DEPARTAMENTO DE GEOLOGÍA, GEOGRAFÍA Y MEDIO AMBIENTE

ESTIMATION OF GAP FRACTION AND CLUMPING INDEX WITH TERRESTRIAL AND AIRBORNE LASER SCANNER DATA

Tesis Doctoral presentada por:

JOHN GAJARDO VALENZUELA

Bajo la dirección de:

Dr. David Riaño Arribas

Dr. Mariano García Alonso

Dr. Javier Salas Rey

Madrid, España

Marzo de 2014

"hay que olvidar a los viejos,
renegar de los enfermos,
odiar al diferente,
expulsar al extranjero,
barrer a los mendigos,
eliminar a los parados,
aborrecer al débil...obedecer al dinero".

Evaristo Páramos

John Gajardo V. is grateful to BECAS CHILE from the Chilean Government for funding this work.

Agradecimientos

En primer lugar quiero dar los agradecimientos a mis directores. Su apoyo y aporte en los ámbitos profesionales y personales ha sido esencial para la finalización de esta empresa. Gracias infinitas por vuestra paciencia y por mantener viva en mí la ilusión de la ciencia. David, no encuentro palabras conocidas para agradecer por todos tus esfuerzos y momentos dedicados. Gracias a Pilar Martin y a su equipo de investigación en el CSIC por acogerme y darme la oportunidad de aprender y participar de sus investigaciones. Son un grupo humano y profesional increíble. A la gente del CSTARS, especialmente a Susan Ustin por brindarme la ocasión de ser partícipe de sus actividades y ampararme en mi estancia. A la Facultad de Ciencias Forestales de la Universidad de Talca por la oportunidad de perfeccionamiento otorgada.

En segundo lugar a todas las personas que he conocido y han compartido conmigo en estos años de aprendizaje, especialmente a los ciudadanos permanentes e itinerantes del sótano y torreón de la UAH.

Agradezco también a mis compañeros de pisos por su compañía, comprensión y apoyo, RL10 y MOFO, ustedes siempre me hicieron sentir en familia. Gracias también a 'my editor for life' Marilú Carter, tu ayuda y compañía al otro lado del río grande fue muy importante.

Al final de este orden, pero siempre en primer lugar de mi corazón, gracias a mi familia. Madre, Hermana y especialmente a mi Padre por haber tenido confianza en mí. Gracias por todo lo que ustedes me han brindado.

CONTENTS

RESUMEN

El dosel forestal es una zona de intercambio de flujos y energía entre la superficie de la tierra y la atmósfera. Su estructura está representada por la organización espacial de todos los elementos vegetales que se encuentran sobre la superficie. La estructura del dosel condiciona una serie de variables microclimáticas en el interior de este espacio, las que influyen en la disponibilidad de los recursos y el comportamiento de las especies que cohabitan en él. Existe una serie de variables que permiten describir la estructura del dosel. Entre las más importantes se encuentran el índice de área foliar, cuyo cálculo y corrección depende de otros parámetros como la fracción de huecos (*gap fraction*, GF) y el índice de agrupamiento foliar (*clumping index*, CI).

En este documento se estudian y desarrollan métodos para la estimación de GF y CI a partir de escáneres láser terrestres y aerotransportados (*Terrestrial* (TLS) *and Airborne* (ALS) *Laser Scanners*). Para lograr esto, se llevaron a cabo mediciones con TLS en Las Majadas del Tiétar (Cáceres, España) en el año 2009 y con ALS en Jasper Ridge (California, EE.UU.) en el 2007. En el caso de la estimación de GF a partir TLS, se desarrolló un nuevo método que calculaba la proporción entre píxeles vacíos y su totalidad a partir de imágenes angulares, una vez que se conocía su resolución. La validación del método fue realizada mediante simulaciones de datos con diversas resoluciones angulares y patrones de huecos en el dosel. El método se comparó también con los resultados de GF a partir de fotografías hemisféricas (*hemispherical photography*, HP), una vez que los datos TLS se reproyectaron para simular HP (TLS-SHP). La estimación del CI se llevó a cabo aplicando la teoría de la distribución del tamaño de los huecos de Chen y Cihlar (1995) sobre las TLS-SHP, que se contrastó con los valores de CI de las HP.

En la zona de Jasper Ridge las estimaciones de GF se realizaron empleando métricas basadas en la ley de transmisividad de Beer-Lambert que miden el porcentaje de retornos láser que llegan al suelo, considerando parcelas circulares de datos ALS con diferentes tamaños de radio, para compararlas con la GF estimado de las HP. Del mismo modo, se probó también con la relación entre las intensidades de los retornos

del suelo y las de todos ellos al interior de las parcelas. El CI se estimó a partir de métricas ALS derivadas de la altura de la vegetación y se relacionaron con el CI de las HP. Además, se adaptó con el mismo propósito el índice de segregación espacial de Pielou (1962), que se aplicó sobre imágenes de GF generadas para parcelas de datos ALS con distintos tamaños de radio y que fueron comparadas con el CI generado desde las HP.

Para los experimentos llevados a cabo con los datos TLS, la GF fue sobreestimada en un 14% respecto a las HP, siendo las correlaciones estadísticamente significativas. El algoritmo desarrollado es operativo siempre y cuando el ruido en los datos angulares sea inferior al 6% de la resolución angular. Por encima de este umbral el método presentó un alto error, especialmente en los datos simulados con una estructura de huecos agrupados (*cluster*). El CI se subestimó en 27% respecto a los valores obtenidos por las HP. Los principales problemas vienen dados por la diferencia en la distribución del tamaño de los huecos registrados por las HP y las TLS-SHP.

Por otra parte, la GF derivada de los datos ALS subestimó en un 3% y sobrestimó en un 43% comparado con las HP, para las parcelas de bosque y matorral, respectivamente. La GF obtenida presentó una clara dependencia del radio de los datos ALS considerados, que varió según el tipo de vegetación. Respecto a las estimaciones del CI, las métricas ALS de las alturas de la vegetación no mostraron buenos resultados. Esta circunstancia es contraria a estudios previos, lo que parece indicar que estas relaciones empíricas sólo funcionarían para el tipo de vegetación y sitio para el que fueron desarrolladas. Sin embargo, la modificación del algoritmo de Pielou subestimó el CI en sólo 6% y 4% para las parcelas de bosques y matorrales, respectivamente. Las posibles causas de estas diferencias radican en las distintas perspectivas y resolución espacial que poseen los datos ALS y HP.

ABSTRACT

Fluxes and energy exchange between the land surface and the atmosphere occur at the forest canopy. Its structure is the detailed spatial organization of all aboveground vegetal elements. The forest canopy structure conditions several microclimatic variables, which influence in the resources availability and the animal behavior that co-habitat in it. There are several variables that allow describing the canopy structure. Among these, one of the most relevant is the leaf area index, whose calculation and correction depends on other parameters like the gap fraction (GF) and the foliar clumping index (CI).

This document studies and develops methods to estimate GF and CI from Terrestrial (TLS) and Airborne (ALS) Laser Scanners. To achieve this, measurements were acquired with a TLS in Las Majadas del Tiétar (Cáceres, Spain) in the year 2009 and with an ALS in Jasper Ridge (California, USA) in 2007. In the case of GF estimated from TLS, a new proposed method computed the ratio between the empty and all pixels from angular images, once their angular resolution was known. The method was validated from data simulations with several angular resolutions and canopy gap patterns. They were also compared to the GF results from hemispherical photography (HP), once the TLS data were reprojected to simulate HP (TLS-SHP). The CI estimation was carried out applying the gap size distribution theory proposed by Chen and Cihlar (1995) over the TLS-SHP that was compared to the CI values from HP.

Based on the transmissivity Beer-Lambert law, the GF estimations at the Jasper Ridge site were done from metrics that measure the percentage of laser returns that reach the ground, considering circular plots of ALS data at different radii size, to compare them with GF estimated with HP. The same way, the relationship between the intensity of the returns from the ground and the ones from all of them was also tested. CI estimation was carried out using ALS metrics derived from vegetation height and they were related to CI from HP. In addition, the spatial segregation index proposed by Pielou (1962) was adapted to be applied over GF images generated from ALS data of different radii size and compared to the CI computed from HP.

For the experiments carried out with TLS data, the GF was overestimated in the order of 14% respect to HP, being the correlations statistically significant. The

algorithm developed is operative as long as the noise in the angular data was below 6% of the angular resolution. Above this threshold the method showed a high error, especially for the simulated data with a cluster gap pattern. The CI was underestimated in 27% respect the values obtained from HP. The main problems came from the difference in gap size distributions registered by the HP and the TLS-SHP.

On the other hand, the GF derived from ALS data underestimated in 3% and overestimated in 43% respect to HP, for the forest and shrub plots, respectively. The GF obtained showed a clear spatial dependence on the radius of the ALS data considered, which varied according to the vegetation type. In relation to the CI estimations, the vegetation height ALS metrics did not perform well. This circumstance is contrary to previous studies and seems to indicate that these empirical relations only would work for the vegetation type and site for which they were developed. However, the modification to Pieolou's algorithm underestimated the CI in only 6% and 4% for the forest and shrub plots, respectively. The possible causes for these dissimilarities rely on the different perspective and spatial resolution that the ALS and HP have.

STRUCTURE OF THE PHD DISSERTATION

The document is divided in four chapters. The first is dedicated to a literature review of the canopy structural indicators, methods and instruments to characterize the canopy and the role of the Terrestrial (TLS) and Airborne (ALS) Laser Scanners to specifically estimate two important structural variables: the canopy gap fraction (GF) and clumping index (CI), also it includes the hypothesis and objectives of this work. The second is the methods section which is organized according to the experiments carried out in two study sites: The first study site was Las Majadas del Tiétar (Spain) where GF was computed with a new method which uses angular images derived from a TLS. Another experiment was based on the estimation of canopy CI, applying the gap size distribution theory after transforming the TLS data in hemispherical images. The second study site was located in Jasper Ridge (USA) and there it was carried out the estimation of canopy GF using height metrics and penetration rates metrics based on frequency and intensity data collecting from ALS. An additional experiment was also conducted in this site to estimate the canopy CI by means of empirical ALS height metrics and the adaptation of Pielou's algorithm to work over ALS point cloud and ground returns images. The third chapter presents the results and discussion for each of the experiments described above. The fourth section provides the conclusions derived from the work and exposes the main factors which influenced the results obtained.

1 INTRODUCTION

1.1 Characterization of the vegetation canopy structure

Branches, twigs, leaves and crown shape arrange to constitute the vegetation canopy (Parker 1995). A detailed organization and arrangement in space and time dimensions at various scales of all above ground vegetation elements form the canopy structure. This description includes information and attributes about the amount, location, connectivity and type of elements (Lefsky et al. 1999). The canopy forms a particular and exclusive sub-system, key to understand the interactions between the vegetation and its surrounding environment. The canopy structure is one of the most influential elements in the microclimate and the habitat that it supports. For example, it modulates the wind speed (Lefsky et al. 2002; Parker 1995) or the amount of irradiance or visible light (Danson et al. 2007; Panferov et al. 2001; Parker 1995) within the canopy. The vertical disposition of the elements inside the canopy controls the reduction and the distribution of the photosynthetic active radiation (Dean et al. 2009), that impacts the growth and mortality rates of seedlings under the canopy (Lhotka and Loewenstein 2008). Precipitation is also affected since the canopy intercepts and redirects the water. According to Parker (1995), between 10 and 30% of the precipitation can be intercepted and evaporated by the canopy.

The presence of multiple vertical vegetation layers also influences the concentration and gradients of particles and gases that move within the canopy (Parker et al. 2004). In the CO_2 case, the upper and middle layers are a sink, especially during the day. Conversely, the layer near the ground surface becomes a source of CO_2 especially during the night as consequence of the roots respiration and decomposition of the ground litter. The detailed knowledge of the several canopy layers, as well as the variability rates of CO_2 contribute to a more accurate simulation of the gases exchange to the atmosphere, as part of the photosynthesis and respiration of the vegetation (Parker 1995; Reich et al. 1990). Moreover, changes in the structure affect the habitat and biodiversity of the ecosystem that the canopy supports. The arrangement of the canopy layers and the distribution and complexity of branches and

foliage, result in changes in space, light, moisture and food availability that limit the resources that species need to survive and develop (Nelson et al. 2003; Pringle et al. 2003; Skirvin 2007). In the context of fire management, canopy structural parameters like fuel loading, vertical profile, canopy bulk density and height of the crown base contribute to model fire behavior (García et al. 2012; Hall et al. 2005; Popescu and Zhao 2008; Riaño et al. 2003). Finally, the crown dimension, and its relation with height and diameter at breast height, is an indicator of competition and forest health in order to design of silvicultural treatments and stand management to increase wood production (Dean et al. 2009; Popescu and Zhao 2008).

The canopy structure can be characterized at different scales, from the identification of individual plants, to the stand and ecosystem level (Van der Zande et al. 2006). The task of producing a complete description of the size, shape and orientations of all the vegetal elements (trunks, branches, leaves, flowers, fruits), is almost impossible considering the canopy is very dynamic (Welles and Cohen 1996). To simply this complexity, a variable such as the maximum tree height, tree density, forest biomass, canopy cover or leaf area index (LAI) can summarize the structure in most circumstances (Parker 1995). The LAI is defined as the half of total leaf area per unit of ground horizontal area (Chen and Black 1992). It is one of the principal input variables to model photosynthesis, evapotranspiration and carbon and water fluxes interchange between the canopy and the atmosphere (Weiss et al. 2004).

Two variables are especially relevant to describe the canopy structure, the canopy gap fraction (GF) and the canopy clumping index (CI), which are the focus of this study. More specifically, the GF is the fraction of sky that is visible under the canopy in any particular direction. The GF ranges from zero to one for a completely obstructed and a completely open sky, respectively (Welles and Cohen 1996). Fraction cover, or sometimes also called canopy cover, is the complementary with a value of one for an obstructed sky and zero for an open one. The GF is a valuable measure for vegetation analysis to monitor phenological fluctuations, evolution and recovery after a disturbance, such as water stress, flood, air and soil contamination or a plague (Walter 2009). Besides animals respond the light availability within the canopy (Endler 1993; Pringle et al. 2003). GF is also an important structural parameter, because its

inversion that theoretically follows a Poisson gap frequency distribution estimates LAI (Campbell and Norman 1989; Miller 1967; Nilson 1971). This LAI estimation, called 'effective LAI', assumes that the canopy elements are randomly distributed. This approach introduces an error because the canopy elements are rarely randomly distributed in nature. To account for this, the effective LAI derived from GF inversion (Kucharik et al. 1997; Myneni et al. 2002), combined with a factor that measures the degree of canopy randomness or clumping index (CI), produces the 'true LAI' (Chen and Cihlar 1995; Gonsamo and Pellikka 2009; Walter 2009). A CI value of zero indicates a very high clumped canopy formed by non-random elements, whereas a value of one means a very low clumped canopy of random spatial elements. Chen et al. (2003) demonstrated the CI relevance in the segmentation of solar radiation distribution between sunlit and shaded leaves to upscale the modelling photosynthesis. In addition, Nilson (1971) indicated the significance of CI to compute the interception and penetration rate of radiation through the canopy.

1.2 Measurement of vegetation canopy structural variables

Canopy complexity, accessibility and time to take the measurements constrain the acquisition of representative direct data of the canopy structure (Fournier et al. 1997; Van der Zande et al. 2006; Welles and Cohen 1996). These are some of the main reasons to limit the measurements to one variable and to find an average over similar spatial features (Parker 1995). In this context, choosing the proper variable that describes suitably the phenomenon to study might be challenging (Campbell and Norman 1989). One of the first characterizations of the canopy structure in the field was just some sketches of the representative trees in the forest (Parker 1995). Furthermore, one of the simplest approaches to describe the structure is to measure the canopy height and cover (Lefsky et al. 1999). More advanced methods have enabled to describe the vertical distribution of the canopy, such as the two dimensional point quadrats by Wilson (1959) first, and later the optical point quadrats by MacArthur and Horn (1969) that use an upward view telephoto.

To get enough, detailed and accurate information, sometimes it is necessary to measure additional characteristics manually. Fournier et al. (1997) carried out a multi-

scale (individual trees, stand and regional) structural characterization to produce a complete digital 3D reconstruction of the canopy elements. They measured manually the tree position, dominant category, tree height, crown extension, diameter at breast height, fractional cover, and a complete characterization of the branches structure that included: amount of leaves, coordinates and positions of the segments, dimensions and spatial organization of the leaves. Another variable critical to summarize the canopy structure such as the LAI is measured directly in two phases: 1) leaf collection by means of destructive or non-destructive techniques (Weiss et al. 2004); 2) the leaf area computation by scanning the leaves (Jonckheere et al. 2004), or based on the dry weight correlation to leaf area (Le Roux et al. 1999).

To deal with some of the difficulties of these direct estimations of the structural parameters, different methods relying on the relation between the canopy structure and light interception were developed (Weiss et al. 2004). The light that passes through the canopy can be estimated by a 3D model that describes the canopy structure (Jonckheere et al. 2004). The light transmission through the canopy can be collected from the GF measurements with linear and hemispherical sensors (Welles and Cohen 1996).

The linear sensors measure in only one direction at a time during direct sun light and require the user to move along a transect to characterize the structure (Leblanc et al. 2002a). The leaves absorb almost 90% of the light in the photosynthetic active radiation (PAR) region (400 to 700 nanometers). The possibility of linear sensors to measure PAR makes these instruments a good indicator of canopy GF (Chen 1996). The GF is obtained by dividing the PAR below the canopy over the one above, at a place with no vegetation. However, a threshold can be set so that the sensor assumes an open sky to avoid the need of taking measurements above the canopy (Jonckheere et al. 2004).

There are several manufactures that build PAR sensors. The AccuPar L-80 from Decagon Devices (Box 2365, Pullman, WA 99163 – USA) is a portable PAR sensor with 80 sensors distributed along a stick every 1 cm and a control unit. The LI-191 is manufactured by LI-COR (Box 4425, Lincoln, NE 68504 – USA) and is integrated by one sensor of 1 m sensitivity length coupled into a probe of 116 cm long and connected to

a data logger with a detachable cable. The Sun Scan SS1 introduced by Delta T Devices (130 Low Road, Burwell, Cambridge - UK) is a ceptometer of 100 cm long by 13 mm wide with an array of 64 PAR sensors embedded into a probe and connected to handheld PDA. This instrument has the possibility to add a reference sensor to measure the light above the canopy and transmit the data via radio frequency. The Tracing Radiation and Architecture of Canopies (TRAC) manufactured by 3rd Wave Engineering (14 Aleutian Rd, Nepean, Ontario – CA) is a system with two PAR sensors looking up and one down, amplifiers, clock and an a I/O circuitry. The PAR sensors are coupled to a stick with bull eye spirit level and connected to a data control and logger unit. In between these linear sensors is also the DEMON instrument built by CSIRO (Locked Bag 10, Clayton South VIC 3169 – AU), which uses a detector facing up to measure both, the transmittance above and below the canopy. The GF is computed as the linear average of transmittance captured along a transect. One of the main problems with this instrument is the need to capture data at least three times per day, in order to get enough information for different sun inclinations (Jonckheere et al. 2004).

Unlike the linear sensors, the hemispherical ones are able to collect measurements from multiple angles at once and for some of them there is no need to use transects (Welles and Cohen 1996). The Digital Plant Canopy Imager (CI-110) manufactured by CID Bio-Science (1554 NE 3rd Ave, Camas, WA 98607 – USA) is a system that includes 24 PAR sensors and a hemispherical fisheye camera (180°) assembled in an arm of 40 cm length. The user can analyze and compute GF at the moment of measurement or store the hemispherical photographs for post-processing in a PC. The LAI 2200C plant canopy analyzer manufactured by LI-COR (Box 4425, Lincoln, NE 68504 – USA) is an instrument with a fisheye light sensor that collects measurements in 5 zenithal bands centered at 7, 23, 38, 53 and 68°. A reference light must be measured above the canopy followed by one or more measurements below the canopy. The ratio between the two measurements gives the GF for each sky sector. The estimations are provided in-situ or stored to be post-processed in a console. The system allows simultaneous above and below canopy measurements using two LAI 2200C instruments, which helps when the light conditions change rapidly

or the canopy is very tall (Welles and Cohen 1996). With this configuration, one of them is fix to collect data at regular intervals in a place with no vegetation and the other moves around to collect data below the canopy at multiple points. In the console, the datasets can be matched by time in a single file that contains readings from both sensors.

The hemispherical photographs (HP) are images captured from cameras with hemispherical (fisheye) lens placed under the canopy and oriented towards the zenith (Jonckheere et al. 2005). For best measurements, the HP needs to be captured when there is a high contrast between the sky and the canopy. This condition is usually given when the sky is cloudy or, during the first or last hours of the day (Welles and Cohen 1996). The HP is a valuable source of information to study the canopy structure, because it provides permanent data about position, size and distribution of the canopy gaps (Gonsamo and Pellikka 2009). The HP provides an extreme field of view, generally 180°, producing a projection of the hemisphere in a plane (Herbert 1986). The projection varies according the lens deformation. The most common geometrical projections used by commercial fisheye lens manufacturers are: polar, orthographic, Lambert equal area and stereographic (Gonsamo et al. 2010). A perfect 180° polar projection results in a circular image. Such image fits completely in the square frame, which leaves blank or black margins. The HP represents the sky view in all azimuthal (θ) and zenithal (φ) directions (azimuth, zenith), with the φ in the center and the horizon at the image borders (Jonckheere et al. 2004). According to fisheye lens deformation, the image is divided into sky segments which form what is called the 'skymap'. The skymap segments are delimited by the intersection of φ rings and θ sectors (Gonsamo et al. 2010).

The GF in the HP can be computed as the ratio of number of pixels in a particular segment that represent the sky to the total number of pixels in this segment (Walter 2009). There are a series of commercial software and freeware to extract the GF from HP images: HemiView (Delta T Devices. 130 Low Road, Burwell, Cambridge - UK), WinSCANOPY (Regent Instruments Quebec – CA), Gap light analyzer (Frazer et al. 1999), CIMES (Walter 2009), CAN-EYE (Weiss and Baret 2010), WINPHOT (Steege 1996) SOLARCALC (Mailly et al. 2013) and LIA32 (Yamamoto 2004).

In order to compute this ratio it is necessary to classify the HP into two classes: canopy and sky, which produces a binary image. This is one of the main problems in the HP processing, because GF computations mainly depend on a manual image threshold to discriminate between the two classes (Jonckheere et al. 2005). The simplest way to segment the HP is that the user defines threshold visually, below or equal which would represent one class, and above which the other class (Frazer et al. 2001). This subjective process introduces a source of variability since the user threshold values cannot be consistently reproduced on other HP (Hale and Edwards 2002). Automatic threshold algorithms can help to partly overcome this problem (Jonckheere et al. 2005), but the changes in illumination conditions on the HP can cause this process to fail. Several studies extracted GF from HP applying either interactive (Frazer et al. 2001; Gonsamo and Pellikka 2009; Gonsamo et al. 2010; Hale and Edwards 2002) or automatic (Gonsamo and Pellikka 2008; Inoue et al. 2004; Jonckheere et al. 2005; Leblanc et al. 2005; Walter et al. 2003) thresholds.

Current HP images that have a high spatial resolution (e.g. 12 megapixels) can be used to derive the foliage CI (Walter 2009). Chen and Cihlar (1995) CI (CCI) is generated from the GF and gap size distributions. According to their theory, the clumped canopy usually shows large gaps combined with small gaps. The probability of appearance of these large gaps within a random canopy can be derived from the gap size distribution (Leblanc et al. 2002a). The algorithm works removing iteratively the large gaps that are not theoretically possible in a random canopy for a given LAI and foliage element width. After the large gap removal process, a new gap size cumulative distribution is computed. The removal process continues until the differences between consecutive computed distributions are insignificant. Finally, the CI is derived comparing the reduced GF closest to a random distribution with the actual measured GF distribution (Walter 2009). This method was originally created for the TRAC instrument, but it was later modified for HP images (Leblanc et al. 2002b).

Other CI was proposed by Lang and Xiang (1986) that is based on the 'logarithmic gap averaging' method (CLX). They established that LAI in a discontinuous and clumped canopy can be estimated by averaging the canopy GF in a logarithmic mode. The average GF is calculated for a sector with a length that is equivalent to ten

times the average size of a determined foliage element (Leblanc et al. 2005). The method assumes that canopy elements are randomly distributed and that each of the segments contains gaps (Gonsamo et al. 2010). The rationale behind this method is that the LAI has logarithmic relation with the GF throughout the Beer Lambert's law, therefore the LAI average should follow the logarithm average of the GF (Lang and Yueqin 1986). The method shows some drawbacks when there are segments with no gaps or with very small GF, because GF values near zero will give infinity high LAI values and in consequence wrong estimations of CI (Lang and Yueqin 1986). Moreover, the randomness assumption of the canopy foliage in the segments could be not accomplished when the algorithm is either applied to segments longer than ten times the foliage elements, because large gaps between tree crowns could be found; or applied over very small segments, because lower φ angles provide short sampling segments. In these situations, the Poisson model assumes an infinite canopy and GF can not be derived (Gonsamo and Pellikka 2009). To solve this problem, the segments with no gaps can be altered artificially inserting a gap of one pixel (Van Gardingen et al. 1999). Another possible solution is to remove the segments with no gaps from the CI calculations (Walter et al. 2003) or merge the null-gap segments with the next found non-null gap segment within the same φ ring to get the average GF value (Gonsamo et al. 2010).

A new CI method, resulting from the combination of both, CCI and CLX, was proposed by Leblanc et al. (2005) to address some of the limitations of the previous methods. It computes the ratio between the logarithmic average GF of all segments and the product of the CCI by the sum of the logarithm GF of each segment. According to Leblanc et al. (2005), the CI will be affected with this method just by the large gaps in the segments.

A hierarchical modification of the CLX was also proposed by Walter et al. (2003) to overcome the problem of empty segments. The GF values are linearly averaged over a sector with a specific θ width, assuming randomness at this specific scale. Subsequently, it computes logarithms on each sector to give an overall logarithmic average (Gonsamo and Pellikka 2009). The process is repeated for the whole θ range and for each φ ring. Finally, the CI is computed as the ratio of average GF and the

overall logarithmic average (Gonsamo and Pellikka 2009). The averaging scale can vary from a full linear with a θ width equal to 360°, to full geometric with a θ width equal to 10°. Between these extreme scales others ones can be evaluated to average the GF and to avoid the empty gaps segments (Walter et al. 2003).

Pielou's spatial segregation coefficient (PCS) (Pielou 1962) can be also considered as a CI when it is applied under dense and clumped canopies (Walter et al. 2003). The method records the absence or presence of one plant species in a field transect. Then the field observations are compared with those that would be expected from an unsegregated population with the same numerical composition. This provides the randomness deviation of each species respect to the other (Pielou 1962). Walter et al. (2003) implemented PCS on HP, in which the black and white pixels represented canopy and gaps, respectively. CI was computed for each φ angle based on the measured gap size distribution and the probability of find a black or white pixel in a theoretical random distribution function (Gonsamo and Pellikka 2009).

1.3 Estimation of vegetation canopy structure using Terrestrial (TLS) and Airborne (ALS) Laser Scanners

The Light Detection And Ranging (LiDAR) is a laser that measures ranges (distances) using the time elapsed between an emitted laser pulse and its return to the sensor after reflecting from an intercepted object (Lim et al. 2003). To measure the distances with LiDAR systems two main principles are used, one is the time of flight that measures the distance based on the speed of the light and the time traveled between the emitted and received pulse, and the other is by measuring the difference in the transmitted and received signal phase that is converted into travel time (Wehr and Lohr 1999).

The type of information that LiDAR systems collect can be grouped into two broad classes. The discrete return devices can measure either one laser pulse return (single-returns) or several returns from each pulse (multiple-returns). On these sensors, the objects are represented as major peaks in the return signal and the systems usually store the distance and the power of each peak (Lefsky et al. 2002). The other LiDAR class are the waveforms devices that record for each pulse return the

time-power variability in small time intervals, almost in a continuous way, giving detailed distribution of all the objects intercepted by laser pulse (García et al. 2009). Due to high amount of information that the waveform records, these devices usually have a large laser pulse footprint size at the ground of a few meters. Conversely, the discrete return devices employ small footprint size of a few centimeters.

A scanner combined with the LiDAR laser allows acquiring data along and across the movement of the device. When the instrument is mounted on a tripod or a plane is called Terrestrial (TLS) or Airborne (ALS) Laser Scanner, respectively. ALS also need to include other instruments like a Global Positioning System (GPS) to collect the platform position, an Inertial Navigation Systems (INS) to record the scanner orientation in relation to the movement of the plane (yaw, pitch and roll) and a control-storage unit (Lim et al. 2003). The combination of all this information provides the three dimensional (3D) coordinates of the objects intercepted by the sensor (García et al. 2009; Lefsky et al. 2002).

TLS and ALS can measure forest structure, because they can provide detailed descriptions of either the horizontal and vertical spatial organization of the elements inside the canopy (Lefsky et al. 1999; Lefsky et al. 2002; Lim et al. 2003; Parker et al. 2004). These instruments overcome some of the problems that direct or semi-direct methods show (Henning and Radtke 2006; Lovell et al. 2003). The sensor records the lasers reflected from one or several surfaces within the canopy, the rest of the pulses continue travelling to intercept with lower vegetation or ground surface. The ability of these instruments to record laser returns at different depths in the canopy or even the ground sets the foundation to measure the gaps within the canopy (Hopkinson and Chasmer 2007). GF follows the Beer Lambert's Law, which can be assumed equivalent to canopy Transmittance (T) from LiDAR:

$$GF = e^{-kLAI};\ T = \frac{I_I}{I_o}; where\ GF \equiv T \tag{1}$$

where I_I is the light intensity after travelling a path length (I) throughout the canopy, I_o is the sky light intensity above canopy and k is the extinction coefficient, which is the fraction of foliage area projected onto a perpendicular plane.

GF has been estimated from TLS and ALS. In the case of TLS, Danson et al. (2007) developed a new model to estimate the GF based on the ratio between the actual number of laser returns that a discrete TLS obtains and the theoretical total number of returns that this instrument would have acquired according to its angular resolution, θ and φ range in the case of a canopy with no gaps. The same approach was applied later by Calders et al. (2011) and Moorthy et al. (2011; 2008), but with a modification using the point density measurements instead of the returns to compute the GF. Similarly, (Garcia et al. 2011) adjusted the data to a grid of occupied/non occupied cells. Those cells occupied by at least one return were codified as 1, whereas the empty cells were considered a canopy gap and codified as 0. The GF was computed as the ratio between the non-occupied pixels and the total pixels in the image.

Parker et al. (2004) used a TLS (Riegl LD90-3100HS), able to record the direction of the non-return laser pulses and computed the GF as the ratio between them and all laser pulses. The same method but using a waveform TLS (ECHIDNA™) was proposed by Lovell et al. (2003), to estimate the GF. They modified the software to repeat pulses when a non-return was detected. According to this configuration if a pulse reached a number of repetitions, this return was classified as a gap.

The use of volumetric pixels (voxels) has been shown useful to summarize the huge data recorded from TLS (Hosoi et al. 2010). The returns are grouped into this voxel structure that has a pre-defined dimension. A voxel can be considered then as an individual element in a 3D array structure (Hosoi and Omasa 2006). Van der Zande et al. (2006) and Henning et al. (2006) used voxels to group the laser returns and subsequently classify them as occupied or empty. The same approach was proposed by Van der Zande et al. (2011) but applying a ray tracing algorithm to identify if a voxel was occupied or not. A modification of the use of voxels was proposed by Seidel et al. (2012). They transformed the voxels into a simulated HP using a polar projection and later compute the GF with the Gap light analyzer software mentioned earlier.

The estimation of GF from the intensity of the TLS laser returns is uncommon. Hopkinson et al. (2013) and Zhao et al. (2011) employed a waveform TLS with the ‘apparent reflectance’ to relate it with GF. This is a target reflectance that would return almost the same intensity value as if it were obtained by the actual target (Zhao

et al. 2011). The integral of this apparent reflectance can be considered as a reduction of the outgoing signal, which can be linked with GF. To compute GF, they estimated some target thresholds that are between the fully and partially intercept beam. Finally, following the same rational of HP classification, all the samples above the thresholds were considered a true gap in the canopy.

In regards to the GF estimation from ALS, one of the simplest methods is to measure the penetration rates based on fraction of laser returns below the canopy to the total of returns (Hopkinson and Chasmer 2009; Lovell et al. 2003; Riaño et al. 2004). Another approach is to derive GF empirically from ALS metrics. In that context, Riaño et al. (2004) tested $P50^{th}$, $P75^{th}$ and $P95^{th}$ height percentiles and Hall et al .(2005) derived metrics like the density of first returns.

