the CS algorithm is approximately 30 times faster than the LM algorithm, the use of the CS method at a sampling frequency of 4 GHz is suggested for a means of waveform processing and intensity calibration. The CS method is also more accurate than the GP method at $f_s \geq 4$ GHz. While the GP method is faster, the CS algorithm can still function in real time with a 2 kHz sampling frequency, taking into account the hardware used to make the timing measurements, and that both the trigger and echo pulses are fitted.

Furthermore, the CS algorithm does not require parametrized equations to produce an accurate fit, unlike both the GP and LM algorithms. This allows the use of the CS method in a measurement settings where the return waveform is not Gaussian. To further study the accuracy of the CS algorithm for non-Gaussian waveforms, the

script (Ilinca 2018) given in the bibliography can be slightly modified.

The CS and GP method hold potential for use in real time material characterization due to their small mean errors in retrieval of the waveform peak amplitude. While the LM algorithm proves to be the most accurate fitting method for pulse parameter retrieval, it requires a considerably larger amount of processing power to be able to retrieve the parameters of interest in the same time period as the CS or GP methods.

Based on the analysis conducted, the sampling or digitization frequency can be optimized relative to the laser pulse width. This, however, demands the use of algorithmic waveform retrieval methods. Based on the methods and algorithms tested, a sampling frequency that is four times larger than the reciprocal of the pulse FWHM is enough to retrieve the peak value with an error of $0.2575 \pm 0.191\%$ of the true peak amplitude of the pulse, while being able to provide real-time pulse amplitude and target reflectance data at a pulse repetition frequency of 2 kHz.

7 Conclusion

The applicability and accuracy of four different algorithms for laser pulse waveform reconstruction and parameter retrieval were tested with the intention to find the most suitable trade-off between intensity retrieval accuracy and processing speed for the FGI-Hyperspectral LiDAR. It was shown that all of the waveform fitting methods analyzed produced mean errors and standard deviations low enough in terms of waveform peak amplitude retrieval, and that they can be implemented for material characterization purposes. Furthermore, two algorithms, namely the cubic spline- and Gaussian parametrization-algorithms, were found to provide real-time waveform assessment capabilities due to their relative mathematical simplicity and low processing power requirements.

The Gaussian parametrization method was found to be the most viable option for real-time material characterization due to it being overwhelmingly the fastest algorithm tested, as it could be theoretically utilized with pulse repetition frequencies larger than 5 MHz. However, the applicability of the GP method relies on the return waveform being Gaussian in shape, as the method utilizes the parameters of the Gaussian equation to reconstruct the waveform. However, when scattering of the emitted pulse occurs from an extended target, the echo waveform maintains its original Gaussian shape and the GP method can be considered to be suitable. Since the emitted pulses of hyperspectral lasers have been shown to be Gaussian, the GP method is suitable in real-world environments such as mines, where the target is relatively flat compared to the size of the laser footprint. However, to absolutely determine the suitability of the GP method for such situations, on-site test measurements and comparisons with well established waveform retrieval algorithms such as the Levenberg-Marquardt algorithm would have to be conducted.

The cubic spline algorithm was found to provide the best trade-off between intensity retrieval accuracy and the time required by the algorithm for full-waveform parameter retrieval. The advantage the CS method has over the GP method is that it can be utilized to characterize waveform shapes differing from the tested Gaussian waveform. Hence the CS method allows research to be conducted in a larger variety of environments, as for example forests and urban areas, where the scattered waveform may exhibit multiple peaks or deformations compared to the trigger pulse. Furthermore, the CS algorithm was shown to produce a smaller mean error and standard deviation in the retrieval of waveform intensity compared to the GP method. A sampling frequency of 4 GHz is suggested to be used with CS algorithm, since the peak amplitude mean error is only 0.0069% larger than that attained with a 5GHz sampling frequency. Furthermore the CS algorithm provides a RSTD value that is 1.06% smaller at sampling frequency of 4 GHz than at 5 GHz. The processing time required for performing the CS fit is also improved by $2\mu s$ compared to the 5 GHz sampling frequency. Taking all of the above into consideration, the CS algorithm is suggested to be used for the FGI-HSL instrument.

