Estimators and confidence intervals for plant area density at voxel scale with T-LiDAR

François Pimont, Denis Allard, Maxime Soma, Jean-Luc Dupuy

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Estimators and confidence intervals for plant area density at voxel scale with T-LiDAR

François Pimont[1,*], Denis Allard[2], Maxime Soma[1], Jean-Luc Dupuy[1]

[1] URFM, INRA, 84914, Avignon, France

[2] BioSP, INRA, 84914, Avignon, France

[*] Corresponding author. francois.pimont@inra.fr

Abstract:

Terrestrial LiDAR becomes more and more popular to estimate leaf and plant area density. Voxel-based approaches account for this vegetation heterogeneity and significant work has been done in this recent research field, but no general theoretical analysis is available. Although estimators have been proposed and several causes of biases have been identified, their consistency and efficiency have not been evaluated. Also, confidence intervals are almost never provided.

In the present paper, we solve the transmittance equation and use the Maximum Likelihood Estimation (MLE), to derive unbiased estimators and confidence intervals for the attenuation coefficient, which is proportional to leaf area density. The new estimators and confidence intervals are defined at voxel scale, and account for the number of beams crossing the voxel, the inequality of path lengths in voxel, the size of vegetation elements, as well as for the variability of element positions between vegetation samples. They are completed by numerous numerical simulations for the evaluation of estimator consistency and efficiency, as well as the assessment of the coverage probabilities of confidence intervals.

• Although commonly used when the beam number is low, the usual estimators are strongly biased and the 95% confidence intervals can be ≈ ±100% of the estimate.

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

• Our unbiased estimators are consistent in a wider range of validity than the usual ones, especially for the unbiased MLE, which is consistent when the beam number is as low as 5. The unbiased MLE is efficient, meaning it reaches the lowest residual errors that can be expected (for an unbiased estimator). Also the unbiased MLE does not require any bias correction when path lengths are unequal.

• When elements are small (or voxel is large), 10^3 beams entering the voxel leads to some confidence intervals ≈ ±10%, but when elements are larger (or voxel smaller), it can remain wider than ±50%, even for a large beam number. This is explained by the variability of element positions between vegetation samples. Such a result shows that a significant part of residual error can be explained by random effects.

• Confidence intervals are much smaller (±5 to 10%) when LAD estimates are averaged over several small voxels, typically within a horizontal layer or in the crown of individual plants. In this context, our unbiased estimators show a reduction of 50% of the radius of confidence intervals, in comparison to usual estimators.

Our study provides some new ready-to-use estimators and confidence intervals for attenuation coefficients, which are consistent and efficient within a fairly large range of parameter values. The consistency is achieved for a low beam number, which is promising for application to airborne LiDAR data. They entail to raise the level of understanding and confidence on LAD estimation. Among other applications, their usage should help determine the most suitable voxel size, for given vegetation types and scanning density, whereas existing guidelines are highly variable among studies, probably because of differences in vegetation, scanning design and estimators.

Keyword: terrestrial LiDAR; TLS; LAI; LAD; element size; bias; consistency; efficiency

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Highlights:

• Voxel-based estimations of LAD/PAD may lack of consistency and efficiency

• We propose new estimators based on theoretical derivation and numerical simulations

• Estimators for confidence intervals are also provided

• New estimators should help determine the most appropriate voxel resolution

1. Introduction

The amount and spatial distribution of foliage in a tree canopy have a fundamental function in ecosystems by affecting energy and mass fluxes through photosynthesis and transpiration. Moreover, canopy structure may reveal plant adaptation strategies to their physical or biotic environment (Norman and Campbell, 1989). Canopy foliage has other important ecological functions since it constitutes the crown fuels involved in high intensity forest fires (Keane, 1995) and its spatial structure may determine the habitat quality for animal species (Vierling et al., 2008). Terrestrial LiDAR (Light Detection And Ranging), referred to hereinafter as TLS (Terrestrial LiDAR System) recently emerged as a promising tool to estimate leaf/plant area density (LAD/PAD) distribution for individual plants and forest plots. Although similar traversal algorithms have recently been used with high resolution airborne data, acquisitions still suffer from substantial occlusion. This occlusion could be reduced with large flight strip overlaps (Kükenbrik et al., 2017), which would lead to a promising application of methods initially developed for TLS. Two classes of methods are commonly applied to derive LAD distributions with TLS. First, the leaf area density profile can be measured through a gap fraction approach (Jupp et al., 2009; Zhao et al., 2011). Rigorous statistical analysis using maximum likelihood estimator (MLE) has been applied to the gap fraction equation inversion, leading to robust estimates of LAD and leaf angulation profiles (Zhao et al., 2015).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Unfortunately, the gap fraction approach does not explicitly account for spatial correlation in vegetation distribution (Zhao et al., 2015), whereas spatial correlation in heterogeneous media are known to modify transmission laws and free path distribution (Davis and Marshak, 2004; Pimont et al., 2009; Larsen and Clark, 2014). A clumping factor is thus required (Chen and Cihlar, 1995; Zhao et al., 2011). Stochastic geometry entails to explicitly account for such clumping, but only to determine the leaf area index, LAI (Allard et al., 2013), which is the integral of the LAD over the vertical. The second class of methods is voxel-based and explicitly account for clumping at scales larger than voxel size. They entail to assess not only the vegetation vertical profile, but the full 3D distribution of area or mass density. Several approaches have been developed: the voxel-based profiling (Hosoi and Omasa, 2006 & 2007; Bailey and Mahafee, 2017a), the relative density index (Durrieu et al., 2008; Pimont et al., 2015), the modified contact frequency (Béland et al., 2011) and the Beer-Lambert approach (Béland et al., 2014b; Grau et al., 2017; Bailey and Mahafee, 2017a). These theoretical indices can be readily applied or combined with field measurements through a calibration phase (e.g., in Pimont et al., 2015).

The application of physical principles such as turbid media and contact frequency to voxelized-TLS data raises several problems that folds in two categories. The first one deals with departure from ideal measurements due to TLS "flaws". An idealized TLS would send an infinite number of infinitely thin beams on any voxel. The actual diameter of the beam (on the order of a few mm) is responsible for partial hits (Hebert and Koktov, 1992; Béland et al., 2011; Grau et al., 2017). There is also uncertainty regarding beam intensity, due to the noise instrument gain, that affects the detection (Grau et al., 2017). Béland et al., (2011) proposed an approach that accounts for partial hits and intensity through a calibration of intensity and view factors. Another aspect that has received little attention until now is the number of beams entering the voxel (sampling size). The beam number depends on the distance to

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

scanner, the direction and scanner resolution, as well as the interaction with vegetation which limits the number of beams reaching a given background voxel (occlusion). A basic rule of the thumb is to only consider measurements with beam number larger than 10 (Béland et al., 2014a) or to compute indices in large voxels, which leads to a fairly large beam number in most cases (Bailey and Mahafee, 2017a).

The second cause of departure from ideal measurement is vegetation "flaws". An idealized vegetation would be made of leaf only, assumed to be infinitely small elements with random distribution of position and orientation. The actual orientation and size of elements alter transmission laws (Larsen and Clark, 2014) and can be accounted for as in Béland et al. (2011), where leaf orientation is separately measured and the interaction between a single leaf and a beam is modelled. Element and branch orientations have been reported to be of secondary importance in comparison to other sources of errors (Grau et al., 2017; Seielstadt et al., 2011; Pimont et al., 2015). However, the assumption of spherical leaf inclination is not valid in many cases (Pisek et al., 2013), which suggests that the assessment of the proper angle distribution is likely to reduce errors. A recent method based on triangulation entails to estimate the orientation factor with a TLS, provided that leaves are large enough to be individually sampled by several beams (Bailey and Mahafee, 2017b). Regarding element size, Grau et al. (2017) reports little effects when elements are much smaller than grid size. This effect has been demonstrated to vary with voxel size (Béland et al., 2014b). Finally, several methods based on return intensity have been proposed to separate leaf from wood returns and to account for it (Béland et al., 2011, 2014a), even though such a method can not be applied to all TLS (Pimont et al., 2015). Despite these known issues, a detailed analysis of the consistency is still missing.

The determination of confidence intervals on LAD estimates has received little attention until now. If such estimators are known in the context of gap fraction approaches (Zhao et al.,

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

2015), confidence interval for voxel-based approaches are seldom proposed (with the exception of Pimont et al., 2015). Most error evaluations are based on simple comparison to experimental data, in which various sources of bias and dispersion may interact. This might explain why there is no consensus about the selection of voxel size among studies (e.g., Béland et al., 2011, 2014a; Bailey and Mahafee, 2017a; Grau et al., 2017). More generally, a rigorous statistical analysis of estimators such as in Zhao et al. (2015) is still missing for voxel-based approaches.

In the present study, we focus on some of the vegetation "flaws". We set our approach in the framework of random set theory, stochastic geometry and stereology (Stoyan et al., 1987; Schneider and Weil, 2008). In stochastic geometry, random distributions of geometrical objects such as points, segments and disks are analyzed and analytical expressions are derived for geometrical characteristics such as mean volume and area, specific surface etc. Stereology is concerned with the estimation of those quantities with limited probing, in particular in lower dimensions, such as beams probing a canopy voxel. We develop generalized estimators towards two different approaches: i) the resolution of the transmittance equation (also called Beer-Lambert law), ii) the maximum likelihood. Our developments are theoretical and validated through numerical simulations. They include bias corrections for the beam number, the element size, as well as for the variability of element positions between vegetation samples. Estimators for variance are also provided and can be used to compute confidence intervals. In order to facilitate the reading of the manuscript, most of the mathematical development are detailed in supplementary materials for reference and only the main equations are presented in the manuscript. Numerical simulations are used to compare the new estimators to usual ones (Beer-Lambert, Modified Contact Frequency), through the analysis of their consistency (i.e. bias size) and of their efficiency at the scale of a single voxel or a group of voxels (i.e. 95% error). The application of the new estimators and their

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

confidence interval are then discussed, especially in the context of the determination of the most appropriate voxel size.

2. Background regarding the estimation of PAD/LAD through the attenuation coefficient

This section summarizes the existing knowledge regarding the estimation of attenuation coefficient in voxel-based approaches and defines a few notations.

2.1. Beer-Lambert law formulation for TLS

The transmittance T in small and randomly distributed vegetation elements with no scattering follows an exponential attenuation along a path of length δ, known as the Beer-Lambert law (Nilson, 1971; Ross, 1981):

$$T = e^{-\int_0^{\delta} \lambda(z)dz} \quad (1)$$

with λ the attenuation coefficient of the medium (m^{-1})

The plant area density (PAD, m^{-1}) is related to the attenuation coefficient (λ, m^{-1}):

$$PAD = \lambda/G \quad (2)$$

where G is the plant projection function, which is frequently assumed to be equal to 0.5.

The complementary to one of the transmittance is the absorbance, A. TLS can be used to estimate the absorbance of a vegetation sample with the relative density index I (also denoted RDI in text), defined for a volume of vegetation, further referred to as the voxel V with volume denoted $|V|$. The RDI is the ratio between the number of hits within the voxel (N_i), to the number of beams that reaches the voxel (N):

$$1 - T(\lambda) = A(\lambda) \approx I = \frac{N_i}{N} \quad (3)$$

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

When beams are aligned with one cell face for cubic voxels or when the geometry of voxels is spherical (Durrieu et al., 2008), the lengths of the different paths are equal. This is generally not true and a first order approximation of Eq. 1 can be obtained using (Béland et al., 2014b; Grau et al., 2017):

$$T \approx e^{-\lambda\bar{\delta}} \tag{4}$$

where $\bar{\delta}$ is the mean path length within the voxel.

Taking the logarithm of the transmittance equation (Eq. 4) and combining with Eq. 3 leads to the usual estimator of the attenuation coefficient:

$$\hat{\lambda} = -\frac{\log(1 - I)}{\bar{\delta}} \tag{5}$$

This estimator (later referred to as the usual Beer-Lambert estimator) assumes that the attenuation coefficient is constant in the voxel, that the vegetation elements are infinitely small and that path length variations are negligible. Unequal path lengths involve the variance of path length within voxels (Grau et al., 2017). An empirical correction that depends on voxel orientation is described in Béland et al. (2014b) for cubic voxels. Another approach is to use the secant method to solve the exponentially-weighted transmittance equation, since such an equation does not have an explicit solution (Bailey and Mahafee, 2017a).

$\hat{\lambda}$ is not defined when I = 1, i.e. when no beam travels beyond the voxel. This occurs with probability $\left(1 - e^{-\lambda\delta}\right)^N$ which is, for example, equal to 0.01 when $\lambda\delta = 1$ and N = 10. Although this probability is very low when N is high, such an event may happen quite often in any large voxelized scene, especially when the vegetation is dense. These cases can simply be ignored, as proposed in Béland et al. (2014a), considering these cases as "occluded", but it will be shown later that it leads to biases and loss of efficiency.

2.2. Modified Contact frequency formulation for TLS

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

The contact frequency of vegetation elements CF is the number of contacts per unit length of point quadrat (probe) (Warren Wilson, 1960):

$$\mathrm{CF} = \frac{\sum_{j=1}^{N} C_j^l}{N\delta} \tag{6}$$

where C_j^l is the number of leaf contacts for the j^{th} probing and δ the probe length, and assuming that the probing number N is large. It is related to plant area density in a similar manner as the attenuation coefficient:

$$\mathrm{PAD} \approx \mathrm{CF}/G \tag{7}$$

With a TLS, the laser represents a virtual probe that is intercepted by vegetation. However, the contact number cannot exceed one and only a fraction of the volume is explored by the beam. A direct application of this method with TLS thus leads to an underestimation of the attenuation coefficient (Bailey and Mahafee, 2017a). This method is adapted in Béland et al., (2011) for TLS data, thank to the volume fraction concept and leads to:

$$\tilde{\lambda} = \frac{\sum_{j=1}^{N} \mathbf{1}_{z_j<\delta_j}}{\sum_{j=1}^{N} z_j} = \frac{N_i}{N\bar{z}} = \frac{I}{\bar{z}} \tag{8}$$

where z_j is the length of the path actually explored (free path) by the j^{th} beam and $\mathbf{1}_B$ is the indicator function of event B ($\mathbf{1}_{z_j<\delta_j} = 1$ if the j^{th} beam hits vegetation inside the voxel and 0 otherwise).

This formulation assumes that the explored volume is statistically representative of the unexplored volumes. However, it does not assume the equality of path lengths (contrary to Beer-Lambert estimator).

2.3. Accounting for the size of vegetation elements

The finite size of elements (ie. the size of elements is larger than zero in the real world) induces a bias in the above estimators when the element size is not negligible when compared to voxel size (Béland et al., 2014a). Assuming that the beams are parallel, let S_1 and S be

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

respectively the cross sections of the element size and voxel volume, which will be assumed constant for simplicity. The probability that a given beam crosses the voxel containing p elements is $\left(1-\frac{S_1}{S}\right)^p$, as shown in Campbell and Norman (1998, chapter 15).

The volume cross section is given by $S=\frac{|V|}{\delta}$, where δ is the path length. The contribution of a single leaf to the attenuation coefficient of the voxel V is:

$$\lambda_1=\frac{S_1}{|V|}=\frac{S_1}{S}\frac{1}{\delta} \tag{9}$$

Since $\lambda = p\lambda_1$, the transmittance of the voxel is:

$$T\approx\left(1-\frac{S_1}{S}\right)^p=(1-\lambda_1\delta)^{\frac{\lambda}{\lambda_1}} \tag{10}$$

When the element size is not neglected is, the Beer-Lambert estimator is modified as follow:

$$\widehat{\lambda_p}=\frac{\lambda_1\log\left(1-I\right)}{\log\left(1-\lambda_1\delta\right)} \tag{11}$$

As show with slightly different notations in Béland et al. (2014a), $\widehat{\lambda_p}=\frac{\hat{\lambda}}{-\frac{1}{\lambda_1\delta}\log\left(1-\lambda_1\delta\right)}$, which converges to $\hat{\lambda}$ when $R=\frac{S}{S_1}=\frac{1}{\lambda_1\delta}$ is large, or equivalently, when $\lambda_1\delta$ tends to 0.

In the discussion section of Béland et al. (2014a), the same correction factor is suggested to apply to the modified contact frequency, so that the modified contact frequency for finite size element would be:

$$\widetilde{\lambda_p}=-\frac{\lambda_1\bar{\delta}}{\log\left(1-\lambda_1\bar{\delta}\right)}\frac{I}{\bar{z}} \tag{12}$$

Depending whether elements are small needles or broad leaves, $\lambda_1\delta$ typically range between $2\ 10^{-5}\delta^{-2}$ and $5\ 10^{-3}\delta^{-2}$ (See supplementary S1). This means that $\lambda_1\delta$ ranges between 0.002 and 0.5 when the voxel size is about 10 cm, and between 0.0002 and 0.05 when the voxel size is 30 cm. $\lambda_1\delta$ is smaller than 0.005 when the voxel size is on the order of 1 m.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

3. Mathematical formulation

In this section, we develop the mathematical framework leading to unbiased estimators for point, variance and confidence intervals of the attenuation coefficient. The proofs are given in Supplementaries S2 and S3.