When the sensors are able to record more than one return per pulse (first, last, intermediate or single), it is possible to test more complex models that define the penetration rate for which is it is necessary to account for the nature of the return type and the possibility that a pulse produces more than one return. To correct pulses than can produce first and last returns, Solberg et al. (2009, 2010 #60) and Korhonen et al. (2011) chose 0, 0.5 and 1 weight factors for canopy-single, canopy-first plus ground-last or ground single, respectively. There are some problems of GF under and overestimation for models based only on first or last returns (Lovell et al. 2003; Morsdorf et al. 2006). For those cases, it is possible to look for the better choice than reduces the bias caused by the interactions between return types, footprint dimension, gap size and return thresholds. The combined use of first, last and intensity return seems to be a good alternative to reduce at least the issue of footprint size (Lovell et al. 2003). To account for this, Hopkinson and Chasmer (2009) preferred models that employed first returns. Morsdorf et al. (2006) tested models with first, last, and single returns to find the best agreement. Sasaki et al. (2008) used ground single returns because these returns ensure the pulse fully hits either the canopy or the ground, avoiding mixed returns.

The estimation of GF from the intensity of the ALS laser returns is also less common than the methods above. Solberg et al. (2010) proposed the ratio between the intensity of ground and all laser returns for first and first and last returns,

considering that all the returns had the same weight. Using the same rationale, Hopkinson and Chasmer (2007; 2009) considered that there are some returns (intermediate or last) that according to its nature modify the transmission lost, especially when they passed throughout the canopy. Therefore, they considered the square root factor for these returns type. Similarly to frequencies models, the GF can be estimated from the fraction of the ground energy to the total energy recorded by full waveform sensors. Lovell et al. (2003), Lefsky et al. (1999) and Hall et al. (2005) applied the fraction of a waveform return at a set height threshold and incorporated a factor to correct the reflectance difference between canopy and ground, using a factor of one for canopy and two for ground.

The number of studies that estimate CI is fewer than GF from either TLS or ALS. In the case of TLS sensors, the CI was estimated by Moorthy et al. (2011; 2008) based on the theory of gap size distribution developed by Chen and Cihlar (1995). They computed the CI by comparing the GF for a tree crown based on (Danson et al. 2007) and for a simulated spatial random distribution of the laser returns for the same area. In addition, Zhao et al. (2012) estimated the CI also based on Chen and Cihlar (1995) theory applied over the gap size distribution data derived from a nominal spatial index obtained from a waveform TLS ECHIDNA™ sensor. Respect to ALS, Thomas et al. (2011) applied an one empirical approach based on several ALS metrics that included mean, median and standard deviation which were derived from the point cloud and from different height percentiles for first, last returns and for both of them.

1.4 Hypothesis and objectives

1.4.1 Hypothesis

The laser pulse returns and their intensity from TLS and ALS can provide an estimate of the vegetation canopy structural parameters GF and CI and they can be considered an alternative to reduce the limitations of other methods like HP.

1.4.2 Objectives

1. To develop a new approach to estimate canopy GF at individual tree scale using angular images from TLS data.

2. To apply the gap size distribution theory adapted to HP to estimate the canopy CI from TLS data.
3. To apply height empirical metrics and canopy penetration models based on the Beer Lambert´s law to estimate the canopy GF from ALS data.
4. To apply height empirical metrics to estimate the canopy CI and to develop a new approach based on the adaptation of the spatial segregation algorithm to ALS data.

2 METHODS

The methods section is organized according the experiments done in the two study sites. The first site was Majadas del Tiétar and the second was Jasper Ridge, for both sites the estimation of vegetation canopy GF and CI was carried out using TLS and ALS data for the first and second site respectively.

2.1 Description of the study sites

2.1.1 Majadas del Tiétar

The Majadas study site is located in the Tiétar valley, Cáceres, Spain (39°56'27" N latitude and 5°46'27" W longitude) (Figure 1). On the place there is an eddy covariance flux tower part of the FLUXNET network (http://fluxnet.ornl.gov/site/440, last accessed on December 21st, 2013). The site belongs to an ecological and biodiversity corridor that was declared in 2003 because of the high diversity of bird species associated to forest ecosystem. The climate is Continental with few rains (400-800 mm annually) concentrated in spring and autumn seasons and with hot and dry summer. The site is a savanna type (Dehesa) managed for grazing, comprised by *Pyro bourgeaneae* Decne., *Querceto rotundifoliae* quercetum mesomediterranean series (Rivas Martínez and Gandullo 1987), which has been transformed into a wooded (*Quercus ilex* L.), with few individuals of *Quercus suber* L., grassland (herbaceous species) and containing some shrub species (*Cistus ladanifer* L., and *Lavandula pedunculata* Mill.). Tree density is about 20 trees/ha, with an average height of 8 m and a breast height diameter (DBH) 0.4 m.

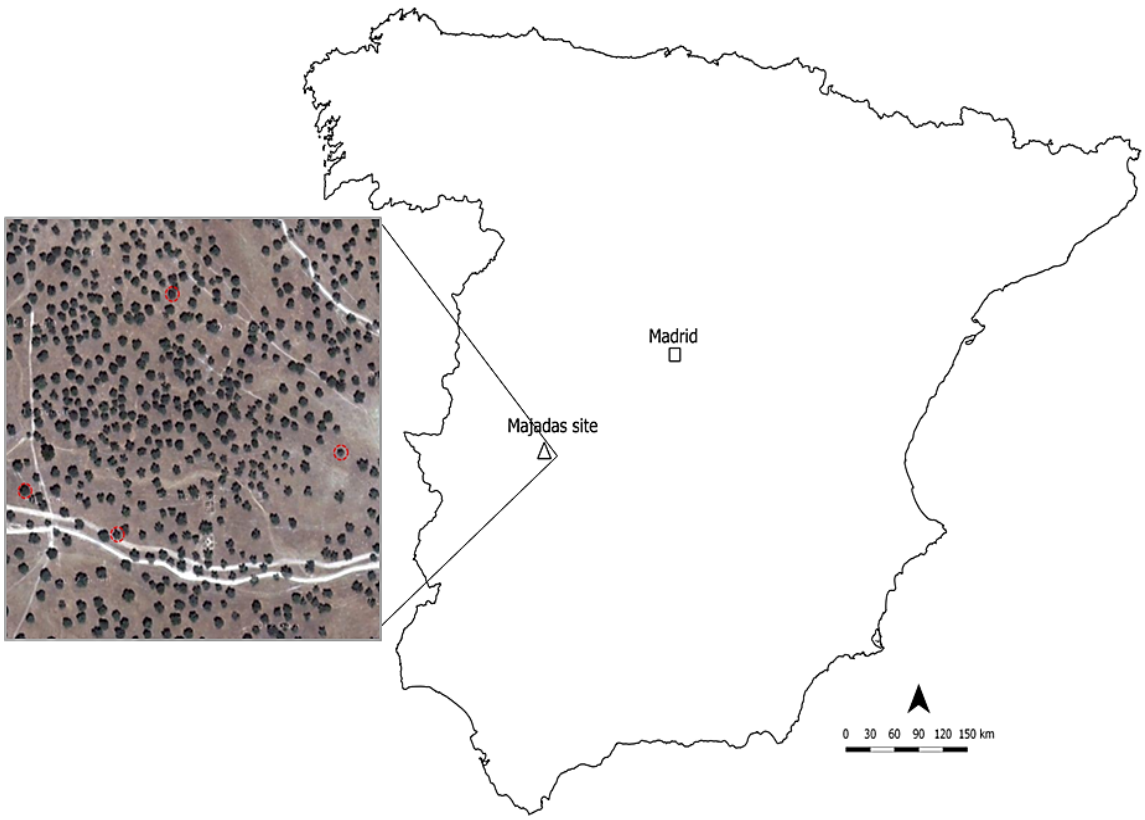

Figure 1. Majadas del Tiétar study site and scanned tree distribution in red circles.

2.1.2 Jasper Ridge

Jasper Ridge study site is located on the northeastern foothills of the Santa Cruz Mountains in Portola Valley (37°24'N latitude and 122°13'30"W longitude), San Mateo County, California, USA (Garcia and Ustin 2001) (Figure 2). The site is a Biological Preserve of about 500 ha owned by Stanford University (http://jrbp.stanford.edu/, last accessed December 21st, 2013). It provides a natural laboratory for researchers from all over the world. This site has a Mediterranean climate with hot and dry summers and an annual average precipitation is about 650 mm (Dahlin et al. 2011). In the sampling plots, the elevation ranges from 52 to 161 m above sea level and the slope varies from a minimum of 2% to a maximum of 34% with an average of 14%. According to Garcia and Ustin (2001) the site is integrated mainly by five vegetation types: shrubland, evergreen and deciduous forest, herbaceous perennial wetlands and annual grasslands. The shrub vegetation type comprises a wide range of Californian

chaparral species like woodbalm (*Lepechinia calycina* (Benth.) Epling ex Munz); buckbrush (*Ceanothus cuneatus* (Hook.) Nutt.); coyotebrush (*Baccharis pilularis* DC.); chamise (*Adenostoma fasciculatum* Hook. & Arn); California yerba santa (*Eriodictyon californicum* (Hook. & Arn.) Torr.); toyon (*Heteromeles arbutifolia* (Lindl.) M. Roem.); coastal sagebrush (*Artemisia californica* Less.); and orange bush monkey flower (*Mimulus aurantiacus* W. Curtis). The forest vegetation type includes blue oak (*Quercus douglasii* Hook. & Arn., also present in the shrubs plots), California live oak (*Quercus agrifolia* Née); and valley oak (*Quercus lobata* Née) (Casas et al. In review).

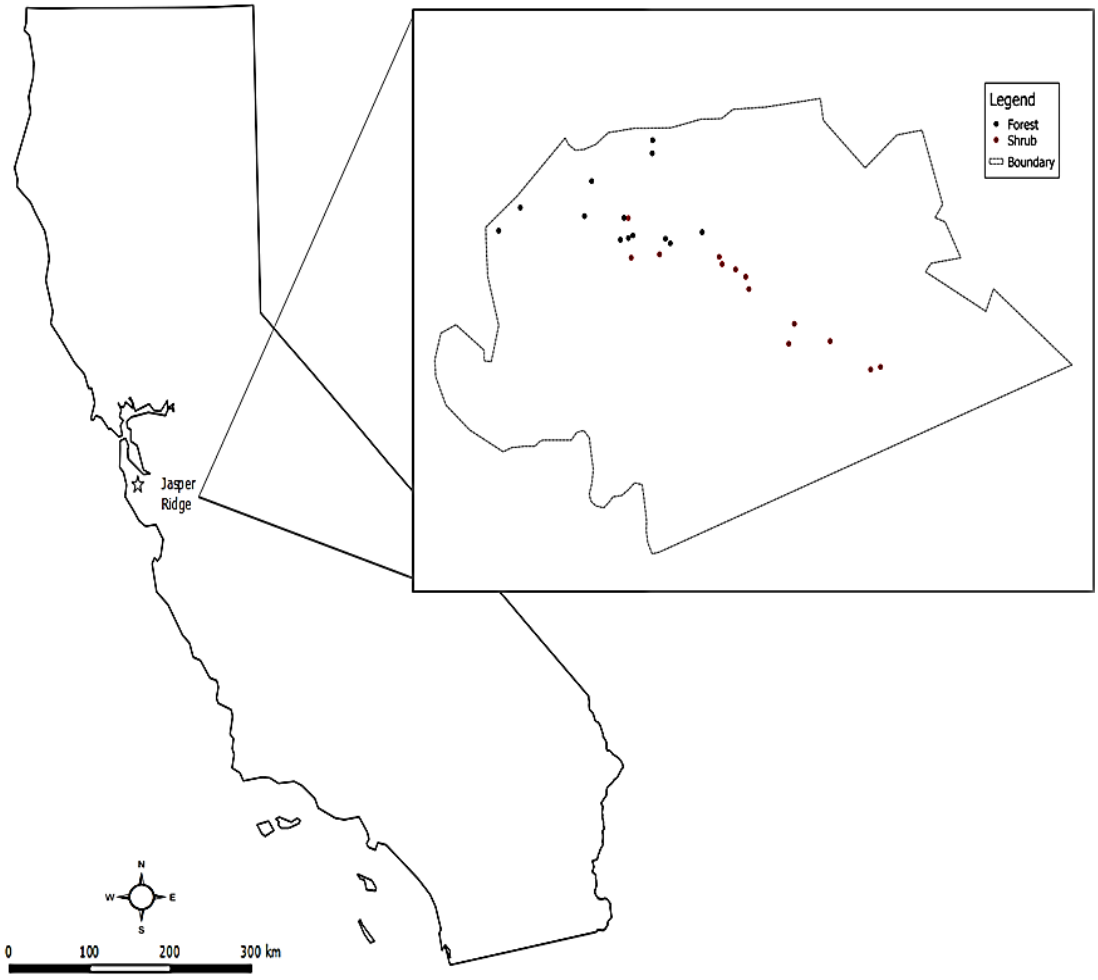

Figure 2. Jasper Ridge study site and plot distributions.

2.2 Vegetation canopy GF from TLS

2.2.1 Data collection

Four holm oak *(Quercus ilex L.)* trees (treeId: Z1P5, Z1P7, Z2P6 and Z2P11) were sampled on October 9, 2009 with a TLS Leica HDS-6000 (Leica Geosystems, Switzerland, www.leica-geosystems.com) (Table 1). The instrument has fully integrated a scanner, controller, data storage and battery. Its laser is a 3R class, which is considered safe if handled carefully (Voegtle and Wakaluk 2009). The 3mm laser footprint at the sensor exit and a 0.22 mrad beam divergence, which translates into 5 mm footprint at 8 m from the sensor. Based on the phase difference between the transmitted and received signal, the TLS measures the time delay to determine the position hit at the object. The system records only the first pulse returns as XYZ positions and the intensity of such returns.

Table 1. Technical specifications of the HDS-6000 scanner (Leica-Geosystems 2008).

Measurement principle	Phase shift
Dynamic scanning range	1-79 m
Data sampling rate	Up to 508.000 pulses/s
Footprint size	3 mm at exit (based on Gaussian)
Laser divergence angle	0.22 mrad
Minimum spot step	1.6 x 1.6 mm@10m
Accuracy Positioning Range	 +- 6 mm (25 m) +- 4 mm (90% albedo, 25 m) +- 5 mm (18% albedo, 25 m)
Laser wavelength	670 nm
Laser class	Class 3R
Scan field of view	360° x 310°
Control interface	Side touch panel, optional PDA or laptop
Power supply	24 V DC
Dimensions	190mm D x 244mm W x 351mm H
Weight	14 kg
Data storage	Integrated hard drive 60 GB
Digital camera	Externally mounted.

The TLS system was deployed approximately at the midpoint between the tree trunk and the edge of the canopy to capture a horizontal perspective of 0°-360° in azimuth (θ) and 0°-154° in zenith (φ) with an angular sampling interval of approximately 6.3 mm at 10 m distance (Figure 3). To reduce the occlusion effect, a second TLS data collection occurred in a diametrically opposite position (180°) to the first one in relation to the tree trunk (Moorthy et al. 2011). Cyclone software (Leica

Geosystems, Switzerland, www.leica-geosystems.com) was used for post processing and to generate the point cloud. Following Seidel et al. (2012) to remove the noisy returns, only were considered data with intensity values between 0.01 and 1.0, range under 80 m and laser returns at least separated by 1.6 mm. Moreover, eight trunk sections from the TLS models were extracted in order to have data without gaps and with similar range and laser pulse return density. Subsequently, the TLS point cloud from the tree canopy was isolated from the rest of the TLS scene and exported as XYZ ASCII files. The ASCII XYZ data were transformed to compute spherical coordinates (R,θ,φ) from Cartesian (X,Y,Z) coordinates using the following equations:

$$R = \sqrt{(X^2 + Y^2 + Z^2)} \tag{2}$$

$$S = \sqrt{(X^2 + Y^2)} \tag{3}$$

$$\theta = \begin{cases} \sin^{-1}\left(\frac{Y}{S}\right), X \geq 0 \\ \pi - \sin^{-1}\left(\frac{Y}{S}\right), X < 0 \end{cases} \tag{4}$$

$$\varphi = \cos^{-1}\left(\frac{Z}{R}\right) \tag{5}$$

where X, Y, Z are the components for the Cartesian coordinates; R is the point´s distance from the origin, S is the distance projected on the plane XY.

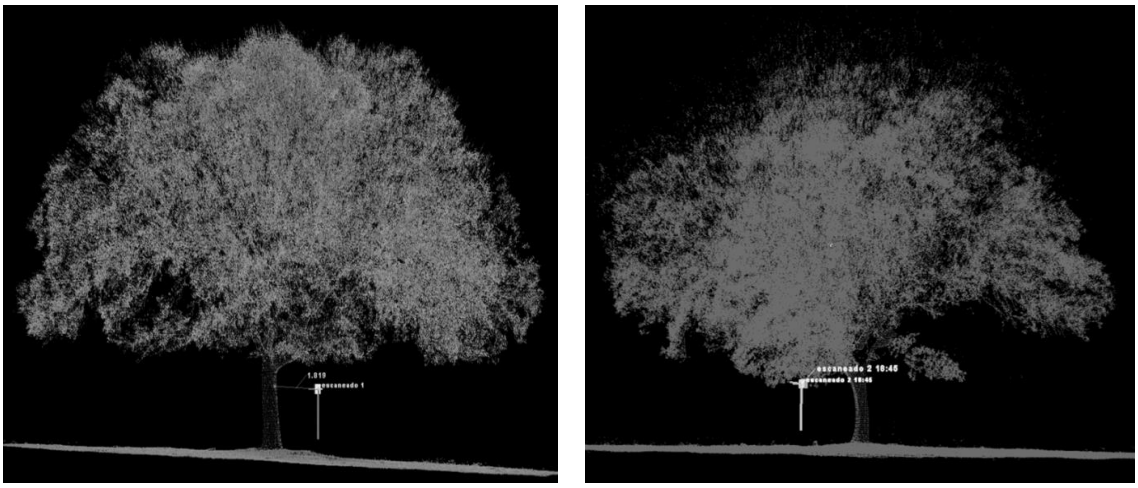

Figure 3. TLS data tree Z2P6-SCAN1 (left), Z2P6-SCAN2 (right) and position of the TLS instrument.

The TLS dataset was complemented with two HP per tree that were taken using a Canon camera EOS 400D with Nikkor 8-mm f/2.8 fisheye lens adaptor coupled to the TLS head. The HP were taken at the same time of TLS data collection and stored in .JPEG format with an image size of 3888 x 2592 pixels (Figure 4).

Figure 4. HP taken from a camera coupled to TLS head for first scan position (left) and second scan position (right).

2.2.2 The angular resolution algorithm

We designed an algorithm to compute the angular resolution of the TLS data (Figure 5). Knowing the angular resolution allows determining the theoretical direction of the laser pulses. Therefore, if each and all missing returns can be identified, this provides the direction of all gaps in the canopy. Angular resolution refers to the sampling interval angle in the θ and φ directions. This definition considers the angular distance between the centers of two adjacent laser footprints, without taking into account its beam width as proposed by Lichti (2004). The angular distance between two adjacent laser pulses was used instead of the Euclidean distance since the former is constant and independent of distance between the TLS and the objects hit by the laser, i.e. the sensor-object range. In order to compute the angular resolution, only were used returns corresponding to adjacent pulses, both horizontally and vertically. The process to calculate the resolution is iterative. Firstly, the algorithm sets an initial large value for θ and φ resolution. For each return the algorithm looks for its four nearest neighbors and their angular distance between them is calculated. Secondly, the angular distance between the return and either of its neighbors is recorded only if 1) the slope between the two returns is less than 10° respective to any of the four

cardinal directions N, S, E or W; and 2) its angular distance is less than 1.5 the initial resolution in either θ or φ. These thresholds were empirically adopted after analyzing a subset of returns over an area with no gaps. To conclude the first iteration step, the algorithm computes the average of all the angular distances recorded in the N-S and E-W direction to obtain the θ and φ resolution, respectively. The iterations continue until the difference between the angular resolutions calculated in two consecutive steps is close to zero. A convergence threshold was established to 1.0e-07 to stop the iterations (Figure 6).

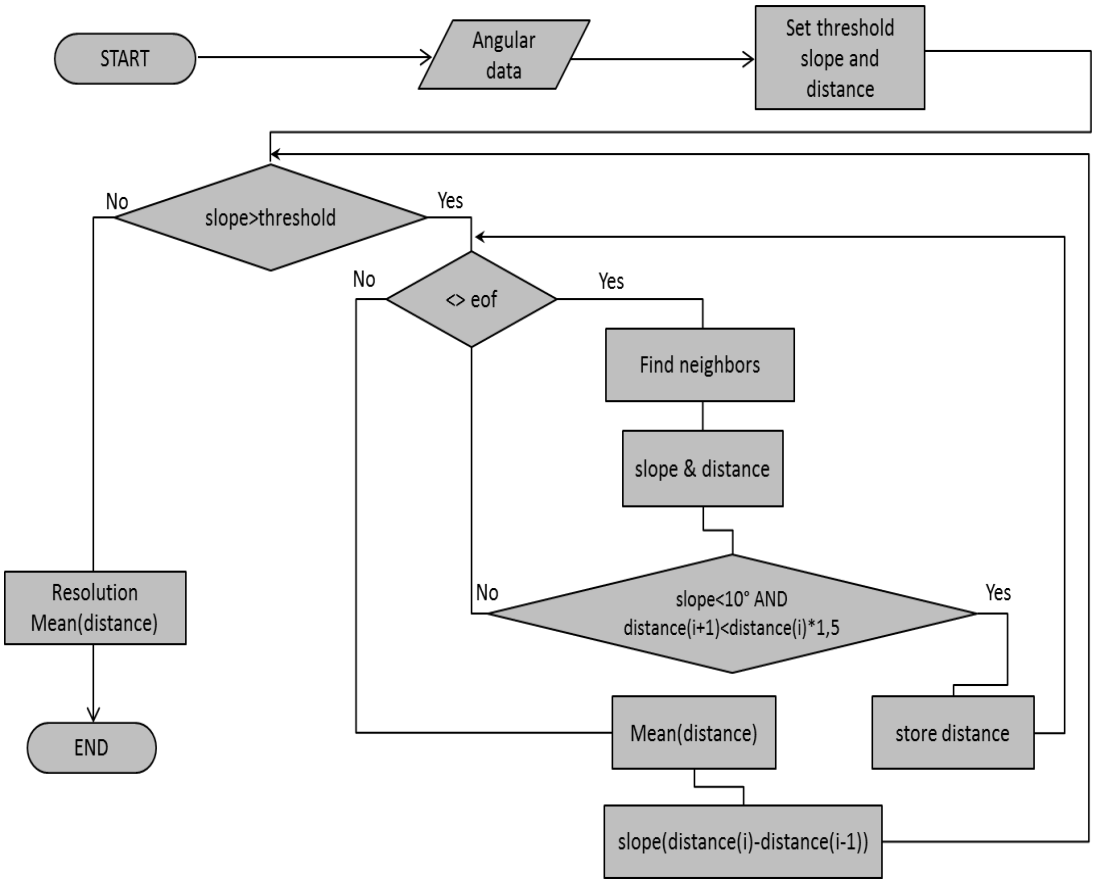

Figure 5. Flow chart of the angular resolution algorithm.

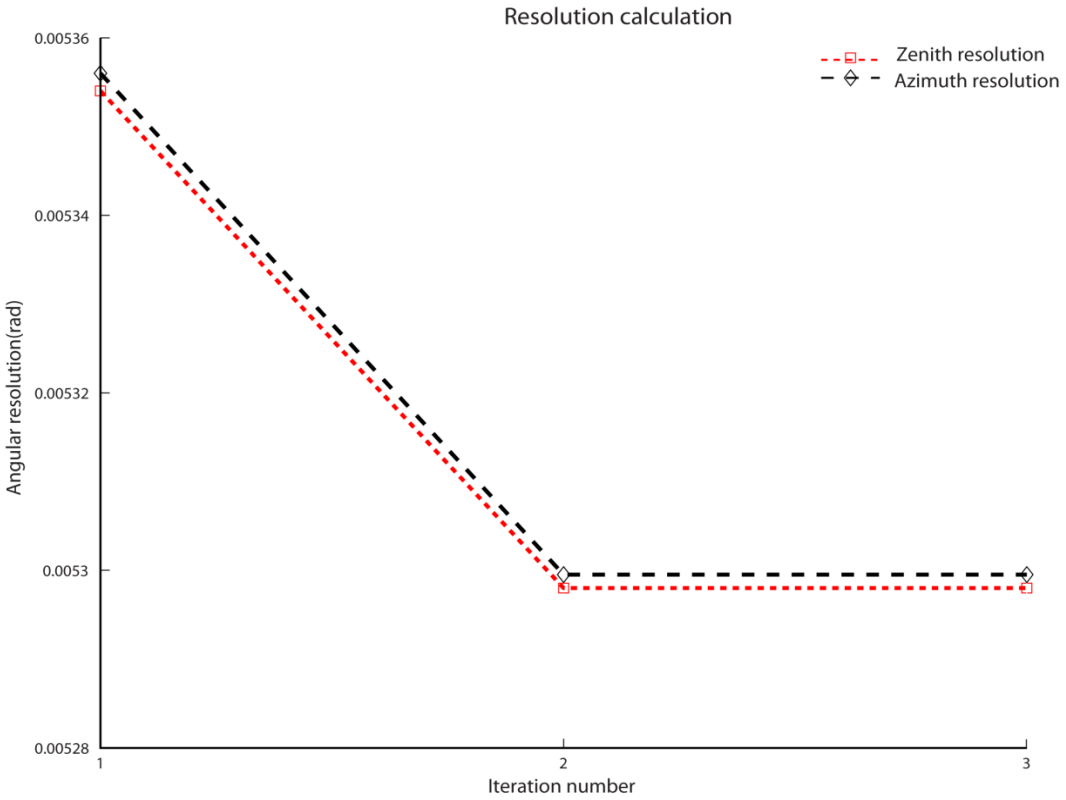

Figure 6. Example of resolution convergence after three iterations.

2.2.3 Computing angular GF images

Once the average angular resolution is determined it is possible to generate an angular grid with cells containing the angular location of the actual gaps, in addition to the already known location of the laser returns that hit the canopy. Under ideal circumstances, the angular location of a laser return should correspond to the center of a cell. However, this is not the case since TLS data is subject to angular noise. The factors that cause this noise are related to the distance to the object, the object surface itself and the mechanical components of the scanner (Kremen et al. 2006). Most TLSs use a scanner with a servomotor or rotational mirrors to produce the angular displacement in both horizontal and vertical directions (Reshetyuk 2006). Therefore, to reduce the difference between the position of the grid and the actual data, the location of the grid was shifted to make the distance between the center of each laser pulse and its corresponding cell the shortest. The grid was tested in twenty-five different positions at -1/2, -1/4, 0, 1/4 and 1/2 of the angular resolution in both θ and φ directions. Finally, a binary image was created taking into account the number of returns that lay within each cell. Thus, if a cell did not contain any laser returns, it

was marked as a gap, whereas if it contained one return, it was considered occupied by a canopy element (Figure 7). GF images were computed as the ratio of cells identified as gaps to the total cells in the image.

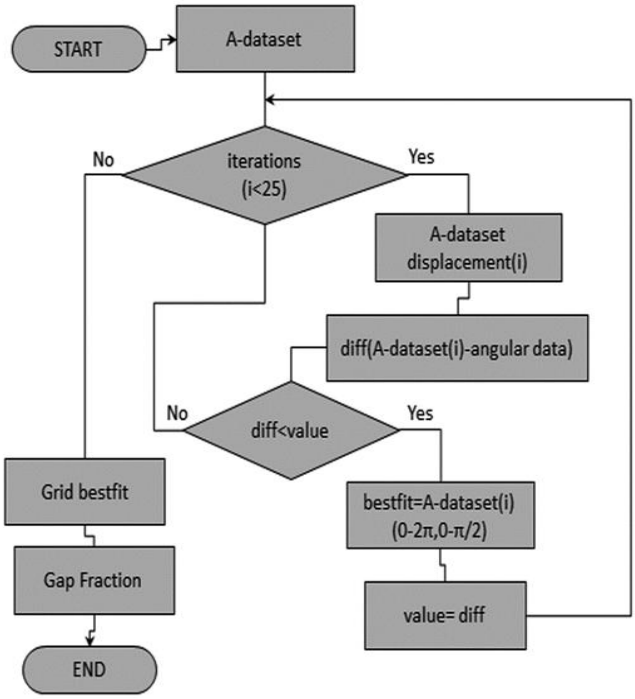

Figure 7. Algorithm to compute the angular GF images.

2.2.4 Validation of the angular resolution algorithm and GF images

A sensibility analysis was carried out using simulated TLS data to evaluate the performance of the algorithm to retrieve the angular resolution and the GF images derived from it. We simulated TLS data with known GF and compared it to the one calculated by the algorithm. In order to simulate different situations found in the actual TLS data, several sources of variation were considered: 1) angular noise; 2) the angular spatial pattern distribution of the gaps; and 3) the amount of GF itself. A multivariate normal distribution available in Matlab (The Mathworks; Natick, MA, USA; www.mathworks.com) was used to mimic the angular noise. Such noise is a function of 1) the amplitude of the angular noise (sigma) that was modeled as a percentage of the angular resolution, 2) its variance ($sigma^2$) and 3) its covariance. The second source of variation consisted in modeling several canopy gap spatial pattern distributions, such

as random gaps (R), clusters of gaps (C) and a combination of both of them, which is closer to the actual vegetation canopies (RC) (Figure 8).

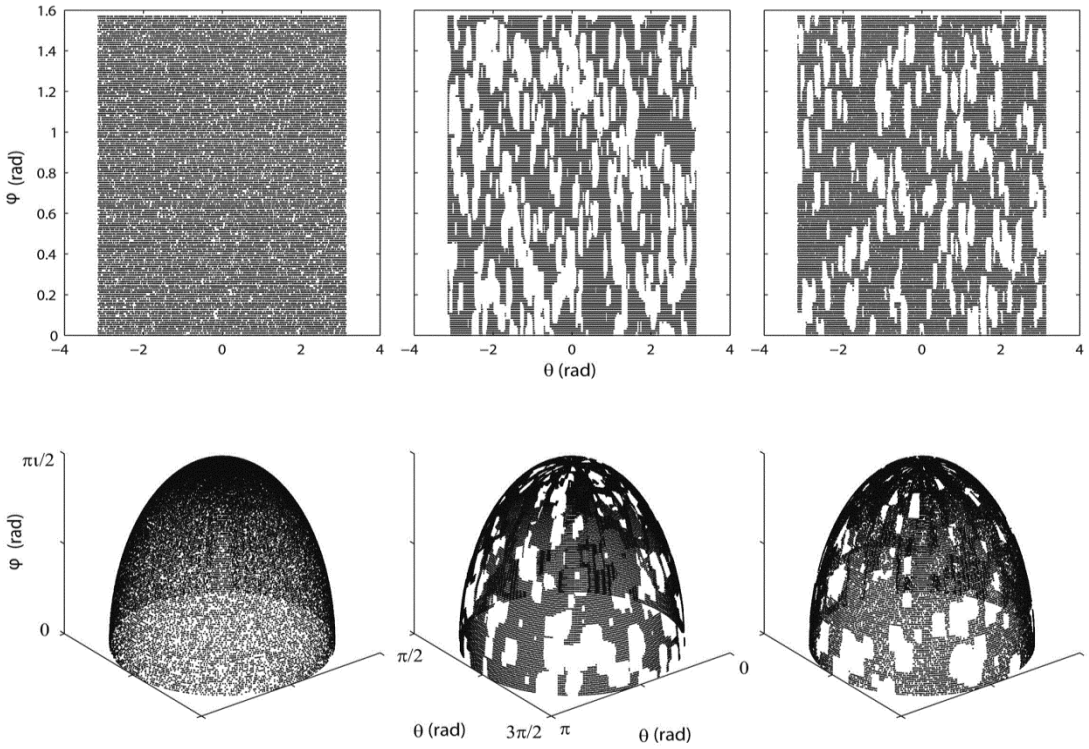

Figure 8. GF images (series above) and hemispheric 3D views (bottom series) of simulated angular data with 50% GF and with either R (left), C (center) or RC (right) gaps.