Further research is needed for the absolute validation for the use of both the CS and GP waveform retrieval methods. In terms of further simulations, the next natural step would be to modify the script used for the analysis to account for

multi-peak waveforms. Furthermore, field measurements for testing the validity of the CS algorithm along with the material characterization capabilities of the FGI-HSL instrument have been scheduled for the near future. A journal article is also being written about the effect of sampling frequency on terrestrial LiDAR waveform processing by the author in collaboration with other FGI researchers.

In conclusion, the results show that complicated, time-consuming, and iterative fitting approaches (such as the LM algorithm) often used for material characterization provide only marginally more accurate results compared to mathematically straightforward approaches such as the CS and GP methods. This implies that the speed at which material characterization can be performed can be vastly increased with the combined use of widely available high speed A/D converters and the GP or CS algorithm. Furthermore, data storage management of collected pulse data can be improved as the digitization speed can be determined based on the FWHM of the trigger pulse. An increase in processing speed along with maintained accuracy could potentially be efficiently used in the mining industry for finding ore in tunnels or separating ore from gangue in general. The improved speed in material characterization capabilities could also be utilized by autonomous mining vehicles or drones providing localized crop health information to farmers in real time reducing working hours and expenses. Furthermore, the technology holds potential for retrieving data for remote hazardous situation management, such as analyzing the extent of chemical spills.

8 Bibliography

Beinert, A. et al. (2006). "Application of Terrestrial Laser Scanners for the Determinatio of Forest Inventory Parameters". In: *ISPRS Commission V Symposium on Image Engineering and Vision Metrology* 36, part 5.

Chauve, A. et al. (2009). "Advanced full-waveform LiDAR data echo detection: Assessing quality of derived terrain and tree height models in alpine coniferous forest". In: *International Journal of Remote Sensing* 30(19), pp. 5211–5228. DOI: 10.1080/01431160903023009.

Dassot, M. et al. (2011). "The use of Terrestrial LiDAR thechnology in forest science: application fields, benefits and challenges". In: *Annals of Forest Science* 68.5, pp. 959–974. DOI: 10.1007/s13595-011-0102-2.

Douglas, E.S et al. (2012). "Dwell: A dual Wavelength Echidna LiDAR for Ground-based Forest Scanning". In: *Proceedings of International Geoscience and Remote-sensing Symposium 2012* Munich, Germany, pp. 1–4.

Du, L. et al. (2016). "Estimation of Rice Leaf Nitrogen Contents based on Hyperspectral LIDAR". In: *Internation Journal of Applied Earth Observation and Geoinformation* 44, pp. 136–143. DOI: 10.1016/j.jag.2015.08.008.

Dudley, J.M et al. (2006). "Supercontinuum Generation in Photonic Crystal Fiber". In: *Reviews of Modern Physics* 78, pp. 1135–1184. DOI: :10.1103/RevModPhys.78.1135.

Eitel, J.U.H et al. (2011). "Early Season Remote Sensing of Wheat Nitrogen Status Using a Green Scanning Laser". In: *Agricultural and Forest Meteorology* 151, pp. 1338–1345. DOI: 10.1016/j.agrformet.2011.05.015.

Gao, Y. et al. (2015). "INS/GPS/LiDAR Integrated Navigation System for Urban and Indoor Environments Using Hybrid Scan Matching Algorithm". In: *Sensors* 15, pp. 23286–23302. DOI: 10.3390/s150923286.

Gaulton, R. et al. (2010). "The Salford Adavance Laser Canopy Analyzer (SALCA): A Multispectal Waveform LiDAR for Improved Vegetation Characterization". In: *Proceedings of the Remote Sensinding and Photogrammetry Society Conference* Burlington House, London, pp. 1–4.

Gaulton, R. et al. (2013). "The potential of dual wavelength laser scanning for estimating vegetation moisture content". In: *Remote Sensing Environment* 132, pp. 32–39. DOI: https://doi.org/10.1016/j.rse.2013.01.001.