3.1. Set-up and notations

We assume a finite number of elements and we rely on the notations defined in 2.3. For a given quantity A, such as an estimator, a variance or a confidence interval radius, $\hat{A}$ denotes the quantity as derived from the Beer-Lambert law, whereas $\tilde{A}$ is derived from the Maximum Likelihood Estimator approach. As shown below, the MLE generalizes the Modified Contact Frequency introduced in Béland et al. (2011), so that these symbols are consistent with section 2 (Eq. 8). Furthermore, the use of upper case letters, such as Λ, refers to our new estimators presented below, whereas lower case letters, such as λ, refers to the usual ones presented in the background section.

We briefly present some stochastic geometry material and refer to the literature for a more detailed exposition (Stoyan et al., 1987; Schneider and Weil, 2008). p elementary objects identical in shape and size are located at random within the voxel V. If one representative of these objects is denoted B, the vegetation elements, denoted X, corresponds to the union of all objects:

$$X = \bigcup_{k=1}^{p} B(x_k) \tag{13}$$

where $B(x_k)$ denotes the element B shifted to the random location $x_k \in V$.

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

One special case of interest is the Boolean model, for which the number of objects *p* is distributed according to a Poisson distribution (Stoyan et al., 1987). A remarkable feature of Boolean models is that the intersection of a Boolean model by a random line is also a Boolean model with intensity $\lambda = S_1 \frac{p}{|V|}$, where S_1 is the cross-section of *B* perpendicular to the line. As a consequence, the lengths *Y* of segments with no intersection with *X*, called free paths in the present context, are distributed according to an exponential random variable with parameter λ:

$$P(Y > y) = e^{-\lambda y}, y > 0 \quad (14)$$

When the object number is not assumed to be large, the distribution becomes:

$$P(Y > y) = (1 - \lambda_1 y)^{\lambda/\lambda_1} \quad (15)$$

with $\lambda_1 = \frac{S_1}{|V|}$.

Let *M* be the segment corresponding to the intersection between a beam and the voxel *V*. Depending on voxel shape and size, its length δ follows a distribution Δ. The distance actually traveled in voxel Z by a beam corresponds to:

$$Z = \min(Y, \Delta) \quad (16)$$

The probability distribution of Z is derived from Eq. 15:

$$f_Z(z; \delta) = \begin{cases} \lambda(1 - \lambda_1 z)^{\lambda/\lambda_1 - 1} \text{ when } z < \delta \\ (1 - \lambda_1 \delta)^{\lambda/\lambda_1} \text{ when } z = \delta \end{cases} \quad (17)$$

In the Z distribution, the density of Y for Y> δ is cumulated at z = δ due to the “min” operator in Eq. 16. Note that there are two components of randomness in this set-up: a random realization of a Boolean model *X*, on the one hand (i.e. element positions for a vegetation sample) and a random beam *M* over the cross-section of the voxel *S*, on the other hand (i.e. instrument sampling).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Let us consider a given realization of the Boolean model X and N beams $\{m_j\}_{j\leq \mathrm{N}}$ distributed over the voxel. We can define the RDI (Eq. 3) as the fraction of beams hitting the canopy elements as:

$$\mathrm{I}(X,\{m_j\}_{j\leq \mathrm{N}}) = \frac{\mathrm{N_i}(X)}{\mathrm{N}} = \frac{\sum_{\mathrm{j}=1}^{\mathrm{N}} \mathbf{1}_{\mathrm{z_j}(X,m_j)<\delta_\mathrm{j}(m_j)}}{\mathrm{N}} \tag{18}$$

Let us now denote $\mathrm{I}_\infty(X)$ the asymptotic RDI, which is the expectation of $\mathrm{I}(X)$ with respect to the instrument sampling ($i.e.$ $\mathrm{N} \to \infty$):

$$\mathrm{I}_\infty(\mathrm{X}) = \mathrm{E}_M\left[\mathbf{1}_{\mathrm{Y}(X)<\Delta}\right] \tag{19}$$

3.2. Point, Variance and Confidence Interval of the relative density index (RDI)

The RDI defined in Eq. (18) for N beams has the same expectation as the asymptotic RDI over all configurations X, since beams are drawn randomly across *S*.

This expectation is, according to Eq. (15):

$$E_X[\mathrm{I}_\infty(\mathrm{X})] = \mathrm{E}_{M,X}\left[\mathbf{1}_{\mathrm{Y}(X)<\Delta}\right] = \frac{1}{\mathrm{S}}\int_S \mathrm{P}\big(Y<\delta(s)\big)ds = 1-\frac{1}{\mathrm{S}}\int_S (1-\lambda_1\delta(s))^{\lambda/\lambda_1}ds \tag{20}$$

As pointed out above, the variance of I has two components. The conditional variance formula provides:

$$\begin{aligned}\mathrm{Var}(\mathrm{I}(\mathrm{X})) &= E_X[\mathrm{Var}_M(\mathrm{I}(\mathrm{X}))] + \mathrm{Var}_X(E_M[\mathrm{I}(\mathrm{X})]) \\ &= E_X[\mathrm{Var}_M(\mathrm{I}(\mathrm{X})|\mathrm{I}_\infty(X))] + \mathrm{Var}_X(\mathrm{I}_\infty(\mathrm{X}))\end{aligned} \tag{21}$$

Now, assuming that the N beams are independent and identically distributed, one gets $\mathrm{Var}_M(\mathrm{I})(X) = \frac{\mathrm{I}_\infty(\mathrm{X})(1-\mathrm{I}_\infty(\mathrm{X}))}{\mathrm{N}}$, since I is simply a proportion estimated on a sample of size N.

The variance $\mathrm{Var}_X(\mathrm{I}_\infty(X))$ is due to the variability of element positions in a vegetation sample. This variance becomes negligible when the vegetation sample is made of a large number of small area elements, but it cannot be evaluated in closed form for actual configurations of *X*. It is instead approximated by numerical simulations, as described in section 4 (for $\lambda_1\overline{\delta} < 0.3$):

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

$$\mathrm{Var}_X(\mathrm{I}_\infty(X)) \approx \sigma^2_{\mathrm{I}_\infty}\big(\mathrm{I}_\infty(X)\big) = 0.230\lambda_1\bar{\delta}\mathrm{I}_\infty(X)^{1.903-2.30\lambda_1\bar{\delta}}\big(1-\mathrm{I}_\infty(X)\big) \tag{22}$$

From now on, we drop the dependence to X and M for the ease of notations. Putting these results together leads to the following estimator for the variance of I:

$$\sigma_I^2 = \frac{\mathrm{I}(1-\mathrm{I})}{\mathrm{N}} + \sigma^2_{\mathrm{I}_\infty}(\mathrm{I}, \lambda_1\bar{\delta}) \tag{23}$$

Hence to the following Wald confidence interval for the expectation of the asymptotic RDI, I_∞:

$$\mathrm{I} \pm z_{\alpha/2}\sqrt{\frac{\mathrm{I}(1-\mathrm{I})}{\mathrm{N}} + \sigma^2_{\mathrm{I}_\infty}(\mathrm{I}, \lambda_1\bar{\delta})} \tag{24}$$

where $z_{\alpha/2}$is the usual $1-\frac{\alpha}{2}$ quantile of the standard Gaussian distribution.

The Wald interval is known to have a lower-than-expected coverage probability when the empirical proportion (here, the RDI) is close to 0 or 1. It means that the actual value of λ is less frequently within the estimated interval than expected (Brown et al., 2001). As an example, it is obviously the case when I=0 (or 1), since the true value is supposed to be 0 (or 1) at 100 %. This is problematic since both cases are quite frequent in TLS data. Among others, the Agresti-Coull interval is a simple alternative to the Wald interval recommended in Brown et al. (2001). Its formulation is similar to that of the Wald interval:

$$\mathrm{I}_c \pm z_{\alpha/2}\sqrt{\frac{\mathrm{I}_c(1-\mathrm{I}_c)}{\mathrm{N}_c} + \sigma^2_{\mathrm{I}_\infty}(\mathrm{I}_c, \lambda_1\bar{\delta})} \tag{25}$$

with corrected values of "I" and "N" defined as follow:

$$\begin{cases} \mathrm{I}_c = \dfrac{\mathrm{I}+\frac{z^2_{\alpha/2}}{2\mathrm{N}}}{1+\frac{z^2_{\alpha/2}}{\mathrm{N}}} \\ \mathrm{N}_c = \mathrm{N} + z^2_{\alpha/2} \end{cases} \tag{26}$$

This correction leads to confidence intervals that are not centered on I and wider than the Wald interval. Agresti and Coull interval is known to have a higher-than-expected coverage

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

probability when N is small, which is not fully satisfactory, but safer than the Wald interval (Brown et al., 2001).

3.3. Point and Variance Estimation of the unbiased Beer-Lambert estimator

The Beer-Lambert estimator derives from solving the transmittance of the voxel medium (section 2.1) and thus rely on the empirical relative density index. In supplementary S2, we extend this approach to define the unbiased Beer-Lambert estimator $\hat{\Lambda}$, valid for close-to-equal path lengths:

$$\hat{\Lambda} = \begin{cases} -\frac{1}{\overline{\delta_e}}\left(log(1-I) + \frac{I}{2N(1-I)}\right) & \text{when } I < 1 \\ \frac{log(2N+2)}{\overline{\delta_e}} & \text{when } I = 1 \end{cases} \quad (27)$$

with the mean “effective” path length:

$$\overline{\delta_e} = mean\left(-\frac{\log(1-\lambda_1\delta_j)}{\lambda_1}\right) \quad (28)$$

When $\lambda_1 \ll \frac{1}{\bar{\delta}}$, it simplifies to $\overline{\delta_e} \approx \bar{\delta}$ (mean path length).

The first term in Eq. (27) when I<1 accounts for the size of elements (Eq. (11)). The second term is a bias correction for the instrument sampling, that compensates a systematic bias caused by the convexity of the log function (See Supplementary S2 for details). Such a bias has never been reported before, to the best of our knowledge.

When I=1, the above formulation is derived from the application of the Beer-Lambert law to the center of the Agresti-Coull interval, which is more robust than I.

For unequal path lengths, the transmittance equation can be approximated as a second order polynomial in λ, which leads to the following unbiased estimator:

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

$$\hat{\Lambda}_2 = \frac{1}{a_e}\left(1 - \sqrt{1 - 2a_e\hat{\Lambda}}\right) \tag{29}$$

with $\hat{\Lambda}$ in Eq. (27) and a_e the ratio between empirical variance to mean of the effective path length:

$$a_e = var\left(-\frac{\log(1-\lambda_1\delta_j)}{\lambda_1}\right) \Big/ mean\left(-\frac{\log(1-\lambda_1\delta_j)}{\lambda_1}\right) \tag{30}$$

Notice that when $\lambda_1 \ll \frac{1}{\bar{\delta}}$, $a_e \approx \frac{\sigma_\delta^2}{\bar{\delta}}$, with σ_δ^2 the empirical variance of path lengths.

The variances of both unbiased estimators $\widehat{\Lambda}$ and $\widehat{\Lambda}_2$ can be derived from the variance of the RDI (Eq. 23), as shown in Supplementary S2:

$$\sigma_{\hat{\Lambda}}^2 = \begin{cases} \dfrac{1}{\overline{\delta_e}^2}\left(\dfrac{I}{N(1-I)} + \dfrac{\sigma_{\mathrm{I}_\infty}^2(\mathrm{I},\lambda_1\bar{\delta})}{(1-I)^2}\right)\left(1 - \dfrac{1}{2N(1-I)}\right)^2 & \text{when } I < 1 \\ \dfrac{1}{\overline{\delta_e}^2}\left(2 + \dfrac{1}{N} + (2N+2)^2\sigma_{\mathrm{I}_\infty}^2\left(\dfrac{2N+1}{2N+2},\lambda_1\bar{\delta}\right)\right) & \text{when } I = 1 \end{cases} \tag{31}$$

and

$$\sigma_{\hat{\Lambda}_2}^2 = \sigma_{\hat{\Lambda}}^2\left(1 + 2\mathrm{a}_e\hat{\Lambda} + 4\left(\mathrm{a}_e\hat{\Lambda}\right)^2\right) \tag{32}$$

These estimators account for the instrument sampling (with the 1/N term when I<1), the asymptotic variability of element positions between vegetation samples (terms with $\sigma_{\mathrm{I}_\infty}^2$) and the convexity of the log function (third factor when I<1). As above, the case corresponding to I=1 is based on the center of the Agresti-Coull interval.

3.4. Point and Variance Estimation from Maximum Likelihood Estimation

The following estimator is derived from Maximum Likelihood (Kay, 1993, Chapter 7), that uses the full information provided by the TLS, namely the actual distribution of free paths in the voxel $\{z_j\}_{j=1,N}$ the N free paths. In supplementary S3, we compute the analytical MLE from the expected free path distribution (Eq. (17)). These derivations show that the modified

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

contact frequency proposed in Béland et al. (2011) is indeed the MLE, which demonstrates its asymptotic consistency when N is large. This formulation is extended to the case of elements of finite size, thanks to the "effective" mean free path:

$$\overline{z_e} = \mathrm{mean}\left(-\frac{\log(1-\lambda_1 z_j)}{\lambda_1}\right) \tag{33}$$

The MLE is asymptotically-normal, meaning that its residuals become normal when N is large (Kai, 1993). However, the MLE is biased when the number of beams is finite. In Supplementary S3, we account for this bias, which leads to the following unbiased MLE:

$$\widetilde{\Lambda} = \frac{\mathrm{I}}{\overline{z_e}} - \frac{\overline{\mathbf{1}_{z<\delta} z_e}}{\mathrm{N}\overline{z_e}^2} \tag{34}$$

with

$$\overline{\mathbf{1}_{z<\delta} z_e} = \mathrm{mean}\left(-\frac{\mathbf{1}_{z_j<\delta_j}\log(1-\lambda_1 z_j)}{\lambda_1}\right) \tag{35}$$

Compared to the Beer-Lambert estimator, this approach does not require any correction for unequal path lengths.

In supplementary S3, we rigorously compute the variance of $\widetilde{\Lambda}$ with bias correction for instrument sampling and the variability of element positions between vegetation samples:

$$\sigma_{\widetilde{\Lambda}}^2 = \frac{\mathrm{I}}{\mathrm{N}\overline{z_e}^2}\left(1 - \frac{\overline{\mathbf{1}_{z<\delta} z_e}}{NI\overline{z_e}}\right)^2 + \frac{\sigma_{I_\infty}^2(\mathrm{I_b}, \lambda_1\bar{\delta})}{\bar{\delta}_e^2(1-\mathrm{I_b})^2} \tag{36}$$

With

$$\mathrm{I_b} = \min\left(\mathrm{I}, 1 - \frac{1}{2N+2}\right) \tag{37}$$

The factor involving $\overline{\mathbf{1}_{z<\delta} z_e}$ expresses the bias correction for the instrument sampling and the term with $\sigma_{I_\infty}^2$ derives from the variability of vegetation samples, as for the variance of the unbiased Beer-Lambert estimator.

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

3.5. Cramer-Rao bound for variance

The Cramer-Rao bound is the theoretical lower bound of the variance of unbiased estimators (Kay, 1993, Chapter 3), meaning that an unbiased estimator with variance as small as the Cramer-Rao bound is optimal. In Supplementary S3, we show that when the vegetation samples are not fixed (i.e. when TLS beams are shot on variable vegetation samples), the Cramer-Rao bound is:

$$\mathrm{CRB}_\lambda = \frac{\lambda^2}{N I_\infty(\lambda)} \tag{38}$$

This analytical formulation is helpful, since the value of the Cramer-Rao bound can be analytically computed (integrating Eq. (20) to compute $I_\infty(\lambda)$), when both attenuation coefficient and voxel geometry are known. This is the case in the numerical simulations developed in sections 4 and 5, which thus provides a way to evaluate the efficiency of the unbiased estimators, that are expected to have empirical variances as close as possible to CRB_λ. A perfect match corresponds to the most efficient estimator.

It is important to notice however, that this theoretical bound can never be achieved when the variability of vegetation samples has a significant contribution to the variance of the RDI, since this variability is not accounted for in this theoretical bound (see Supplementary S3 for more details).