The GF was increased for each gap pattern by removing progressively simulated laser returns. The R laser pulse extraction relied on the extraction of random returns individually without replacement. Random numbers were produced using a Matlab function (The Mathworks; Natick, MA, USA; www.mathworks.com). These random numbers pointed to the identifiers of the laser returns to be extracted. The C gap pattern combined the random selection described above with a convex hull function to extract all the laser returns within a dynamic random radius size from the random point, which could range from one to ten times the angular resolution. The RC gap pattern mixed the two algorithms described before creating 30% R and 70% C gaps. These percentages were selected arbitrarily; nevertheless, the main objective was to simulate laser returns corresponding to a canopy with a known GF and pattern, therefore, the actual value of the percentage used was not relevant.

The algorithm was validated over this simulated dataset that contained a pre-defined known GF. A total of 1890 files were generated with an angular noise ranging

from 2%-14%, the three R, C and RC gap patterns, and a GF ranging between 10-90%. A Kruskal-Wallis test (KW) evaluated the algorithm sensitivity to the sigma noise in relation to the GF estimations (Hollander and Wolfe 1999). The KW test is a non-parametric test and therefore not subject to the assumption of normal distribution of the residuals. This test, which is the equivalent to ANOVA parametric test, is an extension of the Mann-Whitney U test to more than 2 groups, and compares the medians of two or more samples to determine if they come from the same population (null hypothesis) or from different populations (alternative hypothesis).

In order to validate the results an independent technique was applied to the simulated dataset. The analysis of HP is one of the most widely accepted indirect methods used to estimate GF (Bréda 2003; Welles and Cohen 1996). Considering that, the simulated TLS data was converted into simulated HP (SimTLS-SHP) for additional testing on Hemiview (Delta-T Devices, UK, www.delta-t.co.uk), a commercial software that analyzes HP. A low angular noise from 2% to 6% was used to match the noise quantified in previous studies of similar TLS instruments (Kersten et al. 2008; Kremen et al. 2006; Reshetyuk 2006). SimTLS-SHP in essence produces a planar projection of a hemisphere. The simplest and most common hemispherical lens geometries are the polar or also called equiangular projection (Gonsamo et al. 2010; Jonckheere et al. 2004). In this case, the following polar transformation was applied:

$$l = 2R(\varphi)/\pi \tag{6}$$

where R is the radius of the whole field of view and l is the radial distance on the HP image plane. Seidel et al. (2012) also used this projection to estimate GF from simulated HP produced from TLS voxels.

Using as pixel size the TLS's angular resolution, SimTLS-SHP were generated as the ratio between the laser pulse canopy returns and the sum of returns and gaps. Due to the change in projection, the number of returns/gaps per pixel is redistributed, becoming more heterogeneous (Figure 9). A pixel was considered a canopy return (0, black) if the number of laser returns was above 50% or a gap (255, white) if below this threshold, producing a SimTLS-SHP that was saved in BMP format, so that it could be imported on Hemiview. Each SimTLS-SHP was then divided in a skymap with 18 θ and 8 φ angular sections (Figure 10). Finally, the GF was computed in Hemiview from the

SimTLS-SHP using a 128 fixed threshold value that separated the canopy returns from the gaps.

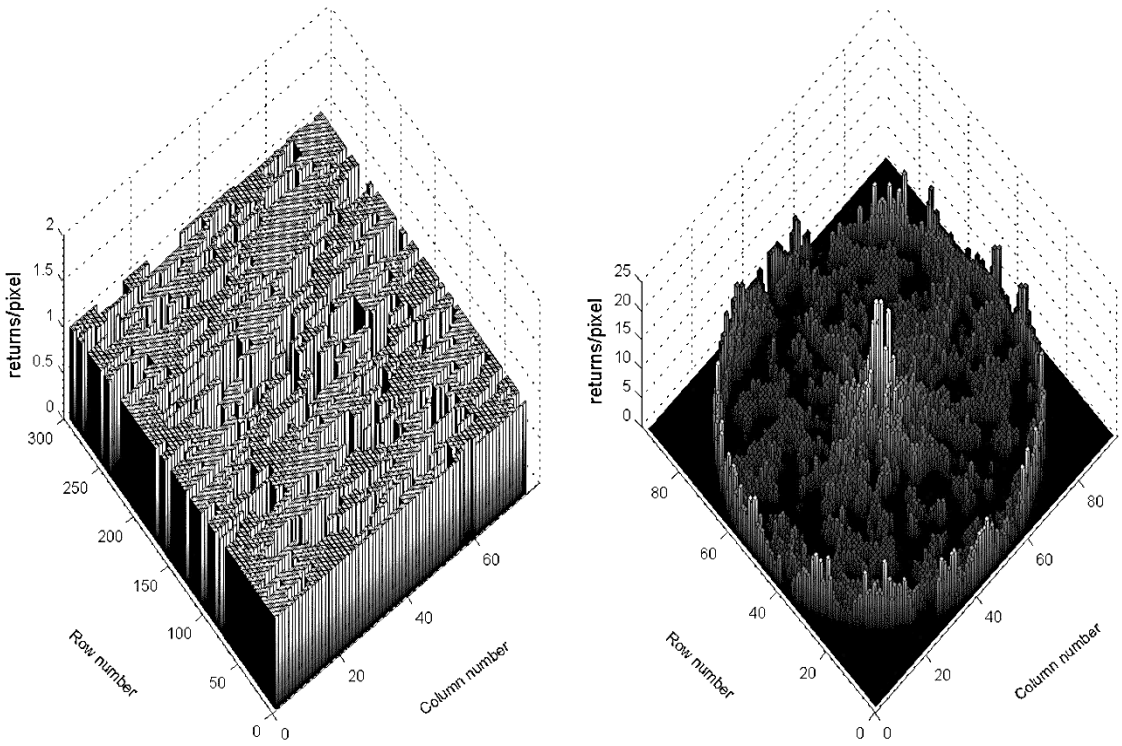

Figure 9. Number of laser returns per pixel for a simulated C gap pattern GF image (left) and its correspondent SimTLS-SHP (right).

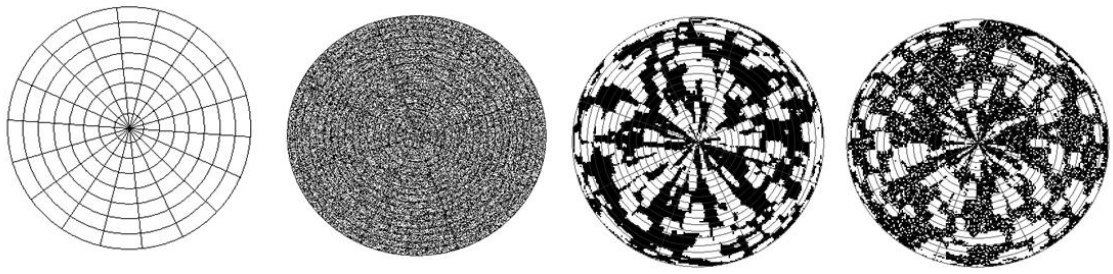

Figure 10. From left to right: the skymap divisions, and SimTLS-SHP with R, C and RC gap patterns.

2.2.5 Estimation of GF on real TLS data

Given the large TLS data volume, the TLS acquisition for each tree was sliced in 2° zenithal widths ($\Delta\varphi$) similarly to previous studies (Moorthy et al. 2008; Zhen et al. 2011). The average angular resolution of all φ slices was calculated using the algorithm described in 2.2.2. This average excluded below 4° φ slices that contained a lower laser pulse density and showed a higher angular noise, due to the lower accuracy of these

canopy returns, since they are coming from more remote locations (Antonarakis et al. 2010; Seidel et al. 2012). In addition, angular resolution was also estimated the same way for tree trunk sections, where gaps rarely occur. As it will be shown in the results and discussion 3.1 section, this latter angular resolution value was selected for further analysis since it was more accurate. An angular grid was adjusted based on this angular resolution to transform it into a GF image and then into a HP (TLS-SHP) as described above in sections 2.2.3 and 2.2.4.

The GF was computed not only for the TLS-SHP but also for the HP taken from the camera coupled to the TLS head. For the HP, only the blue channel was chosen, since it gives the maximum contrast between canopy objects and sky (Zhao et al. 2011). As Jonckheere et al. (2005) described, manual thresholds are a subjective and hard task to reproduce consistently, which can introduce an additional error into the GF computation. For that reason, an automatic threshold method developed by Ridler and Calvard (1978) was implemented to classify the pixels in the HP as canopy (black) or sky/gaps (white) and are described in deep in section 2.4.1. For both type of images, HP and TLS-SHP (Figure 11), the GF computation was constrained to the 0°-60° φ range in order to avoid fuzzy pixels that would comprise a mix of vegetation canopy and sky information (Jonckheere et al. 2004) and also to consider only the φ range common in the HP and the TLS-SHP .

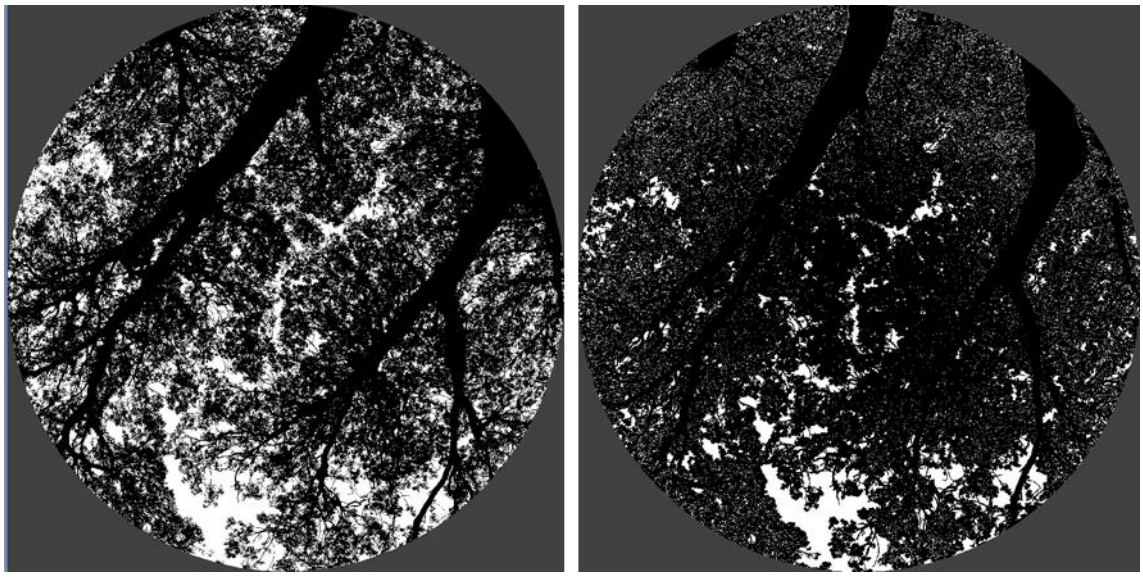

Figure 11. Tree ID Z2P6 seen from the HP (left) and TLS-SHP (right).

2.3 Vegetation canopy CI from TLS data

The Chen and Cihlar´s clumping index (CI) that quantifies the gap size accumulation was computed for the eight TLS data that were transformed into HP (TLS-SHP) (for details see section 2.2.4 and 2.2.5). This method needs the images has been already segmented in two categories, sky (white pixels) and canopy (black pixels). However, as mentioned in section 2.2.4 the TLS-SHP images no needs this pre-processing step. Due to the size of the HP camera sensor is smaller than the fisheye lens, the final HP image was cropped. That resulted in HP images with an angular φ range that reached until the 55° for effective analysis. Since the estimation of the clumping degree of foliage elements using indirect methods is affected by the φ angles used (Walter 2009), CI was estimated considering the following different φ ranges: 5-55°; 20-55°; 30-55°; 40-55°; 50-55° and 54-55°.

2.3.1 Chen & Cihlar's clumping index (CI)

Chen and Cihlar (1995) developed a method to estimate CI based on the gap size distribution. This method, which is an improvement of the one developed by Lang and Xiang (1986), can be used for heterogeneous canopies. Although it was originally developed for the TRAC instrument, it has also been used on HP (Gonsamo and Pellikka 2009; Walter et al. 2003). For the exact same LAI, a clumped canopy presents larger canopy GF than a random one, in addition to changing its gap size distribution. Thus, the method estimates the gap size distribution of the canopy assuming its randomness, by removing large gaps iteratively, until there are no significant differences when comparing with the gap size distribution of the real canopy (Figure 12). Finally, the distribution function of GF closest to a random distribution (compacted canopy) is compared with the actual distribution to derive the CI (equation 7)

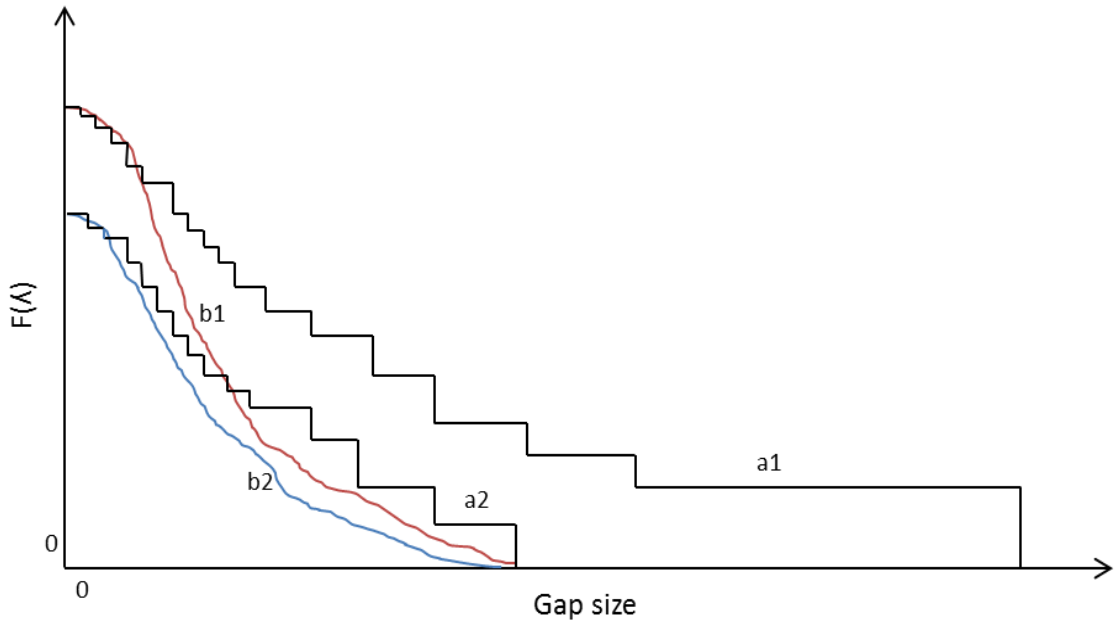

Figure 12. The measured gap size distribution (a1) and the first distribution estimate for a random canopy (b1). After the two largest gaps are removed, its gap size is redistributed (a2) and a second random canopy is estimated (b2) adapted from (Chen and Cihlar 1995).

$$CI = \frac{\ln[F_m(0,\theta)]}{\ln[F_{mr}(0,\theta)]} \quad (7)$$

where, $F_m(0,\theta)$ is the measured GF, $F_{mr}(0,\theta)$ is the GF for an imaginary canopy with a random spatial distribution.

This approach was subsequently modified by Leblanc et al. (2002a) to consider the effect of canopy compactness when large gaps are removed:

$$CI = \frac{\ln[F_m(0,\theta)]}{\ln[F_{mr}(0,\theta)]} \cdot \frac{[1 - F_{mr}(0,\theta)]}{[1 - F_m(0,\theta)]} \quad (8)$$

The final CI estimation that incorporated the modification by Leblanc et al. (2002a) integrated all of the CI values for each φ interval.

To validate all of the CI estimations from TLS-SHP, the CI was also calculated from real HP processed in CIMES package (program CLMPML) using the same φ ranges as above. The CLMPML needs the images binarized to compute the gap size distribution and then the CI. Unlike the TLS-SHP images, these HP were pre-processed to segment them according to methods described in the section 2.2.5.

2.4 Vegetation canopy GF from ALS

2.4.1 Description of the HP equipment, data collection and pre-processing

The HP data were supplied by the Center for Spatial Technologies and Remote Sensing (CSTARS) from the University of California, Davis (http://cstars.metro.ucdavis.edu). A total of 347 HP were collected on thirteen forest and thirteen shrub plots between May 12^{th} – 17^{th} of 2006, September 20^{th} – 21^{th} of 2006 and August 13^{th} – 14^{th} of 2007 in Jasper Ridge (Figure 13). To acquire the HP, a digital camera Nikon Coolpix 4300 was attached to a Nikon FC-E8 fisheye lens converter. The Coolpix 4300 has a focal length of 8-24 mm and effective resolution of 4 megapixels in a charge-coupled-device (CCD) array. The FC-E8 has a maximum FOV of 180° with a polar projection (Frazer et al. 2001). The HP were overexposed, two stops more exposure than the automatic reference exposure. This way sky appear whiter and the contrast between sky and leaves will be greater, following Zhang et al. (2005). The images were stored at the maximum resolution (2272x1704 pixels .JPEG format). To avoid the understory vegetation, a tripod elevated the camera approximately 0.7 m and 0.2 m above the ground for the forest and shrub plots, respectively. The lens was leveled to point the φ and was oriented so that the magnetic north was located at the top of the photograph (Walter et al. 2003). Each plot center was geo-located using survey grade differentially corrected Trimble GPS receiver. With a compass and measuring tape, HP were taken at the center (not always) and 2.5 m away from the center along the four cardinal directions (N, S, E, W). All HP were taken under indirect sunlight conditions, to minimize glare from direct sunlight; that is at predawn, after sunset, or when the sky was overcast (Figure 14).

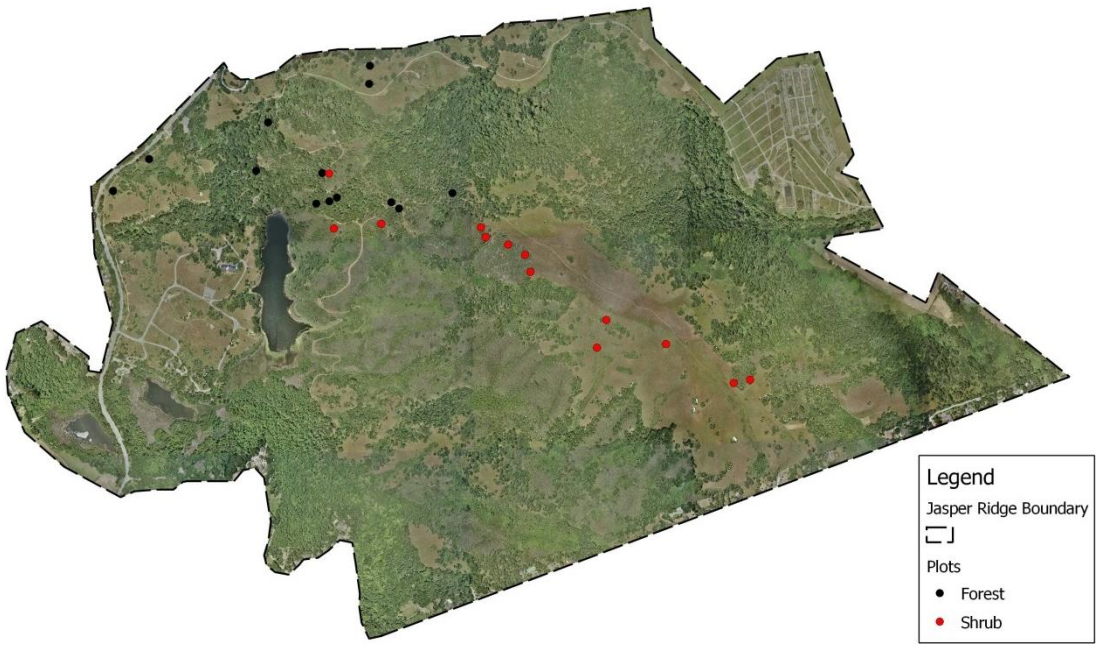

Figure 13. Plot distribution in Jasper Ridge site.

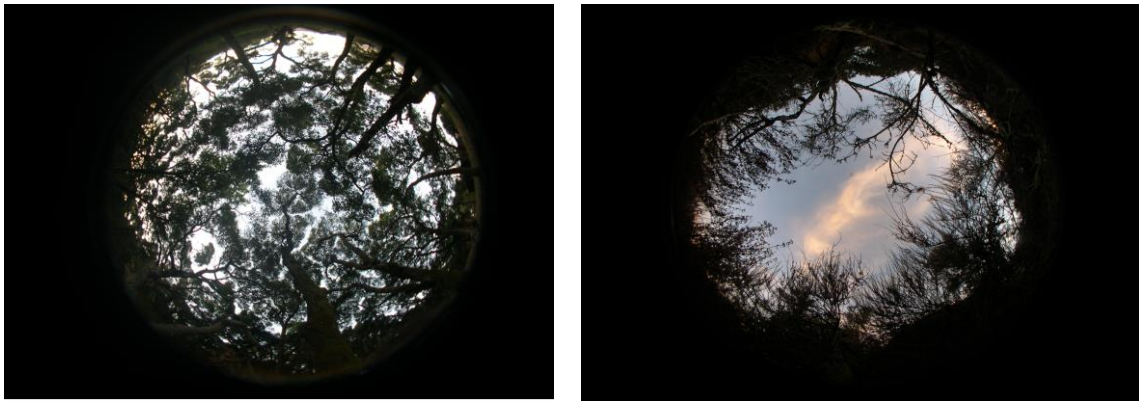

Figure 14. HP (RGB composite) from a forest (left) and shrub plot (right).

Before computing the GF, The Fiji freeware software (http://fiji.sc/Fiji) served to carry out a series of pre-processing steps in the HP. The circular limit in the HP marked the horizon beyond which the HP was black and therefore excluded from the analysis. From the true color HP, only the blue band was selected since the vegetation canopy in the blue appears darker than in the red or green bands, and hence it has a better contrast with the sky (Frazer et al. 2001; Gonsamo et al. 2010). In addition, large canopy openings cause a blooming effect with very high digital values in the red and green that mostly disappear in the blue (Leblanc et al. 2005).

One of the backwards of the digital HP is the poor sharpness in comparison with film cameras according to Frazer et al. (2001). For this reason, digital HP appears blurred, and many of the fine canopy structural details are poorly defined. This was especially evident in zones with high contrast: outside edges of stems, branches, shoots and the inside edges of canopy gaps. To minimize this effect and enhance small gaps within dense foliage, we applied an edge enhancement 3x3 filter [-1 -1 -1;-1 12 -1;-1 -1 -1] which was proposed by Kucharik et al. (1997).

The final pre-processing step is to define the optimal threshold value in the HP to produce a binary image separating the vegetation canopy from the sky, which is critical to calculate GF (Jonckheere et al. 2004). Jonckheere et al. (2005) tested several automatic threshold algorithms and concluded that clustering performed the best compared to histogram, entropy, object-attribute, spatial and local based methods. Bearing this in mind, we chose as optimal the average threshold value from the four clustering algorithms within Fiji software: Isodata (Ridler and Calvard 1978), Otsu (Otsu 1979), Intermode (Prewitt and Mendelsohn 1966) and IsodataMod (http://fiji.sc/Auto_Threshold#Default) (Figure 15). Clustering methods initially segments the histogram of the color image into two parts, using a starting threshold value for the vegetation canopy and sky classes ($T^0=2^{p-1}$; i.e. half of the maximum dynamic range). The distances to the center of each class are computed and each pixel is assigned to its closest class based on any of the four algorithms above (Jonckheere et al. 2005). A new threshold (T^1) is then computed and the process is repeated interactively, until the threshold value remains constant. After the clustering process, vegetation canopy/sky pixels were converted to black/white (0/1) if the digital value was below/above the optimal threshold and then, the binary HP were exported to .bmp files. All this process was carried out automatically using a mixed Java and Fiji script. The resulting HP were all visually inspected and compared to its original HP to search for any cases of anomalous clustering classifications.

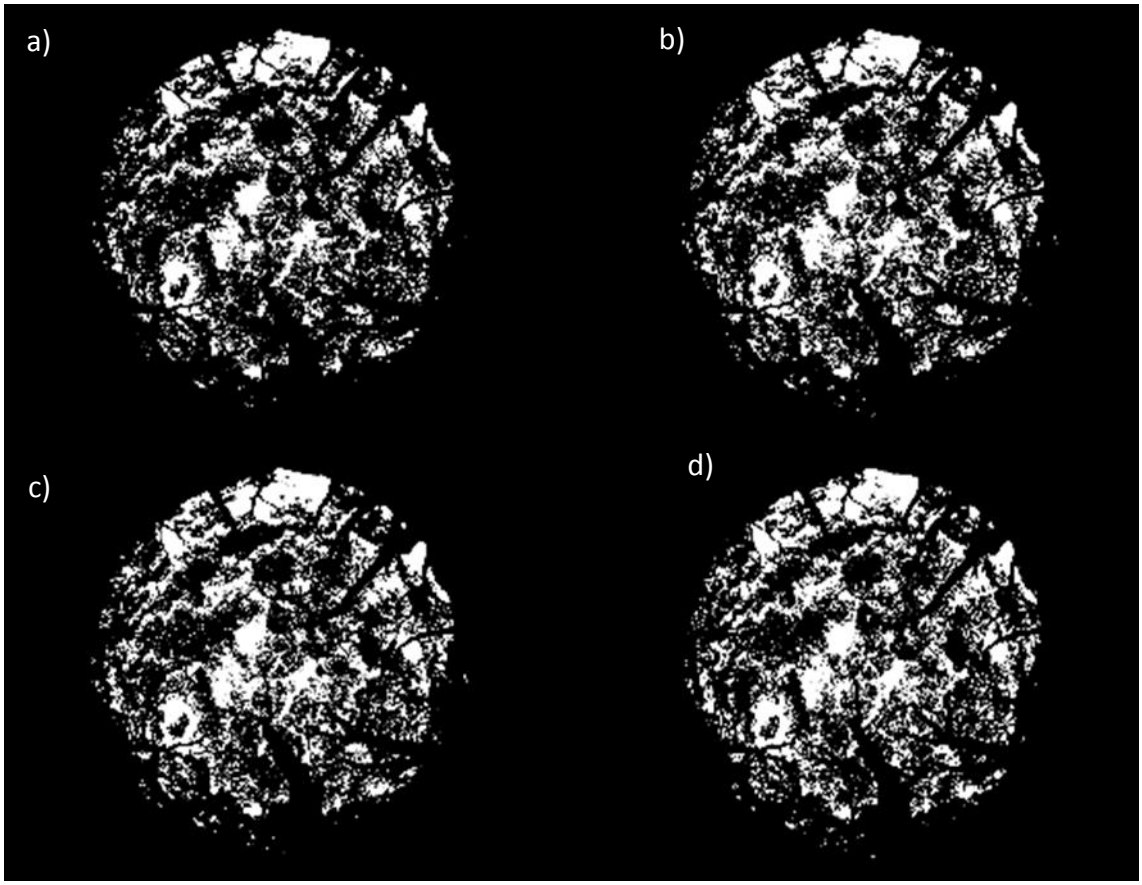

Figure 15. Binary HP after applying Isodata(a), Otsu (b), Intermodes (c) and IsodataMod (d) clustering threshold algorithms to HP in Figure 14 left.

2.4.2 Description of the ALS sensor, data collection and pre-processing

The ALS data was supplied by the Carnegie Airborne Observatory (CAO) from Stanford University (http://cao.ciw.edu/). The airborne system flown, called CAO Beta, combined three instruments into a single package: 1) the Jet Propulsion Laboratory´s Airborne Visible/Infrared Spectrometer (AVIRIS); 2) a small footprint full waveform ALS; and 3) a global positioning system-inertial measurement unit (GPS-IMU) (Dahlin et al. 2011). The AVIRIS, which was not used in this study, samples reflected sunlight in 10 nm increment across the 380-2510 nm range. The GPS-IMU subsystem provides the 3D positioning and altitude data for the sensor package on board the aircraft, allowing for highly precise and accurate projection of ALS (and AVIRIS) observations to the ground. The ALS instrument was a customized version of the Optech ALTM-3100EA (http://www.optech.ca/prodlatm.htm) with a new waveform digitization system (Table 2). This ALS can provide the data in two modes: discrete-return and full waveform (Asner et al. 2007; Mallet and Bretar 2009).

Table 2. The ALS instrument, ALTM-3100EA, in the CAO Beta.

Wavelength	1064 nm
Pulse repetition frequency	programmable up to 100 kHz
Scan angle	programmable up to 44 °
Scan frequency	programmable up to 70 Hz
Laser beam divergence	0.56 mrad (1/e)
Intensity capture	12 bit dynamic range
Waveform digitization	up to 44 slices or elevation per laser shot

The CAO Beta flew over Jasper Ridge and the surrounding area in August of 2007 at an average 2700 m above ground level, capturing eighteen flight lines and covering 5.93 km^2. The ALS data was provided as discrete-return laser points and each pulse was labeled either first, intermediate, last or single. It operated at 33 kHz pulse repetition frequency and 36° scan angle. This configuration rendered a laser spot spacing of 0.5 m, a footprint of 1.5 m and an average laser pulse density of 5 points/m^2. The integration of the GPS-IMU and the ALS determined the 3-D location of the laser returns that was projected into UTM WGS84 (Dahlin et al. 2011).

An exploratory analysis of the ALS data demonstrated the existence of replicated laser returns elements in the point cloud with the exact same 3D position and intensity values. This issue happened systematically along the border of perfect tiles. More than probably CAO pre-processed the data subdividing them in such tiles with an overlapping buffer for each one and then merged all tiles back in a single dataset. Since this issue would affect the ALS metrics that relate to GF, a Matlab script identified these identical replicated laser returns and removed them from the point cloud. This filtered ALS dataset contained a final average laser pulse density of 4 points/m^2.

The absolute and relative (between flight lines) accuracies of the ALS 3D-models depend on several error sources: 1) the alignment between the laser, the aircraft and navigation system; 2) the accuracy in range determination; 3) vibration and oscillation on the scanner mirror; and 4) other problems in the GPS-IMU system (Huising and Gomes Pereira 1998). These errors cause planimetry (X and Y) and altimetry (Z) displacements that impact the extraction of meaningful information and hence the quality of the final product (Lee et al. 2007). Several methods have been proposed to quantify these displacements in order to match the overlapping areas between flight

lines: 1) least squares between point clouds (Maas 2002); 2) point cloud object segmentation of lines or surfaces (Pfeifer et al. 2005); and 3) point cloud intensity images (only X and Y) (Vosselman 2002).

An adaptation of the methods proposed by Pfeifer et al. (2005) quantified the Z displacement between flight lines in this study. Thirty-two flat surfaces, such as cross roads or flat roofs, of at least 25 m^2 and a minimum of 50 laser pulses that felt on overlapping areas between flight lines were extracted from the point cloud. The average all Z laser pulses within each surface for each flight line was computed and their difference determined their Z displacement (Table 3). The flight lines demonstrated to be inter-calibrated since the absolute average Z difference between flight lines of all surfaces was 0.02 m. In addition, the standard deviation of all surfaces in Table 3 provided an average of 0.09 m that verified that all the surfaces were actually flat in either flight line.

Table 3. Surface identifier (ID), flight line number (FL), Z average ($\bar{Z}$), Z standard deviation (σ) and $\bar{Z}$ absolute difference between FL.