Hakala, T. et al. (2012). "Full waveform hyperspectral LiDAR for terrestrial laser scanning". In: *OSA, Optics Express* 20(7), pp. 7119–7127. DOI: https://doi.org/10.1364/OE.20.007119.

Hofton, M.A. et al. (2000). "Decomposition of Laser Altimeter Waveforms". In: *IEEE Transactions on Geoscience and Remote Sensing* 38(4), pp. 1989–1996. DOI: 10.1109/36.851780.

Ilinca, J.N (2018). *Simulation Script for Assessing the Accuracy of Gaussian Waveform parameter retrieval*. URL: https://www.dropbox.com/sh/wvgpatnn4asvbvu/AAAkQxCx4-DoUNyKgIIWCE_za?dl=0 (visited on 05/05/2018).

Jackson, J.D (1998). *Classical Electrodynamics*. 3rd ed., p. 457. ISBN: 047130932X.

Jutzi, B. et al. (2003). "Estimation and measurement of backscattered signals from pulsed laser radar". In: *Image and Signal Processing for Remote Sensing VIII, Proceedings of SPIE* 4885, pp. 256–267. DOI: 10.1117/12.463086.

Jutzi, B. et al. (2006). "Range Determination with Waveform Recording Laser System Using a Wiener Filter". In: *ISPRS Journal od Phortogrammetry and Remote Sensing* 61, pp. 95–107. DOI: 10.1016/j.isprsjprs.2006.09.001.

Kaasalainen, S. et al. (2005). "Study of Surface Brightness From Backscattered Laser Intensity: Calibration of Laser Data". In: *IEEE Geoscience and Remote Sensing Letters* 2(3), pp. 255–259. DOI: 10.1109/LGRS.2005.850534.

Kaasalainen, S. et al. (2008). "Brightness Measurements and Calibration With Airborne and Terrestrial Laser Scanners". In: *IEEE Transactions on Geoscience and Remote Sensing* 46(2), pp. 528–534. DOI: 10.1111/j.1467-8667.2006.00466.x.

Kaasalainen, S. et al. (2016). "Incidence Angle Dependence of Leaf Vegetation Indices form Hyperspectral LiDAR Measurements". In: *Journal of photogrammetry, remote sensing and geoinformation processing* 2, pp. 75–84. DOI: 10.1127/pfg/2016/0287.

Krooks, A. et al. (2013). "Correction of Intensity Incidence Angle Effect in Terrestrial Laser Scanning". In: *ISPRS Annals of the Photogrammetry, Remote Sensing ans Spatial Information Sciences* II-5/W2, pp. 145–150. DOI: 10.5194/isprsannals-II-5-W2-145-2013.

Leader, J.C (1978). "Analysis and prediction of laser scattering from rough-surface materials". In: *Journal of the Optical Society of America* 69 (4), pp. 610–628.

Levenberg, K. (1944). "A Method for Solution of Certain Non-Linear Problems in Least Squares". In: *Quarterly of Applied Mathematics* 2, pp. 164–168.

Li, Z. et al. (2001). "A Multispectral Canopy LiDAR Demonstrator Project". In: *IEEE Geoscience and Remote Sensing aussiLetters* 8(5), pp. 839–843. DOI: 10.1109/LGRS.2011.2113312.

Lim, K. et al. (2003). "LIDAR remote sensing of forest structure". In: *Progress in Physical Geography* 27, pp. 88–106. DOI: 10.1191/0309133303pp360ra.

Mallet, C. et al. (2009). "Full-waveforom topographic LiDAR: State-of-the-art". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 64, pp. 1–16. DOI: 10.1016/j.isprsjprs.2008.09.007.

Marquardt, D. W (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". In: *J. Soc. Indust. Appl. Math* 11 (2), pp. 431–441.

Nicodemus, F.E (1965). "Directional Reflectance and Emissivity of an Opaque Surface". In: *Applied Optics* 4.7, pp. 767–773.

Nicodemus, F.E et al. (1977). "Geometrical Considerations and Nomenclature for Reflectance". In: *National Bureau of Standards*, pp. 3–9.

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes*. 2nd ed., pp. 147–148. ISBN: 0-07-048477-5.