3.6. Estimating confidence intervals for a voxel or a group of voxel

From unbiased estimators $\widetilde{\Lambda}$ and $\sigma^2_{\widetilde{\Lambda}}$, the confidence interval at a risk level α can naturally be estimated as:

$$\widetilde{\Lambda} \pm z_{\alpha/2}\sqrt{\sigma^2_{\widetilde{\Lambda}}} \tag{39}$$

However, such a formulation is expected to have the same limitations as the Wald interval for the RDI, when the probabilities of interception (RDI) are low or high. This interval is thus

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

expected to lead to lower-than-expected coverage probabilities in voxels with low and high density. As for the Agresti-Coull interval, an alternative is to replace I by I_c and N by N_c in estimation of $\widetilde{\Lambda}$ and $\sigma^2_{\widetilde{\Lambda}}$, leading to:

$$\widetilde{\Lambda}_c \pm z_{\alpha/2}\sqrt{\sigma^2_{\widetilde{\Lambda}_c}} \tag{40}$$

If estimations at voxel scale is a key outcome of TLS, the scale of interest is often larger, typically the individual plants, the horizontal slice of vegetation, or the forestry plot. In this case, the variable of interest is not the single voxel estimation, but the average attenuation coefficient in a group of voxels. For a group of n_v voxels and assuming independence between voxels, the confidence interval on the mean of attenuation coefficient estimators is:

$$\frac{1}{n_v}\sum_{n_v}\widetilde{\Lambda} \pm \frac{z_{\alpha/2}}{n_v}\sqrt{\sum_{n_v}\sigma_{\widetilde{\Lambda}}{}^2} \tag{41}$$

The 95% errors, defined as the radius of the confidence interval at 95%, for a single voxel or a group of voxel are thus:

$$E95_{\widetilde{\Lambda}} = 1.96\sigma_{\widetilde{\Lambda}} \tag{42}$$

And

$$E95^{n_v}_{\widetilde{\Lambda}} = \frac{1.96}{n_v}\sqrt{\sum_{n_v}\sigma_{\widetilde{\Lambda}}{}^2} \tag{43}$$

Similar quantities can be defined for the unbiased Beer-Lambert estimators $\widehat{\Lambda}$ and $\widehat{\Lambda}_2$, as well as the bound of the 95% error, based on the Cramer-Rao bound for variance, which is the lower bound of 95% error for an unbiased estimator (i.e. no unbiased estimator can lead to smaller errors).

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

4. Design of the numerical experiments

4.1. Overview

The aim of our numerical simulations is to compare the estimates of attenuation coefficients to their true values. Simple configurations are generated to simulate replicates of virtual TLS point clouds in voxels filled with idealized vegetation of known properties. The numerous replicates enable to compute various statistics through a MonteCarlo approach, to evaluate the consistency and efficiency of each estimator, as well as the consistency of variance and confidence interval estimators. For each estimator, we compute its expectation, variance and 95 % errors. The 95 % errors are estimated as the 95th percentile of the absolute residuals and are evaluated for a single voxel and a group of voxels (here $n_v = 100$). We also compute the expectation of variance estimators, as well as the coverage probability of the estimated confident intervals, which is the empirical frequency at the true value of the attenuation coefficient belongs to the estimated confidence interval.

These statistics are computed for various values of attenuation coefficients, element sizes, voxel sizes, and beam numbers. Simulations are run for two different configurations described below. The first configuration, described in details in subsection 4.2, assumes finite element size (meaning that $\lambda_1 > 0$) and equal path lengths. Equal path lengths imply that the second order correction for the Beer-Lambert estimators (i.e. $\widehat{\Lambda}^2$) is not required. The computation of RDI and distance travelled is done over the actual distributions of vegetation elements (generated with random positions), each corresponding to a fixed vegetation sample *X*. Simulations are run for a large number of vegetation samples, which entails to compute the asymptotic variance of the RDI, $\sigma^2_{I_\infty}$, (when the beam number is infinite) for each value of attenuation coefficient and element size. The second set of simulations (subsection 4.3) is specifically designed to evaluate the correction for unequal path lengths involved in unbiased

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Beer-Lambert estimator $\widehat{\Lambda}_2$. For simplicity, we assume that vegetation elements are infinitely small ($\lambda_1 = 0$) and that the voxel is a sphere, so that the free-path distribution and the Cramer-Rao bound can be analytically solved (See Supplementary S4 for details). We separate both sets of simulations to facilitate the presentation of results.

In order to simplify the numerical experiment design and the presentation of the results, we build dimensionless quantities, namely the beam depth ($L_j = \lambda\delta_j$) and the voxel depth ($L = \lambda\bar{\delta}$). Likewise, $y_j = \lambda z_j$ is the free depth for beam j. We can notice that I can be computed from the distribution of depth $\{y_j\}$ since $\mathbf{1}(z_j < \delta_j) = \mathbf{1}(y_j < L_j)$. The element depth is $L_1 = \lambda_1\bar{\delta}$. More generally, dimensionless quantities can be derived for all quantities of interest developed above, as shown in Supplementary S5. The practical interest of this substitution is that computations can be done for a series of voxel depth values, and easily extrapolated to λ by simply dividing results by $\bar{\delta}$, instead of running simulations for series of $(\lambda, \bar{\delta})$ values.

4.2. Numerical simulations for finite-size elements and equal path lengths

Simulated vegetation samples correspond to flat square elements that are randomly distributed in a voxel, parallel to one face of the voxel. The virtual beams are sent perpendicular to elements, so that the path lengths are equal to cube size (Figure 1). The voxel and element depths are $L = \lambda\delta$ and $L_1 = \lambda_1\delta$, respectively.

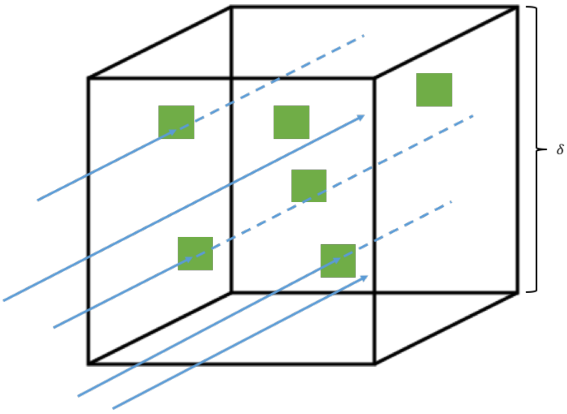


Fig. 1. Illustration of a numerical simulation of TLS beams over finite-size elements and equal path length δ. Each square element has a size equal to $\sqrt{S_1} = \sqrt{\lambda_1 \delta^3} = \sqrt{L_1}\delta$.

For each value of L and L_1 tested (Table 1), we simulate 10000 vegetation samples on which we shoot M batches of N virtual beams, with N between 5 and 10000. Batches serve as replicates of TLS shooting, to compute the different statistics (estimator expectation, variance, 95% error, variance and confidence interval estimator, confidence interval coverage probability). The batch numbers are selected so that the total number of beams MN is constant, equal to $= 10^8$. This number is large enough for the convergence of the different statistics, despite the replicate number M decreases with N, since the variance of the estimates sharply decays with N.

Table 1. Parameter values in numerical simulations

Parameter	Values
Voxel depth (L)	0.05, 0.1, 0.5, 1,1.5, 2, 2.5, 3
Beam number (N)	3,5,7,10,15,20,30,40,50,75,100,150,200,300,400, 500,750 1000,5000,10000
Element depth (L_1)	$\begin{cases} Subsection\ 4.2: 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.5 \\ Subsection\ 4.3: 0 \end{cases}$

Nb: in subsection 4.3, elements are assumed infinitely small so that L_1=0.

We use simulations with the largest beam number ($N = 10000 \approx \infty$) to estimate the asymptotic variance $\sigma^2_{I_\infty}$, for various values of L_1 and L. We remind that this variance is due to the variability of element position between vegetation samples *X*. When $L_1 < 0.3$, which is the case for most vegetation when voxels are greater than 10 cm (See Supplementary S1), $\sigma^2_{I_\infty}$ can be estimated from a simple empirical function of L_1 and RDI (Fig. 2 and Eq. (22)).

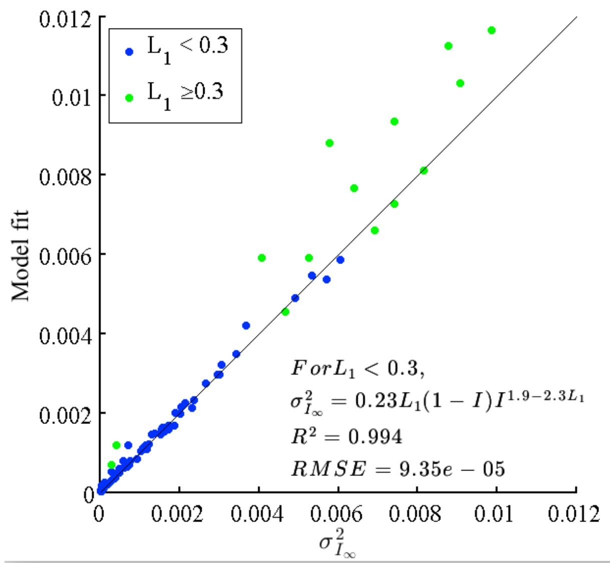


Fig. 2. Empirical model for the variance of the asymptotic relative density index I_∞. $\sigma^2_{I_\infty}$ is caused by the variability of element positions between vegetation samples.

4.3. Numerical simulations for unequal path lengths

In a sphere with radius R, the voxel depth is:

$$L = \lambda\overline{\delta} = \lambda\frac{V}{S} = \lambda\frac{\frac{4}{3}\pi R^3}{\pi R^2} = \frac{4}{3}\lambda R \tag{44}$$

The distribution of dimensionless optical depth L_j is, with u between 0 and 1 (Supplementary S4):

$$\mathrm{PDF}\left(L_j = \frac{3}{2}L\sqrt{1-u^2}\right) = 2u \tag{45}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

For each value of L (Table 1), we simulate a total of 10^8 virtual beams with lengths $\{L_j\}$ and the corresponding free path lengths $\{y_j\}$, that respectively follow Eq. (45) and the exponential law. As in subsection 4.2, virtual beams are grouped in batches of N beams to simulate M replicates of TLS shooting and to compute the same statistics as above.

5. Numerical simulation results

In this section, we show the statistics described in section 4. Subsection 5.1. corresponds to finite size element simulations (described in subsection 4.2), whereas subsection 5.2. corresponds to unequal path simulation (described in subsection 4.3). Expectation, variance and 95% error enable to compare the consistency and efficiency of the usual estimators ($\hat{\lambda}$ and $\tilde{\lambda}$) and the new ones ($\widehat{\Lambda}$, $\widehat{\Lambda}_2$ and $\widetilde{\Lambda}$). We also show the expectation of the variance estimators ($\sigma^2_{\widehat{\Lambda}}$, $\sigma^2_{\widehat{\Lambda}_2}$ and $\sigma^2_{\widetilde{\Lambda}}$), the confidence interval radiuses ($E95_{\widehat{\Lambda}}$, $E95_{\widehat{\Lambda}_2}$ and $E95_{\widetilde{\Lambda}}$), and the coverage probabilities of estimated confidence intervals.

5.1. Estimator performance for finite size elements

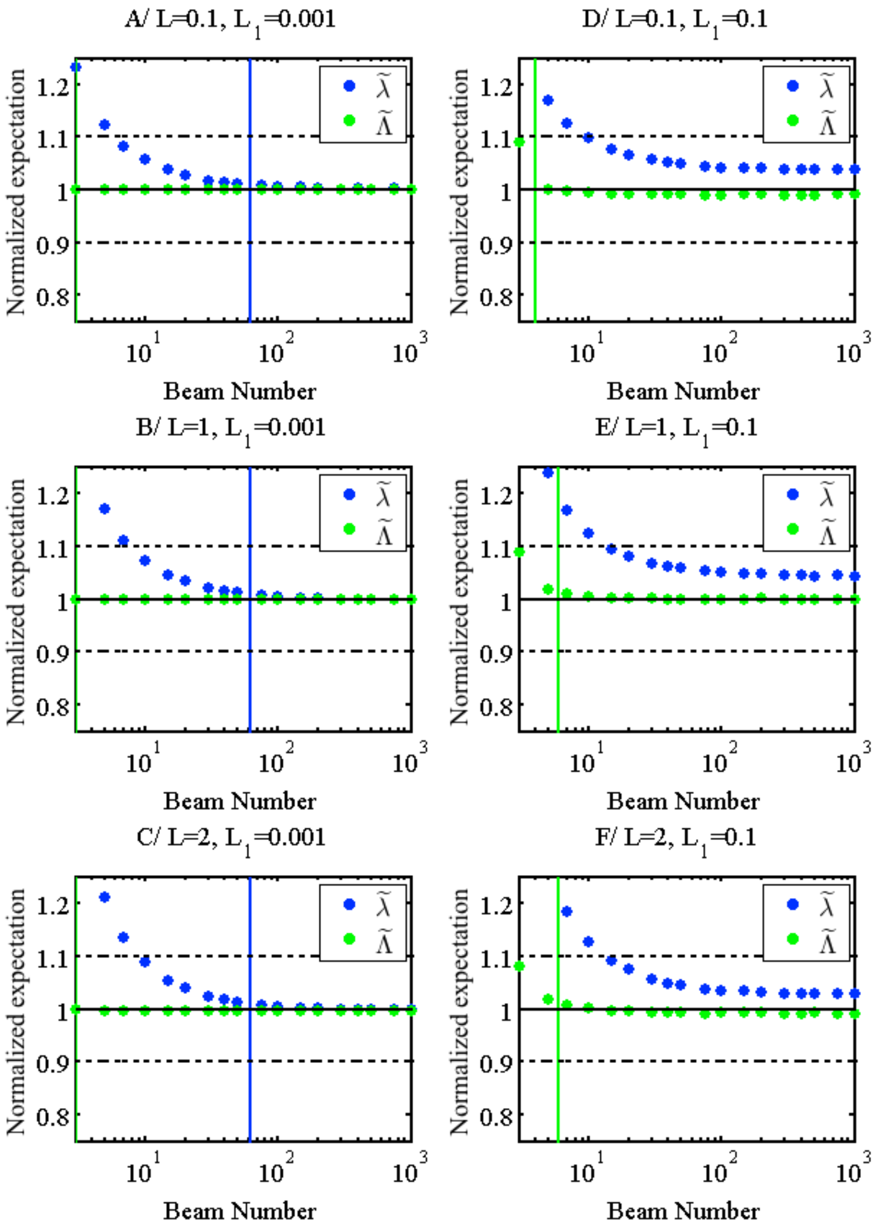
5.1.1 Estimator consistency

Figure 3 shows the expectation of estimator derived from the MLE, as a function of the beam number. Blue dots corresponds to the modified contact frequency estimator $\tilde{\lambda}$ (Eq. 8), which is the biased MLE for infinitely small elements. Green dots corresponds to the unbiased MLE $\widetilde{\Lambda}$ (Eq. 34), that accounts for element size and beam number. Since expectations are normalized by the true value of λ, estimators are consistent when expectations equal one, and deviations from 1 quantifies the bias. Subplots A, B and C correspond to small elements (L_l=0.001) for three voxel depths L, whereas subplots D, E and F correspond to "larger" elements compared to voxel size (L_l=0.1) for the same values of voxel depths.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Even when elements are small, the modified contact frequency is positively biased when N is small (subplots A, B, C). Vertical blue lines show the thresholds of N for which the bias of $\tilde{\lambda}$ is larger than 1%. These thresholds range between N=30 and N=75, depending on L. When elements are larger (subplots D, E, F), the positive biases remain for large values of N, so that the 1% threshold is not reached. On the contrary, the unbiased MLE $\tilde{\Lambda}$ (green dots) shows biases always lower than 1% (green lines) when elements are small, even when N is as small as 3 for L_1=0.001 (subplots A, B, C) and 5 when L_1=0.1 (subplots D, E, F).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Fig. 3. Expectations of the attenuation coefficient estimators derived from the MLE, as a function of the beam number. Blue dots corresponds to the modified contact frequency estimator $\tilde{\lambda}$ (Eq. 8), which is the biased MLE for infinitely small elements. Green dots corresponds to the unbiased MLE $\tilde{\Lambda}$ (Eq. 34), that accounts for element size and beam number. Estimators are normalized by their true value λ, so that they are consistent when the expectation equals to one. The vertical lines correspond to the lowest values of N leading to a bias smaller than 1% in blue and green for respectively the biased and unbiased estimators.