ID	FL_1	$\bar{Z}_1$(m)	$\sigma(Z)_1$(m)	FL_2	$\bar{Z}_2$(m)	$\sigma(Z)_2$(m)	$\lvert\bar{Z}_1 - \bar{Z}_2\rvert$(m)
01	26	69.02	0.10	28	69.09	0.11	0.07
02	26	69.17	0.11	28	69.24	0.12	0.07
03	28	73.55	0.09	29	73.52	0.11	0.03
04	28	73.41	0.10	29	73.38	0.09	0.02
05	28	73.33	0.09	29	73.33	0.09	0.00
06	28	73.18	0.10	29	73.17	0.10	0.01
07	28	73.09	0.09	29	73.10	0.09	0.01
08	28	72.98	0.09	29	73.00	0.10	0.02
09	28	72.89	0.10	29	72.90	0.10	0.01
10	28	72.77	0.10	29	72.77	0.07	0.00
11	29	71.94	0.09	30	71.94	0.11	0.00
12	29	71.85	0.10	30	71.85	0.10	0.01
13	29	71.70	0.08	30	71.74	0.10	0.04
14	29	71.62	0.11	30	71.63	0.10	0.02
15	29	71.52	0.10	30	71.54	0.10	0.02
16	30	70.79	0.08	31	70.78	0.11	0.01
17	30	70.68	0.09	31	70.65	0.08	0.02
18	30	70.59	0.09	31	70.59	0.08	0.00
19	30	70.52	0.09	31	70.50	0.09	0.02
20	30	70.39	0.09	31	70.38	0.09	0.02
21	30	70.29	0.09	31	70.31	0.12	0.02
22	30	70.17	0.11	31	70.20	0.11	0.04
23	30	70.04	0.08	31	70.08	0.11	0.04
24	30	69.94	0.08	31	69.98	0.11	0.04
25	30	69.85	0.09	31	69.88	0.10	0.03
26	30	69.73	0.11	31	69.78	0.10	0.05

27	30	69.62	0.10	31	69.65	0.08	0.03
28	31	69.13	0.10	33	69.07	0.10	0.06
29	31	68.93	0.08	33	68.89	0.09	0.04
30	31	68.80	0.10	33	68.80	0.08	0.00
31	31	68.70	0.10	33	68.69	0.10	0.01
32	31	68.62	0.08	33	68.59	0.09	0.03

A method based on Lee et al. (2007) evaluated in this work the X and Y displacement between objects sampled in overlapping flight lines. Triangular shaped roofs were manually identified and isolated from the rest of the point cloud (Figure 16). Only three roofs were found for this validation exercise due to the absence of these human-made elements in the study site. For each roof flank, a planar surface was fitted using a least square approach to obtain its equation and its vector's director (Figure 17). This process was done for the two flight lines data in each selected roof. The intersection of these two planes provided the equation which describes the position of the roof top. Therefore, the distance between the roof top in each flight line measured the X and Y displacement (Table 4). This distance averaged 0.15 m which agreed with expected error presented by the ALS system manufacturer (Optech 2013), as also did the 0.02 m error in Z. In addition, an ALS intensity image with 0.5 m pixel size was generated using the nearest neighbor algorithm and later a filter proposed by Lee (1980) was applied to correct the noise. Coordinates XY from points in road cross and corners were collected from an orthophoto and the intensity image to ensure co-registration between the ALS and the plot locations. The Euclidian distance was computed between coordinates and the average distance obtained. This difference was considered the displacement between the ALS and the orthophoto. The RMSE was 2.07 m for 17 ground control points identified in both images (Table 5).

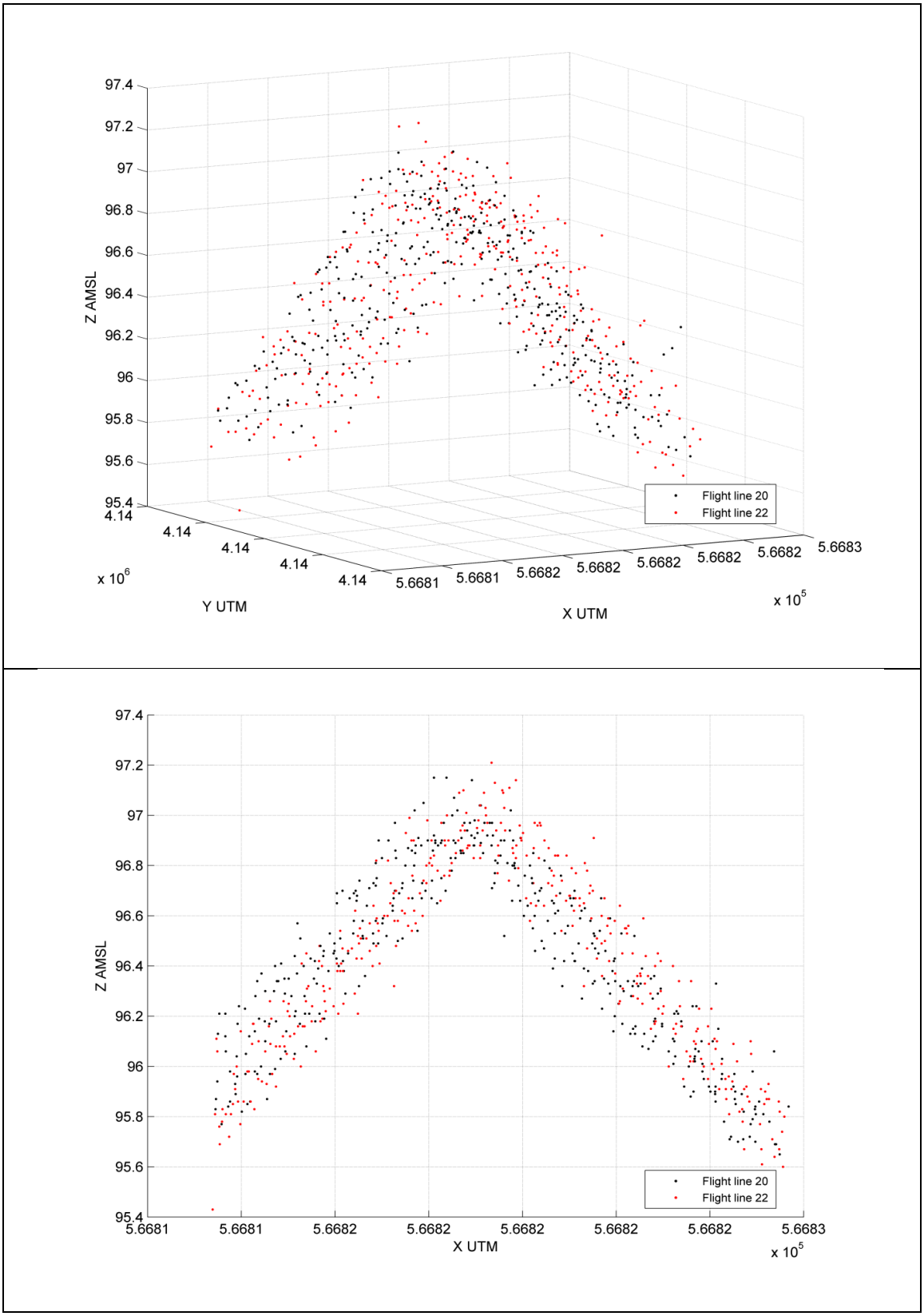

Figure 16. One of the roofs isolated from the point cloud (first row), XZ projection for the roof (second row).

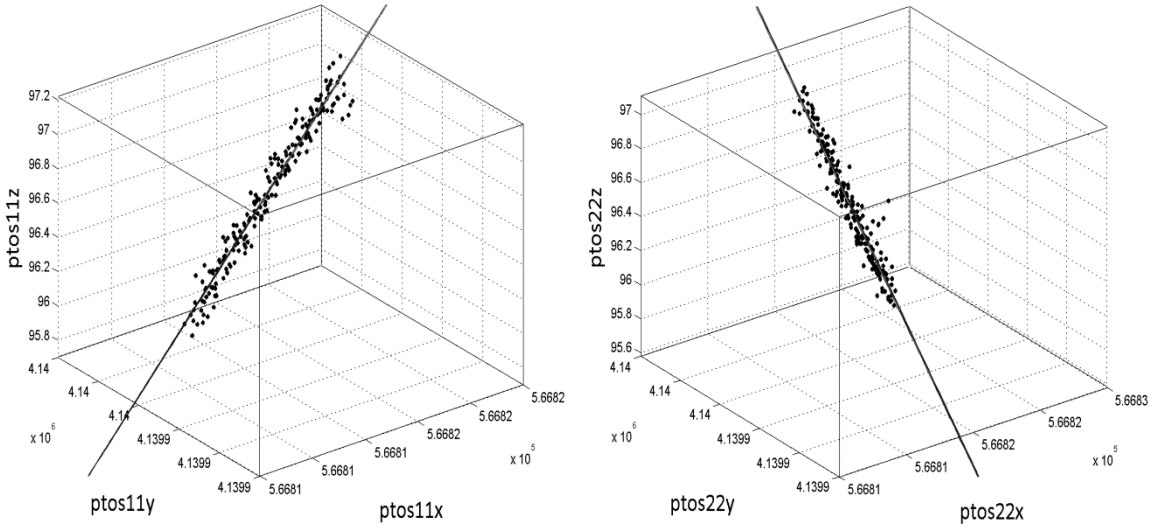

Figure 17. Planar surface fitted to laser pulses on the East and West side of the roof in Figure 16.

Table 4. Roof identifier (ID), flight lines number (FL), angle between vectors (α) and distance between intersection lines (XY).

ID	FL	α (°)	XY (m)
1	20 & 22	4.24	0.23
2	26 & 28	4.47	0.07
3	20 & 22	0.51	0.13
$\overline{XY}$			0.15

Table 5. Coordinates taken from orthophoto and ALS intensity image.

ID	Orthophoto X	Orthophoto Y	Intensity X	Intensity Y	Distance (pixel)
1	4077.3283	592.53950	4072.1743	590.90750	5.4062
2	1338.3894	1092.9101	1332.4983	1092.4959	5.9056
3	1376.9107	2484.1009	1373.5022	2483.0013	3.5815
4	1512.9478	2554.2244	1508.9037	2553.2985	4.1487
5	1470.7508	2597.8711	1467.5238	2596.0837	3.6889
6	1360.3390	2466.9829	1355.6331	2465.5315	4.9246
7	5272.3446	2739.5181	5270.5407	2737.7142	2.5511
8	5646.9932	3026.7469	5647.2355	3026.4016	0.4218
9	5997.9831	3310.8618	5995.5098	3309.3478	2.8999
10	7182.2854	2155.5778	7177.5610	2155.3631	4.7293
11	7262.5266	2130.2138	7256.9431	2129.3548	5.6492
12	4082.4929	590.42350	4078.3698	587.05630	5.3233
13	1758.0738	540.99150	1754.4230	539.81040	3.8371
14	2224.7918	437.88800	2223.5302	436.57990	1.8173
15	5652.9904	1343.1522	5646.9774	1344.8702	6.2536
16	1654.4971	2376.7907	1649.0211	2377.2202	5.4928
17	1835.2868	2545.5954	1831.4280	2545.2599	3.8734
	Average				4.1422

Once the replicated ALS laser pulses were removed and the data accuracy was verified, the algorithm proposed by Streutker and Glenn (2006) and implemented in the BCAL LiDAR Tools software (BCAL 2013) classified the laser pulses in two categories: ground and vegetation canopy (non-ground). It was assumed that at least one laser pulse every 4 m fully penetrated de canopy with which an initial ground surface was created using natural neighbor interpolation. All laser pulses on or below 0 m were reclassified as ground and a maximum 50 m height was allowed. With these additional ground points, a new ground surface was re-interpolated and the process continued for fifteen interactions. Ninety nine percent of the laser pulses converged after six iterations. Outliers were found near the edges of the dataset which belonged to electrical power lines. Heights above this ground surface were calculated for all remaining non-ground points.

In addition to the X, Y and Z position of the laser pulses, this ALS system recorded the intensity of the returns, which is a measure of the amount of energy reflected back to the sensor (García et al. 2010). Intensity varies according to the sensor-target distance (range) among other factors (Coren and Sterzai 2006). As demonstrated in several studies before (Donoghue et al. 2007; García et al. 2010; Starek et al. 2006), this effect can be easily eliminated to provide values equivalent to the intensity that would have been recorded if all points were sampled at the same range (equation 9).

$$I' = I \frac{R^2}{Rs^2} \tag{9}$$

where I' is the normalized intensity, I is the raw intensity value, R is the range and Rs is the standard 1000 m range used here.

2.4.3 Calculation of GF from HP

The Gap Fraction Analysis tool within the CIMES freeware package (http://jmnw.free.fr/) processed GF from the binary HP using a batch script in MS-DOS. The program needed 1) the coordinates (X,Y) of three points on the horizon in order to determine the limits of the HP; 2) the geometric distortion of the fisheye lens; 3) the magnetic declination to rotate the image to its true North; 4) the number of skymap divisions in the φ and θ angles required to build the GF statistics; and 5) the φ range

needed to calculate GF. The number of skymap divisions in the HP must be a compromise of the following criteria for a reliable GF estimate: 1) low enough so that GF statistics are meaningful and also invertible; 2) high enough so that the assumption of randomness of leaf distribution within a skymap segment remains valid. Taking these two factors into account, the minimum number of pixels in any skymap segment should contain at least ten pixels (Lang and Yueqin 1986). Therefore, we decided on 9 φ rings and thirty six θ radial lines to divide the skymap in a total of 324 segments, which assured at least 700 pixels for each one (Figure 18). Instead of the full hemispherical range, 30-60° φ angles were preferred to calculate GF following Gonsamo et al. (2010) and Leblanc et al. (2005). The intense scattering of diffuse light at very low φ angles specially affects digital HP (Frazer et al. 2001). In addition, very low φ angles present higher errors due to the foliage being poorly sampled and very high ones also do due to being close to the edge of the lens (Jonckheere et al. 2012). The non- parametric Kruskal Wallis test evaluated if there was a statistical difference in terms of GF between the three dates in which the HP data were acquired.

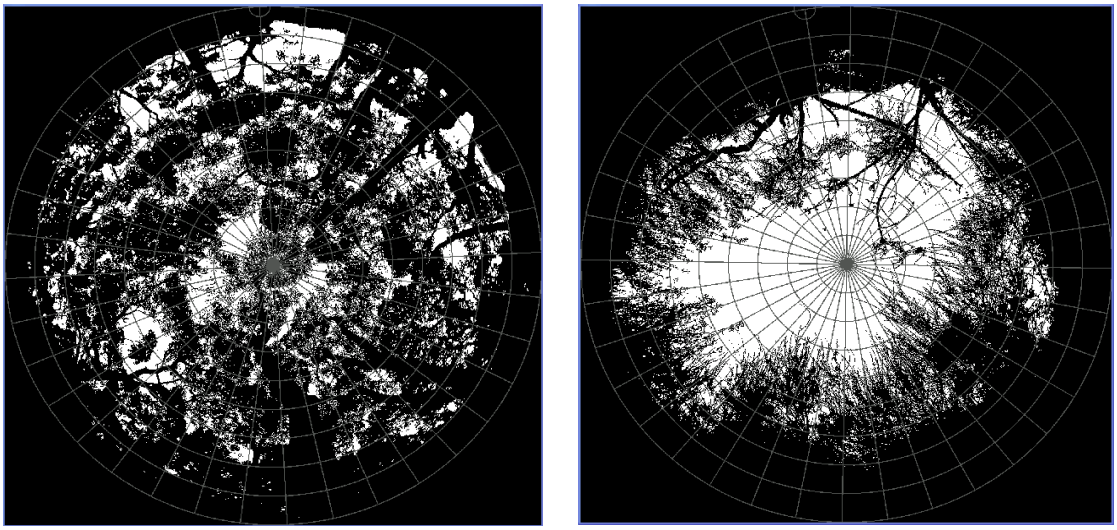

Figure 18. HP skymap grid sampling composed by 9 φ annuli and 36 θ segments for forest plot (left) and shrub plot (right).

2.4.4 Calculation of ALS metrics to estimate GF

The plot radius size of the ALS data that needs to be considered to fit the GF calculated from HP depends on the canopy structure (Riaño et al. 2004). Several authors have extracted ALS data with different plot radii in diverse ecosystems (Morsdorf et al. 2006; Riaño et al. 2004; Solberg 2010; Solberg et al. 2009), but no one analyzed a mixture of forest and shrub cover like our study site in Jasper Ridge. For this reason, we tested 10-20 m radius sizes, at 1 m increments. For each plot radius size, all the laser pulses were extracted and separated into ground and canopy returns at a 1.3 m threshold as suggested by Jonckheere (2005), which ensures that all laser returns above such height come from the vegetation canopy and avoid ground pulses coming from steep slopes (Figure 19).

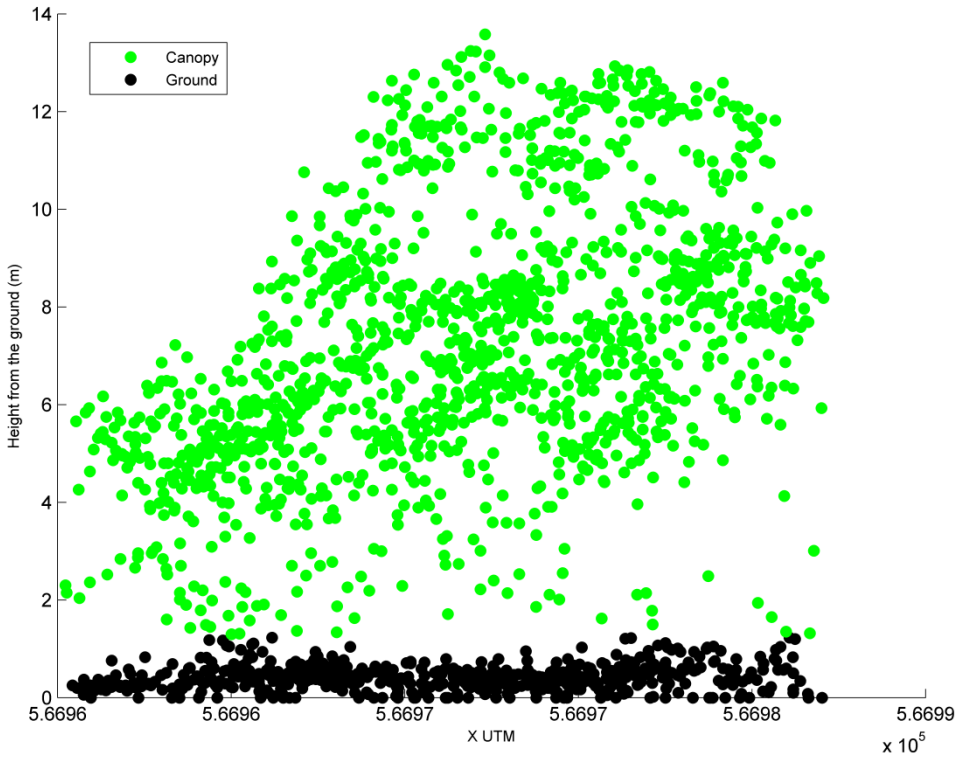

Figure 19. Forest plot point cloud classified according height from the ground.

Several ALS metrics were computed for each specific plot radius size to relate with GF, which is complementary of FC: 1) laser returns percentiles; 2) laser penetration rates; and 3) intensity ratios. The premise of these metrics is that the interaction of the vegetation canopy with the laser pulses emitted from a ALS system

can be considered in some ways analogous to the one with the direct beam solar radiation (Hopkinson and Chasmer 2007). Riaño et al. (2004) assumed and tested an empirical relationship between GF and the 50^{th} 75^{th} and 95^{th} percentiles of all laser pulse returns, their average (AverageH) and the maximum (MaxH) plot height. ALS metrics based on laser penetration rate assume that GF may be directly inferred by a pulse return ratio of the number of canopy to total returns (Hopkinson and Chasmer 2009; Morsdorf et al. 2006; Solberg 2010) (Equation 10). Moreover, Morsdorf et al. (2006) also tested last and single returns (Equations 11 and 12). However, these methods do not account for the fact that an emitted pulse can encounter only a small area of the canopy foliage and there might be insufficient reflected energy to actually record a canopy first return. Bearing these this in mind, all returns types provide an increased in sampling density to alternatively predict GF (Equation 13) (Hopkinson and Chasmer 2007; Hopkinson and Chasmer 2009; Morsdorf et al. 2006; Riaño et al. 2004; Sasaki et al. 2008). Meanwhile, Sasaki et al. (2008) proposed to combine first and single returns (Equation 14) while Solberg et al. (2009) weighted the first and last returns in a different way than the single ones (Equation 15).

$$GF_F = \frac{\sum RGround_F}{\sum RTotal_F} \tag{10}$$

$$GF_L = \frac{\sum RGround_L}{\sum RTotal_L} \tag{11}$$

$$GF_S = \frac{\sum RGround_S}{\sum RTotal_S} \tag{12}$$

$$GF_A = \frac{\sum RGround_A}{\sum RTotal_A} \tag{13}$$

$$GF_{c1} = \frac{\sum RGround_A}{\sum RTotal_F + \sum RTotal_S} \tag{14}$$

$$GF_{c2} = \frac{\sum RGround_S + 0.5(\sum RGround_F + \sum RGround_L)}{\sum RTotal_S + 0.5(\sum RTotal_F + \sum RTotal_L)} \tag{15}$$

where RGround is the frequency of the laser first (F), last (L), single (S) or all (A) pulse returns classified as ground while RTotal is the frequency of vegetation canopy plus ground F, L, S or A ones.

ALS metrics based on laser pulse intensity have implicitly estimated GF from full waveform ALS data (Lefsky et al. 1999). In this situation, the strength of the returned signal from within or below the vegetation canopy is related to the transmittance through it (Lefsky et al. 2002). At near nadir scan angles, we can assume that the total reflected energy from the ground and vegetation canopy is a function of the total emitted energy, whereas the reflected energy from the ground is a similar function of the transmitted energy (Hopkinson and Chasmer 2007) (Equation 16). However, this model does not explicitly account for potentially different probabilities associated with the transmission loss in and out of the canopy. For discrete laser return data, first and single returns generally incurs in a similar proportion of transmission loss, assuming uniform transmission losses per unit path length travelled. Instead, intermediate or last returns are a reflected component of the residual energy left over after a previous return was reflected from a surface encountered earlier in the travel path of the emitted pulse. According to Beer's law, Hopkinson and Chasmer (2007) proposed the square root of these intermediate or last returns to account for the likelihood of two way transmission losses and the combination of all return types to build Equation 17.

$$GF_I = \frac{\sum IGround_A}{\sum ITotal_A} \tag{16}$$

$$GF_{IC} = \frac{\frac{\sum IGround_S}{\sum ITotal_A} + \sqrt{\frac{IGround_L}{ITotal_A}}}{\frac{\sum ITotal_F + \sum ITotal_S}{\sum ITotal_A} + \sqrt{\frac{\sum ITotal_I + \sum ITotal_L}{ITotal_A}}} \tag{17}$$

where IGround is the intensity of the laser, first (F), intermediate (I), last (L), single (S), or all (A) pulse returns classified as ground while ITotal is the intensity of vegetation canopy plus ground pulse returns.

2.4.5 Computing the GF models from ALS metrics

Bivariate linear regression technique evaluated the relationships between the predictor ALS metrics and GF from HP. This approach was preferred over multivariate to avoid the risk of multicollinearity and inflated correlations from a high number of predictor variables, especially when all these metrics are derived from the same source (Thomas et al. 2011). In addition, the goal was to find metrics that were directly related to GF, instead of choosing complex combinations which could render a very site specific model (Xu and Zhang 2001). The regressions models were tested they satisfied the bivariate linear regression assumptions. For this reason, the Shapiro-Wilk test was applied to the models residuals. Residuals are normally distributed when the Shapiro-Wilk statistic (W) is near to 1 and p-value is not significant (Royston 1995). Besides this, the root mean square error (RMSE equation 18) goes down as R^2 goes up when comparing regression models, as in our case, with the same dependent variable and estimation period. Hence, this fact points to the model with the highest adjusted R^2, which will have the lowest RMSE and finally preferred because of it (Korhonen et al. 2011).

$$RMSE=\sqrt{\frac{\sum\left(GF_i y_i - \widehat{GF_i}\right)^2}{n}} \tag{18}$$

where n is the number of samples, and GF_i and $\widehat{GF_i}$ are each measured and estimated GF by the model.

2.5 Vegetation canopy CI from ALS

2.5.1 Calculation of CI from HP

In order to validate CI estimates from ALS, CI was computed with CIMES package using the steps described in the section 2.3.1 from the same binary HP images pre-processed with Fiji freeware in section 2.4.1. Since HP were acquired at three different times of the year (see section 2.4.1), we tested if there was a significant difference in CI in between dates.

The whole process to generate CI was automatized running a batch MS-DOS script. CIMES calculates first the gap size distribution for each binary HP and saves it as an ASCII file. This file contains a series of three lines: (1) each 1° φ interval; (2) the length in pixel units of the sequence of white (sky) pixels; and (3) length also in pixel units of the black (canopy) pixels. The CLMPML program within CIMES generates a CI value integrating the lengths of white/black pixels for the selected φ range. A 30-60° φ range for the HP was applied here since this range captures better the overall average CI according to Gonsamo et al. (2010) and Gonsamo and Pellikka (2009). In addition, avoiding pixels in the lower and upper φ angles reduces the errors in the estimation of gap size distributions (Frazer et al. 2001). More specifically, lower φ pixels represent a very tiny section of the canopy and upper ones occupy a very broad section with mixed values due to the light scattering (Gonsamo and Pellikka 2008).

2.5.2 Estimation of CI from ALS metrics and ground laser returns ALS images (GRI)

Most of the few efforts to estimate CI from ALS data relate empirically the non-random leaf spatial distribution with simple ALS metrics. For example, Thomas et al. (2011) predicted CI in a boreal mixed wood forest testing the mean, standard deviation, and 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of either all or only vegetation canopy laser returns within a plot. In addition, Thomas et al. (2011) also calculated the mean, median and standard deviation of only the selection of laser pulses higher than the mentioned percentiles in the vegetation canopy ones. Bivariate regression models were proposed to estimate CI using each of these ALS metrics as predictor variables and their performance was evaluated in terms of the best RMSE.

Alternatively to ALS metrics, ground laser returns ALS images (GRI) served to apply here the current methods available that compute CI from sequences of canopy/sky pixels using HP or others optical instruments like TRAC (Chen and Cihlar 1995; Pielou 1962). We generated circular binary GRI of 10-20 m radii by classifying the laser returns into canopy (0) or ground (1) based on 1.3 m height threshold discussed earlier in section 2.4.4 (Figure 20). Several pixels sizes were tested to produce the GRI, starting from the ALS 0.5 m nominal point spacing to 1 m at 0.10 m steps. A pixel size of 0.80 m was chosen, which minimizes the number of pixels with multiple laser returns and the ones with no laser returns. In case a pixel contained more than one

laser return, it was considered ground as long as there was at least one ground laser return. Once the GRI were generated we computed the CI applying the spatial segregation coefficient (PCS).

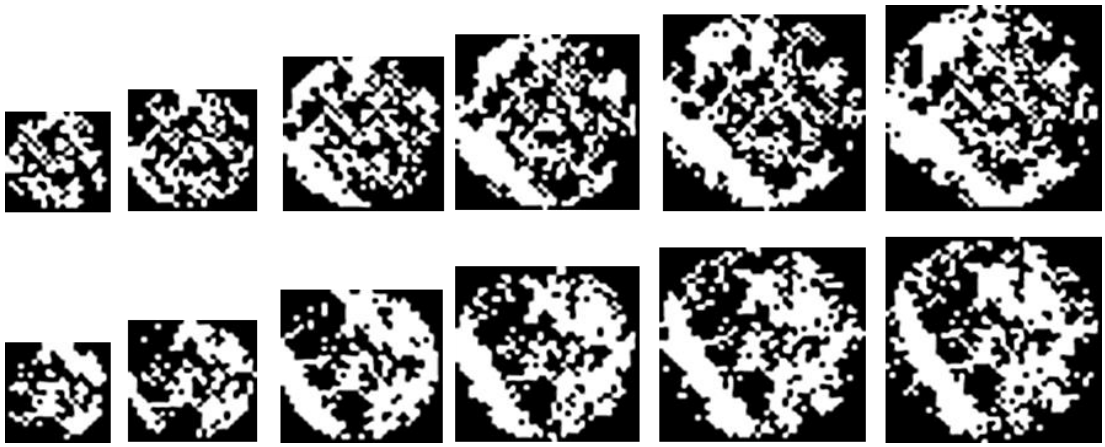

Figure 20. GRI from 10 to 20 radii step 2 m. First row, forest plots. Second row, shrub plots.

2.5.3 Pielou's coefficient of segregation (PCS)

Pielou (1962) proposed a coefficient to estimate the degree of segregation to determine the relationship between two plant species growing together. The method is based on recording the sequences in which a given species occurs in a sampling transect for a two-species population. Subsequently, the observations are compared to those to be expected from an unsegregated population of the same numerical composition, providing a measure of the amount of randomness of each species in respect to the other (Pielou 1962). The probability of encountering species A and B are given by a and b respectively, and for a two-species population it is clear that a+b=1. Provided that the two populations are unsegregated, that is, randomly distributed; it follows that with 95% probability:

$$PCS = \hat{a} + \hat{b} = \frac{1}{\mu_a} + \frac{1}{\mu_b} = 1 \pm 1.96\sqrt{s_a^2 + s_b^2} \quad (19)$$

where $\hat{a}$ and $\hat{b}$ are the maximum likelihood estimates of A and B, μ_a and μ_b are their mean length of occurrence and, s_a and s_b are their variances. For PCS<1, PCS=1 or PCS>1 the distribution of the species is clumped, random or uniform, respectively.

This method was adjusted to be applied to HP by Walter et al. (2003) and later by Gonsamo and Pellikka (2009), where the segregation of black and white pixels was computed for each circle with constant φ angle. Similarly to Walter et al. (2003), we adapted the PCS method to the GRI data by computing the segregation of occupied/unoccupied pixels. We applied three variations of the PCS method to the GRI: (1) an analysis of the GRI row by row; (2) a direct application of the PCS-CIMES algorithm to the GRI and (3) the PCS-CIMES algorithm restricted to certain φ angles.

For the analysis of the GRI row by row, a Matlab script was developed to compute an average PCS index from the ones calculated for each row based on the length of sequences of 1 and 0 pixels throughout the GRI, avoiding the outer corner pixels. The second option applied directly the PCS-CIMES algorithm developed for HP to the GRI even though pixel resolution, view perspective and projection of HP and GRI are different. This method applied to GRI therefore assumes a linear relation between φ angle distribution and the plots radii. The GRI were segmented by annular φ ring instead of the row by row analysis of the first case (Figure 21). The PCS-CIMES calculates the gap size distribution files according to details in section 2.5.1 and integrates PCS index value for the whole GRI. The third method takes into account that the maximum ALS scans angle value of 36° to adjust the φ angles in PCS-CIMES. The gap size distribution files were modified so that the φ between 0-90° were replaced and rescaled for a value between 0-18°. This range maintained the order between φ rings for a closer match to the ALS φ angles, but it was fictitious since the actual φ was not given by the data provider. Similarly to ALS metrics, bivariate regression models estimated CI using as predictor variables each of the three proposed PCS methods and their performance was evaluated in terms of the best RMSE.

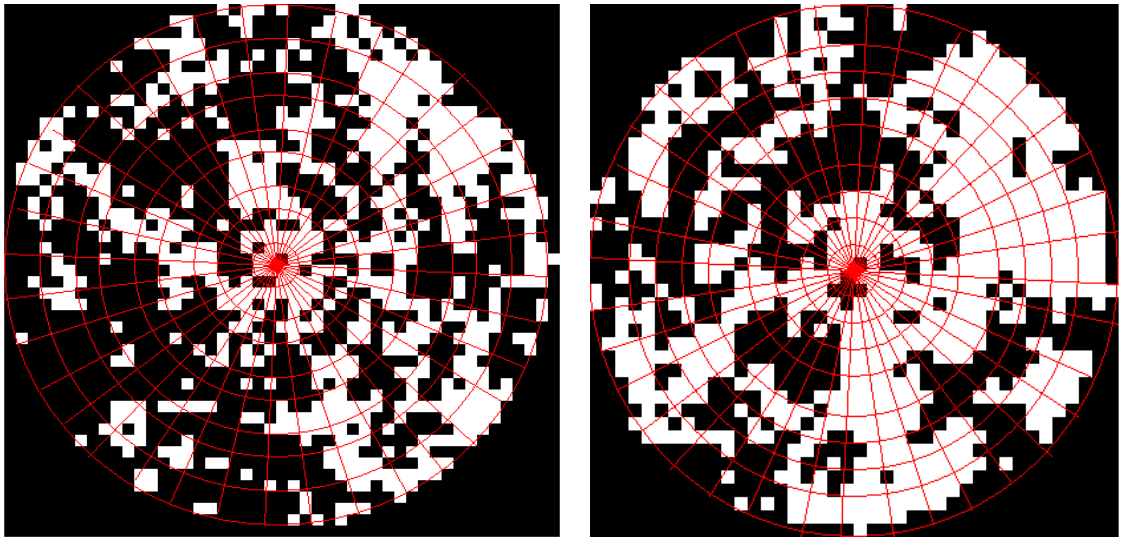

Figure 21. GRI 0.80 m pixel size and 20 m radius for a forest plot (left) and shrub plot (right).