Park, H.S et al. (2007). "A new approach for health monitoring of structures: Terrestrial Laser Scanning". In: *Computer-Aided Civil and Infrastructure Engineering* 22, pp. 19–30. DOI: 10.1111/j.1467-8667.2006.00466.x.

Parrish, C.E et al. (2011). "Empirical Comparison of Full-Waveform LiDAR Algorithms: Range Extraction and Discrimination Performance". In: *American Society fo Photogrammetry and Remote Sensing* 8, pp. 825–838. DOI: `10.14358/PERS.77.8.825`.

Persson, A. et al. (2005). "Visualization and Analysis of Full-Waveform Airborne LiDAR Scanner Data". In: *International Archives of Photogrammetry and Remote Sensing* 59 (Part3/W19), pp. 103–108.

Pfeifer, N. et al. (2007). "Investigating Terrestrial Laser Scanning Intensity Data: Quality and Functional Relations". In: *8th Conf. on O3D, Terrestrial Laser Scanning I* Zurich, Switzerland, pp. 328–337.

Puttonen, E. et al. (2015). "Artificial Target Detection with a Hyperspectral LiDAR over 26-h Measurement". In: *SPIE, Optical Engineering* 54(1), pp. 1–15. DOI: `10.1117/1.OE.54.1.013105`.

Reitberger, J. et al. (2006). "Full-Waveform Analysis of Small Footprint airborne Laser Scanning Data in the Bavarian Forest National Park for Tree Species Classification". In: *Workshop on 3D Remote Sensing in Forestry*, pp. 218–227.

Schaepman-Strub, G. et al. (2006). "Reflectance quantities in optical remote sensing - definitions and case studies". In: *Remote Sensing of Environment* 103, pp. 27–42. DOI: `10.1016/j.rse.2006.03.002`.

Seber, George A. F. et al. (2003). *Linear Regression Analysis*. 2nd ed., pp. 1–16. ISBN: 978-0-471-41540-4.

Shen, X. et al. (2017). "Decomposition of LiDAR waveforms by B-Spline based modeling". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 128, pp. 182–191. DOI: `10.1016/j.isprsjprs.2017.03.006`.

Wagner, W. (2010). "Radiometric calibration of small footprint full waveform airborne laser scanner measurements: Basic Physical concepts". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.6, pp. 505–513. DOI: `10.1016/j.isprsjprs.2010.06.007`.

Wagner, W. et al. (2006). "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitizing airborne laser scanner". In: *ISPRS Journal of Photogrammetry & Remote Sensing* 60, pp. 100–12. DOI: `10.1016/j.isprsjprs.2005.12.001`.

Wagner, W. et al. (2008). "Radiometric Calibration of Full-Waveform Small-Footprint Airboren Laser Scanners". In: *The International Archives of the Photogrammetry, Remote Sensing ans Spatial Information Sciences* 37, pp. 163–168.

Wallace, A.M et al. (2014). "Design and Evaluation of Multispectral LiDAR for Recovery of Arboreal Parameters". In: *IEEE Transactions on Geoscienca and Remote Sensing* 52.8, pp. 4942–4954. DOI: `10.1109/TGRS.2013.2285942`.

Wei, G. et al. (2012). "Multi-Wavelength Canopy LiDAR for Remote sensing of vegetation: Design and System Performance". In: *IISPRS Journal of Photogrammetry and Remote Sensing* 69, pp. 1–9. DOI: `10.1016/j.isprsjprs.2012.02.001`.

Woodhouse, I.H et al. (2001). "A Multispectral Canopy LiDAR Demonstrator Project". In: *IEEE Geoscience and Remote Sensing Letters* 8(5), pp. 839–843. DOI: `10.1109/LGRS.2011.2113312`.

Young, P. (2014). "Least Squares Fitting". In: *Physics 115/242, University of California Santa Cruz* 115, pp. 1–11.

Zhu, X. et al. (2017). "Canopy Leaf Water Content Estimated using Terrestrial LiDAR". In: *Agricultural and Forest Meteorology* 232, pp. 152–163. DOI: http://dx.doi.org/10.1016/j.agrformet.2016.08.016.