Figure 4 is similar to figure 3, but for the usual ($\hat{\lambda}$) and unbiased ($\hat{\Lambda}$) Beer-Lambert estimators. Trends are similar with two main differences. First, the biases are slightly larger with Beer-Lambert estimators than with the MLE when the vegetation density is low to moderate (L<=1, subplot A, B, D and E), whether bias are corrected or not. Second, the bias of the usual Beer-Lambert estimator decreases until becoming negative for small values of N when the vegetation density is high (subplots C and F). Such a decay is attributed to the occurrence of cases in which all beams are intercepted inside the voxel (I=1), referred to as "occluded" in Béland et al. (2014a). In this cases, the usual Beer-Lambert estimator is $+\infty$, but is ignored in expectation computation to avoid divergence. Attenuation coefficient estimates are thus bounded by $\frac{\log(N)}{\delta}$, leading to this negative bias of increasing magnitude when N is small. This trend is also visible for $\hat{\Lambda}$, but it is far less pronounced and it occurs for lower values of N. This demonstrates the benefit of the definition of $\hat{\Lambda}$, which is extended when I=1 with Agresti-Coull interval centers. The range of consistency of $\hat{\Lambda}$, however, is clearly narrower than the one of the unbiased MLE $\tilde{\Lambda}$.

 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

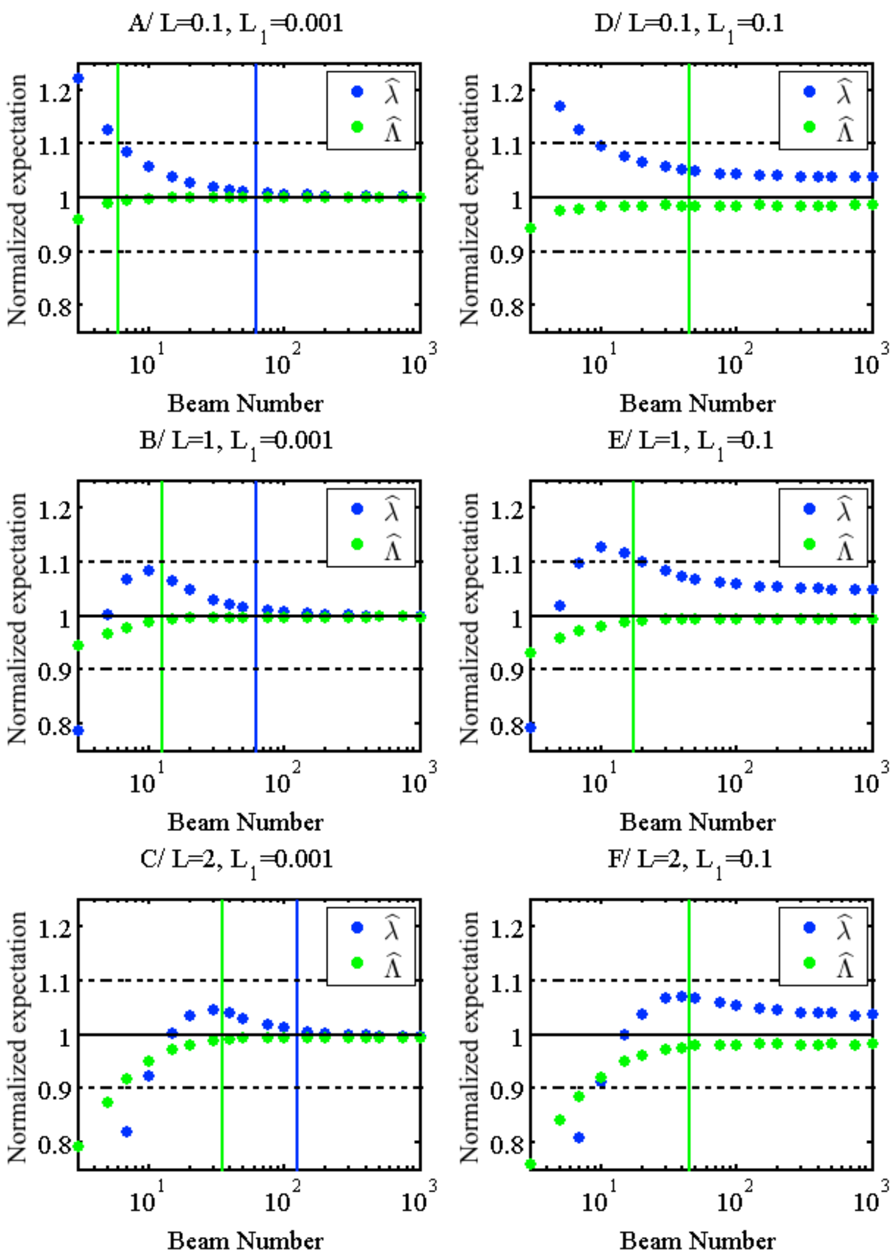


Fig. 4. Same as Figure 3 for the usual and unbiased Beer-Lambert estimators.

Expectations of the Beer-Lambert-attenuation-coefficient estimators, as a function of the beam number. Blue dots corresponds the usual Beer-Lambert estimator ($\hat{\lambda}$). Green dots corresponds to the unbiased Beer-Lambert estimator ($\widehat{\Lambda}$), that accounts for element size and beam number and extended definition when I=1. Estimators are normalized by their true value λ, so that they are consistent when the expectation equals to one. The vertical lines correspond to the lowest values of N leading to a bias smaller than 1% in blue and green, for respectively the biased and unbiased estimators.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

The computations of biases are done for all values of parameters in Table 1 and lead to the range of validity for three levels of consistency (1%, 5%, 10%) for the four estimators summarized in Table 2. As expected from Figures 3 and 4, the usual Beer-Lambert and modified contact frequency are biased in a much wider range than the corrected indices introduced in section 3, and generally requires smaller elements and a larger beam number to be consistent. Also, the biases of the Beer-Lambert estimators increase with density, which leads to less straightforward formulations of range of consistency, since their definition vary with both L_1 and L. MLE thus has wider range of validity than Beer-Lambert estimators.

Table 2. Range of consistency of the four estimators of attenuation coefficient for three consistency thresholds (biases smaller than 1, 5 and 10 %). NB: According to numerical simulations, the bias thresholds are never reached for values of L_1 and L that are out of the ranges provided below (even when N =10000).

Index	Consistency (1%)	Consistency (5%)	Consistency (10%)
$\tilde{\lambda}$	$L_1 \leq 0.01$ and $N \geq 100$	$\begin{cases} L_1 \leq 0.01 \text{ and } N \geq 20 \\ L_1 \leq 0.05 \text{ and } N \geq 30 \end{cases}$	$\begin{cases} L_1 \leq 0.01 \text{ and } N \geq 10 \\ L_1 \leq 0.1 \text{ and } N \geq 20 \end{cases}$
$\tilde{\Lambda}$	$\begin{cases} L_1 \leq 0.01 \text{ and } N \geq 3 \\ L_1 \leq 0.1 \text{ and } N \geq 5 \\ L_1 \leq 0.2 \text{ and } N \geq 15 \\ L_1 \leq 0.3 \text{ and } N \geq 30 \end{cases}$	$\begin{cases} L_1 \leq 0.05 \text{ and } N \geq 3 \\ L_1 \leq 0.1 \text{ and } N \geq 5 \\ L_1 \leq 0.3 \text{ and } N \geq 10 \end{cases}$	$\begin{cases} L_1 \leq 0.1 \text{ and } N \geq 3 \\ L_1 \leq 0.2 \text{ and } N \geq 5 \\ L_1 \leq 0.3 \text{ and } N \geq 7 \\ L_1 \leq 0.5 \text{ and } N \geq 10 \end{cases}$
$\hat{\lambda}$	$L \leq 2$ and $L_1 \leq 0.01$ and $N \geq 100$	$L_1 \leq 0.05$ and $N \geq 40$	$\begin{cases} L \leq 2 \text{ and } L_1 \leq 0.01 \text{ and } N \geq 10 \\ L \leq 2.5 \text{ and } L_1 \leq 0.1 \text{ and } N \geq 20 \\ L \leq 3 \text{ and } L_1 \leq 0.1 \text{ and } N \geq 30 \end{cases}$
$\hat{\Lambda}$	$\begin{cases} L \leq 0.5 \text{ and } L_1 \leq 0.2 \text{ and } N \geq 7 \\ L \leq 1 \text{ and } L_1 \leq 0.2 \text{ and } N \geq 10 \\ L \leq 1.5 \text{ and } L_1 \leq 0.2 \text{ and } N \geq 15 \\ L \leq 2 \text{ and } L_1 \leq 0.05 \text{ and } N \geq 40 \\ L \leq 2.5 \text{ and } L_1 \leq 0.005 \text{ and } N \geq 75 \\ L \leq 3 \text{ and } L_1 \leq 0.001 \text{ and } N \geq 75 \end{cases}$	$\begin{cases} L \leq 1 \text{ and } L_1 \leq 0.1 \text{ and } N \geq 5 \\ L \leq 1.5 \text{ and } L_1 \leq 0.2 \text{ and } N \geq 10 \\ L \leq 2 \text{ and } L_1 \leq 0.2 \text{ and } N \geq 15 \\ L \leq 2.5 \text{ and } L_1 \leq 0.2 \text{ and } N \geq 40 \\ L \leq 3 \text{ and } L_1 \leq 0.1 \text{ and } N \geq 75 \end{cases}$	$\begin{cases} L \leq 1.5 \text{ and } L_1 \leq 0.2 \text{ and } N \geq 5 \\ L \leq 2 \text{ and } L_1 \leq 0.3 \text{ and } N \geq 10 \\ L \leq 2.5 \text{ and } L_1 \leq 0.3 \text{ and } N \geq 20 \\ L \leq 3 \text{ and } L_1 \leq 0.3 \text{ and } N \geq 40 \end{cases}$

5.1.2 Estimator efficiency

Figure 5 shows the empirical variances (multiplied by δ^2 for the generality of results) of estimator derived from the MLE, similarly to Figure 3. As expected, the variances decay with the beam number. When the elements are large and the density is moderate to high (subplots E and F), variances remain significantly larger than zero, even for large beam numbers, because of the variability between vegetation samples. The variances of the biased and unbiased estimators are similar in magnitude, the former being slightly larger when the beam number is small. Both variances are very close to the theoretical Cramer-Rao bound (in black), when L_1 is small (subplots A, B and C). Since $\tilde{\Lambda}$ is unbiased when L_1 is small, it can thus be considered as efficient. When the elements are large and the vegetation is dense (subplots E and F), the variance of $\tilde{\Lambda}$ is much larger than the Cramer-Rao bound, even when N is large. This is because the Cramer-Rao bound does not account for asymptotic variability due to the variability of vegetation samples.

The green dashed lines corresponds to the expectation of the estimator of the variance of $\tilde{\Lambda}$, namely $\sigma_{\tilde{\Lambda}}^2$ (Eq. 36). Its expectation is very close to the empirical expectation of the variance of $\tilde{\Lambda}$ (green dots), demonstrating the consistency of our variance estimator when the beam number is larger than 5.

 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

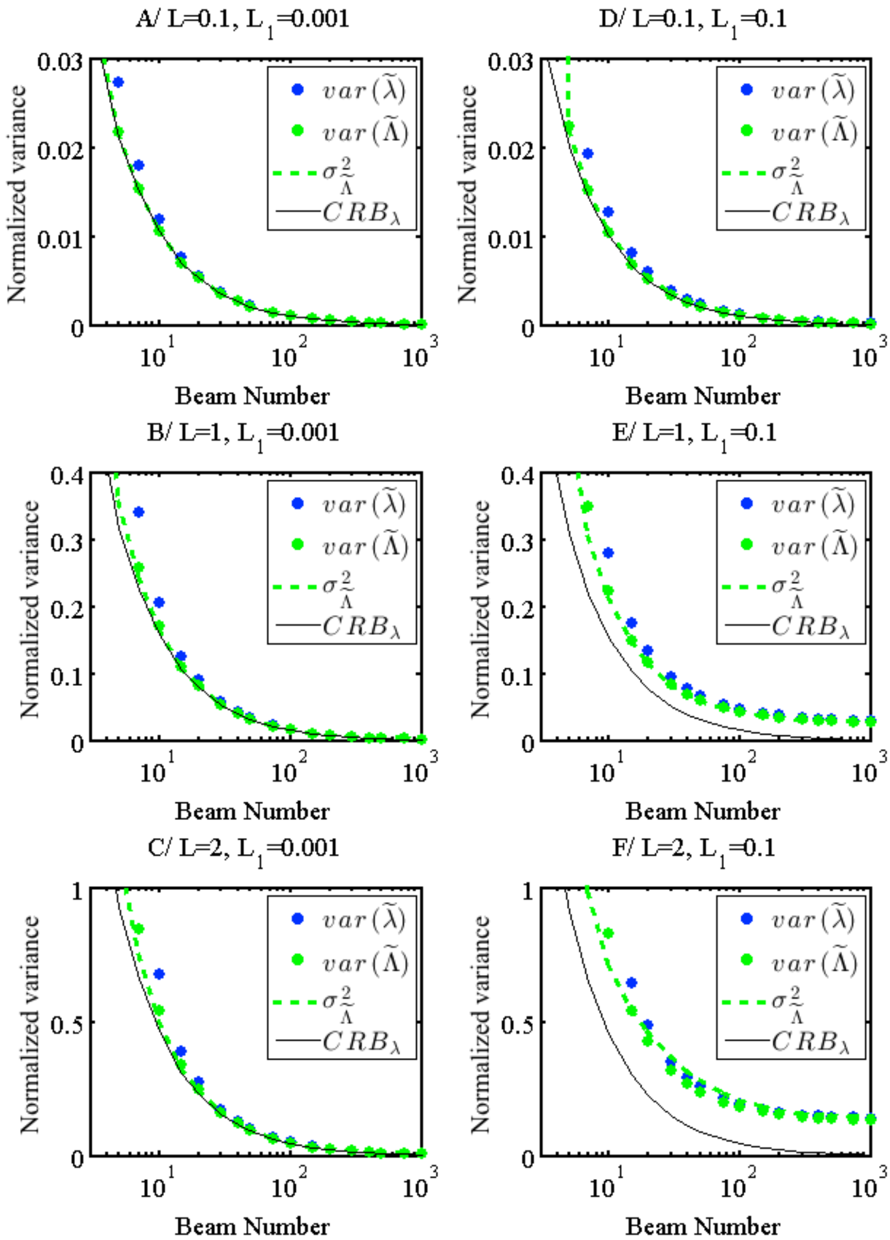


Fig. 5. Normalized variances of attenuation coefficient estimators derived from the MLE, as a function of the beam number. Normalized variances correspond to variances multiplied by δ^2. Blue dots corresponds to the variance of the modified contact frequency estimator $\tilde{\lambda}$. Green dots corresponds to the variance of the unbiased MLE $\tilde{\Lambda}$. Green dashed lines correspond to the dimensionless expectation of the variance estimator $\sigma^2_{\tilde{\Lambda}}$. The black line corresponds to Cramer-Rao bound for the variance of unbiased estimator.

 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Figure 6 is similar to Figure 5 for the variances of Beer-Lambert estimators. Although trends are similar, it is worth noting that the variance of $\widehat{\Lambda}$ is greater than the Cramer-Rao bound for small elements when vegetation is dense (Fig. 6C), showing that $\widehat{\Lambda}$ is suboptimal and less efficient than the MLE. Also, the expectation of the variance estimator $\sigma^2_{\widehat{\Lambda}}$ can significantly overestimate the empirical variance, showing a lack of consistency for this estimator. This situation mostly occurs in range of data where $\widehat{\Lambda}$ itself is biased (dense vegetation, low number of beams). The variance of the basic Beer-Lambert law can often be lower than the Cramer-Rao bound. This is simply another evidence that this estimator is strongly biased (due to the I=1 cases), since it would otherwise be greater than the Cramer-Rao bound.