3 RESULTS AND DISCUSSIONS

3.1 Vegetation canopy GF from TLS

Figure 22 shows the differences between the reference θ and φ angular resolution and the computed ones for the R, C and RC gap patterns with GF from 10 to 80% and for noise sigma values ranging from 0 to 14%. Results showed that R and RC gap patterns kept resolution values closer to the reference until sigma reached 6% and over. Instead, C gap pattern showed good resolutions values only for a sigma less than 4%. Above 6% sigma, the R and RC gap patterns had coarser resolutions than the reference values, except for GF data between 10-20%, whereas C gap pattern was the opposite, with finer resolution than its reference.

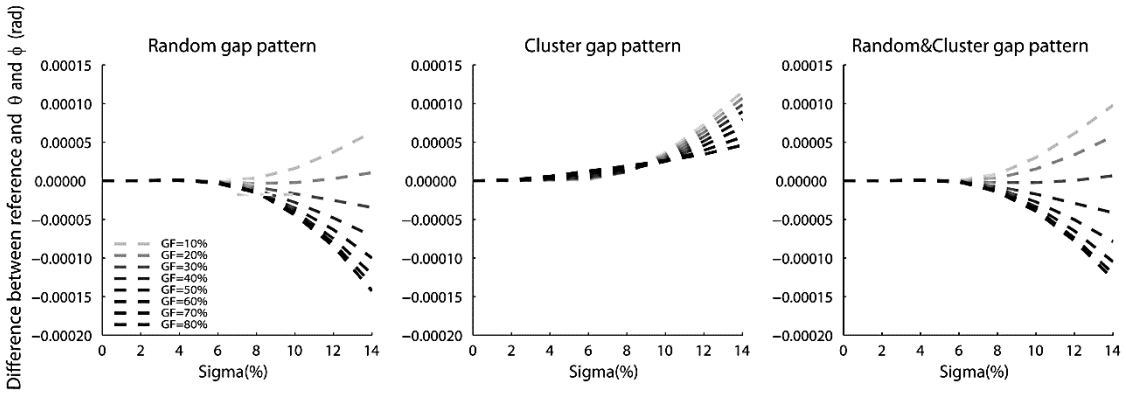

Figure 22. Difference between reference θ and φ angular resolution and the computed ones.

Computed GF showed an average difference with the reference that was close to zero for noise sigma values ranging from 2 to 6% for all the gap patterns simulated. For sigma values between 8 to 14%, average differences were from 2% to a maximum of 5% for R and RC gap patterns, respectively. As for the C gap pattern for the same sigma values, it systematically showed higher differences with values ranging between 2% and 7% (Figure 23). The KW test demonstrated that GF estimates are sensitive to the angular noise of the data for all the gap patterns and GF amounts (p-value<0.05). The test was also carried out to assess the effect of the gap patterns in the amount of GF estimated. The results demonstrated that this variable does not influence the GF calculation as long as the noise remains below 6% of the angular resolution. For higher

sigma noise, the test showed that there were significant differences between GF obtained from different gap patterns (Table 6).

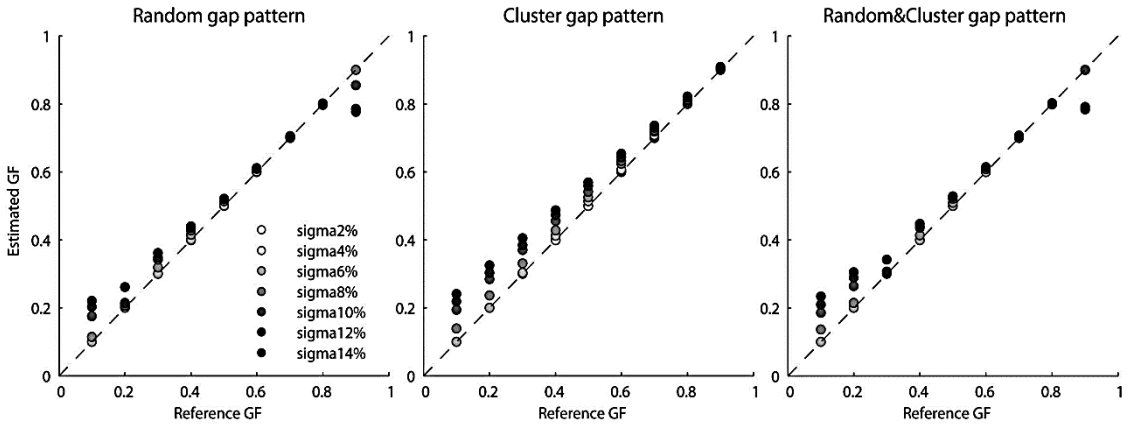

Figure 23. GF comparison for R, C and RC gap patterns simulated with several sigma noises.

Figure 24 shows a comparison between GF obtained from the angular grids and its SimTLS-SHP. Comparison was made only for simulated TLS data with 2-6% sigma values, since this level of noise did not affect GF estimates. The results showed that the best agreement was for the C gap pattern with an average 1% difference, followed by the R and RC gap patterns, presenting around 7% and 8% difference, respectively.

Table 6. P-values significance between R/C/RC gap patterns for KW test carried out over angular GF images with sigma noise from 2 to 14%.

GF10%	GF20%	GF30%	GF40%	GF50%	GF60%	GF70%	GF80%	GF90%
2%(*)	2%(*)	2%(*)	2%(*)	2%(*)	2%(*)	2%(*)	2%(*)	2%(*)
4%(*)	4%(*)	4%(*)	4%(*)	4%(*)	4%(*)	4%(*)	4%(*)	4%(*)
6%(*)	6%(*)	6%(*)	6%(*)	6%(*)	6%(*)	6%(*)	6%(*)	6%(*)
8%(**)	8%(**)	8%(**)	8%(**)	8%(**)	8%(**)	8%(**)	8%(**)	8%(**)
10%(**)	10%(**)	10%(**)	10%(**)	10%(**)	10%(**)	10%(**)	10%(**)	10%(**)
12%(**)	12%(**)	12%(**)	12%(**)	12%(**)	12%(**)	12%(**)	12%(**)	12%(**)
14%(**)	14%(**)	14%(**)	14%(**)	14%(**)	14%(**)	14%(**)	14%(**)	14%(**)

(*) There is no significant difference.

(**) There is significant difference.

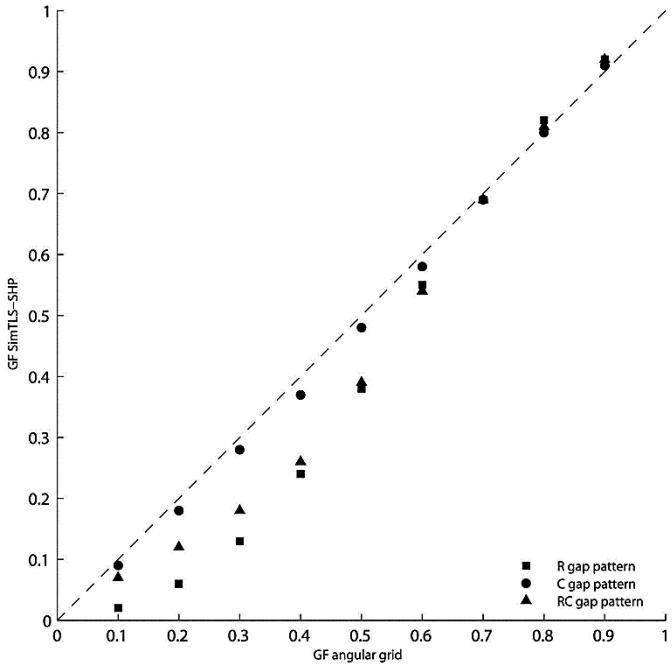

Figure 24.Comparison between GF estimated from an angular grid and SimTLS-SHP with simulations with 6% sigma.

Computation of the angular resolution for the actual TLS tree models yielded an average resolution of 6.3e-04 rad for both the θ and φ, whereas the sigma noise was 5.81% and 3.29% of the angular resolution, respectively (Table 7). GF ranged between 27% and 39%, with an average value of 33.5%. Figure 25 shows an example of an angular grid and its transformation to TLS-SHP. For the subset TLS data from the selected trunks, the results showed a very similar angular resolution of 6.28e-04 rad for both the θ and φ, whereas the sigma noise decreased to 4.43% and 1.27% of the angular resolution, respectively (Table 8). On the other hand, lower φ TLS slices (<4°) showed different resolutions 5.3e-04 rad and 4.4e-04 rad for θ and φ and higher sigma noise of 42.54% and 64.76% of the angular resolution, respectively (Table 9). Moreover, some noisier sectors in the canopy showed angular resolutions which deviated from the average computed values (Figure 26).

Table 7. Computed angular resolution for the TLS data.

Tree ID	Resolution (θ)	Sigma (θ)(%)	Resolution (φ)	Sigma (φ)(%)	GF
Z1P5_SCAN1	6.30e-04	6.66	6.30e-04	3.22	34
Z1P5_SCAN2	6.30e-04	5.81	6.30e-04	3.16	36
Z1P7_SCAN1	6.30e-04	5.65	6.30e-04	2.67	39
Z1P7_SCAN2	6.20e-04	6.26	6.30e-04	3.50	37
Z2P6_SCAN1	6.30e-04	5.29	6.30e-04	3.06	27
Z2P6_SCAN2	6.30e-04	5.64	6.30e-04	2.95	27
Z2P11_SCAN1	6.30e-04	5.38	6.30e-04	3.17	30
Z2P11_SCAN2	6.30e-04	5.81	6.20e-04	4.60	38
AVERAGE	6.30e-04	5.81	6.30e-04	3.29	33.5

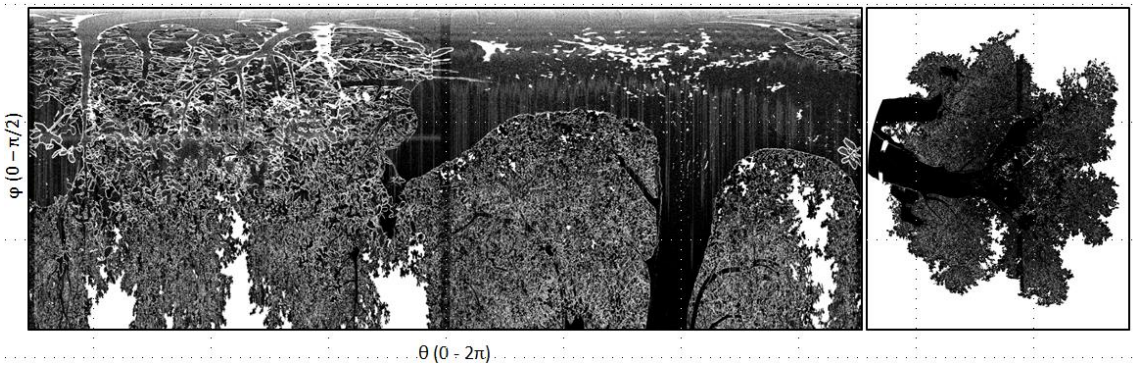

Figure 25. Angular grid GF image (left) and its transformation into TLS-SHP (right).

Table 8. Computed angular resolution for the reference TLS data (tree trunks).

TrunkID	Resolution (θ)	Sigma (θ)(%)	Resolution (φ)	Sigma (φ)(%)
Z1P5_S1	6.27e-04	4.43	6.29e-04	1.29
Z1P5_S2	6.28e-04	4.22	6.29e-04	1.18
Z1P7_S1	6.27e-04	5.07	6.28e-04	1.29
Z1P7_S2	6.28e-04	4.99	6.28e-04	1.27
Z2P6_S1	6.28e-04	4.11	6.29e-04	1.24
Z2P6_S2	6.28e-04	4.63	6.28e-04	1.27
Z2P11_S1	6.27e-04	3.57	6.29e-04	1.53
Z2P11_S2	6.28e-04	4.39	6.28e-04	1.11
AVERAGE	6.28e-04	4.43	6.28e-04	1.27

Table 9. Computed angular resolution for upper φ slices of the tree models.

Tree ID	Resolution (θ)	Sigma (θ)(%)	Resolution (φ)	Sigma (φ)(%)
Z1P5_SCAN1	4.6e-04	47.83	5.2e-04	44.42
Z1P5_SCAN2	5.2e-04	45.24	3.4e-04	91.18
Z1P7_SCAN1	5.2e-04	45.24	4.2e-04	70.67
Z1P7_SCAN2	5.2e-04	43.33	4.0e-04	73.50
Z2P6_SCAN1	5.7e-04	31.27	5.5e-04	37.41
Z2P6_SCAN2	5.6e-04	35.91	4.8e-04	54.89

Z2P11_SCAN1	5.4e-04	56.46	3.7e-04	83.38
Z2P11_SCAN2	5.6e-04	35.06	4.5e-04	62.64
AVERAGE	5.3e-04	42.54	4.4e-04	64.76

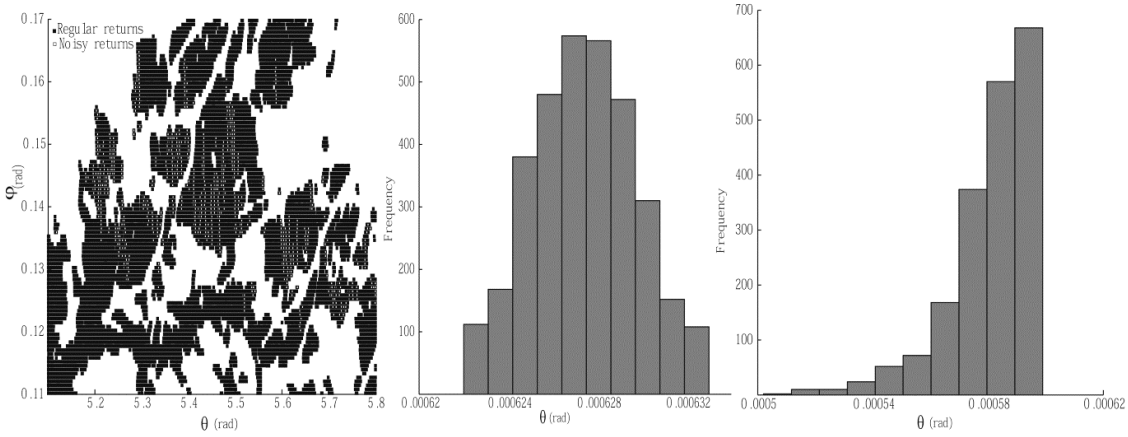

Figure 26. Canopy section (left), histogram for regular points (middle), histogram for noisy points (right).

Figure 27 on the left shows comparison between GF computed from HP and TLS-SHP. The TLS-SHP underestimated GF by an average 4% compared to HP. Figure 27 on the right shows a comparison between GF values computed from the angular grids and values derived from HP and TLS-SHP. The GF from angular grid showed 14% and 17% values higher than computed by HP and TLS-SHP respectively.

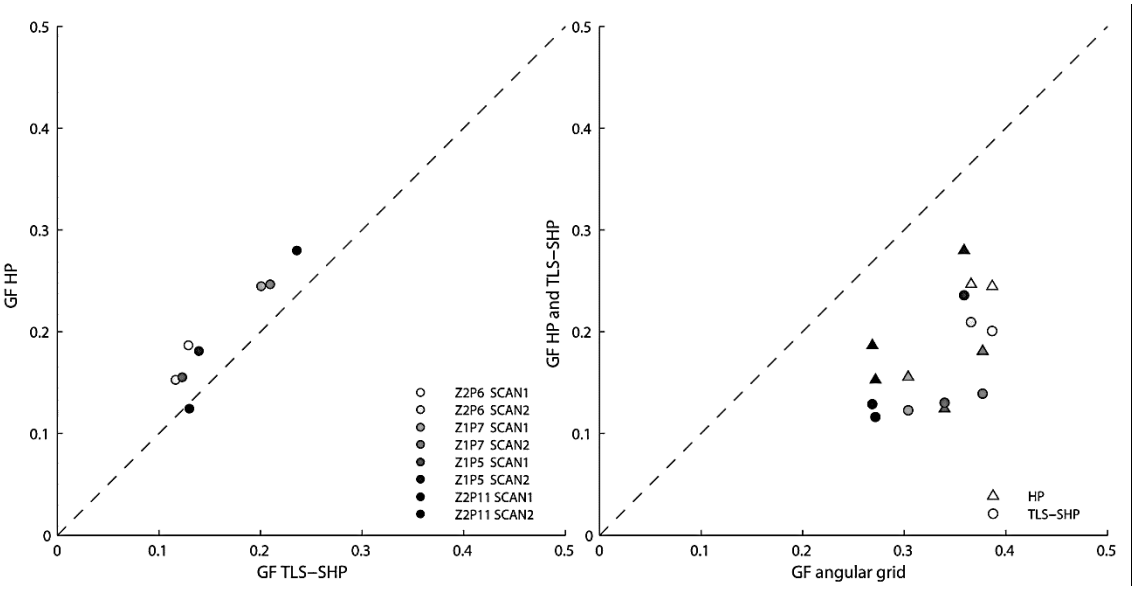

Figure 27. GF comparison between TLS-SHP and HP (left), GF comparison for angular grid and HP and TLS-SHP (right).

The figure 28 shows the GF distribution according the φ range for the eight TLS tree models. The average absolute differences between TLS-SHP and HP was a

maximum of 12% in the lower φ ranges (2.5°), a minimum of 2% in the medium part of the φ range (32.5°) and an average value of 5% for all the angular range. The figure 29 shows the comparison for GF computed from TLS-SHP and HP. The R^2 were 0.88 and 0.89, and statistically significant (p-value<0.05) for the 0-60° and 30-60° φ range respectively. The TLS-SHP underestimated the GF in about 4% and 2% respect to HP as both as in the 0-60° and 30-60° φ range.

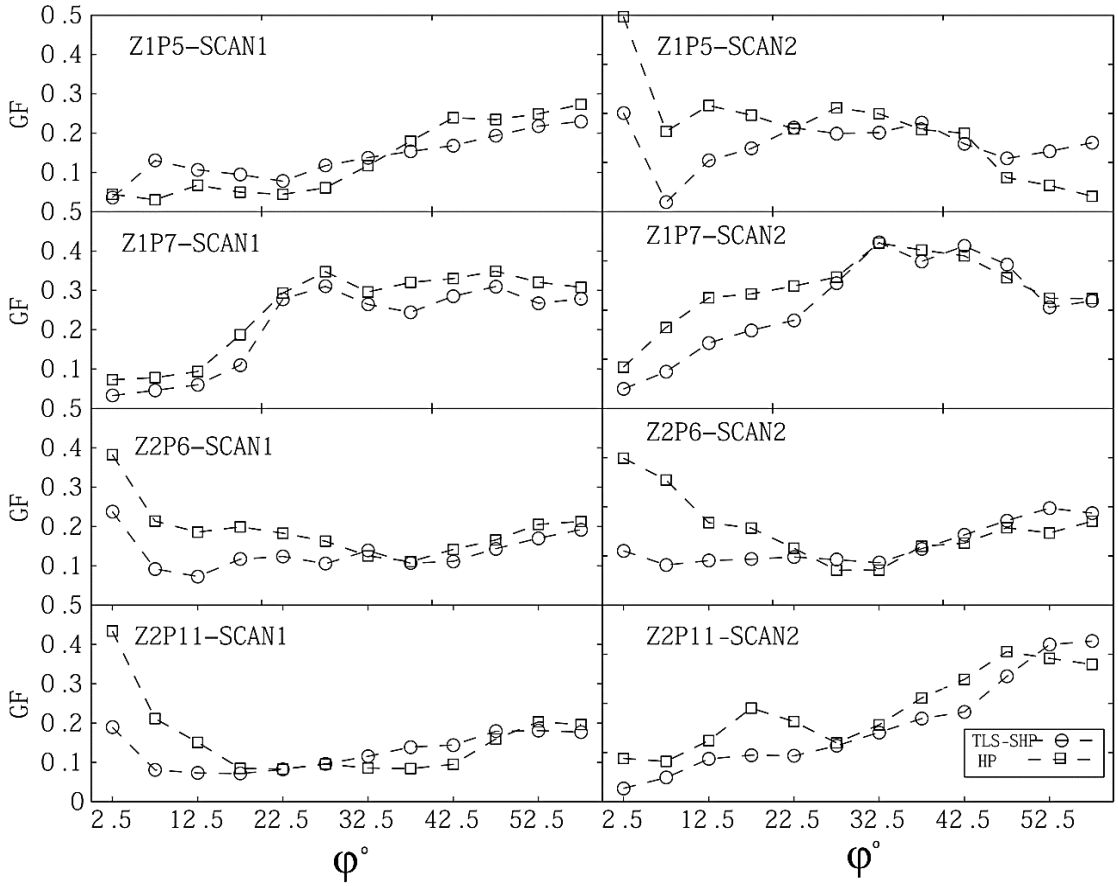

Figure 28. Distribution of GF for zenithal angles for the TLS models.

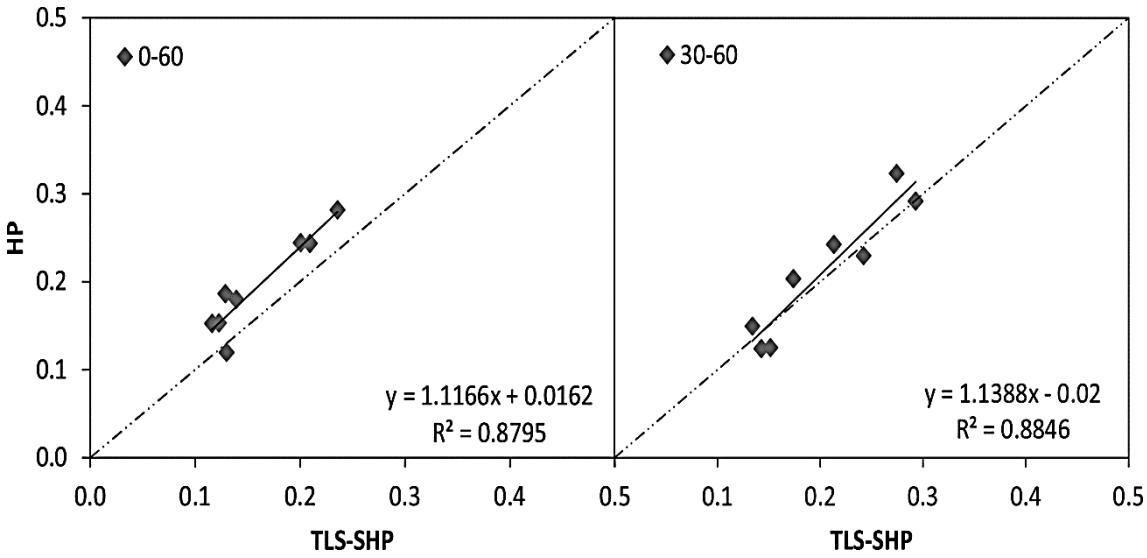

Figure 29. GF comparison for TLS-SHP and HP, considering (0-60) φ range (left), considering (30-60) φ range (right).

Angular grids estimate accurately the angular resolution on simulated TLS data which included variations in sigma, gap patterns and GF. However, when the noise level used to simulate TLS data exceeds the 6% threshold, the computed angular resolutions become unreliable (Table 6). This error in the angular resolution estimation causes an error in the GF estimation, as it determines the size of the grid cell as so the level of detail that can be captured. For example, the angular resolution estimated is smaller than the reference for high sigma in the case of the C gap pattern. This error increases as sigma increases (Figure 22). Its smaller calculated angular resolution causes an increase of cells in the angular grid, appearing as false gaps or as extra rows or columns, hence overestimating GF, especially when lower GF is simulated (Figure 23). For this C gap pattern, there will always be a higher chance of selecting groups of points with adjacent neighbors, closer to each other and, therefore, smaller angular resolution even for low sigma.

The probability of selecting points coming from non-consecutive returns increases when simulated GF higher than 80% for R and RC gap patterns. Therefore, the distance between returns augments and so does its computed angular resolution. As a result, an angular grid with fewer cells reduces the chance of having gaps and thus underestimating GF, but only for extreme 12-14% sigma values (Figure 23). On the other hand, the same problems of overestimation as earlier described for C gap patterns were observed for R and RC, selecting returns closer to each other than

expected, but in this case only for low GF (10-40%) and only when sigma was higher than 6%.

Regarding the real TLS data, the manufacturer of the TLS Leica HDS-6000 system used in this study reported *a priori* angular noise sigma value of 1.25e-04 rad, but the actual sigma value calculated here was much lower, 2.8e-05 rad (4.43%) in θ and 8.0e-06 rad (1.27%) in φ (Table 8). This angular noise found in the TLS data is optimal for the estimation of GF using angular grids since for all the conditions tested on simulated TLS, the algorithm performed consistently well as long as sigma is below 6% (Figures 22 and 23). Several other studies have demonstrated also lower sigma values than reported *a priori* by the manufacturer. For example, Kersten et al. (2008) and Kremen et al. (2006) estimated 4.7e-05 rad deviations in both θ and φ; and Reshetyuk (2006) found 2.79e-05 and 1.57e-05 rad in θ and φ with a Leica HDS-3000 TLS system. The difference between horizontal and vertical deviations can be attributed to the mechanism used in the scanner where vertical precision is about twice as good as horizontal precision, because the scanning mirror shows lower inertia than the servomotor which is used to get horizontal angles (Reshetyuk 2006).

The TLS Leica HDS-6000 system produced an oversample of about 0.07 rad in the θ direction (Figure 30) that was confirmed by Leica Geosystems technical support in a personal communication. This problem could be identified with the angular grid method since due to the higher point density in those areas, the estimated angular resolution was smaller than the average value, generating false gaps and therefore higher GF. Danson et al. (2007) calculated GF as a ratio of the pulse return to the total number of lasers emitted from the theoretical resolution of the instrument without knowing the exact location of the gaps. In this case, the oversampling error would be overlooked since their method cannot determine exactly where these gaps occur in the vegetation canopy, causing a higher ratio between pulse returns and total number of pulses emitted and, as a consequence, lower GF.

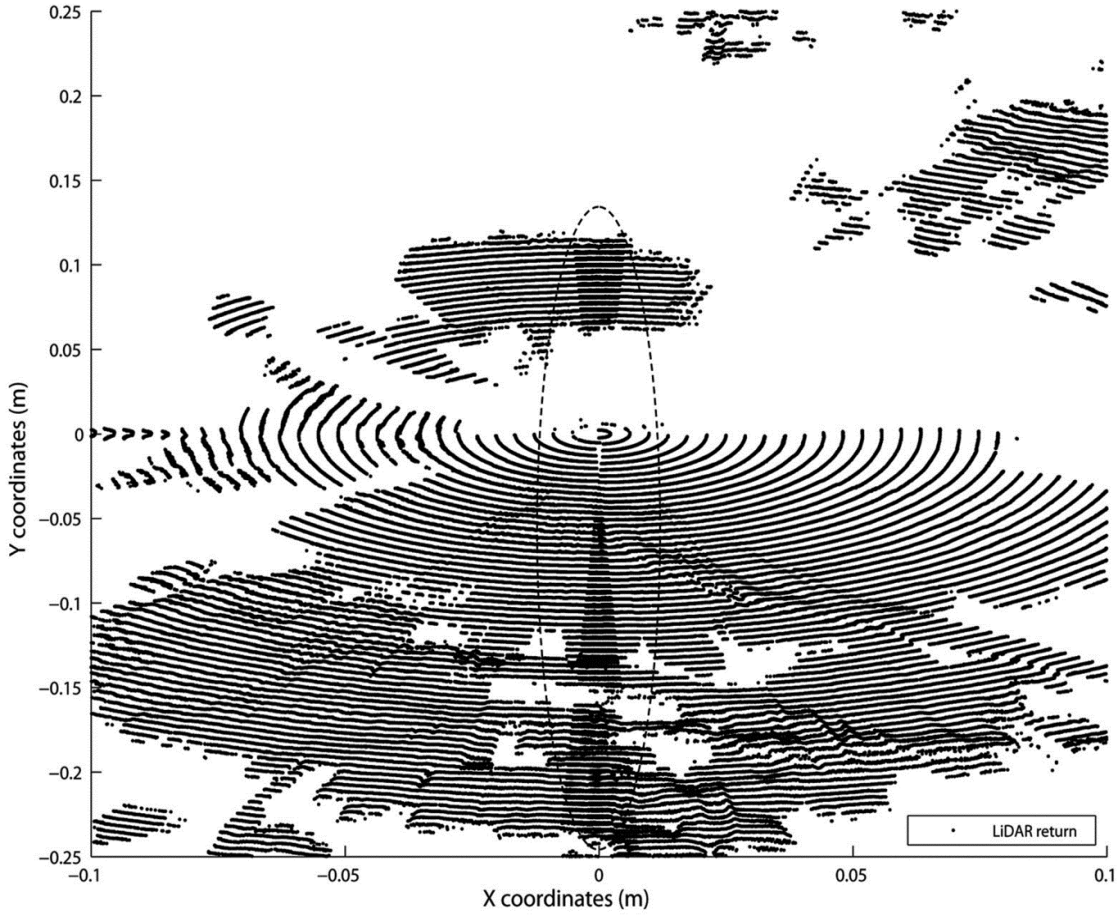

Figure 30. Oversampled scanned zone (dashed line) shown in Cartesian XY plane.

The low φ slices in the TLS tree data contained high noise and their angular resolution was biased from the average (Table 9 and Figure 26). Such noisier returns can occur due to movement of the canopy objects during the TLS data scanning produced by the presence of mild to moderate wind conditions (Cote et al. 2009; Seidel et al. 2012). Moreover, the higher noise obtained for this section is due to the occlusion effect which prevent the laser beam from reaching the upper canopy, therefore the algorithm fails to find suitable neighbor pixels. That occlusion effect was also described by Van der Zande et al. (2006) and causes that the number of laser returns decreases as the distance from the TLS system to the canopy increases. As a consequence, objects in the upper part of the canopy are usually represented by fewer returns and with less accuracy than the ones closer to the TLS system (Seidel et al. 2012). Moreover, when a laser beam partially hits a canopy object, only a fraction of the beam reaches back to the TLS system, causing a return only if the intensity is strong enough. The other part of the laser beam continues travelling, but it is generally

too weak to trigger a return (Eitel et al. 2010). As a result, the edge of the canopy objects will lack laser returns (Figure 31). Vegetation canopies contain complex discontinuous rough elements to scan, such as trunks, branches and leaves, resulting in a noisier laser return that can affect subsequent data processes (Van der Zande et al. 2011).

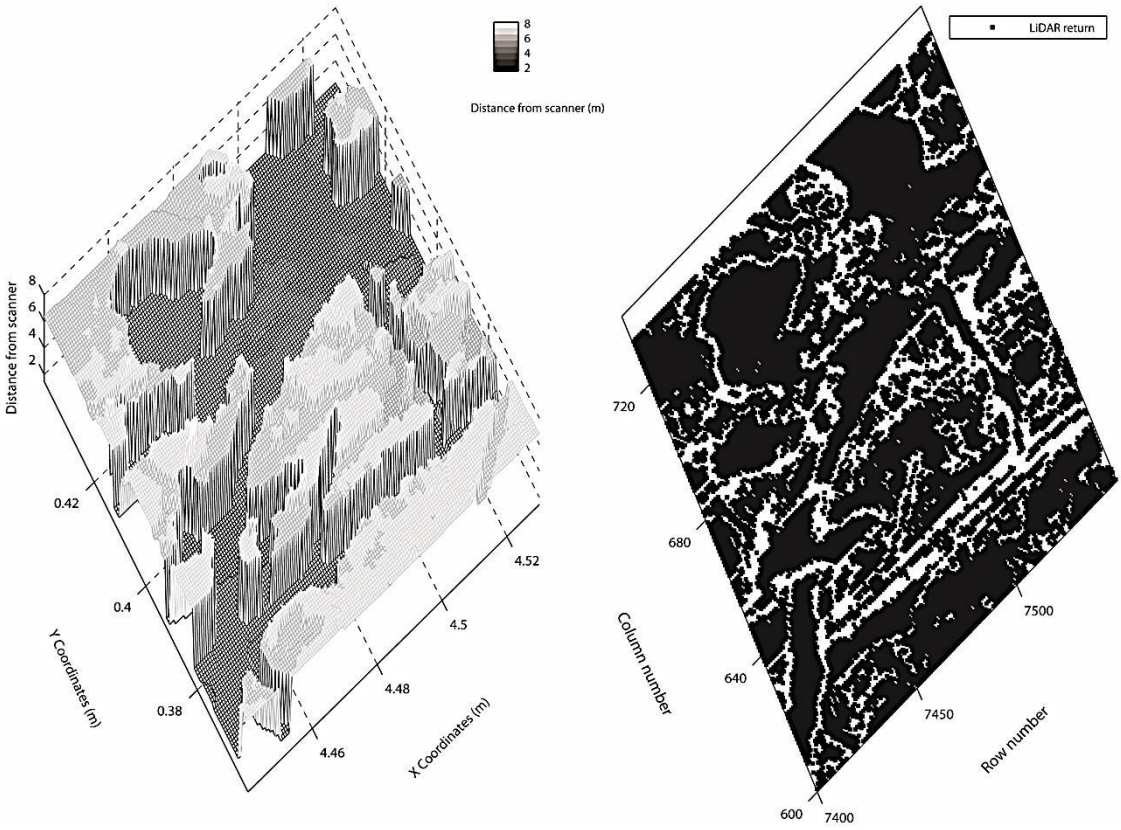

Figure 31. Edge effect on TLS data (XYZ subset) (left), GF angular image from XYZ subset (right).