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

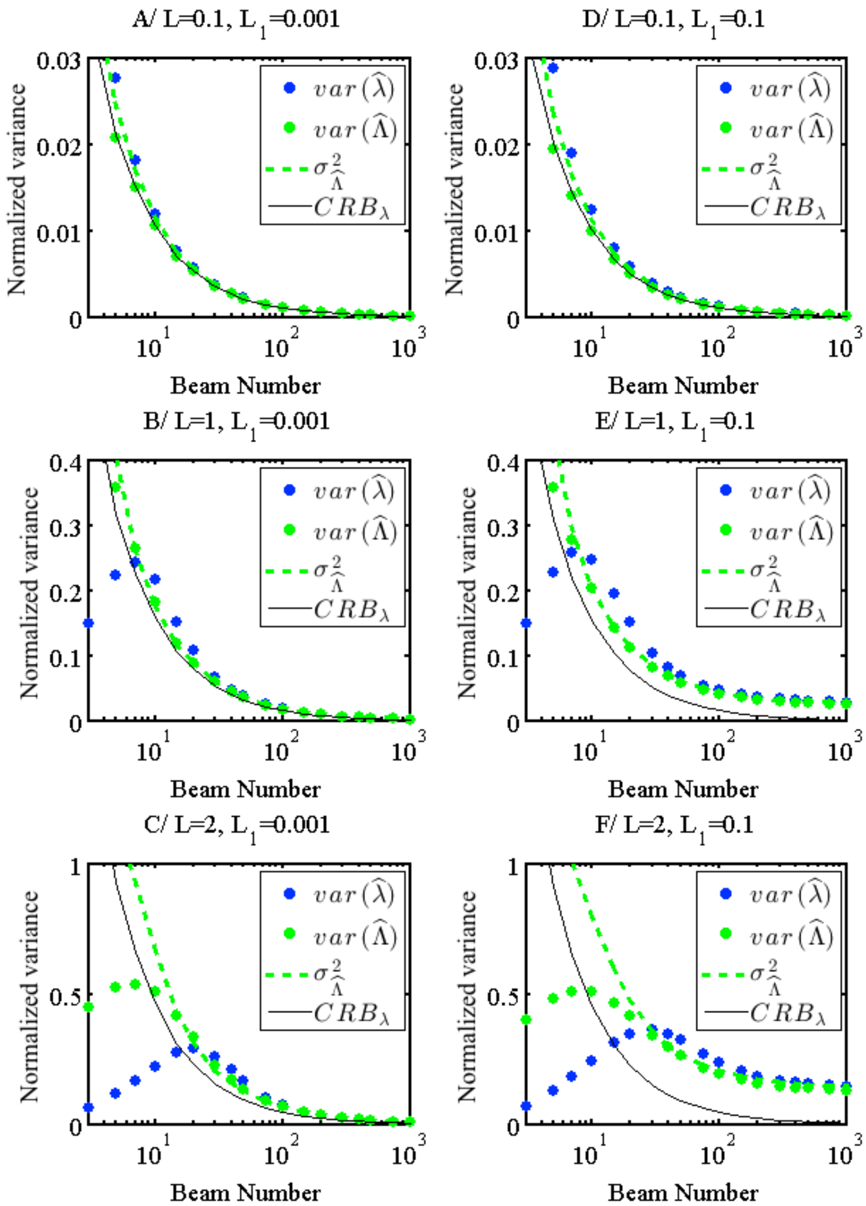


Fig. 6. Same as Figure 5 for Beer-Lambert estimators. Normalized variances of Beer-Lambert attenuation coefficient estimators, as a function of the beam number. Normalized variances correspond to variances multiplied by δ^2. Blue dots corresponds to the variance of the usual Beer-Lambert estimator $\widehat{\lambda}$. Green dots corresponds to the variance of the unbiased Beer-Lambert estimator $\widehat{\Lambda}$. Green dashed lines correspond to the dimensionless expectation of the variance estimator $\sigma^2_{\widehat{\Lambda}}$. The black line corresponds to Cramer-Rao bound for the variance of unbiased estimator.

5.1.3 Coverage probabilities of the estimated confidence intervals

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Figure 7 shows the coverage probabilities of the estimated confidence interval based on unbiased MLE, $\tilde{\Lambda} \pm z_{\alpha/2}\sqrt{\sigma^2_{\tilde{\Lambda}}}$ for three confidence levels (50, 90 and 95%). When the confidence intervals are correctly estimated, empirical coverage probabilities should match the confidence level (dashed lines). Estimated confidence intervals are satisfactory in most cases, with the exception of low density when the beam number is low (subplots A and B), for which the true value is less frequently in the confidence interval than expected.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

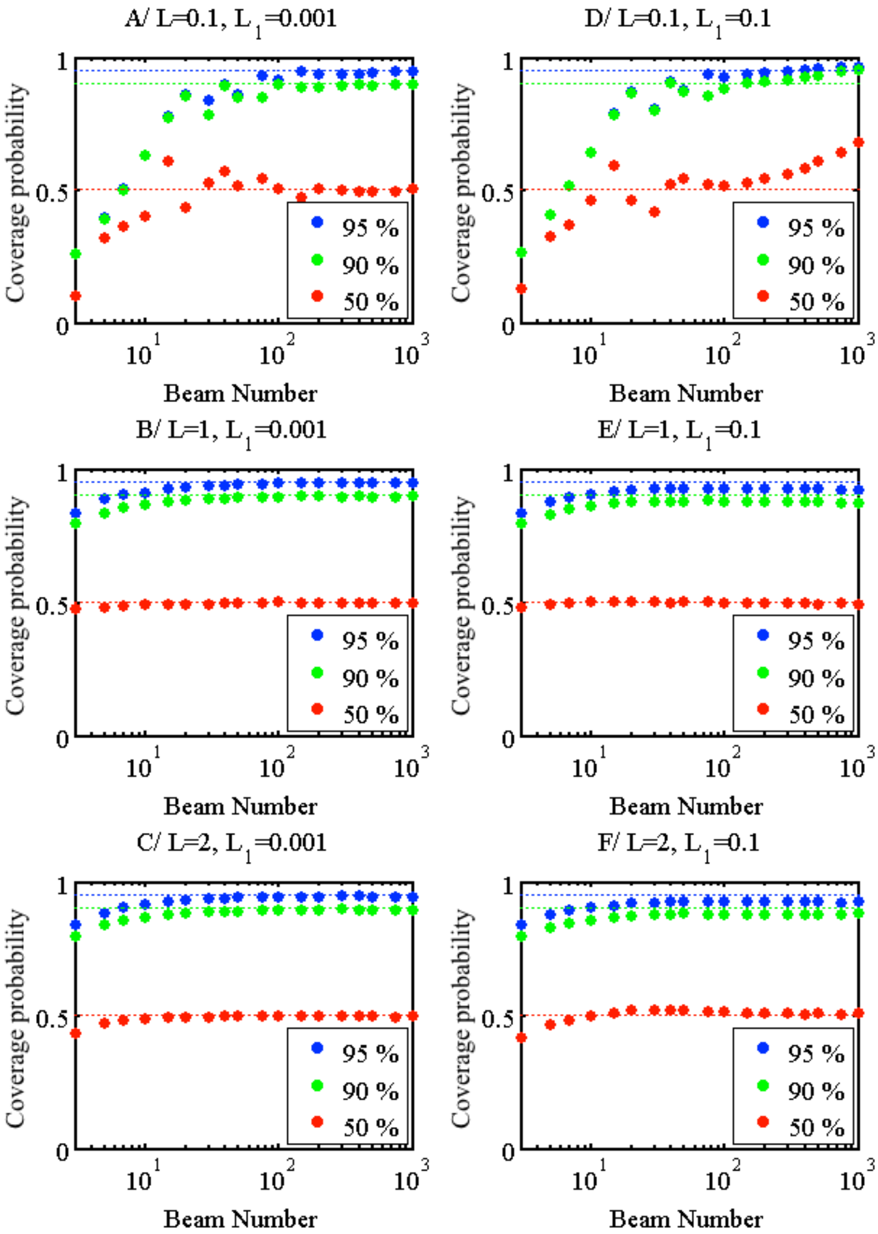


Fig. 7. Coverage probabilities of the estimated confidence interval $\widetilde{\Lambda} \pm z_{\alpha/2}\sqrt{\sigma^2_{\widetilde{\Lambda}}}$ (computed with I), function of the beam number, for 3 levels of confidence (50, 90 and 95%).

As explained in section 3.5, the alternative interval estimation based on Agresti-Coull correction (Eq. (40)) leads to higher-than-expected coverage rates, as shown in Figure 8, which is safer when density is low. Very similar intervals can be obtained for the unbiased Beer-Lambert $\widehat{\Lambda}$ (not shown).

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

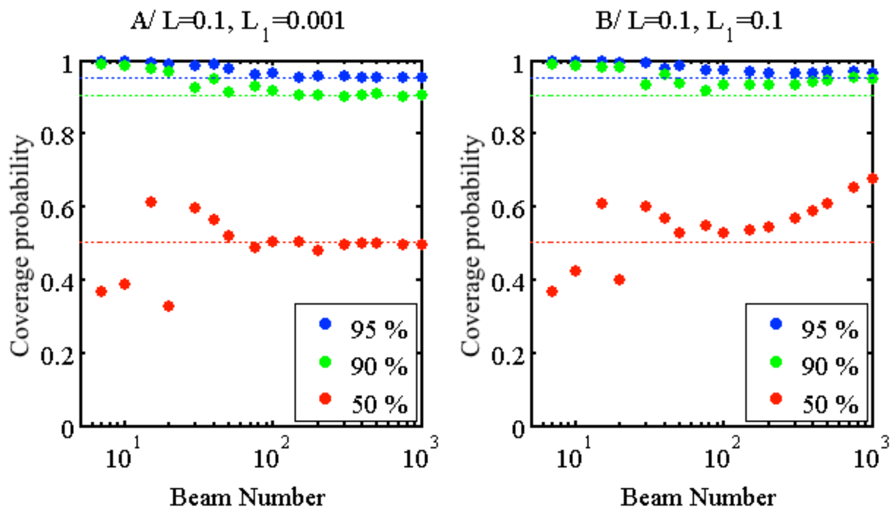


Fig. 8. Coverage probabilities of the estimated confidence interval $\tilde{\Lambda}_c \pm z_{\alpha/2}\sqrt{\sigma^2_{\tilde{\Lambda}_c}}$ (based on the Agresti-Coull values I_c and N_c), function of the beam number, for 3 confidence levels (50, 90 and 95%).

Coverage probabilities are computed for all simulated cases. For a given confidence level (90%, 95%), we can determine the range of parameter values (beam number, element and voxel depths) for which the coverage probabilities match the expected value (0.9, 0.95), within 5% for both the usual formulation of confidence interval (Eq. (39)) and for the "Agresti-Coull" interval (Eq. (40)). We find that using the "Agresti-Coull" interval increases the range of validity, when L is estimated lower than 0.5, but that the usual formulation performs better for higher densities. We thus adopt the following partical rule

$$\begin{cases} \tilde{\Lambda}_c \pm z_{\frac{\alpha}{2}}\sqrt{\sigma^2_{\tilde{\Lambda}_c}}, & \text{when } \tilde{L} \le 0.5 \\ \tilde{\Lambda} \pm z_{\frac{\alpha}{2}}\sqrt{\sigma^2_{\tilde{\Lambda}}}, & \text{when } \tilde{L} > 0.5 \end{cases} \tag{46}$$

We summarize the ranges of validity of confidence intervals defined as in Eq. 46 in Table 3. Confidence intervals are consistent in a fairly large range of parameters. As in the results presented above, the range of validity of the unbiased "Beer-Lambert" confidence interval is

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

not as wide as the one of the unbiased MLE, especially when the voxel depth is larger than 2, for which more than 100 beams are required.

Table 3. Range of validity of the confidence intervals at rate $\alpha = 0.90$ and 0.95. We consider that the confidence interval is consistent, when the empirical probability reaches the expected level within 5%. When L is estimated lower than 0.5, we use the "Agresti-Coull" confidence interval, whereas the usual formulation is used otherwise (Eq. 46).

Index	$\alpha = 0.90$ and Coverage probability within 0.9 ± 5 %	$\alpha = 0.95$ and coverage probability within 0.95 ± 5 %
$\widetilde{\Lambda}$	$\begin{cases} L \geq 0.1 \text{ and } L_1 \leq 0.1 \text{ and } \mathrm{N} \geq 10 \\ L_1 \leq 0.1 \text{ and } \mathrm{N} \geq 100 \end{cases}$	$\begin{cases} L \geq 0.1 \text{ and } L_1 \leq 0.1 \text{ and } \mathrm{N} \geq 10 \\ L_1 \leq 0.1 \text{ and } \mathrm{N} \geq 20 \end{cases}$
$\widehat{\Lambda}$	$\begin{cases} L \in [0.5; 2] \text{ and } L_1 \leq 0.05 \text{ and } \mathrm{N} \geq 40 \\ L_1 \leq 0.01 \text{ and } \mathrm{N} \geq 100 \\ L_1 \leq 0.05 \text{ and } \mathrm{N} \geq 200 \end{cases}$	$\begin{cases} L \leq 2 \text{ and } L_1 \leq 0.05 \text{ and } \mathrm{N} \geq 30 \\ L \leq 2.5 \text{ and } L_1 \leq 0.01 \text{ and } \mathrm{N} \geq 100 \\ L_1 \leq 0.05 \text{ and } \mathrm{N} \geq 150 \end{cases}$

5.1.4. 95% errors for a single voxel and a group of voxels

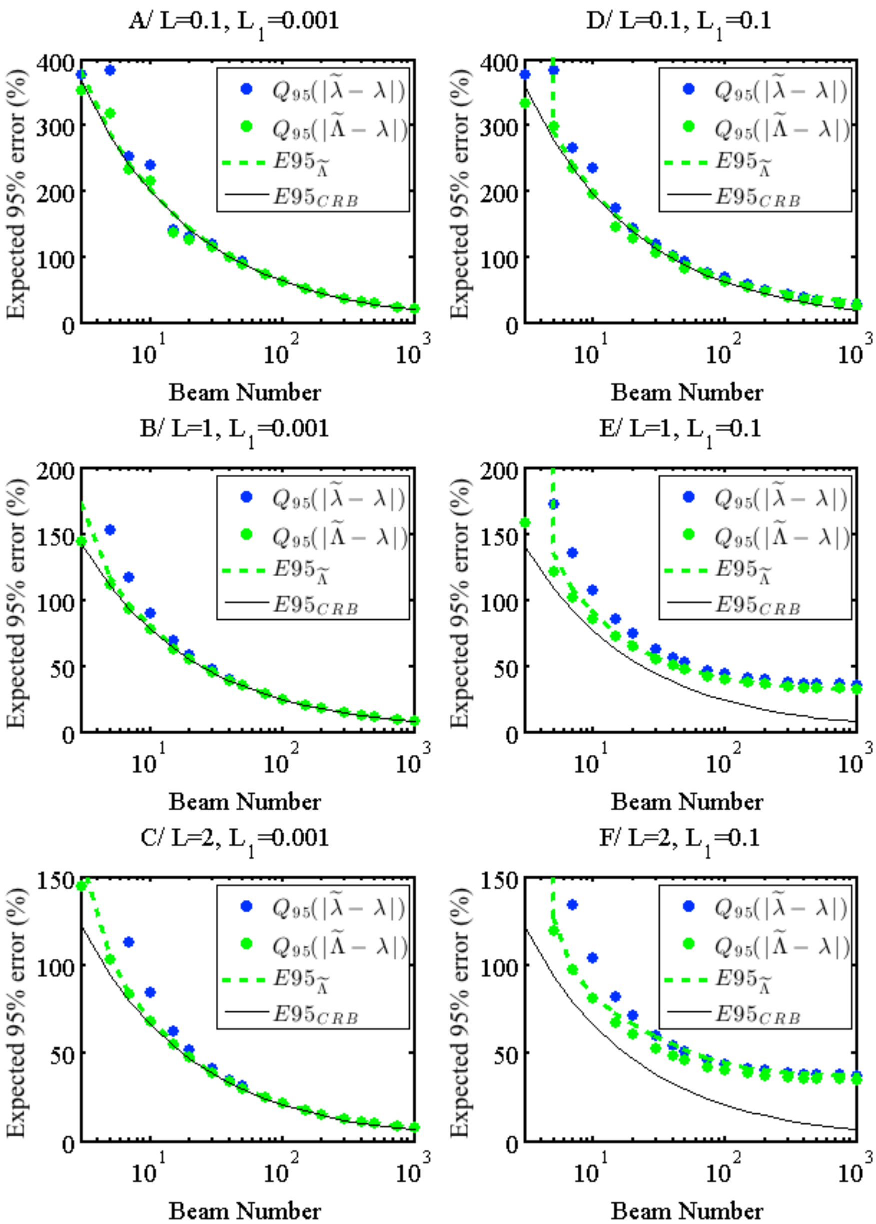
Figure 9 shows the expectation of the 95% errors for the MLE estimators in the same setting as before. When the beam number is small and the density is low, this percentage can largely exceed 100%. In these cases, the estimates remain very uncertain, although close to optimal (Cramer-Rao-95%-error bound in black). The accuracy increases with vegetation density and beam number. However, the 95% errors remain very high even for large beam number, when elements are large because of the variability of vegetation samples (subplots E and F). At the scale of a single voxel, using $\tilde{\lambda}$ or $\widetilde{\Lambda}$ leads to similar errors, which may be disappointing. This is explained by the fact that the bias corrections accounted for in $\widetilde{\Lambda}$ are significant in a range of parameter values for which variances are fairly large. Results are very different, however, when errors are computed after averaging over several voxels (here, 100 voxels), which leads to much smaller errors (Figure 10): using the unbiased estimator ($\widetilde{\Lambda}$) rather than the usual

 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

modified contact frequency ($\tilde{\lambda}$) leads to a reduction of the error on the order of 50%, typically in cases with low beam number or large elements, demonstrating the interest of bias corrections. When elements are large, 95%-errors are below 10% with $\tilde{\Lambda}$, when the beam number is greater than 100.

The expectation of the radiuses of the confidence interval $E95_{\tilde{\Lambda}}$ and $E95_{\tilde{\Lambda}}^{n_v}$ (green dashed line in Figure 9 and 10) are very close to the expectation of the 95% error, showing again that the estimated confidence intervals are consistent.



Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Fig. 9. Expectations of the 95% error, expressed in percentage of λ, as the 95th percentile of the absolute residual to λ for $\tilde{\lambda}$ (blue dots) and $\widetilde{\Lambda}$ (green dots). We show for comparison the estimated radius of the confidence interval $E95_{\widetilde{\Lambda}}$ (green dashed line) and the radius bound derived from the Cramer Rao bound $E95_{\mathrm{CRB}}$ (black line).

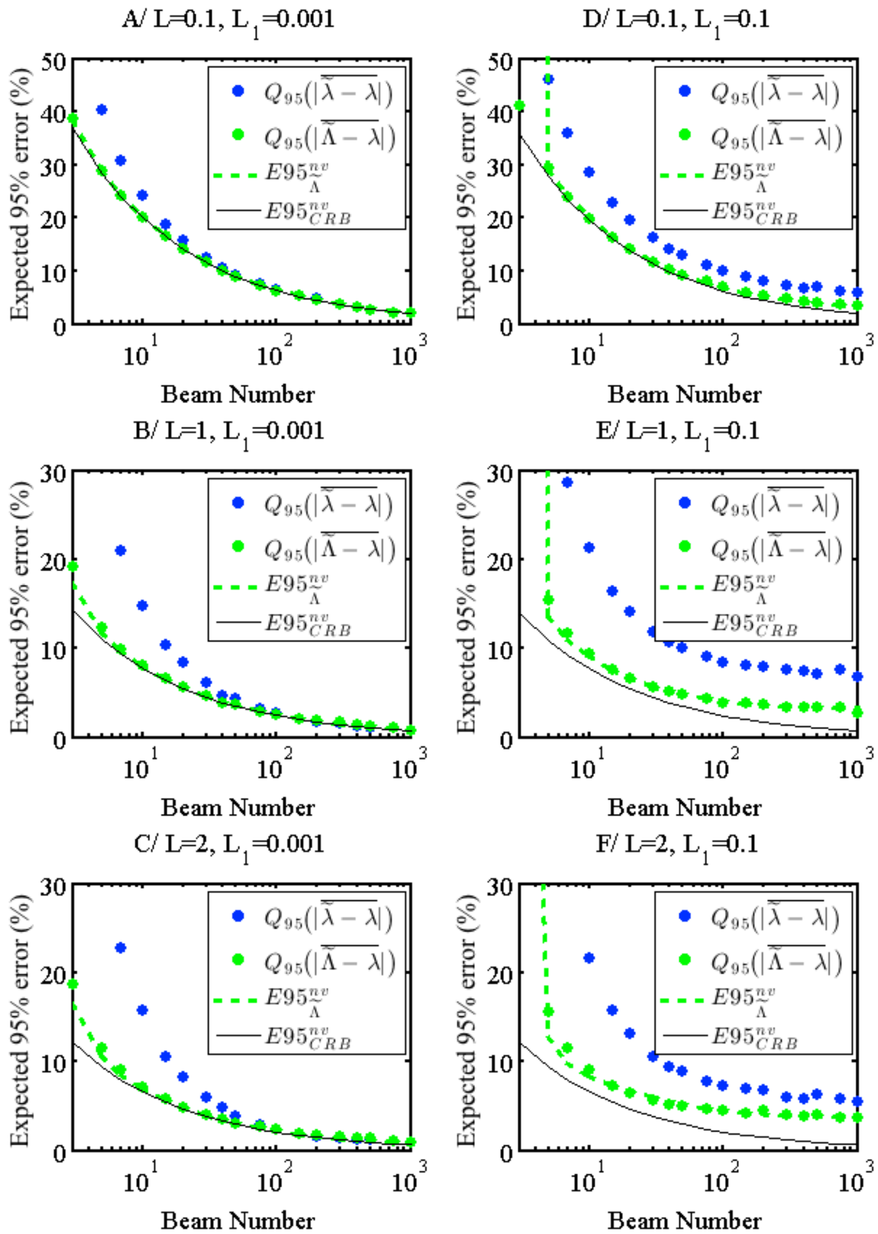


Fig. 10. Same as Fig. 9 for an average over 100 voxels. Expectations of the 95% error over nv=100 voxels, expressed in percentage of λ, as the 95th percentile of the absolute averaged residual to λ for $\tilde{\lambda}$ (blue dots) and $\widetilde{\Lambda}$ (green dots). We show for comparison the estimated radius of the confidence

interval $E95_{\hat{\Lambda}}^{nv}$ (green dashed line) and the radius bound derived from the Cramer Rao bound $E95_{\mathrm{CRB}}^{nv}$ (black line).

Figures 11 and 12 show similar trends to Figures 9 and 10, for both biased and unbiased Beer-Lambert estimators. When blue dots are missing (usual Beer-Lambert), they correspond to cases where I=1 in more than 5% of the voxel, so that the 95% error is in this case infinite. As expected from previous results, the estimator of the radius of the 95% confident interval (green dashed lines) is not consistent for the average over 100 voxels, in ranges where when $\hat{\Lambda}$ is biased (typically L and L_1 large, Fig. 12F).

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

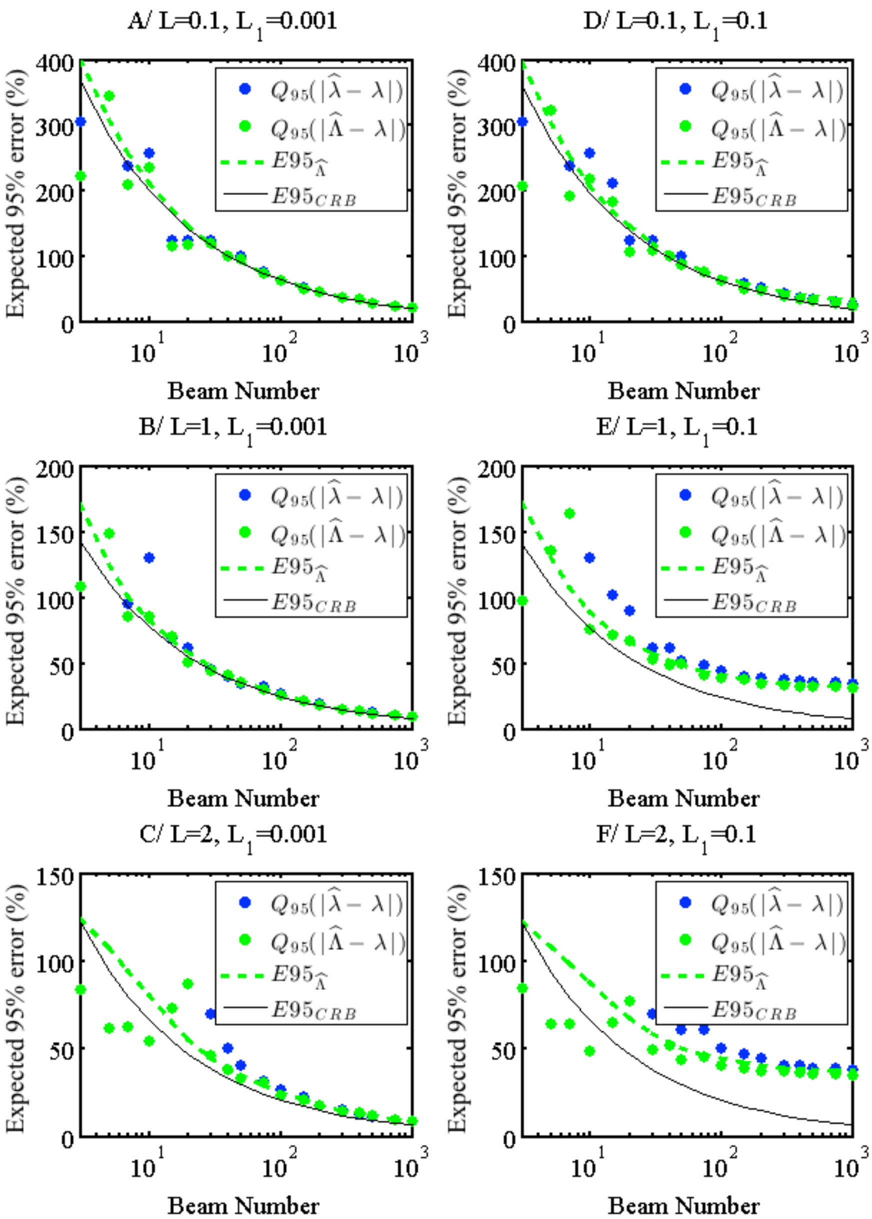


Fig. 11. Same as Fig. 9 for the Beer-Lambert estimators. Expectations of the 95% error, expressed in percentage of λ, as the 95th percentile of the absolute residual to λ for $\hat{\lambda}$ (blue dots) and $\widehat{\Lambda}$ (green dots). We show for comparison the estimated radius of the confidence interval $E95_{\widehat{\Lambda}}$ (green dashed line) and the radius bound derived from the Cramer Rao bound $E95_{\mathrm{CRB}}$ (black line).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

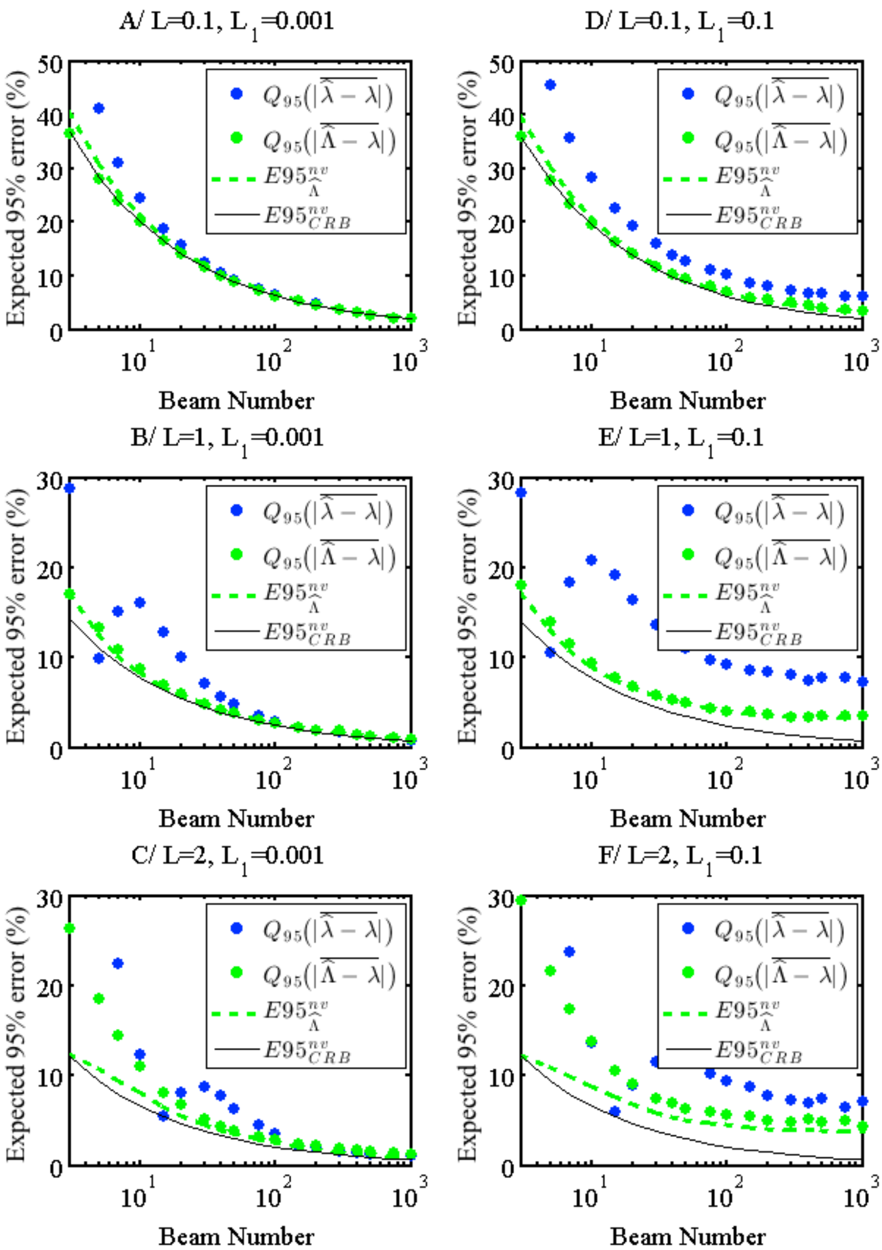


Fig. 12. Same as Fig. 10 for the Beer-Lambert estimators. Expectations of the 95% error over nv=100 voxels, expressed in percentage of λ, as the 95th percentile of the absolute averaged residual to λ for $\hat{\lambda}$ (blue dots) and $\hat{\Lambda}$ (green dots). We show for comparison the estimated radius of the confidence interval $E95_{\hat{\Lambda}}^{nv}$ (green dashed line) and the radius bound derived from the Cramer Rao bound $E95_{\mathrm{CRB}}^{nv}$ (black line).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

5.2. Estimator performance for unequal path lengths

In this subsection, we show the statistics of estimators computed with simulations described in subsection 4.3, in the context of a spherical voxel (unequal path lengths). Since MLE performance is similar to the results shown in the previous section, it is not shown again. Here, we focus on the comparison between Beer-Lambert estimators $\hat{\lambda}$ and $\widehat{\Lambda}_2$ (similar to $\widehat{\Lambda}$, but which includes the correction for unequal path lengths). This is of major importance, since Beer-Lambert law is mostly applied to cubic voxels, for which path lengths are generally not equal. In this subsection, we assume that elements are infinitely small for simplicity (L_1=0).

5.2.1 Estimator consistency

Figure 13 shows the expectations of $\hat{\lambda}$ and $\widehat{\Lambda}_2$ similarly to Fig. 4. For low density (Fig. 13A), the expectation of $\hat{\lambda}$ and $\widehat{\Lambda}_2$ are similar to those obtained with equal path lengths (Fig. 4A). When density is higher (Fig. 13 B and C), the basic Beer-Lambert estimator $\hat{\lambda}$ is negatively biased, and the bias does not tend to zero when the beam number is large.

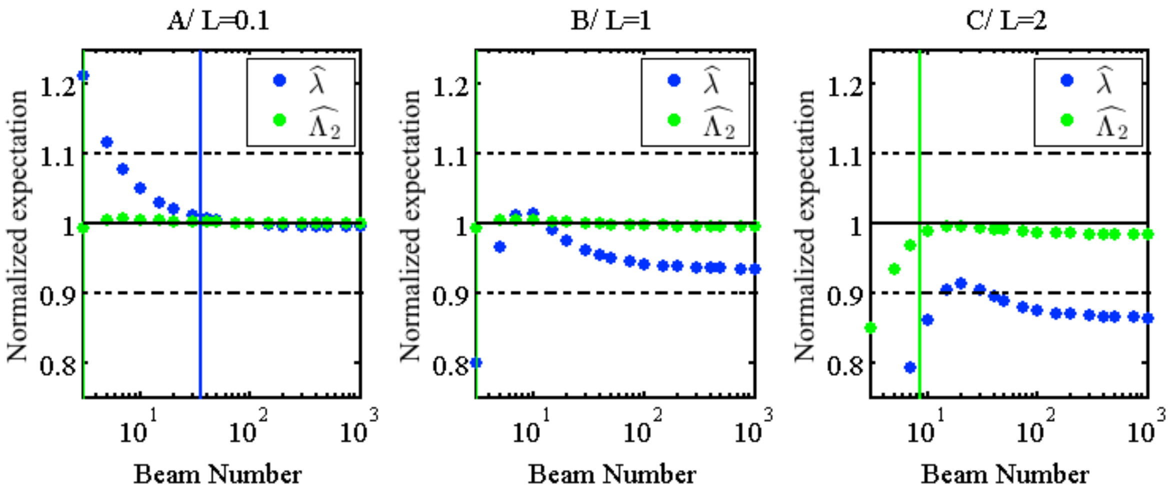


Fig. 13. Same as Figure 4 for the unequal path lengths. Expectations of the Beer-Lambert-attenuation-coefficient estimators, as a function of the beam number. Blue dots corresponds the usual Beer-Lambert estimator ($\hat{\lambda}$). Green dots corresponds to the unbiased Beer-Lambert estimator ($\widehat{\Lambda}_2$), that

doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

accounts for unequal path length, element size and beam number and extended definition when I=1. Estimators are normalized by their true value λ, so that they are consistent when the expectation equals to one. The vertical lines correspond to the lowest values of N leading to a bias smaller than 1% in blue and green, for respectively the biased and unbiased estimators.