SimTLS-SHP underestimated GF, especially for data between 10-50% GF (Figure 24). Average differences on this range reached 9% and 14% for RC and R gap patterns respectively. The lower GF values obtained by SimTLS-SHP can be explained in part by the deformation caused by this projection. Polar projection generates a transformation in which each return is shown in its true direction and exact distance from projection center point, whereas all other directions and distances are deformed. In consequence, area distortion and shapes increase as it moves away from the center (Maling 1992). As a result of the deformation, pixels in the center zone and in the outer border of the image had a higher return frequency producing, therefore, an underestimation in GF values. Since the method used to classify pixels as occupied or empty is based on the proportion of returns to the total number of gaps and returns,

variation in the return point frequency within each pixel (Figure 9), greatly affects the generation of the binary image and so, the GF computation. This problem is exacerbated for the R and RC gap patterns, because small random gaps are part of pixels with several returns, masking such gaps.

The angular grid method showed an average GF overestimation of 14% when it compared with HP. (Figure 27 right). Danson et al. (2007) also found a light overestimation of 9% on average respect to HP. These differences can be attributed to processing doing over the HP that can be influenced by the different light conditions present in the images, the lack of returns in the scanned objects edges and also the position of the HP relative to the position of the TLS system. However, Calders et al. (2011) and Lovell et al. (2003) obtained that TLS underestimated the GF from HP. These discrepancies could be related with the different forest height, crown cover, scanner characteristics and HP camera position (Danson et al. 2007). The last factor can be supported also by results of Calders et al. (2011), because they did not took the HP at the same position of the TLS, instead of that, they took several HP in the plot and computed the average, but their HP GF values greatly varied within the plot.

GF derived from the TLS-SHP was about 17% lower than the angular grid GF (Figure 27 left). This difference could be explained in part by deformations associated to the hemispherical projection of the TLS-SHP and the lack of returns in the objects edge that is noticeable in the angular grids. As it was seen for the SimTLS-SHP, the effect of re-projecting caused a systematic underestimation of GF, which is noticeable in R and RC gap patterns with lower GF values (Figure 24).

The GF showed to be a φ dependent structural variable, highly variable near lower φ angles for both approaches, HP and TLS-SHP (Figure 28). The main differences between the HP and TLS-SHP were between 2.5-22.5° with a better agreement near 32°. Similar trends were found by Fournier et al. (1997) for HP and simulated HP. Their poor agreement was between 0-20°, but with better results for the range between 25-70°. In our case, one possible explanation of these differences could be in the method we used to produce the TLS-SHP images and the number of pixels per φ ring (Figure 32). This method measures the TLS returns frequencies within each pixel to assign a gap or return class to it. Considering this fact and that the number of samples for lower

φ ranges is lower, it could explain the higher differences for these φ ranges. Besides, GF from TLS-SHP was an average 4% smaller than the one from HP (Figure 29 right), and the correlation (R^2 =0.89) between them is in concordance with values obtained by Seidel et al. (2012), who found R^2 =0.88 in a forest with large gaps. An explanation could be the threshold selection when classifying the HP which could still be erroneous, even though an automatic method was applied (Jonckheere et al. 2005). In some cases the sky light illumination was too bright causing blooming effects in parts of the HP. Another reason that could influence was that the small and thin branches were not visible in the classified HP images, even when they showed up in the original HP (Seidel et al. 2012). Moreover, no filtering was applied to the HP to improve contrast between sky and foliage, particularly for the small gaps and for the foliage in the periphery (Walter et al. 2003). Even so, canopy geometry seems to be well represented in both approaches, although small gaps which are visible in HP seemed to be smaller or even absent in the TLS-SHP (Figure 11). This effect could be related to the approximately 5 mm footprint at 8 m from the TLS system, which impeded the detection of smaller gaps than this size. In that context, Lovell et al. (2003) indicated that this effect is cause for the beam divergence, because the laser beam increased with the range, and the presence of small canopy gaps will not be detected if there is an object within the beam that produces a return. In this respect, the existence of canopy scenes with gaps larger than the laser footprint seems to shows better agreements with TLS measurements than canopy mainly conformed by small gaps (Seidel et al. 2012). In addition, since the exact position for the coordinates system and the field of view of the TLS scanner and the HP were not the same, it was necessary to select a common φ angular range for match both sensors. The matching of both images was carried out manually, aiming at obtaining all processed φ sectors coincident in the canopy. This selection causes also additional uncertainty, that is especially noticeable at the lower φ angles (Calders et al. 2011), which contributing to the differences between GF values obtained from each instrument.

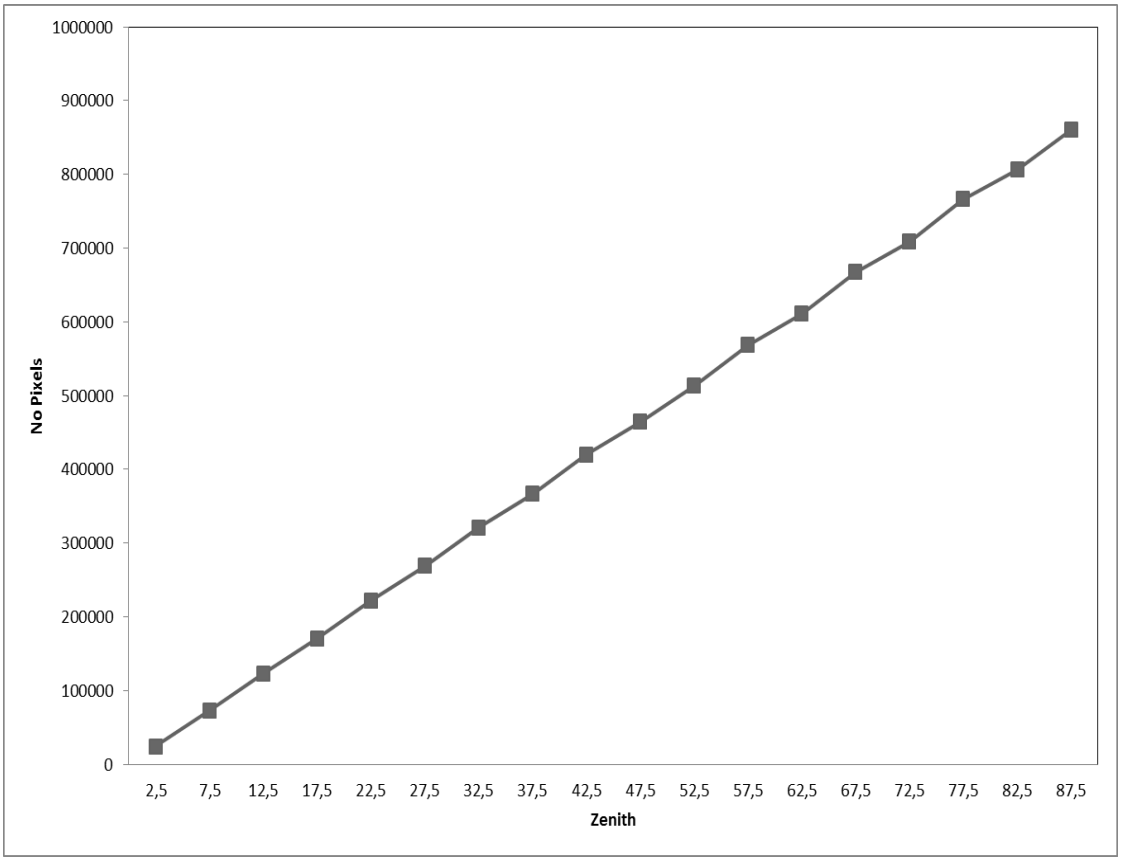

Figure 32. Distribution for the number of pixels in the φ rings for a TLS-SHP.

3.2 Vegetation canopy CI from TLS

Figure 33 shows the CI computed using the CI from HP and TLS-SHP for the eight trees and for different φ angles ranging from 5-55°, at 1° steps. The lower the CI means higher clumping in the canopy. A general underestimation of the CI values was observed for the TLS-SHP versus the real HP, which was considered the reference value. Differences ranged from 3 to 51%, with an average underestimation of 27%. The regressions between the CI computed from TLS-SHP and HP were statistically significant (p-value<0.05) for all φ ranges except for 5-55° which yielded a p-value of 0.145 (Figure 34a). For the significant regressions, the R^2 values ranged from 0.54 to 0.92, and the RMSE from 0.07 to 0.04 for the φ range comprised between 20-55° and 54-55°, respectively (Figure 34b).

Regarding to the number of gaps, the closest match between methods after varying the φ range occurred on Figure 35c, with 217 and 142 gap segments for HP and TLS-SHP, respectively. For the worst case (Figure 35d), these differences were higher,

with 123 versus 30 gap segments, respectively. The average gap run-length showed values very similar for the best case (Figure 35a), with 4 and 3 pixels, but high differences for the worst one (Figure 35b), with values of 5 and 70 pixels for HP and TLS-SHP, respectively.

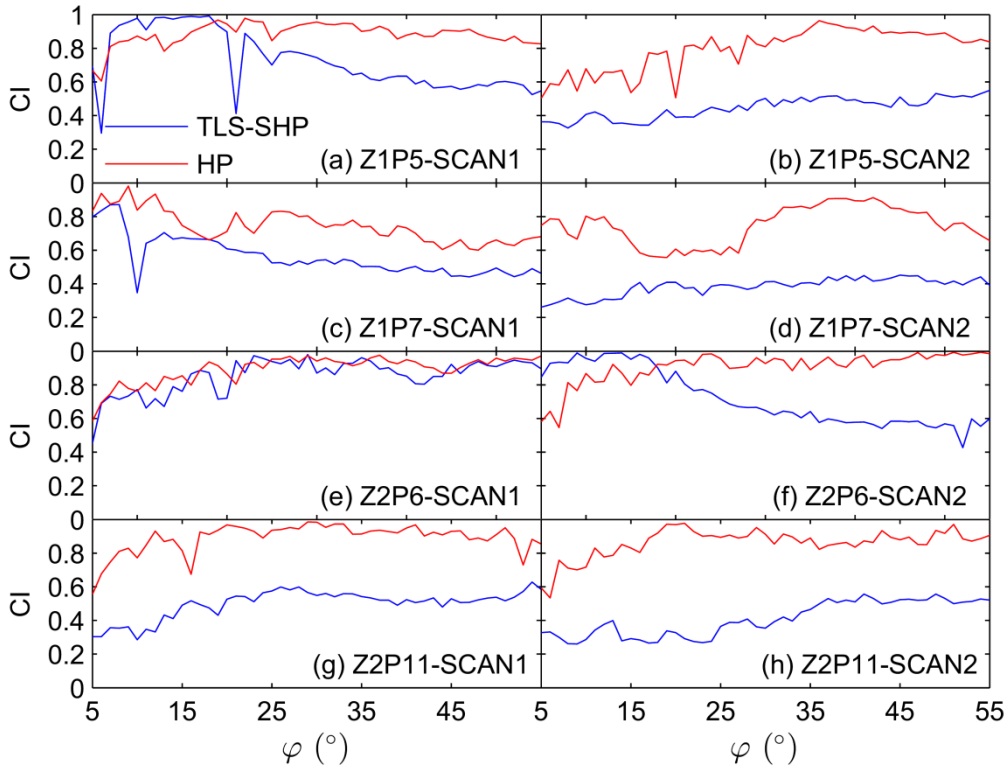

Figure 33. CI derived from HP and TLS-SHP.

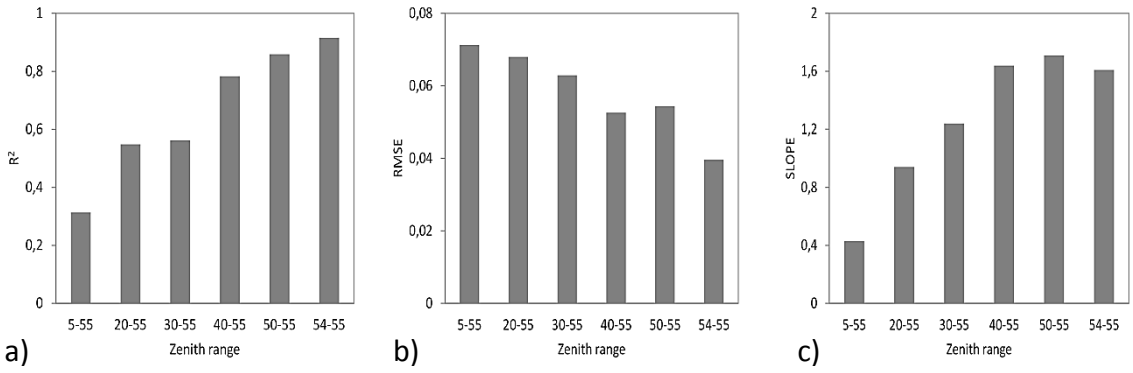

Figure 34. R^2 (a), RMSE (b) and slope (c) between CI from the TLS-SHP and HP for each φ range.

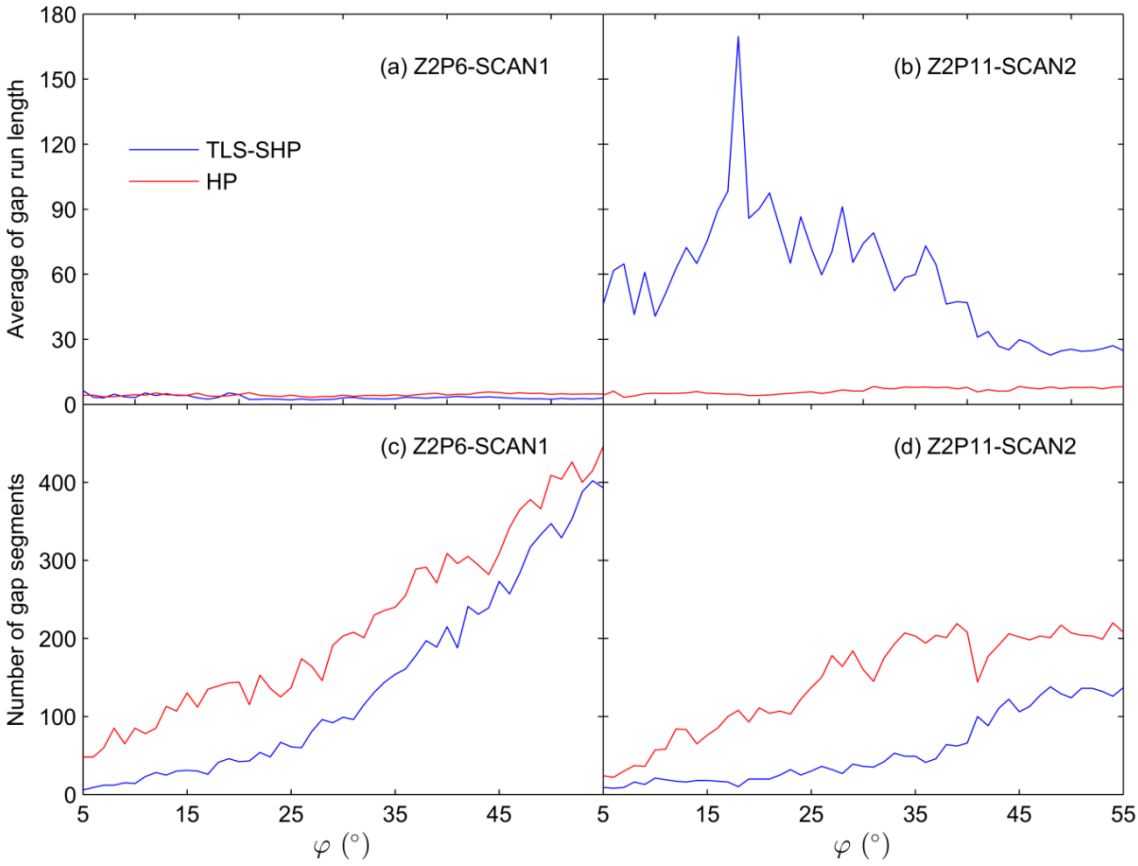

Figure 35. Average gap size and gap number for best (a,c) and worst (b,d) cases.

The CI underestimation for the TLS-SHP compared to the HP can be attributed to differences in the gap size distributions. Furthermore, the size of the gaps in the TLS-SHP, accounted as a run-length in pixel units, was larger and therefore, the number of gap sequences was lower than the HP. Precisely, larger but fewer gaps are characteristic of clumped canopies, with lower CI values (Chen and Cihlar 1995), which happens to coincide with the tree with highest disagreement between the two methods (Figure 33h). On the other hand, the best agreement occurs for the tree with shorter gaps segments but more numerous (Figure 33e).

Differences in the gap size distributions between HP and TLS-SHP can be attributed at least to three sources. The first one is the lack of returns from the edge of the canopy objects on the TLS-SHP, an effect earlier described in the discussion section 3.1. This issue is related to the TLS beam divergence, which measures the laser beam widening with the distance travelled and therefore, influences the position uncertainty of the measured canopy objects (Van Genechten et al. 2008). When a laser pulse hits an object edge, the pulse is split in two or more parts. One part of the pulse reflects

here, while the other portion continues travelling to intercept an object further away in the canopy. This laser returns towards the TLS from two different locations and with less intensity. As a consequence, the coordinates for that point might not be recorded by the TLS or at most the sensor will average the returns and their position will be computed in the wrong place (Rosell et al. 2009; Van Genechten et al. 2008). The laser footprint size at the canopy object is driven by the TLS-object distance and the instrument's beam divergence. For the same laser footprint size, the probability to hit an edge increases with the number of laser pulses, namely higher TLS resolution (Lichti 2004). Our resolution was very high with one laser point every 6.3 mm at a 10 m distance, increasing the chances of creating gaps at the border of the canopy objects. This fact contributes to a bias in the gap size distribution, with larger gaps of fewer segments, and as a consequence lower CI values. It seems that this effect is more pronounced in the dense areas of the canopy, especially for the larger φ angles (Figure 36). In addition, the presence of woody material with cylindrical shapes, such twigs and branches, also difficulties the representation the canopy as also shown by Runions et al. (2007).

The second factor to be considered is the multipath effect that laser returns experience in complex scenarios such as the canopy (Gatziolis and Andersen 2008). The problem is the change in direction that sometimes occurs when a laser pulse hits objects of appreciable mass, such as trunks, wide branches or fruits. For those pulses, the actual path is unknown and the registered laser returns are recorded further away or are even missing, increasing the final sensor noise. Vaaja et al. (2013) mentioned that these noisy returns especially occur on phase difference instruments, like our TLS, rather than on the ones that measure time-of-flight, mainly because their higher laser frequency tend to cause sometimes a slight fluctuation on the phase. A TLS system able of recording first and last returns and with smaller beam width could minimize the described problems (Van Genechten et al. 2008).

The third factor affecting the results is the different resolution and field of view of the HP and TLS sensors. A HP contained ~10M pixels whereas the TLS-SHP included the information of approximately ~15M laser returns and it was 3D. This effect was also pointed out by Zhao et al. (2012) as a contribute factor for the differences. As it

was stated by Frazer et al. (2001), digital HP produces blurred images. Pre-processing steps including channel selection were carried out in order to improve the contrast and find the best threshold to binarize the image. In addition, this binarization process is required and subjective (even if we used an automatic thresholding approach) for HP but it is unnecessary for TLS-SHP, causing additional differences between the two methods. Finally, perfect angular matching between the two sensors is challenging and slight changes in the field of view leads to differences in the observed gaps and in consequence their CI, even though the HP were taken supposedly in the exact same position as the TLS instrument.

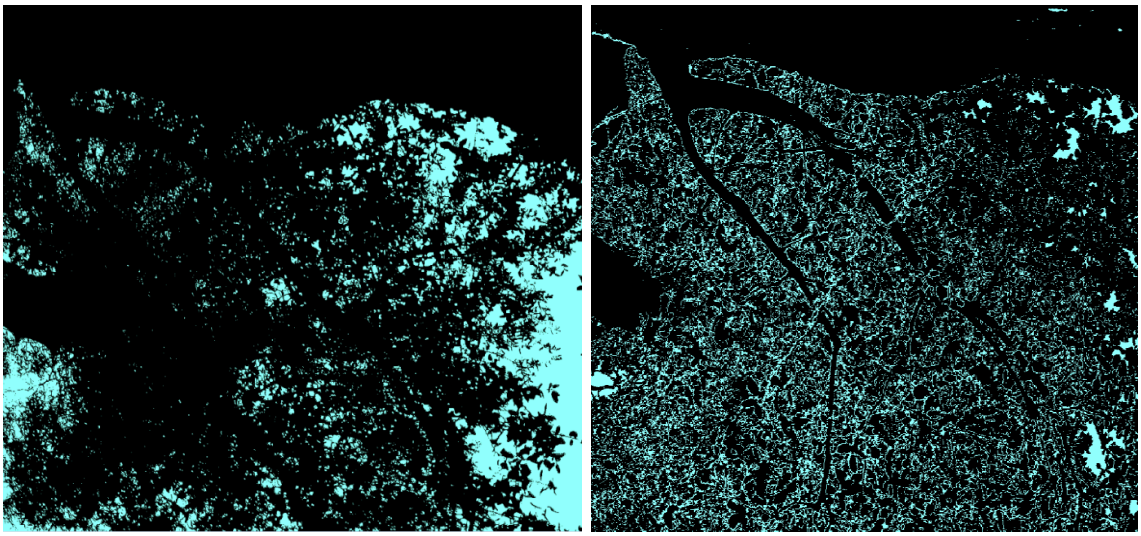

Figure 36. Canopy section from HP (left) and TLS-SHP (right).

As for the effect of the φ angle on the CI estimates, TLS-SHP showed more variability in the very low (<30°) φ values. This is related to the fact that the GF and gap size distributions are highly variable for extreme φ and the resolution worsens (Fournier et al. 1997), which affects the CI computation. Moreover, the CI algorithm did not converge to provide a CI value for the lower φ angles (0-5°), probably because the number of gap segments is so small, less than five, which is the minimum the program CIMES needs to compute CI. In general, this translates into lower CI values (higher clumping) closer to the φ, which agreed with the results mentioned by Kucharik et al. (1997). Walter et al. (2003) also pointed out that as the view direction changes from the horizon to the φ, the clumping effect in the canopy becomes more evident.

A comparison between TLS-SHP and HP was only possible for φ<60°, due to the observation angle limitation in the camera that acquired the HP. The CI relationships between these two instruments that included the lower φ (5-55°) was not statistical significant (p-value>0.05), but the middle φ range (20-55° and 30-55°) showed a better fit with higher R^2 and slope near to 1 (Figure 34a and Figure 34c). This agreed with the results showed by Leblanc and Cheng (2001) where measurements in the φ range between 30-60° indicated small variations, because the CI in this range represents the mean CI of all the angles. Similar variations in CI by φ angle were also mentioned by Leblanc et al. (2005). However, the determination coefficient increased by constraining the φ range to contain only higher values, especially in the φ range 54-55° (Figure 34). Nevertheless, further research is needed to demonstrate if this φ range performs well on other datasets or if it is just the result of an over fitting to our reduced number of samples.

3.3 Vegetation canopy GF from ALS

An automatic threshold was able to compute GF for most of the HP images. After visual inspection, 12 out of 347 HP (~3%) had to be reprocessed with a manual threshold. The affected HP happened on 05 and 13 forest plots and 06 shrub plot for September 2006. After re-processing these HP, their GF per plot increased 5, 9 and 9%, respectively. We compared the GF estimated with this automatic threshold for all HP versus a manual one calculated in Casas et al. (In review). Figure 37 shows that there is a statistically significant agreement between the two methods (R^2 = 0.77, p-value<0.01), where the automatic threshold underestimated GF overall in 12%. When separating by cover type, forest demonstrated a better performance than shrubs, with an R^2 of 0.89 versus 0.62 and an average 0.6% overestimation versus a 20% underestimation, respectively. When comparing the GF computed for the three different dates, the box-plot obtained for the KW test in Figure 38 showed no significant difference between dates.

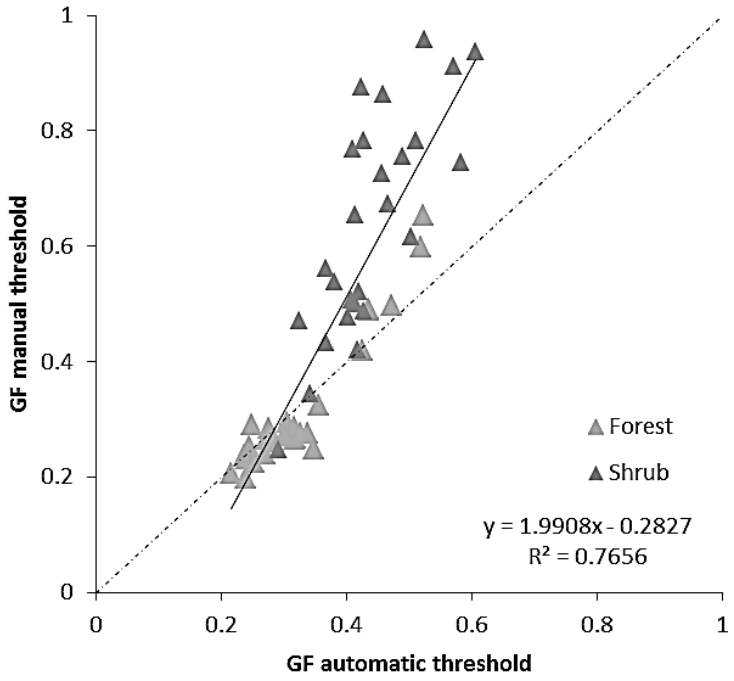

Figure 37. Comparison between the GF data derived with an automatic versus a manual threshold.

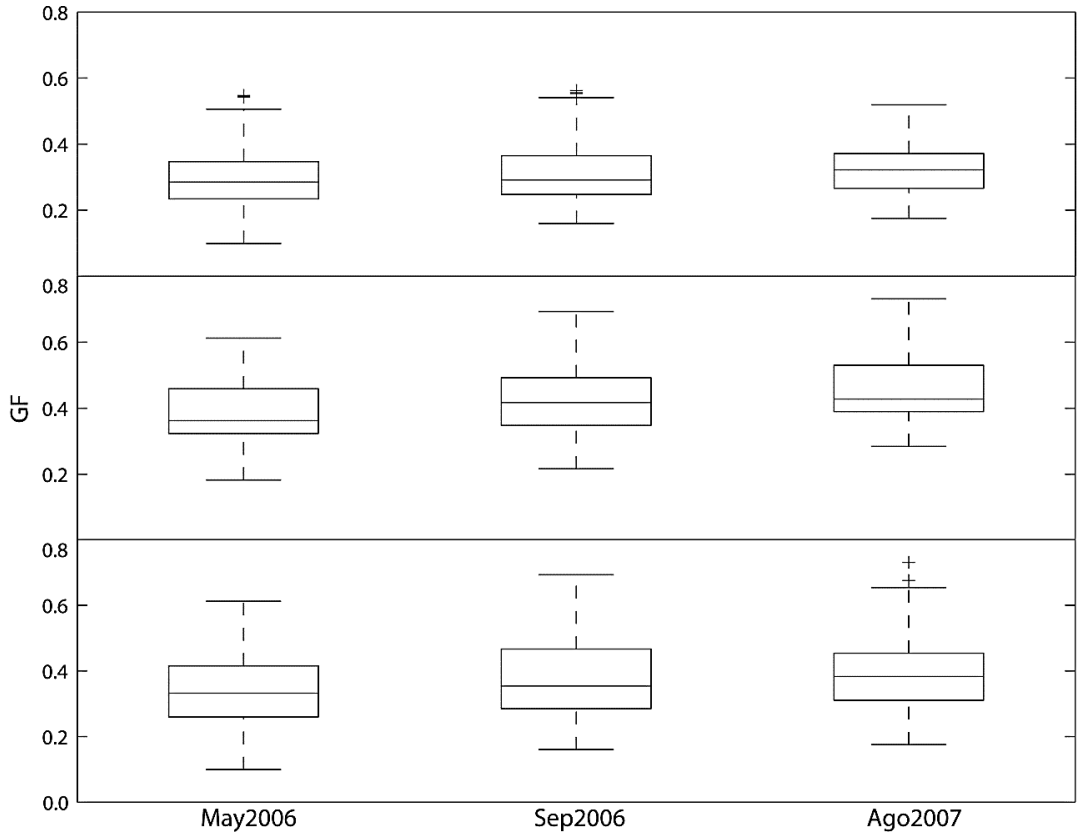

Figure 38. KW test box-plot for forest (top), shrub (middle) and forest and shrub plots together (bottom).

Figure 39 shows the relationship between the GF calculated from HP and the ALS metrics for different ALS radii size. In the case of the forest plots, the R^2 values ranged from 0.1 to 0.9 for the GF_L and GF_S metrics, respectively. The highest correlations for most of the ALS metrics happened at 11 m radius; however for the GF_{c1} metric it was at 15 m. The R^2 values in the case of the shrubs ranged from 0.007 (MaxH) to 0.65 (GF_F) with the highest correlations also at 11 m. When considering forest and shrub data together, R^2 values ranged from 0.007 (GF_F) to 0.80 (GF_S), remaining the highest correlations also at 11 m. Table 10 shows the best models obtained from GF_S and GF_I that tend to underestimate the forest and overestimate the shrubs GF when compared to HP. The W-statistic and its p-value demonstrated that all the models residuals were normally distributed.

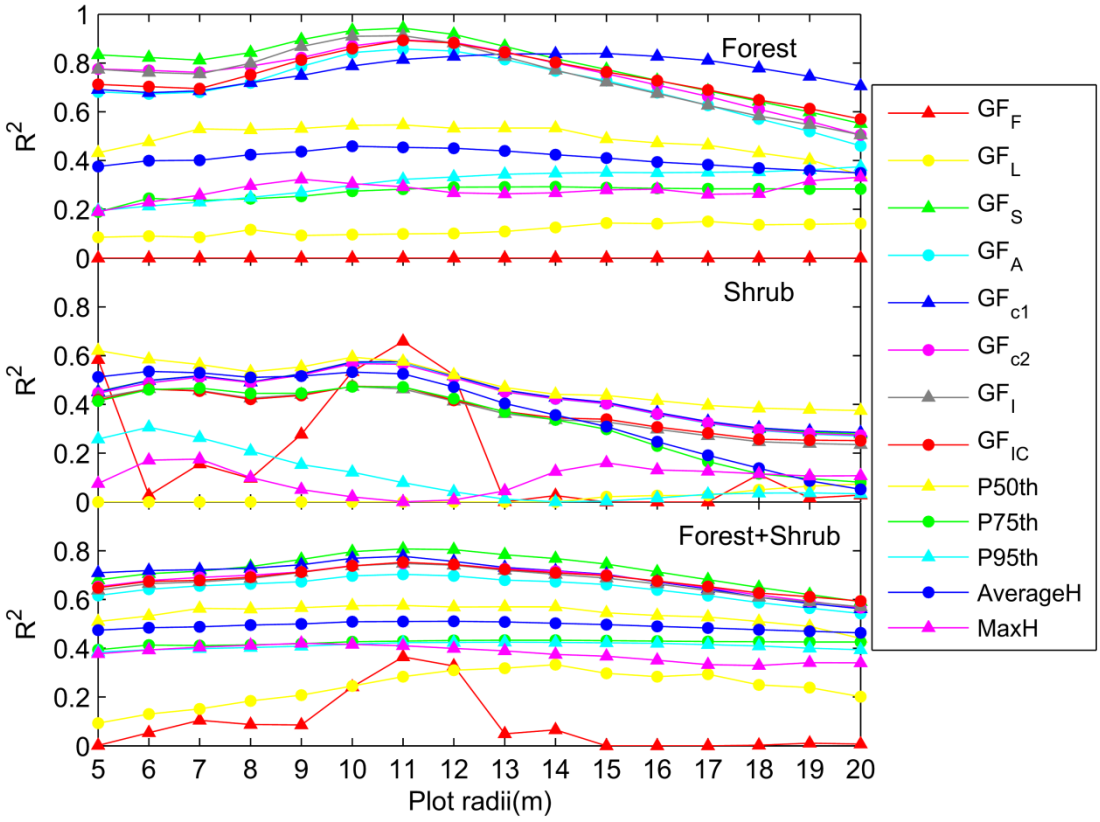

Figure 39. R^2 between GF from HP and all the ALS metrics for different ALS radii on the forest, shrub and both plots together.

Table 10. Selected models for the GF estimation from frequency (f) and intensity (I) ALS metrics with 11 m radius.

	Model (GF)	R2	RMSE	Bias (%)	W-statistic	W-statistc p-value
Forest(f)	0.445xGFS+0.230	0.94	0.022	-36	0.912	0.196
Forest(I)	0.465xGFI+0.176	0.92	0.027	-3	0.917	0.230
Shrub(f)	0.297xGFS+0.237	0.57	0.042	+43	0.907	0.119
Shrub(I)	0.257xGFI+0.228	0.46	0.047	+74	0.908	0.150
Forest+Shrub(f)	0.302xGFS+0.246	0.81	0.039	+8	0.968	0.576
Forest+Shrub(I)	0.276xGFI+0.227	0.73	0.045	+29	0.945	0.156

Several reasons could explain the differences between the automatic and manual thresholds to compute GF. Casas et al. (In review) chose the whole φ range (0-90°) to compute the manual thresholds and masked out the woody materials. In contrast, this work selected 30-60° to calculate the automatic threshold but did not mask anything. Shrub plots showed a higher proportion of woody materials than forest ones in the HP, leading to more gaps in Casas et al. (In review) after applying the mask and hence the 20% GF underestimation on this work. Other possible reason for discrepancy is that the manual threshold is usually biased, since an operator is in charge of finding the edge between the foliage and sky to set a threshold value, which sometimes is only useful for some areas in the image (Jonckheere et al. 2005). Instead of that, clustering is applied to set an automatic threshold that differentiates foliage from sky in the HP. Finally, Casas et al. (In review) preferred the three RGB channels in the HP, whereas here we chose only the blue channel, because it has a better contrast between foliage and sky elements (Gonsamo et al. 2010).

Despite of the presented differences in the HP pre-processing, the algorithm applied here to automatically classify the HP in two clusters successfully estimated GF, and this option was a good choice compared to other automatic thresholds. For example, choosing as threshold the DN with the lowest frequency in the HP histogram is not recommend, because there is a large range of DN with very similar low frequencies (Figure 40). Using this criterion would cause masking small gaps in the canopy (Walter 2009). With the clustering automatic threshold, only 3% of the HP needed to be re-processed after visual inspection. These HP were taken when it was too dark, and even if the exposure was modified, the resulting GF was still too low. In

addition, Jonckheere et al. (2005) pointed out that clustering algorithms perform better in lighter than darker conditions.

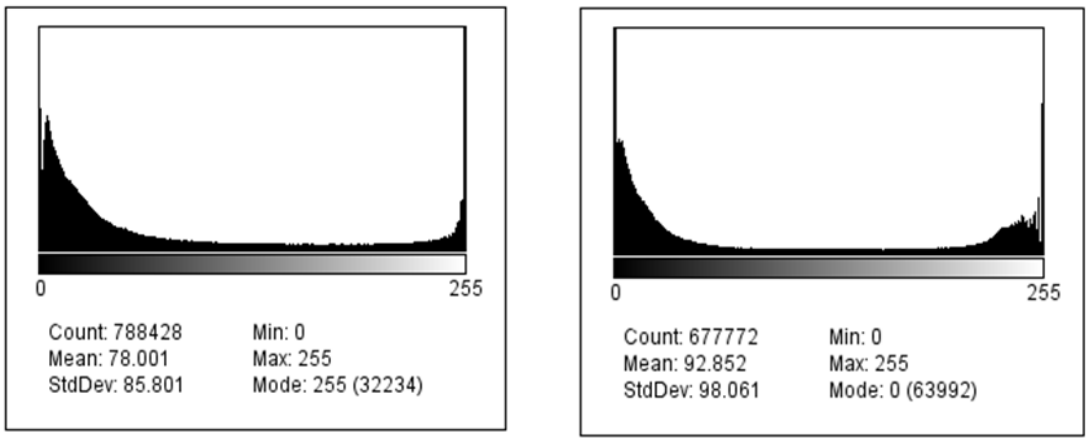

Figure 40. Histograms for a forest (left) and a shrub plot (right).

HP and ALS observe the vegetation canopy from a different perspective which could explain at least partially the differences in the GF estimates (Morsdorf et al. 2006; Riaño et al. 2004). The ALS captures the canopy from the top facing down with 18° scan angle, whereas HP does it from the bottom facing up with 30-60° φ range. In consequence, there will be some areas that would be visible from the HP and invisible from the ALS and vice versa. Besides, Morsdorf et al. (2006) mentioned that HP cannot discriminate the gaps within and between the crowns whereas ALS could. In regards to the plot radii, Morsdorf et al. (2006) found the best agreement at 15 m with 7.15° scan angle, but 11 m was the optimal radius in this study, like reported also by Hopkinson et al. (2007) (Figure 39). Despite of these discrepancies, we chose an adequate radii range that contained a maximum correlation that decreased in either direction (Figure 39). The optimal radius size was the same for the forest and shrub canopies. Contrary, Riaño et al. (2004) found a different optimal radius size for a pine and an oak forest that authors explained due to the size and distribution of the elements within canopy.

ALS metrics, GF_S and GF_I, estimated GF accurately but results differed by canopy type. For the forest plots, GF_S provided a very high R^2=0.94 (Table 10). This simple ratio of only the single laser returns from either the canopy or the ground avoids the divergence generated by those returns that hit two different surfaces to produce a first and a last return (Sasaki et al. 2008). A similar agreement for GF_S

(R^2=0.83) was found on an evergreen and deciduous forest by Sasaki et al. (2008). These authors demonstrated that laser point density influences the GF estimation, which was 1.13 points/m^2 in their case whereas ours was 4.0 points/m^2. Slighter lower performances (R^2=0.73 and 0.70) were found by Morsdorf et al. (2006) and Hopkinson et al. (2009) with points density of 20.0 and 1.0 points/m^2 respectively. Regarding to metrics which considered different weight for return types, the GF_{c2} showed a better performance of 0.89 compared with 0.61 obtained by Solberg et al. (2009), however it was still lower than GFs. According to Solberg et al. (2009) the rate of penetration of GF_{c2} metric is conditioned by the weight factors, which were assigned arbitrarily.

The GF_I metric based on the intensity of the laser returns performed well (R^2=0.91 Table 10), similarly to Hopkinson et al. (2007) who reached an R^2 of 0.89 with a density of 3.0 points/m^2 and for the same radius of 11 m as our study. Similarly to GF_I, GF_{IC} provided also a good GF estimate (R^2=0.89). This metric simulates the power loss transmission by the intermediate and last returns according to the Beer's Law. Hopkinson et al. (2007) also found a good agreement, but in their case GF_{IC} slightly improved over GF_I, which could be related to how this metric accounts for the power loss in relation to the difference in canopy structure between the two works. Hopkinson et al. (2009) indicated that GF_I underestimated GF in 6% and our study only did in 3% (Table 10).

Shrub plots showed weaker correlations with GF_S and GF_I than the forest ones (Table 10). In addition, these metrics overestimated GF on shrub plots whereas the opposite occurred for forest plots (Table 10). These differences could be related to canopy structure and height according to Hopkinson et al. (2009). Forest GF estimated from HP is only 9 % lower than the shrub GF, with 32% versus 41%, but GF_S predicted a 39% difference, with 20% versus 59%, respectively. Relatively fewer ground returns are expected on taller canopies where laser pulses cannot penetrate as much, and as consequence lower GF values (Sasaki et al. 2008). Contrary, shorter canopies such as the shrub allow relatively more ground returns, overestimating GF. Furthermore, Hopkinson et al. (2009) reported that ALS metrics based on the penetration rates like GF_S are more dependent on canopy height than metrics based on the intensity of the laser returns like GF_I. The intensity metrics would be more sensitive to the 'peak pulse

power concentration' parameter, calculated as the ratio between the peak pulse power and the footprint area (Hopkinson 2007). For our case, the trends found by Hopkinson et al. (2009) did not perpetuate, because both kind of metrics are similarly dependent on canopy height (Figure 41).

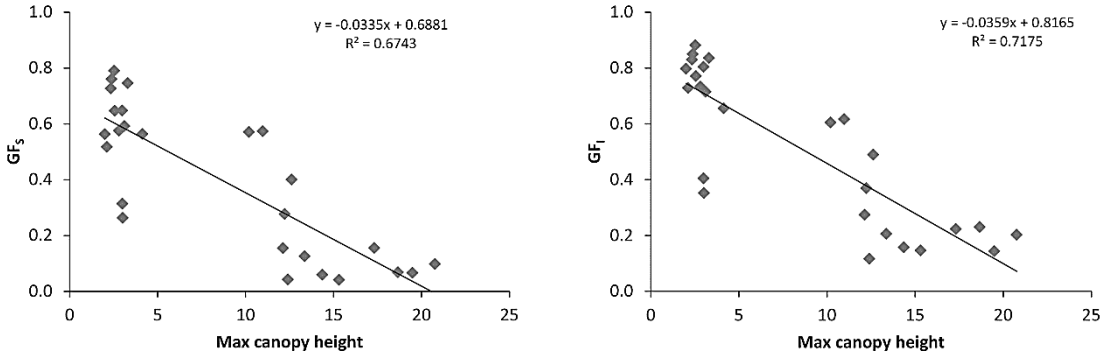

Figure 41. Correlation between GFS (left) or GFI (right) with maximum canopy height.

The lack of first and last ground returns on the forest and shrub plots causes the failure of GF_F and GF_L, respectively (Figure 39). Although the GF_F showed some peaks in the shrubs plots, they were not significant (p-value>0.05), because it still presented several cases for which it was not possible to compute it. For the GF_L, this return type is residual from reaching multiple surfaces and therefore rare or inexistent in either canopy for the plot radii considered. However, Hopkinson et al. (2009) applied the GF_F metric and got an agreement of R^2=0.65 with an underestimation of 30% respect to HP GF measurements. Besides, Morsdorf et al. (2006) got a R^2=0.73 for the same metric with also an underestimating trend. Hopkinson et al. (2009) mentioned that this metric is influenced by the canopy structure and height, giving a better result in a mixed wood forest than in a closed conifer forest. In addition, Lovell et al. (2003) indicated that these metrics from first and last returns are biased by the interaction between the canopy structure in term of its gaps size and the footprint size. The empirical metrics derived from vegetation height (P50,th P75,th P95,th AverageH and MaxH) did not provide results close to GF_S for forest plots but did so for shrub ones. The P50th offered an R^2=0.62 on shrub plots. Similar correlation magnitude were found by Riaño et al. (2004) for the same metric and radii size on an oak forest.

Moreover, Riaño et al. (2004) found that regression parameters with these metrics are canopy type specific like it happens in our case (Figure 39).

3.4 Vegetation canopy CI from ALS

The statistical analysis carried out over the CI derived from HP indicated that there was no significant difference between the three sampling dates (Figure 42). The CI models computed from the ALS height metrics are in Figure 43. For the forest plots, the best three were StdP50plot, MedianP75plot and P50plot, all of them with R^2 of 0.47. For the shrub plots, the best three were Stdup95c, MedianP25plot, Stdup75c with R^2 0.36, 0.35 and 0.33, respectively. For both plots together, the best three were MedianP25plot, MeanP25plot and MeanP50plot with R^2 0.35, 0.31 and 0.30 respectively.

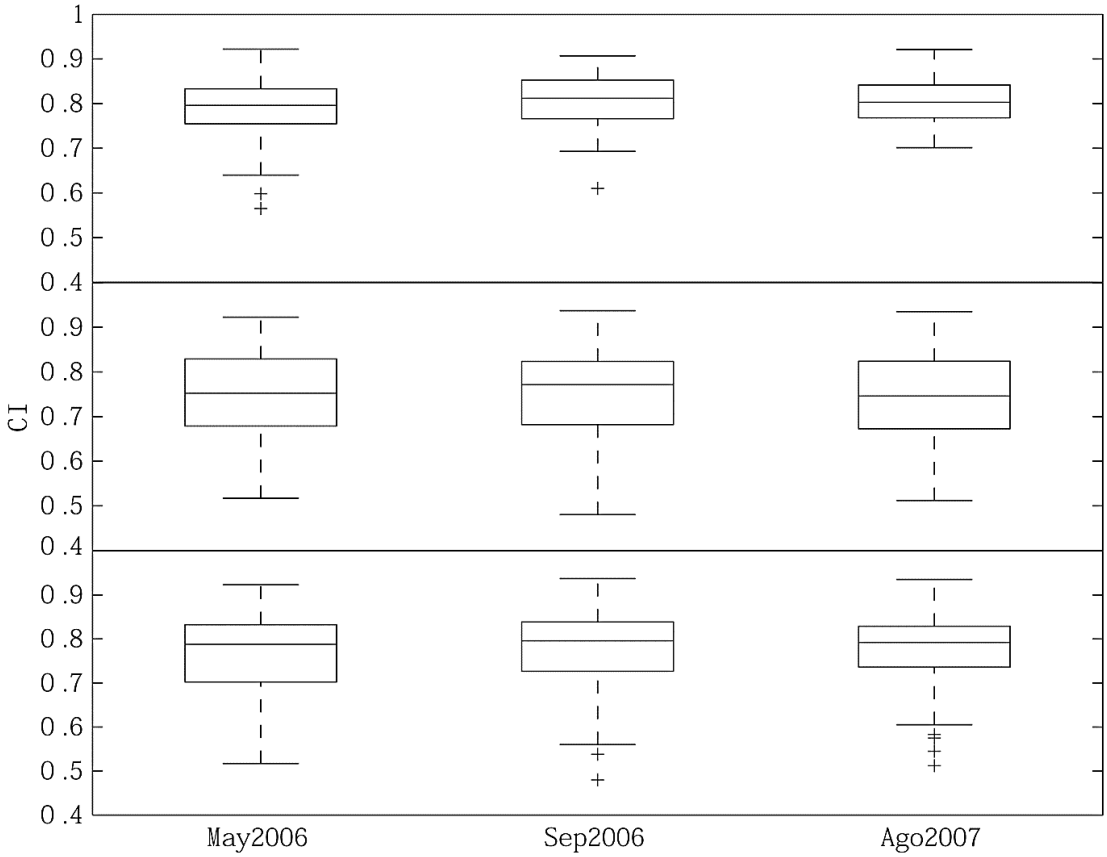

Figure 42. KW test box-plot for the CI calculated by date for forest (top), shrub (middle), and both plots together (bottom).

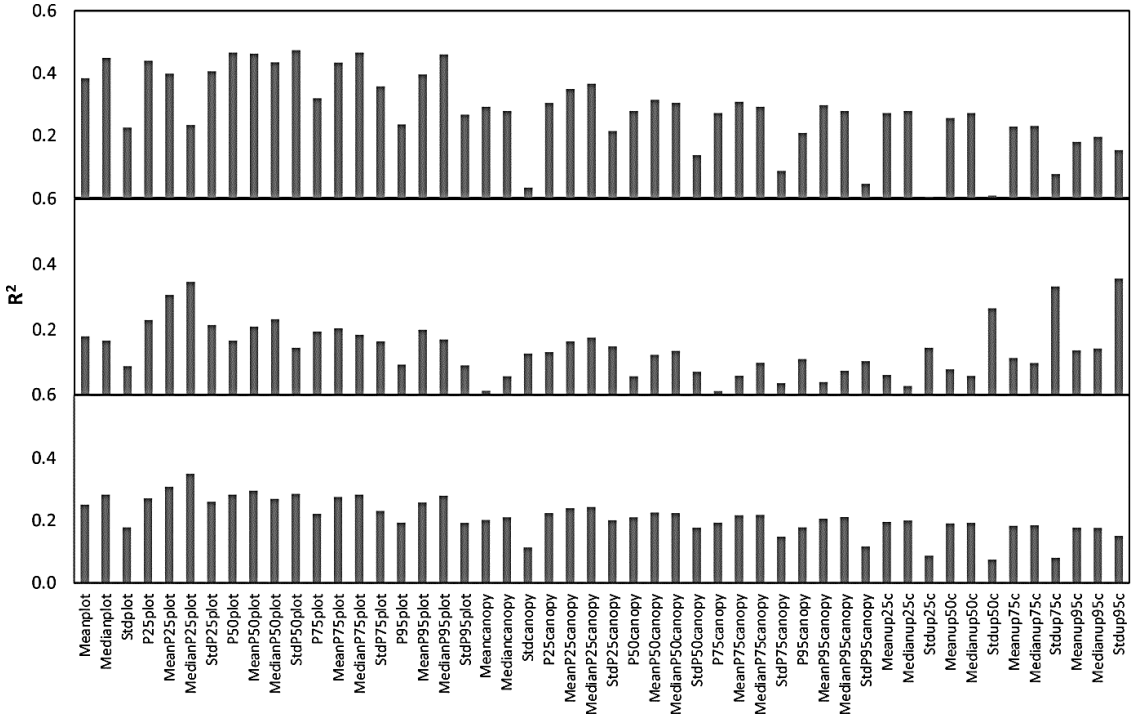

Figure 43. R^2 for the CI regression models computed using ALS height metrics for forest (top), shrub (middle), and both plots together (bottom).

The PCS algorithm applied row by row to the GRI estimated the CI in the best case scenario with an R^2 of 0.57, 0.32 and 0.45 for the forest, shrub and both plots together, respectively (Figure 44 and 45). A different plot radius size accomplished these best results for each of the plots considered (Figure 44). Presuming GRI were angular data coming from HP, a direct application of the PCS-CIMES algorithm provided higher R^2 values that reached 0.65, 0.61 and 0.50 for the same three cases as above. When the PCS-CIMES algorithm was restricted to $\varphi<18°$, the R^2 values reached a higher value than the previous case for the forest (0.72), lower for the shrub (0.50) and the same for both plots together (0.50). Finally, Table 11 shows the selected models in each case after the application of the best PCS algorithm, for which the residuals were always normally distributed according to their W statistic and p-value.

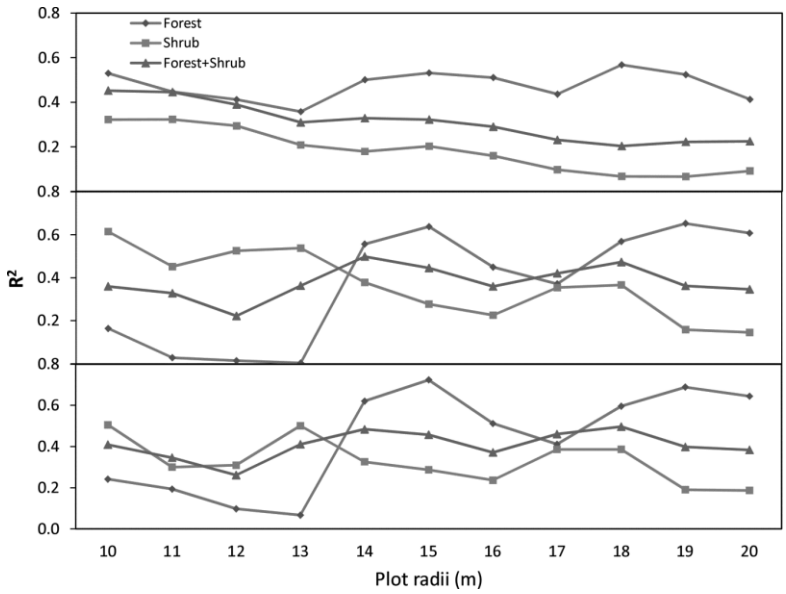

Figure 44. R^2 for the CI regression models from the PCS by plot radii using a row average (top), φ angle average (middle) and φ <18 (bottom).

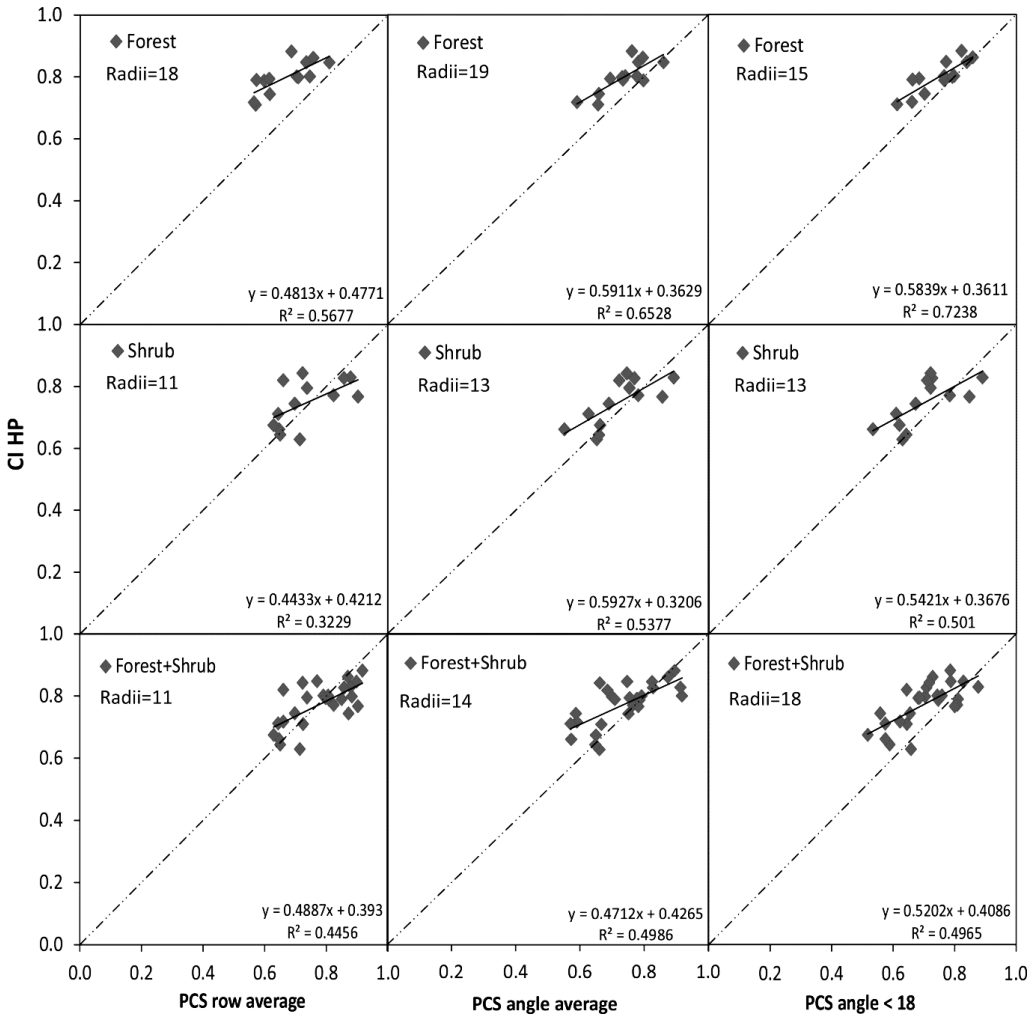

Figure 45. Best CI models PCS algorithm using row average (left), φ average (middle) and φ <18 (right).

Table 11. Selected CI regression models for the PCS algorithm.

	Model (CI)	Radius (m)	R^2	RMSE	Bias (%)	W statistic	W p-value
Forest	0.584xPCSangle18+0.361	15	0.7238	0.0289	-6	0.8769	0.0648
Shrub	0.593xPCSangleaverage+0.321	13	0.5377	0.0539	-4	0.9558	0.6888
Forest+Shrub	0.471xPCSangleaverage+0.426	14	0.4986	0.0499	-5	0.9875	0.9822

The CI obtained from HP were on average 0.799 and 0.746 for the forest and shrub plots, respectively (Figure 42). The forest was a mixture of semi open broadleaf with dense understory and the shrub was chaparral. Gonsamo et al. (2009) got a similar average CI value (0.769) for several forest types that included: natural tropical cloud forest with dense under and overestory and plantations of Cupressus, Eucalyptus and Pinus. In another study, CI oscillated between 0.82–1.02 for a forest of sugar maple, black spruce, European beech and Scots pine (Walter et al. 2003). According to Leblanc et al. (2005), the CI can be quite similar across multiple species, which is supported by the comparison above between the three studies, where the vegetation significantly differed.

The main source of error and variation in CI estimated from HP according to Leblanc et al. (2005) comes from the type of threshold applied to process the HP. The threshold method particularly influences and changes the amount of GF and gap size distribution than later impacts the CI computation. This work preferred an automatic one based on clustering. Gonsamo et al. (2009) chose a visual threshold to separate sky from the foliage. Instead, Walter et al. (2003) computed the threshold by means of a linear interpolation for the frequency peaks between a full covered canopy and a clear sky, a method developed by Olsson et al. (1982).

Another source of variation between these studies is the φ range selected to compute the CI. In our study we computed the CI for a 30-60° φ range, Gonsamo et al. (2009) chose 0-60° and Walter et al. (2003) decided on 5-75°. According to Walter et al. (2009), the CI is an angular dependent variable that increases with φ, so a different φ range can lead to dissimilar CI values. Figure 42 shows a higher CI value for the forest than for the shrub plots. This has a relation with the GF values that were lower for the forest than for the shrub plots (Figure 37). In general, the lower the GF the higher the

CI values and vice versa, according to the gap size theory developed by Chen and Cihlar (1995).

The CI derived from empirical ALS height metrics (Figures 43), contradicts the good results obtained by Thomas et al. (2011) in which the median of the 25th percentile and higher (Medianup25c) rendered R^2=0.81. However in our case, the same metric performed poorly for either forest (0.28) or shrub (0.02) plots. Instead of applying linear relations like in our study, Thomas et al. (2011) chose more complex power or polynomial equations that could be potentially more site specific. In addition, their study comprised a boreal forest whose canopy structure is more homogeneous than our mixed Mediterranean forest and chaparral.

Despite of the differences in perspective and resolution between the HP and GRI images, the application of the PCS algorithm to the GRI seems to be a better approach to estimate the CI than the empirical ALS height metrics (Figure 44). The best model of Thomas et al. (2011) showed a higher R^2 than our PCS results, but their model residuals were not normally distributed, an assumption accomplished for all our PCS models (Table 11). In order to figure out if more complex relations could improve our results, we decided to test the same height metric chose by Thomas et al. (2011) in a form of polynomial regression. The results improved substantially for the forest plots (R^2=0.71, p-value<0.05), but not for the shrub ones (R^2=0.06 p-value>0.05) (Figure 46). That clarifies that, rather than an empirical fit like the one based on Medianup25c metric, our models based on the gap size distribution have the potential to be applied to other vegetation conditions.

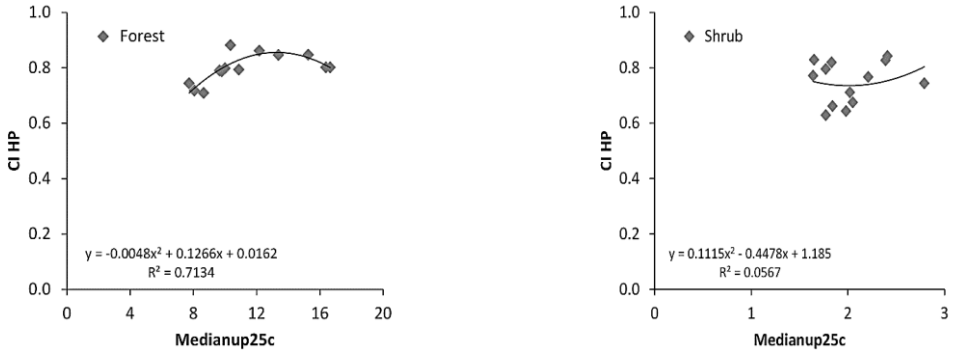

Figure 46. Regressions for CI derived from Medianup25c for forest plots (left) and shrub plots (right).

The CI from the PCS algorithm was tested for multiple plot radii and validated with the CI from HP (Figure 44). The best plot radius to estimate CI was not as clear as for the other structural variable (GF), where we found an optimal and strong relationship for a radius size decreasing in both directions (see Figure 39 in section 3.3). For the CI, forest plots always performed better for larger radius size than shrub ones. This fact could be related to the gap size distribution which is very dependent on the footprint size (Lovell et al. 2003). Shrub plots presented higher GF values (see Figure 38 in section 3.3), so the ALS pulses could penetrate more often the canopy to reach the ground and describe better the gap size distribution to determine the right CI for smaller radii. However, the denser forest plots needed larger ALS radii to allow more chances to penetrate to the ground and capture better the variability in the gap size distribution.

The PCS algorithm applied to the GRI performed better when GRI were treated as angular data than when processed row by row (Figure 45). This improvement could be related to the way the directions of the gap sequences are processed. Despite the difference in perspective between sensors, angular GRI sequences are likely to match the HP ones since they are also treated by φ angle. When the φ was restricted <18°, the results only improved for the forest plots (Figure 45). This issue addresses the need to match the viewing geometry between the observed angles by the ALS sensor and the HP images to find better agreement between sensors (Korhonen et al. 2011).

All the selected ALS PCS models (Table 11) presented an underestimation of the CI value when compared to HP (Figure 45). Gonsamo and Pellikka (2009) and Walter et al. (2003) also observed this and indicated that it could be related to the sensibility of PCS algorithm to the small variances in the gap size distribution of the ALS data. The dissimilarities between CI from HP and ALS PCS could also be in part due to the different resolution of these sensors. ALS it has a much lower resolution, with only ~1700 laser pulses for 15 m radius plot, than the HP, with ~4M pixels per image that prevents the ALS from capturing more canopy details. Another factor that affects the CI comparison between sensors is related to their difference in perspective and geometry. Firstly, the ALS captures the gaps data almost at nadir (18° scan angle), meanwhile HP does for almost the whole range of φ angles, which affects the captured

gap distribution and therefore the CI estimation. This difference in the view perspective and its consequences was mentioned by Demarez et al. (2008) who found significant differences in LAI and CI estimation from HP images when they took them facing upwards and downwards over crops.

4 CONCLUSIONS

In this dissertation we have tested, adapted and developed new methods based on the gap frequency, the Beer Lambert's law and the gap size theory in order to estimate GF and CI from TLS and ALS data. These two essential canopy structural variables play an important role in the estimation of LAI through indirect optical methods and in the correction of LAI to account for the fact that the foliage is not always randomly distributed (Chen and Cihlar 1995; Walter et al. 2003). They also are critical in the study and modelling of the microclimate that exists within the vegetation canopies, which in turn affects growth, photosynthesis, evapotranspiration, nutrients distribution and conditions the animal behavior (Gonsamo et al. 2010; Leblanc et al. 2005; Panferov et al. 2001).

These conclusions section is organized according to the experimental cases described in the methods chapter. The results exposed here confirm that the methods developed from TLS and ALS can help describing GF and CI, becoming an alternative that can reduce the limitations of suitable illumination conditions, unbiased image thresholds and more comprehensive spatially than traditional HP methods.

4.1 Vegetation canopy GF from TLS

A new method to compute GF from TLS angular data was developed. This method computes the proportion of empty cells in binary angular grid generated after calculating the angular resolution of the TLS data. By working with angular data instead of projecting the data into a plane with Cartesian coordinates, the distortion due to the increase in distance between laser pulses as they move away from the TLS scanner was removed. The angular grid method predicted accurately GF for simulated angular data of known GF value, as long as the sigma noise did not exceed 6% of the angular resolution. Above this threshold the algorithm failed to compute the angular resolution accurately and, as a consequence, to obtain adequate GF values. The angular resolution was more difficult to predict for the simulated C gap pattern than the R and RC ones.

The angular grid GF obtained over actual experimental TLS data is also affected by noisy angular returns, which are far from the average behavior, causing false gaps and hence a GF overestimation. Another noticeable problem is the lack of returns at the edge of the canopy objects, which also contributes to the GF overestimation. The actual TLS data demonstrated to have lower noise than the manufacturer indicated a priori. This noise level was below 6%, within which the angular grid algorithm performs accurately. The transformation of the simulated angular data into a hemispherical projection (Sim-TLS-SHP) causes a deformation that underestimates GF which is also more critical for the R gap pattern. An advantage of the proposed method when applied to actual TLS data (TLS-SHP) is that is independent from the illumination conditions unlike the HP. Furthermore, TLS could identify the location of the gaps in the canopy to accomplish a characterization of the gap size distribution and the CI in the 3D space, for which further investigation is needed.