5.2.2 Estimator efficiency

Figure 14 show the variances of $\hat{\lambda}$ and $\widehat{\Lambda}_2$ similarly to Fig. 6. For low density (Fig. 14A), the variances of $\hat{\lambda}$ and $\widehat{\Lambda}_2$ are similar to those obtained with equal path lengths (Fig. 6A). When density is higher (Fig. 14 B and C), the variance of $\widehat{\Lambda}_2$ (Green dots) is much larger, which is mostly explained by the variability of the empirical path lengths. $\sigma^2_{\widehat{\Lambda}_2}$ (green dashed line) provides consistent estimators for the variance of $\widehat{\Lambda}_2$, at least when the beam number is larger than 10. This variance, however, is significantly larger than the Cramer-Rao bound (black line), showing that $\widehat{\Lambda}_2$ is suboptimal when N is smaller than 100, contrary to the MLE, that reaches the Cramer-Rao bound (not shown). Again, $\hat{\lambda}$ being biased, its variance cannot be assessed against the Cramer-Rao bound to evaluate its efficiency.

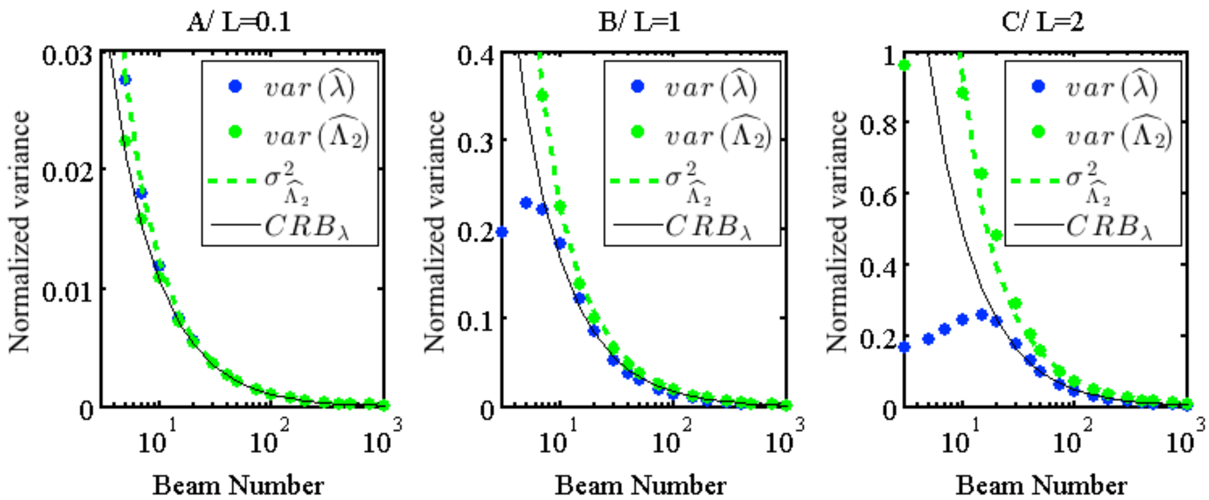


Fig. 14. Same as Figure 6 for unequal path lengths. Normalized variances of Beer-Lambert attenuation coefficient estimators, as a function of the beam number. Normalized variances correspond to variances multiplied by δ^2. Blue dots corresponds to the variance of the usual Beer-Lambert estimator

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

$\hat{\lambda}$. Green dots corresponds to the variance of the unbiased Beer-Lambert estimator $\widehat{\Lambda}_2$. Green dashed lines correspond to the dimensionless expectation of the variance estimator $\sigma^2_{\widehat{\Lambda}_2}$. The black line corresponds to Cramer-Rao bound for the variance of unbiased estimator.

5.2.3. 95% error of estimators for a single and a group of voxels

The coverage probabilities of the estimated confidence interval $\widehat{\Lambda}_2 \pm z_{\alpha/2}\sqrt{\sigma^2_{\widehat{\Lambda}_2}}$ are similar to those shown for $\widehat{\Lambda} \pm z_{\alpha/2}\sqrt{\sigma^2_{\widehat{\Lambda}}}$ in Figure 8 (and thus not shown). More interestingly, Figures 15 and 16 show the expectation of the 95% error, as in Figure 11 and 12. The 95% errors are significantly reduced at the scale of a single voxel when the density is high and N is large (Fig. 15C). As for the other bias correction, the error reduction is limited in other cases since estimators are too uncertain. When averaged at the scale of several voxels, the benefit of the correction for unequal path lengths is clearly visible when the beam number is moderate and large (Fig. 16B and C). In these cases, 95% errors are always greater than 7 and 12%, even when the beam number is large. Bias correction leads to an important reduction of the error, that becomes close to the Cramer-Rao bound. However, contrary to unbiased MLE for which the Cramer Rao bound is reached with unequal path length (not shown here, but logical since the formulation is not affected by the path length), the Beer-Lambert estimator is not optimal even after bias corrections. This is especially true when the beam number is small, because of the variability induced by the empirical correction factor a_e. This demonstrates that the unbiased MLE is clearly more efficient than the (unbiased) Beer-Lambert estimator, since those estimators are mostly computed in cubic voxels.

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

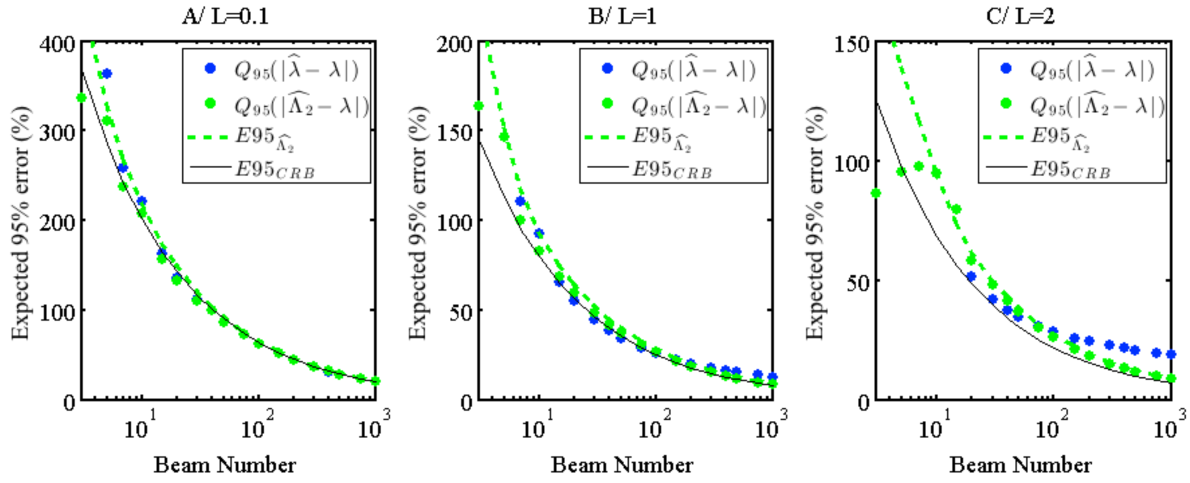


Fig. 15. Same as Fig. 11 for the unequal path lengths. Expectations of the 95% error, expressed in percentage of λ, as the 95th percentile of the absolute residual to λ for $\hat{\lambda}$ (blue dots) and $\widehat{\Lambda}_2$ (green dots). We show for comparison the estimated radius of the confidence interval $E95_{\widehat{\Lambda}_2}$ (green dashed line) and the radius bound derived from the Cramer Rao bound $E95_{\mathrm{CRB}}$ (black line).

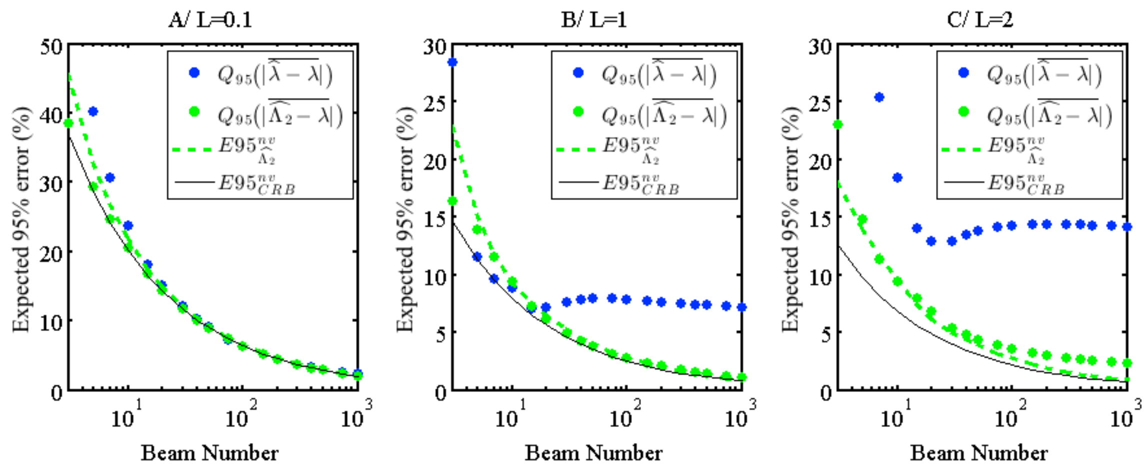


Fig. 16. Same as Fig. 12 for unequal path length. Expectations of the 95% error over nv=100 voxels, expressed in percentage of λ, as the 95th percentile of the absolute averaged residual to λ for $\hat{\lambda}$ (blue dots) and $\widehat{\Lambda}_2$ (green dots). We show for comparison the estimated radius of the confidence interval $E95_{\widehat{\Lambda}_2}^{nv}$ (green dashed line) and the radius bound derived from the Cramer Rao bound $E95_{\mathrm{CRB}}^{nv}$ (black line).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

5. Discussion

The method proposed here is based on a mathematical formulation of the estimation problem in a formal statistical framework. Technical derivations are detailed in several supplementary materials for reference, to facilitate the reading of the manuscript. This theoretical part is completed with numerical simulations, for validation and determination of the range of validity of our new estimators. As every modeling approach, both mathematical formulation and simulations are based on assumptions that are not necessarily valid in the field. Here, we assume that the spatial distribution of beams is random, whereas the actual one has a periodic pattern, potentially altered by occlusion. Also, we assume simple square leaves for our vegetation elements. More realistic patterns for beam shooting and vegetation elements can be simulated (Grau et al., 2017). However, a drawback of this later approach is that it limits both theoretical derivations and simulation number. The benefits of our simplifying assumptions are that the mathematical framework can be deeply explored and that the cost of numerical simulations is very limited, so that a full statistical analysis of estimator performance can be done, over a wide range of parameter values (here element size, voxel size and beam number).

Our derivations entails to propose some new ready-to-use, analytical estimators for the attenuation coefficient, which is proportional to PAD/LAD. These estimators generalize the ones proposed in several pioneering studies. In this sense, our unbiased Beer-Lambert estimator combines the effects of finite-element size and unequal path lengths that are already identified respectively in Béland et al. (2014a), and Béland et al. (2014b) and Grau et al. (2017). Regarding the effect of the element size, we choose to explicitly correct our estimator for the associated bias, rather than to restrict its range of validity to largest voxels, as in Béland et al. (2014a). For unequal path lengths, our formulation is more general than the

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

empirical correction proposed in Béland et al. (2014b), since it does not assume a particular shape for the voxel. Contrary to the secant method, used for example in Bailey and Mahafee (2017b), our formulation is analytical and easy to implement. Also, our Beer-Lambert estimators are defined even when the RDI is equal to one, whereas earlier formulation considered this special case as "occluded" (Béland et al., 2014a), leading to a negative bias as shown above. Our approach also demonstrates that the modified contact frequency, introduced in Béland et al. (2011), is indeed the Maximum Likelihood Estimator of the mathematical problem. It extends the modified contact frequency initially developed for infinitely small elements to the case of finite-size elements in a theoretically-based formulation. This formulation slightly differs from the one proposed in the discussion in Béland et al. (2014a), in which the correction term that accounts for finite-element size is the same as for the Beer-Lambert law (see Eq. (12)). Such a proposition is not supported by the Maximum Likelihood, but numerical consequences are probably limited. More importantly, our formulation includes some bias corrections that depend on the beam number, for both approaches (Beer-Lambert and MLE), whereas usual estimators are shown to be positively biased to more than 20 % when the beam number is small. To the best of our knowledge, such an effect has never been reported before.

Numerical simulations show that the new estimators are consistent for a much wider range of parameter values (element size, attenuation coefficient, beam number), than the usual ones. The range of consistency of the unbiased MLE $\tilde{\Lambda}$ is wider than the one of the unbiased Beer-Lambert estimator. Interestingly, the beam number required to reach consistency of the unbiased MLE depends on the element depth L_1 only. Contrary to the unbiased Beer-Lambert estimator, for which a larger number of beams is required when the vegetation density increases, the unbiased MLE is not affected by the actual value of the attenuation coefficient.

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

This is practically convenient, since the attenuation coefficient is unknown prior to the computation of the estimates when dealing with field data. When $L_1 \leq 0.1$, which is the case for most vegetation when the voxel size is larger than 5 cm, $\tilde{\Lambda}$ is consistent (bias smaller than 1%) as soon as the beam number is larger than 5. This is important when computing the attenuation in small voxels, because the beam number in often low (e.g. Béland et al., 2009), especially when vegetation is dense because of occlusion. Another potential application is airborne LiDAR, for which the point density is much lower than TLS. For comparison, more than 100 beams are required in vegetation with $L_1 \leq 0.01$ and $L \leq 2$ to reach the same consistency with the usual Beer-Lambert estimator as with the unbiased MLE. Another benefit of the MLE when compared to the Beer-Lambert estimator is that it does not require any bias correction when path lengths are unequal. This is all the more important, that the bias correction for unequal path lengths tends to reduce the efficiency of this estimator (in comparison with Cramer-Rao bound). We also demonstrate that the unbiased MLE is efficient, at least when the element size is small, since it reaches the Cramer-Rao bound. This means that no unbiased estimator can have a smaller error than this estimator, so that the unbiased MLE can be considered as optimal. This result is of major importance, since it shows that there is no need for further correction, as long as the assumptions leading to these results are valid.

Our mathematical derivations allow us to derive estimators of the variance and hence, confidence interval for the unbiased estimators. Analyzing their coverage properties shows that they are generally consistent, especially when using the formulation based on both the "Agresti-Coull" interval and the basic interval, depending whether the voxel density is low or high (Eq. 46). Providing such confidence intervals fill a gap for voxel-based approach, as done earlier for gap fraction methods in Zhao et al. (2015). The first outcome of our confidence interval estimates is that the prediction at voxel scale is uncertain, especially when the voxels are small with errors larger than 100%, since the beam number is low and the variability of the vegetation sample is high. When elements are large, uncertainty remains high, even when scanning density is very high, because of the variability of element positions within vegetation samples. The confidence intervals of the mean attenuation coefficient in a larger volume (eventually discretized in small voxels) is much narrower, since the total beam number is larger and the impact of the variability of vegetation samples is dampened. Our numerical results, however, might be affected by our assumptions (random position, square flat leaves, random beams), so that it would be worthwhile in the future to evaluate the asymptotic residual variability in the context of realistic vegetation, for example following the approach detailed in Grau et al. (2017).

Until now, most of the evaluation of the performance of voxel-based estimators was based on the analysis of residual error between estimations and field data. The different sources of bias and dispersion were thus merged. We believe that the applications of the new estimators and their confidence interval should help to choose the appropriate resolution. Small voxels lead to a smaller probability to get larger gaps that invalidate the assumption of random distribution and results in an underestimation of LAD (Béland et al., 2014a). In other words, transmission laws are wider-than-exponential in presence of spatial correlations (Davis and Marshak, 2004; Pimont et al., 2009; Larsen and Clark, 2014). The question of resolution is critical, since the recommendations vary among studies from some millimeters to 2 m (e.g. Hosoi and Asama, 2006; Grau et al. 2017; Béland et al., 2011; Béland et al., 2014a; Pimont et al., 2015; Bailey and Mahafee, 2017b). Among others, these studies deal with various vegetation (various element size from needles to large leaves, various spatial distribution, single tree vs forestry plot, etc.) as well as various scanning density (from single scan to high

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

density scanning). Also, the formulations vary among studies, some biases being corrected in some studies, while not corrected in others. Most of them are affected by the positive biases caused by the beam number and the element size (with the exception of Béland et al., 2014a in which element size sensitivity is evaluated). Such biases are stronger at high resolution since the beam number is lower and the element path is larger. Some of them are affected by the negative biases of the usual Beer-Lambert estimator when path lengths are unequal, or when the RDI is equal to 1. Again, such biases both vary with voxel resolution and orientation. A general use of the unbiased MLE $\widetilde{\Lambda}$, for instance, should cancel these biases and thus gives the opportunity to focus on the remaining sources of bias and dispersion pointed out in the introduction. Among them, we can cite the TLS "flaws" (partial hit and detection threshold) and the remaining vegetation "flaws" (element orientation and clumping, leaf and wood separation). Vegetation heterogeneity is especially concerned by the issue of the voxel size. At the end, the computation of confidence intervals could also help determine the resolution that minimizes errors, since the resolution that minimizes the confidence interval radius of the average attenuation coefficient within a given volume could be selected.