4.2 Vegetation canopy CI from TLS

An implementation of the gap size theory to compute the Chen and Cihlar CI from TLS data was developed. The method requires previous knowledge of the angular resolution of the TLS to determine the right pixel size to transform the TLS data into HP (TLS-SHP). TLS-SHP underestimated the CI values when compared to the one derived from HP. This underestimation is directly related to the difference in the gap size distributions. Two factors can affect the canopy gap distribution of the TLS-SHP: Firstly, the transformation into HP causes the elimination of small canopy gaps. Secondly, false gaps appear at the edge of the canopy objects when the laser return dissipates due to hitting partially one or more objects.

The difference between sensors to estimate CI is also affected by their resolution, meanwhile a HP had about 10M pixels, the TLS-SHP was composed of approximately 15M returns. In addition, the field of view of both sensors might be slightly different, even though the TLS and HP were placed in the same position; a perfect alignment is not always possible. An advantage of TLS-SHP is that the laser returns or no return separates automatically the canopy from the gaps, but HP requires identifying an image threshold to separate between them which is always

subjective. The CI presents high variability for the extreme low φ values due to the lower resolution. A lower variability in the CI values occurs for the middle φ range. The TLS-SHP delivered promising results to be used as a reference data for the CI estimations from ALS.

4.3 Vegetation canopy GF from ALS

Based on the Beer Lambert's law, ALS systems estimate the canopy GF using different metrics built from the ground laser returns frequency and their intensity. These metrics had a different performance depending on the vegetation and the radius size of the ALS data. The best GF predictor was the GF_S followed by the GF_I, for both the forest and shrub plots. The ALS estimations showed a clear underestimation for the forest plots, but an overestimation trend for the shrub ones. There is a relation between the GF metrics and the vegetation height which suggests that the incorporation of height data in the GF models could improve their performance. Contrary to other previous studies (Solberg 2010; Solberg et al. 2009), the GF_{C2} metric that considered diverse weight factors for the different laser return types did not improve compared with the simple GF_S frequency metric. Similarly, the intensity metric GF_{IC}, modified to account for the transmission loss of the intermediate and last returns, did not improve either the simple ratio GF_I metric. This metric considered a spatial randomly canopy and in our case this assumption was not accomplished, especially for the shrub plots. The metrics based on first (GF_F) or last (GF_L) laser returns only did not work due to the minimal amount of these return types in our dataset. The empirical ALS height metrics were not able to estimate the GF, except for the $P50^{th}$ that showed a good result for the shrub plots. The ALS metrics have the advantage to be more spatially extensive than HP or TLS measurements to produce GF maps once the best models have been selected and validated.

4.4 Vegetation canopy CI from ALS

Unlike other previous works (Thomas et al. 2011), the ALS height metrics did not predict well CI, therefore a new method to estimate the canopy CI from ALS data was tested and proposed. This method was an adaptation of the Pielou´s segregation algorithm (PCS) applied to a GRI derived from ALS data. The PCS performed better

when GRI were treated as angular data, probably because the gap sequences were processed also as angular for the validation HP data. This issue also demonstrates the need to match as best as possible the perspective and φ angles of the observation point of view between sensors. The relationship between the PCS performance and the ALS radius size for the plots was not as evident as for the GF. Forest needed larger radius than shrub plots for an accurate estimate probably because the higher and denser the canopy the harder to find enough ground laser returns to characterize properly the gap size distribution.

The PCS presented an underestimation compared to the CI values from HP. The GRI resolution is much smaller than HP which causes that GRI must have a larger size to allow the PCS algorithm to collect enough spatial variability of the gaps to characterize the CI. Despite of the multiples differences between the HP and GRI, the application of PCS on GRI demonstrated an improvement over the traditional approach of empirical height metrics in the CI estimation. Unlike the empirical methods that are restricted to the calibration sites, the PCS method based on the gap size distribution has the potential to be extrapolated to other vegetation conditions.

5 REFERENCES

Antonarakis, A.S., Richards, K.S., Brasington, J., & Muller, E. (2010). Determining leaf area index and leafy tree roughness using terrestrial laser scanning. *Water Resour. Res., 46*, W06510.

Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Jones, M.O., Martin, R.E., Boardman, J., & Field, C.B. (2007). Carnegie airborne observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging for three-dimensional studies of ecosystems. *Journal of Applied Remote Sensing, 1*, 013536-013536-013521.

BCAL (2013). LiDAR BCAL Tools ver 1.5.2. Idaho State University, Department of Geosciences, Boise Center Aerospace Laboratory (BCAL), Boise, Idaho. In. http://bcal.boisestate.edu/tools/lidar/

Bréda, N.J.J. (2003). Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany, 54*, 2403-2417.

Calders, K., Verbesselt, J., Bartholomeus, H.M., & Herold, M. (2011). Applying terrestrial LiDAR to derive gap fraction distribution time series during bud break. In, *Proceedings SilviLaser 2011, 11th international LiDAR forest applications conference, 16-20 Oct. 2011, Tasmania*. Hobart, Tasmania: University of Tasmania.

Campbell, G., & Norman, J. (1989). The description and measurement of plant canopy structure. In Russell G, M. B, & J. PG (Eds.), *Plant Canopies: Their Growth, Form and Function* (pp. 1-19). Cambridge Cambridge University Press.

Casas, A., Riaño, D., Ustin, S.L., Dennison, P., & Salas, J. (In review). Estimation of water-related biochemical and biophysical vegetation properties using multitemporal airborne hyperspectral data and its comparison to MODIS spectral response. *Remote Sensing of Environment.*

Chen, J.M. (1996). Canopy architecture and remote sensing of the fraction of photosynthetically active radiation absorbed by boreal conifer forests. *Geoscience and Remote Sensing, IEEE Transactions on, 34*, 1353-1368.

Chen, J.M., & Black, T.A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell & Environment, 15*, 421-429.

Chen, J.M., & Cihlar, J. (1995). Plant canopy gap-size analysis theory for improving optical measurements of leaf-area index. *Applied Optics, 34*, 6211-6222.

Chen, J.M., Liu, J., Leblanc, S.G., Lacaze, R., & Roujean, J.-L. (2003). Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption. *Remote Sensing of Environment, 84*, 516-525.

Coren, F., & Sterzai, P. (2006). Radiometric correction in laser scanning. *International Journal of Remote Sensing, 27*, 3097-3104.

Cote, J.F., Widlowski, J.L., Fournier, R.A., & Verstraete, M.M. (2009). The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. *Remote Sensing of Environment, 113*, 1067-1081.

Dahlin, K.M., Asner, G.P., & Field, C.B. (2011). Environmental filtering and land-use history drive patterns in biomass accumulation in a mediterranean-type landscape. *Ecological Applications, 22*, 104-118.

Danson, F.M., Hetherington, D., Morsdorf, F., Koetz, B., & Allgower, B. (2007). Forest canopy gap fraction from terrestrial laser scanning. *Ieee Geoscience and Remote Sensing Letters, 4*, 157-160.

Dean, T.J., Cao, Q.V., Roberts, S.D., & Evans, D.L. (2009). Measuring heights to crown base and crown median with LiDAR in a mature, even-aged loblolly pine stand. *Forest Ecology and Management, 257*, 126-133.

Demarez, V., Duthoit, S., Baret, F., Weiss, M., & Dedieu, G. (2008). Estimation of leaf area and clumping indexes of crops with hemispherical photographs. *Agricultural and Forest Meteorology, 148*, 644-655.

Donoghue, D.N., Watt, P.J., Cox, N.J., & Wilson, J. (2007). Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment, 110*, 509-522.

Eitel, J.U.H., Vierling, L.A., & Long, D.S. (2010). Simultaneous measurements of plant structure and chlorophyll content in broadleaf saplings with a terrestrial laser scanner. *Remote Sensing of Environment, 114*, 2229-2237.

Endler, J.A. (1993). The color of light in forests and its implications. *Ecological monographs*, 2-27.

Fournier, R.A., Rich, P.M., & Landry, R. (1997). Hierarchical characterization of canopy architecture for boreal forest. *Journal of Geophysical Research-Atmospheres, 102*, 29445-29454.

Frazer, G.W., Canham, C., & Lertzman, K. (1999). Gap Light Analyzer (GLA), Version 2.0: Imaging software to extract canopy structure and gap light transmission indices from true-colour fisheye photographs, users manual and program documentation. In, *Simon Fraser University, Burnaby, British Columbia, and the Institute of Ecosystem Studies, Millbrook, New York*. http://www.caryinstitute.org/science-program/our-scientists/dr-charles-d-canham/gap-light-analyzer-gla

Frazer, G.W., Fournier, R.A., Trofymow, J., & Hall, R.J. (2001). A comparison of digital and film fisheye photography for analysis of forest canopy structure and gap light transmission. *Agricultural and Forest Meteorology, 109*, 249-263.

Garcia, M., Danson, F.M., Riano, D., Chuvieco, E., Ramirez, F.A., & Bandugula, V. (2011). Terrestrial laser scanning to estimate plot-level forest canopy fuel properties. *International Journal of Applied Earth Observation and Geoinformation, 13*, 636-645.

García, M., Popescu, S., Riaño, D., Zhao, K., Neuenschwander, A., Agca, M., & Chuvieco, E. (2012). Characterization of canopy fuels using ICESat/GLAS data. *Remote Sensing of Environment, 123*, 81-89.

García, M., Prado, E., Riaño, D., Chuvieco, E., & Danson, F.M. (2009). Ajuste planimétrico de datos LiDAR para la estimación de características dasométricas en el Parque Natural del Alto Tajo. *Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, 184-208.

García, M., Riaño, D., Chuvieco, E., & Danson, F.M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment, 114*, 816-830.

Garcia, M., & Ustin, L. (2001). Detection of interannual vegetation responses to climatic variability using AVIRIS data in a coastal savanna in California. *Geoscience and Remote Sensing, IEEE Transactions on, 39*, 1480-1490.

Gatziolis, D., & Andersen, H.-E. (2008). *A guide to LiDAR data acquisition and processing for the forests of the Pacific Northwest*. US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Gonsamo, A., & Pellikka, P. (2008). Methodology comparison for slope correction in canopy leaf area index estimation using hemispherical photography. *Forest Ecology and Management, 256*, 749-759.

Gonsamo, A., & Pellikka, P. (2009). The computation of foliage clumping index using hemispherical photography. *Agricultural and Forest Meteorology, 149*, 1781-1787.

Gonsamo, A., Walter, J.-M.N., & Pellikka, P. (2010). Sampling gap fraction and size for estimating leaf area and clumping indices from hemispherical photographs. *Canadian Journal of Forest Research, 40*, 1588-1603.

Hale, S.E., & Edwards, C. (2002). Comparison of film and digital hemispherical photography across a wide range of canopy densities. *Agricultural and Forest Meteorology, 112*, 51-56.

Hall, S.A., Burke, I.C., Box, D.O., Kaufmann, M.R., & Stoker, J.M. (2005). Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management, 208*, 189-209.

Henning, J.G., & Radtke, P.J. (2006). Ground-based laser imaging for assessing three-dimensional forest canopy structure. *Photogrammetric Engineering and Remote Sensing, 72*, 1349-1358.

Herbert, T.J. (1986). Calibration of fisheye lenses by inversion of area projections. *Applied Optics, 25*, 1875-1876.

Hollander, M., & Wolfe, D.A. (1999). *Nonparametric statistical methods*. (Second ed.). New York: John Wiley ans Sons.

Hopkinson, C. (2007). The influence of flying altitude, beam divergence, and pulse repetition frequency on laser pulse return intensity and canopy frequency distribution. *Canadian Journal of Remote Sensing, 33*, 312-324.

Hopkinson, C., & Chasmer, L. (2007). Modelling canopy gap fraction from lidar intensity. In, *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007* (pp. 190-194).

Hopkinson, C., & Chasmer, L. (2009). Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment, 113*, 275-288.

Hopkinson, C., Lovell, J., Chasmer, L., Jupp, D., Kljun, N., & van Gorsel, E. (2013). Integrating terrestrial and airborne lidar to calibrate a 3D canopy model of effective leaf area index. *Remote Sensing of Environment, 136*, 301-314.

Hosoi, F., Nakai, Y., & Omasa, K. (2010). Estimation and Error Analysis of Woody Canopy Leaf Area Density Profiles Using 3-D Airborne and Ground-Based Scanning Lidar Remote-Sensing Techniques. *Geoscience and Remote Sensing, IEEE Transactions on, 48*, 2215-2223.

Hosoi, F., & Omasa, K. (2006). Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning lidar. *Ieee Transactions on Geoscience and Remote Sensing, 44*, 3610-3618.

Huising, E.J., & Gomes Pereira, L.M. (1998). Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. *ISPRS Journal of Photogrammetry and Remote Sensing, 53*, 245-261.

Inoue, A., Yamamoto, K., Mizoue, N., & Kawahara, Y. (2004). Effects of image quality, size and camera type on forest light environment estimates using digital hemispherical photography. *Agricultural and Forest Meteorology, 126*, 89-97.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., & Baret, F. (2004). Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology, 121*, 19-35.

Jonckheere, I., Macfarlane, C., & Walter, J.-M. (2012). Image Analysis of Hemispherical Photographs. Algorithms and calculations. In R.F.a.R. Hall (Ed.), *Hemispherical Photography in Forest Science: Theory, Methods, Applications.* Sherbrook: Kluwer.

Jonckheere, I., Nackaerts, K., Muys, B., & Coppin, P. (2005). Assessment of automatic gap fraction estimation of forests from digital hemispherical photography. *Agricultural and Forest Meteorology, 132*, 96-114.

Kersten, T., Mechelke, K., Lindstaedt, M., & Sternberg, H. (2008). Geometric accuracy investigations of the latest terrestrial laser scanning systems. In, *Proc. Integrating Generations FIG Working Week* (pp. 1-16). Stockholm, Sweden.

Korhonen, L., Korpela, I., Heiskanen, J., & Maltamo, M. (2011). Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment, 115*, 1065-1080.

Kremen, T., Koska, B., & Pospıšil, J. (2006). Verification of laser scanning systems quality. In, *Proceedings XXIII FIG Congress, Munich, Germany* (pp. 1-16). Munich, Germany.

Kucharik, C., Norman, J., Murdock, L., & Gower, S. (1997). Characterizing canopy nonrandomness with a multiband vegetation imager (MVI). *Journal of Geophysical Research, 102*, 29455-29429,29473.

Lang, A., & Yueqin, X. (1986). Estimation of leaf area index from transmission of direct sunlight in discontinuous canopies. *Agricultural and Forest Meteorology, 37*, 229-243.

Le Roux, X., Sinoquet, H., & Vandame, M. (1999). Spatial distribution of leaf dry weight per area and leaf nitrogen concentration in relation to local radiation regime within an isolated tree crown. *Tree Physiology, 19*, 181-188.

Leblanc, S., Chen, J., & Kwong, M. (2002a). Tracing Radiation and Architecture of Canopies TRAC Manual version 2.1. 3. *Natural Resources Canada, 2*, 1-25.

Leblanc, S.G., & Chen, J.M. (2001). A practical scheme for correcting multiple scattering effects on optical LAI measurements. *Agricultural and Forest Meteorology, 110*, 125-139.

Leblanc, S.G., Chen, J.M., Fernandes, R., Deering, D.W., & Conley, A. (2005). Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agricultural and Forest Meteorology, 129*, 187-207.

Leblanc, S.G., Fernandes, R., & Chen, J.M. (2002b). Recent advancements in optical field leaf area index, foliage heterogeneity, and foliage angular distribution measurements. In, *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International* (pp. 2902-2904 vol.2905).

Lee, J.-S. (1980). Digital Image Enhancement and Noise Filtering by Use of Local Statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-2*, 165-168.

Lee, J., Yu, K., Kim, Y., & Habib, A. (2007). Adjustment of Discrepancies Between LIDAR Data Strips Using Linear Features. *Geoscience and Remote Sensing Letters, IEEE, 4*, 475-479.

Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., & Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment, 70*, 339-361.

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). Lidar Remote Sensing for Ecosystem Studies. *BioScience, 52*, 19-30.

Leica-Geosystems (2008). Leica HDS 6000 Product Specifications. In. http://archive.cyark.org/temp/Leicahds6000datasheet.pdf

Lhotka, J.M., & Loewenstein, E.F. (2008). Influence of canopy structure on the survival and growth of underplanted seedlings. *New Forests, 35*, 89-104.

Lichti, D.D. (2004). A resolution measure for terrestrial laser scanners. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 34*, 6.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography, 27*, 88-106.

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

Maas, H.-G. (2002). Methods for Measuring Height and Planimetry Discrepancies in Airborne Laserscanner Data. *Photogrammetric Engineering & Remote Sensing, 68*, 933-940.

MacArthur, R.H., & Horn, H.S. (1969). Foliage profile by vertical measurements. *Ecology*, 802-804.

Mailly, D., Turbis, S., & Chazdon, R.L. (2013). solarcalc 7.0: An enhanced version of a program for the analysis of hemispherical canopy photographs. *Computers and Electronics in Agriculture, 97*, 15-20.

Maling, D.H. (1992). *Coordinate systems and map projections*. (Second ed.). Oxford - UK: Pergamon Press.

Mallet, C., & Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 1-16.

Miller, J. (1967). A formula for average foliage density. *Australian Journal of Botany, 15*, 141-144.

Moorthy, I., Miller, J.R., Berni, J.A.J., Zarco-Tejada, P., Hu, B., & Chen, J. (2011). Field characterization of olive (Olea europaea L.) tree crown architecture using terrestrial laser scanning data. *Agricultural and Forest Meteorology, 151*, 204-214.

Moorthy, I., Miller, J.R., Hu, B., Chen, J., & Li, Q. (2008). Retrieving crown leaf area index from an individual tree using ground-based lidar data. *Canadian Journal of Remote Sensing, 34*, 320-332.

Morsdorf, F., Kötz, B., Meier, E., Itten, K.I., & Allgöwer, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment, 104*, 50-61.

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., & Running, S.W. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment, 83*, 214-231.

Nelson, R., Valenti, M.A., Short, A., & Keller, C. (2003). A Multiple Resource Inventory of Delaware Using Airborne Laser Data. *BioScience, 53*, 981-992.

Nilson, T. (1971). A theoretical analysis of the frequency of gaps in plant stands. *Agricultural Meteorology, 8*, 25-38.

Olsson, L., Carlsson, K., Grip, H., & Perttu, K. (1982). Evaluation of forest-canopy photographs with diode-array scanner OSIRIS. *Canadian Journal of Forest Research, 12*, 822-828.

Optech (2013). ALTM 3100 EA Product Specifications. In. http://www.optech.ca/pdf/Brochures/ALTM3100EAwspecsfnl.pdf

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *Systems, Man and Cybernetics, IEEE Transactions on, 9*, 62-66.

Panferov, O., Knyazikhin, Y., Myneni, R.B., Szarzynski, J., Engwald, S., Schnitzler, K.G., & Gravenhorst, G. (2001). The role of canopy structure in the spectral variation of transmission and absorption of solar radiation in vegetation canopies. *Ieee Transactions on Geoscience and Remote Sensing, 39*, 241-253.

Parker, G.G. (1995). *Structure and microclimate of forest canopies*. Academic Press, Inc.; Academic Press Ltd.

Parker, G.G., Harding, D.J., & Berger, M.L. (2004). A portable LIDAR system for rapid determination of forest canopy structure. *Journal of Applied Ecology, 41*, 755-767.

Pfeifer, N., Elberink, S.O., & Filin, S. (2005). Automatic tie elements detection for laser scanner strip adjustment. *International Archives of Photogrammetry and Remote Sensing, 36*, 1682-1750.

Pielou, E.C. (1962). Runs of One Species with Respect to Another in Transects through Plant Populations. *Biometrics, 18*, 579-593.

Popescu, S.C., & Zhao, K. (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment, 112*, 767-781.

Prewitt, J.M.S., & Mendelsohn, M.L. (1966). THE ANALYSIS OF CELL IMAGES*. *Annals of the New York Academy of Sciences, 128*, 1035-1053.

Pringle, R.M., Webb, J.K., & Shine, R. (2003). Canopy structure, microclimate, and habitat selection by a nocturnal snake, Hoplocephalus bungaroides. *Ecology, 84*, 2668-2679.

Reich, P.B., Ellsworth, D.S., Kloeppel, B.D., Fownes, J.H., & Gower, S.T. (1990). Vertical variation in canopy structure and CO2 exchange of oak-maple forests: influence of ozone, nitrogen, and other factors on simulated canopy carbon gain. *Tree Physiology, 7*, 329-345.

Reshetyuk, Y. (2006). Calibration of terrestrial laser scanners Callidus 1.1, Leica HDS 3000 and Leica HDS 2500. *Survey Review, 38*, 703-713.

Riaño, D., Meier, E., Allgower, B., Chuvieco, E., & Ustin, S.L. (2003). Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment, 86*, 177-186.

Riaño, D., Valladares, F., Condes, S., & Chuvieco, E. (2004). Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agricultural and Forest Meteorology, 124*, 269-275.

Ridler, T., & Calvard, S. (1978). Picture thresholding using an iterative selection method. *IEEE transactions on Systems, Man and Cybernetics, 8*, 630-632.

Rivas Martínez, S., & Gandullo, J.M. (1987). *Memoria del mapa de series de vegetación de España: 1: 400.000*. Spain: ICONA.

Rosell, J.R., Llorens, J., Sanz, R., Arnó, J., Ribes-Dasi, M., Masip, J., Escolà, A., Camp, F., Solanelles, F., Gràcia, F., Gil, E., Val, L., Planas, S., & Palacín, J. (2009). Obtaining the three-dimensional structure of tree orchards from remote 2D terrestrial LIDAR scanning. *Agricultural and Forest Meteorology, 149*, 1505-1515.

Royston, P. (1995). Remark AS R94: A remark on algorithm AS 181: The W-test for normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 44*, 547-551.

Runions, A., Lane, B., & Prusinkiewicz, P. (2007). Modeling Trees with a Space Colonization Algorithm. In, *NPH* (pp. 63-70).

Sasaki, T., Imanishi, J., Ioki, K., Morimoto, Y., & Kitada, K. (2008). Estimation of leaf area index and canopy openness in broad-leaved forest using an airborne laser scanner in comparison with high-resolution near-infrared digital photography. *Landscape and Ecological Engineering, 4*, 47-55.

Seidel, D., Fleck, S., & Leuschner, C. (2012). Analyzing forest canopies with ground-based laser scanning: A comparison with hemispherical photography. *Agricultural and Forest Meteorology, 154–155*, 1-8.

Skirvin, D.J. (2007). Modelling plant canopies for biocontrol and biodiversity. In J. Vos, L.F.M. Marcelis, P.H.B. DeVisser, P.C. Struik, & J.B. Evers (Eds.), *Functional-Structural Plant Modelling in Crop Production* (pp. 253-264).

Solberg, S. (2010). Mapping gap fraction, LAI and defoliation using various ALS penetration variables. *International Journal of Remote Sensing, 31*, 1227-1244.

Solberg, S., Brunner, A., Hanssen, K.H., Lange, H., Næsset, E., Rautiainen, M., & Stenberg, P. (2009). Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sensing of Environment, 113*, 2317-2327.

Starek, M., Luzum, B., Kumar, R., & Slatton, K. (2006). Normalizing Lidar intensities. *Rep_2006-12-001, Geosensing Engineering and Mapping (GEM), Civil and Coastal Engineering Department, University of Florida.*

Steege, H.t. (1996). WinPhot 5: a Programme to analyze Vegetation Indices, Light and Light Quality from Hemispherical Photograps. In T.G.P. Tropenbos Guyana Reports 95-2, Georgetown Guyana (Ed.). http://testweb.science.uu.nl/Amazon/winphot/wp_index.htm

Streutker, D.R., & Glenn, N.F. (2006). LiDAR measurement of sagebrush steppe vegetation heights. *Remote Sensing of Environment, 102*, 135-145.

Thomas, V., Noland, T., Treitz, P., & McCaughey, J.H. (2011). Leaf area and clumping indices for a boreal mixed-wood forest: lidar, hyperspectral, and Landsat models. *International Journal of Remote Sensing, 32*, 8271-8297.

Vaaja, M., Kukko, A., Kaartinen, H., Kurkela, M., Kasvi, E., Flener, C., Hyyppä, H., Hyyppä, J., Järvelä, J., & Alho, P. (2013). Data processing and quality evaluation of a boat-based mobile laser scanning system. *Sensors, 13*, 12497-12515.

Van der Zande, D., Hoet, W., Jonckheere, I., van Aardt, J., & Coppin, P. (2006). Influence of measurement set-up of ground-based LiDAR for derivation of tree structure. *Agricultural and Forest Meteorology, 141*, 147-160.

Van der Zande, D., Stuckens, J., Verstraeten, W.W., Mereu, S., Muys, B., & Coppin, P. (2011). 3D modeling of light interception in heterogeneous forest canopies using ground-based LiDAR data. *International Journal of Applied Earth Observation and Geoinformation, 13*, 792-800.

Van Gardingen, P., Jackson, G., Hernandez-Daumas, S., Russell, G., & Sharp, L. (1999). Leaf area index estimates obtained for clumped canopies using hemispherical photography. *Agricultural and Forest Meteorology, 94*, 243-257.

Van Genechten, B., Caner, H., Heine, E., Lerma, J.L., Poelman, R., & Santana Quintero, M. (2008). Theory and practice on Terrestrial Laser Scanning: Training material based on practical applications. In (p. 241): Flemish Agency of the European Leonardo Da Vinci program.

Voegtle, T., & Wakaluk, S. (2009). Effects on the measurements of the terrestrial laser scanner HDS 6000 (Leica) caused by different object materials. *Proceedings of ISPRS Work*, 68-74.

Vosselman, G. (2002). On the estimation of planimetric offsets in laser altimetry data. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, 34*, 375-380.

Walter, J.-M. (2009). CIMES-FISHEYE. Hemispherical Photography of Forest Canopies. A Package of Programs for the Determination of Canopy Geometry and Solar Radiation Regimes through Hemispherical Photographs. In. http://jmnw.free.fr/: Université de Strasbourg. France.

Walter, J.-M.N., Fournier, R.A., Soudani, K., & Meyer, E. (2003). Integrating clumping effects in forest canopy structure: an assessment through hemispherical photographs. *Canadian Journal of Remote Sensing, 29*, 388-410.

Wehr, A., & Lohr, U. (1999). Airborne laser scanning—an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing, 54*, 68-82.

Weiss, M., & Baret, F. (2010). CAN-EYE V6. 1 USER MANUAL. In F.N.I. of , & A.R. (INRA) (Eds.). http://www6.paca.inra.fr/can-eye/Download

Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., & Coppin, P. (2004). Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology, 121*, 37-53.

Welles, J.M., & Cohen, S. (1996). Canopy structure measurement by gap fraction analysis using commercial instrumentation. *Journal of Experimental Botany, 47*, 1335-1342.

Wilson, J.W. (1959). ANALYSIS OF THE SPATIAL DISTRIBUTION OF FOLIAGE BY TWO-DIMENSIONAL POINT QUADRATS. *New Phytologist, 58*, 92-99.

Xu, L., & Zhang, W.-J. (2001). Comparison of different methods for variable selection. *Analytica Chimica Acta, 446*, 475-481.

Yamamoto, K. (2004). LIA for Win 32 (LIA 32). In L.o.F.E.a. Resources, G.S.o.B. Sciences, N. University, & Japan (Eds.). http://www.agr.nagoya-u.ac.jp/~shinkan/LIA32/index-e.html

Zhang, Y., Chen, J.M., & Miller, J.R. (2005). Determining digital hemispherical photograph exposure for leaf area index estimation. *Agricultural and Forest Meteorology, 133*, 166-181.

Zhao, F., Strahler, A.H., Schaaf, C.L., Yao, T., Yang, X., Wang, Z., Schull, M.A., Román, M.O., Woodcock, C.E., Olofsson, P., Ni-Meister, W., Jupp, D.L.B., Lovell, J.L., Culvenor, D.S., & Newnham, G.J. (2012). Measuring gap fraction, element clumping index and LAI in Sierra Forest stands using a full-waveform ground-based lidar. *Remote Sensing of Environment, 125*, 73-79.

Zhao, F., Yang, X., Schull, M.A., Román-Colón, M.O., Yao, T., Wang, Z., Zhang, Q., Jupp, D.L.B., Lovell, J.L., Culvenor, D.S., Newnham, G.J., Richardson, A.D., Ni-Meister, W., Schaaf, C.L., Woodcock, C.E., & Strahler, A.H. (2011). Measuring effective leaf area index, foliage profile, and stand height in New England forest stands using a full-waveform ground-based lidar. *Remote Sensing of Environment, 115*, 2954-2964.

Zhen, Z., Wuming, Z., Ling, Z., & Jmg, Z. (2011). Research on different slicing methods of acquiring LAI from terrestrial laser scanner data. In, *Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on* (pp. 295-299)

6 LIST OF FIGURES

7 LIST OF TABLES

LIST OF ACRONYMS

ALS	Airborne laser scanner
CCI	Chen and Cihlar´s clumping index
CI	Canopy clumping index
FC	Fraction cover
GF	Canopy gap fraction
GPS	Global Positioning System
HP	Hemispherical photograph
LAI	Leaf Area Index
LiDAR	Light Detection And Ranging
PAR	Photosynthetic active radiation
PCS	Pielou´s spatial segregation coefficient
SimTLS-SHP	Simulated hemispherical photograph
TLS	Terrestrial laser scanner
TLS-SHP	Terrestrial laser scanner simulated hemispherical photograph
θ	Azimuth angle
φ	Zenith angle

LIST OF ACRONYMS FOR ALS HEIGHT METRICS

Meancanopy	Mean of canopy height
MeanP25canopy	Mean of percentile 25 of canopy height
MeanP25plot	Mean of percentile 25 of plot height
MeanP50canopy	Mean of percentile 50 of canopy height
MeanP50plot	Mean of percentile 50 of plot height
MeanP75canopy	Mean of percentile 75 of canopy height
MeanP75plot	Mean of percentile 75 of plot height
MeanP95canopy	Mean of percentile 95 of canopy height
MeanP95plot	Mean of percentile 95 of plot height
Meanplot	Mean height of plot
Meanup25c	Mean of percentile 25 and higher of canopy height
Meanup50c	Mean of percentile 50 and higher of canopy height
Meanup75c	Mean of percentile 75 and higher of canopy height
Meanup95c	Mean of percentile 95 and higher of canopy height
Mediancanopy	Median of canopy height
MedianP25canopy	Median of percentile 25 of canopy height

MedianP25plot	Median of percentile 25 of plot height
MedianP50canopy	Median of percentile 50 of canopy height
MedianP50plot	Median of percentile 50 of plot height
MedianP75canopy	Median of percentile 75 of canopy height
MedianP75plot	Median of percentile 75 of plot height
MedianP95canopy	Median of percentile 95 of canopy height
MedianP95plot	Median of percentile 95 of plot height
Medianplot	Median height of plot
Medianup25c	Median of percentile 25 and higher of canopy height
Medianup50c	Median of percentile 50 and higher of canopy height
Medianup75c	Median of percentile 75 and higher of canopy height
Medianup95c	Median of percentile 95 and higher of canopy height
P25canopy	Percentil 25 of canopy height
P25plot	Percentil 25 of plot height
P50canopy	Percentil 50 of canopy height
P50plot	Percentil 50 of plot height
P75canopy	Percentil 75 of canopy height
P75plot	Percentil 75 of plot height
P95canopy	Percentil 95 of canopy height
P95plot	Percentil 95 of plot height
Stdcanopy	Standard deviation of canopy height
StdP25canopy	Standard deviation of percentile 25 of canopy height
StdP25plot	Standard deviation of percentile 25 of plot height
StdP50canopy	Standard deviation of percentile 50 of canopy height
StdP50plot	Standard deviation of percentile 50 of plot height
StdP75canopy	Standard deviation of percentile 75 of canopy height
StdP75plot	Standard deviation of percentile 75 of plot height
StdP95canopy	Standard deviation of percentile 95 of canopy height
StdP95plot	Standard deviation of percentile 95 of plot height
Stdplot	Standard deviation of plot height
Stdup25c	Standard deviation of percentile 25 and higher of canopy height
Stdup50c	Standard deviation of percentile 75 and higher of canopy height
Stdup75c	Standard deviation of percentile 75 and higher of canopy height
Stdup95c	Standard deviation of percentile 95 and higher of canopy height