6. *Conclusion*

The present work provides an innovative approach of TLS point clouds, based on both analytical derivations and numerical simulations to propose some new efficient estimators of the attenuation coefficient, which is proportional to the LAD/PAD. These estimators are designed for TLS point clouds of high density, so that they mostly concern TLS, although their consistency with low beam number is also promising for their application to airborne scanner. Among them, the unbiased MLE is consistent and efficient in a wider range of parameter values than the usual estimators. It accounts for statistical biases associated with beam number and element size. Although the biases caused by partial hit and clumping at larger scale are not included, this new estimator should improve the choice of voxel resolution, since it corrects several biases that depends on resolution and that might have been mixed up in some earlier studies. Also, this work provides some estimators for the confidence intervals of the attenuation coefficient within a volume containing one or several voxels, increasing our knowledge of PAD/LAD regarding measurement accuracy by TLS, which is probably lower than expected when voxels are small, and again contributes to the determination of the best resolution.

Funding: The PACA region and INRA funded the PhD of one of the co-authors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

References

Allard, D., Lopez-Lozano, R., Baret, F., 2013 Modeling forest canopies with a hierarchical multi-ring Boolean model for estimating Leaf Area Index. Spat. Stat-Neth 5, 42-56

Bailey, B.N., Mahafee, W.F., 2017a. Rapid, high-resolution measurement of leaf area and leaf orientation using terrestrial LiDAR scanning data. Meas. Sci. Technol. 28(6).

Bailey, B.N., Mahafee, W.F., 2017b. Rapid measurement of the three-dimensional distribution of leaf orientation and the leaf angle probability density function using terrestrial LiDAR scanning. Remote Sens. Environ. 194, 63-76.

Béland, M., Widlowski, J.-L., Fournier, R.A., Côté, J.-F., Verstraete, M., 2011. Estimating leaf area distribution in savanna trees from terrestrial LiDAR measurements. Agric. For. Meteorol. 151, 1252–1266.

Béland, M., Baldocchi, D.D., Widlowski, J.-L., Fournier, R.A., Verstraete, M.M., 2014a. On seing the wood from the leaves and the role of voxel size in determining leaf area distribution of forests with terrestrial LiDAR. Agric. For. Meteorol. 184, 82–97.

Béland, M., Widlowski, J.-L., Fournier, R.A., 2014b. A model for deriving voxel-level tree leaf area density estimates from ground-based LiDAR. Environ. Model. Softw. 51, 184–189.

Brown, L.D., Cai T.T., DasGupta, A., 2001. Interval Estimation for a Binomial Proportion. Stat. Sci. 16(2), 101-133.

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Chen, J.M., Cihlar, J., 1995. Quantifying the effect of canopy architecture on optical measurements of leaf area index using two gap size analysis methods. Geosci. Remote Sens. IEEE Trans. 33, 777–787.

Campbell G.S. & Norman J.N. (1998) An Introduction to Environmental Biophysics 2nd ed. Springer, New York.

Davis, A.B., Marshak, A., 2004. Photon propagation in heterogeneous optical media with spatial correlations: enhanced mean-free-paths and wider-than-exponential free-path distributions. J. Quant. Spectrosc. Ra. 84, 3-34.

Durrieu, S., Allouis, T., Fournier, R.A., Véga, C., Albrech, L., 2008. Spatial quantification of vegetation density from terrestrial laser scanner data for characterization of 3D Forest structure at plot level. In Proceedings of the SilviLaser, Edinburgh, UK, 17–19 September 2008.

Grau., E., Durrieu, S., Fournier, R., Gastellu-Etchegorry, J.P., Yin, T., 2017. Estimation of 3D vegetation density with Terrestrial Laser Scanning data using voxels. A sensitivity analysis of influencing parameters. Remote Sens. Environ. 191, 373 - 388.

Hebert, M., Kroktov, E., 1992. 3-D Measurements from imaging laser radars: How good are they? Int. J. Image Vis. Comput. 10, 170–178.

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Hosoi, F., Omasa, K., 2006. Voxel based 3D modeling of individual trees for estimating leaf area density using high-resolution portable scanning LiDAR. IEEE Trans. Geosci. Remote Sens. 44, 3610–3618.

Hosoi, F., Omasa, K., 2007. Factors contributing to accuracy in the estimation of the woody canopy leaf area density profile using 3D portable LiDAR imaging. J. Exp. Bot. 58, 3463–3473.

Jupp, D.L.B., Culvenor, D.S., Lovell, J.L., Newnham, G.J., Strahler, H., and Woodcock, C.E., 2009. Estimating forest LAI profiles and structural parameters using a ground- based laser called “Echidna”. Tree Physiol. 29, 171–181.

Kay, S.M., 1993. *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 595 p.

Keane, R.E., 2015. Wildland fuel fundamentals and application. Springer International Publishing, Switzerland, 191 p.

Kukenbrink, D., Schneider, F.D., Leiterer, R., Schaepman M.E., Morsdorf F., 2017. Quantification of hidden canopy volume of airborne laser scanning data using a voxel traversal algorithm. Remote Sens. Environ. 194, 424-436.

Nilson, T., 1971. A theoretical analysis of the frequency of gaps in plant stands. Agric. Meteorol. 8 (C), 25–38.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Norman, J.M., Campbell, G.S., 1989. Canopy structure. In: Pearcy, R.W., Ehleringer, J.,Mooney, H.A., Rundel, P.W. (Eds.), Plant Physiological Ecology – Field Methodsand Instrumentation. Chapman and Hall, London, pp. 301–325.

Larsen, M.L., Clark, A.S., 2014. On the link between particle size and deviations from the Beer-Lambert-Bouguer law for direct transmission. J. Quant. Spectrosc. Ra. 133, 646-651.

Pimont. F., Dupuy, J.-L., Caraglio, Y., Morvan, D., 2009. Effect of vegetation heterogeneity on radiative transfer in forest fires. Int. J. Wildland Fire 18, 536–553.

Pimont. F., Dupuy, J.-L., Rigolot, R., Prat, V., Piboule, A., 2015. Estimating Leaf Bulk Density Distribution in a Tree Canopy Using Terrestrial LiDAR and a Straightforward Calibration Procedure. Remote Sens. *7*(6), 7995-8018.

Pisek, J., Sonnentag, O., Richardson, A.D., ottus, M.M., 2013. Is the spherical leaf inclination angle distribution a valid assumption for temperate and boreal broadleaf tree species? Agric. For. Meteorol. 169, 186–194.

Ross, J., 1981. The Radiation Regime and Architecture of Plant Stands. Task for Vegetation Sciences 3; Springer: The Hague, The Netherlands, p. 391.

Ruel, J.J., Ayres, M.P., 1999. Jensen's inequality predicts effects of environmental variation. Trends Ecol. Evol., 14, 361–366.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Schneider, R., Weil, W., 2008. Stochastic and integral geometry. Springer Science & Business Media.

Seielstad, C., Stonesifer, C., Rowell, E., Queen, L., 2011. Deriving Fuel Mass by Size Class in Douglas-fir (Pseudotsuga menziesii) Using Terrestrial Laser Scanning. Remote Sens. 3, 1691-1709.

Stoyan, D., W. Kendall, J. Mecke. 1987. Stochastic Geometry and Its Applications. J. Wiley & Sons/Akademie-Verlag, Berlin 1987, 345 pp.

Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S., Clawges, R.M., 2008. Lidar: shedding new light on habitat characterization and modeling. Frontiers Ecol. Environ. 6, 90–98.

Warren Wilson J., 1960. Inclined point quadrats. New Phytol. 59(1), 1-8.

Zhao, F., Yang, X., Schull, M.A., Román-Colón, M.O., Yao, T., Wang, Z., Zhang, Q., Jupp, D.L., Lovell, J.L., Culvenor, D.S., 2011. Measuring effective leaf area index, foliage profile: and stand height in New England forest stands using a full-waveform ground-based lidar. Remote Sens. Environ. 115, 2954–2964.

Zhao, K., García, M., Liu, S., Guo, Q., Chen, G., Zhang, X., Zhou, Y., Meng, X., 2015. Terrestrial lidar remote sensing of forests: Maximum likelihood estimates of canopy profile, leaf area index, and leaf angle distribution. Agric. For. Meteorol. 209-201, 100-113.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Supplementary Material

Supplementary S1. Expected range of the optical depth of an element of vegetation in a voxel

Following the definition of λ_1 in Eq. (9), the optical depth of an element in a cubic voxel of size δ is

$$L_1 = \lambda_1 \delta = \frac{S_1}{S} \approx \frac{S_1}{\delta^2} \tag{S1-1}$$

For a needle of radius r and length l, this leads to:

$$\lambda_1 \delta \approx \frac{2\pi r l}{4\delta^2} \tag{S1-2}$$

For a (small) needle of diameter 2r = 0.5 mm and length l = 5 cm, we have:

$$\lambda_1 \delta_{min} \approx 2\ 10^{-5} \delta^{-2} \tag{S1-3}$$

For a flat leaf of radius r, this leads to:

$$\lambda_1 \delta \approx \frac{2\pi r^2}{4\delta^2} \tag{S1-4}$$

For a (large) leaf of diameter 2r = 10 cm, we have:

$$\lambda_1 \delta_{max} \approx 5\ 10^{-3} \delta^{-2} \tag{S1-5}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Supplementary S2. Point estimators and their variance based on Beer-Lambert law

The usual Beer-Lambert estimator is based on the RDI. It assumes that (i) a mean path length $\bar{\delta}$ (~V/S) is suitable to account for unequal path length in the voxel and (ii) $-\log(1-\mathrm{I})$ is a good estimator $\hat{\kappa}$ of $\kappa = -log(1-E(I))$ (Note that κ is the optical depth of the voxel). However, in both cases, the non-linearity of the log function limits the validity of such assumptions and makes the standard estimator $\hat{\lambda}$ a biased estimator of λ.

Taking the log(1-x) of Eq. (20), we have:

$$\log\big(1-E(\mathrm{I})\big) = \log\left(\frac{1}{\mathrm{S}}\iint_{s\in\mathrm{S}}(1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}\mathrm{dS}\right) \tag{S2-1}$$

The Lemma proved below enables to approximate the logarithm in (S2-1) and gives a second order approximation of $\log\big(1-E(\mathrm{I})\big)$, as a function of the actual attenuation coefficient λ. Combining (S2-1) with the Lemma leads to:

$$\log\big(1-E(\mathrm{I})\big) \approx -\overline{\delta_e}\lambda + \frac{1}{2}\sigma_{\delta_e}^2\lambda^2 \tag{S2-2}$$

where the effective mean path length $\overline{\delta_e}$ and its variance $\sigma_{\delta_e}^2$ are defined as the mean and variance of the effective path lengths $\delta_{e,j} = -\frac{log(1-\lambda_1\delta_j)}{\lambda_1}$.

We then consider the issue of the bias associated with $\log(1-\mathrm{I})$. A bias correction can be computed applying the approximation S6-5 to the function $\mathrm{g(x)} = \log(1-\mathrm{x})$, which depends on $\mathrm{g''(I)}$ and the variance of I (given by Eq. 23). An unbiased estimator of $\hat{\kappa}$ is thus

$$\hat{\kappa} = -\widehat{\log\big(1-\mathrm{E(I)}\big)} = -\log(1-\mathrm{I}) - \frac{1}{2}\left(\frac{\mathrm{I}(1-\mathrm{I})}{\mathrm{N}} + \sigma_{\mathrm{I}\infty}^2\right)(1-\mathrm{I})^{-2} \tag{S2-3}$$

Combining (S2-2) and (S2-3) leads to a second order polynomial in λ that can be solved to derive the corrected estimator $\widehat{\Lambda}_2$ accounting for unequal path lengths (S2-7).

$$\log(1-\mathrm{I}) - \frac{1}{2}\left(\frac{\mathrm{I}(1-\mathrm{I})}{\mathrm{N}} + \sigma_{\mathrm{I}\infty}^2\right)(1-\mathrm{I})^{-2} = -\overline{\delta_e}\lambda + \frac{1}{2}\sigma_{\delta_e}^2\lambda^2 \tag{S2-4}$$

We first derive the estimator for nearly equal path length $\widehat{\Lambda}$. When path lengths are nearly constant ($\sigma_{\delta_e}^2 \approx 0$), the equation S2-4 in λ leads to:

$$\widehat{\Lambda} = -\frac{1}{\overline{\delta_e}}\left(\log(1-\mathrm{I}) + \frac{\mathrm{I}}{2\mathrm{N}(1-\mathrm{I})} + \frac{\sigma_{\mathrm{I}\infty}^2}{2(1-\mathrm{I})^2}\right) \approx -\frac{1}{\overline{\delta_e}}\left(\log(1-\mathrm{I}) + \frac{\mathrm{I}}{2\mathrm{N}(1-\mathrm{I})}\right) \tag{S2-5}$$

Numerical simulations show that $\frac{\sigma_{I\infty}^2}{2(1-\mathrm{I})^2}$ is always very small compared to log(1-I) in the range of interest. It is thus neglected in the rest of the study.

For unequal path length, the second order polynomial in λ of (S2-4) can be rewritten (since by definition of $\widehat{\Lambda}$, $\log(1-E(\mathrm{I})) = -\overline{\delta_e}\widehat{\Lambda}$):

$$\frac{1}{2}\sigma_{\delta_e}^2\lambda^2 - \overline{\delta_e}\lambda + \overline{\delta_e}\widehat{\Lambda} = 0 \tag{S2-6}$$

Assuming that $2\frac{\sigma_{\delta_e}^2}{\overline{\delta_e}}\widehat{\Lambda}$, is smaller than 1, we can solve the polynomial and keep the smallest root. This leads to the $\widehat{\Lambda}_2$ estimator that accounts for unequal path lengths:

$$\widehat{\Lambda}_2 = \frac{\overline{\delta_e}}{\sigma_{\delta_e}^2}\left(1 - \sqrt{1 - 2\frac{\sigma_{\delta_e}^2}{\overline{\delta_e}}\widehat{\Lambda}}\right) \tag{S2-7}$$

The above indices are not defined when I=1, since the Beer-Lambert approach does not provide any insight regarding the attenuation coefficient rather than "probably high". As explained in section 3.2, the center of the confidence interval can be estimated as a function of N, by the Agresti-Coull interval. With $z_{\alpha/2}^2 = 1$, it is:

$$\mathrm{I}_c = \frac{1+\frac{1}{2\mathrm{N}}}{1+\frac{1}{\mathrm{N}}} = 1 - \frac{\frac{1}{2\mathrm{N}}}{1+\frac{1}{\mathrm{N}}} = 1 - \frac{1}{2N+2} \tag{S2-8}$$

Since $\mathrm{I_c}$ is at the center of the confidence interval, $-\frac{\log(1-\mathrm{I}_c)}{\overline{\delta_e}}$ is a more robust estimator for λ in this context:

$$\widehat{\Lambda} = \frac{\log(2\mathrm{N}+2)}{\overline{\delta_e}} \tag{S2-9}$$

The estimator of the variance of $\widehat{\Lambda}$ is derived from (S6-2). Let $\mathrm{g(x)} = -\left(\log(1-\mathrm{x}) + \frac{\mathrm{x}}{2\mathrm{N}(1-\mathrm{x})}\right)$

$$\mathrm{g}'(\mathrm{x}) = \frac{1}{1-\mathrm{x}} - \frac{(1-\mathrm{x}) - \mathrm{x}(-1)}{2\mathrm{N}(1-\mathrm{x})^2} = \frac{1}{1-\mathrm{x}}\left(1 - \frac{1}{2\mathrm{N}(1-\mathrm{x})}\right) \tag{S2-10}$$

We can thus define the estimator of variance of $\widehat{\Lambda}$ as:

$$\sigma_{\widehat{\Lambda}}^2 = \frac{\sigma_I^2}{\overline{\delta_e}(1-\mathrm{I})^2}\left(1 - \frac{1}{2\mathrm{N}(1-\mathrm{E(I)})}\right)^2 \tag{S2-11}$$

Estimating the variance of $\widehat{\Lambda}_2$ as defined in (S2-7) can be done using (S6-2):

$$\sigma_{\widehat{\Lambda}_2}^2 = \sigma_{\widehat{\Lambda}}^2\left(\frac{\overline{\delta_e}}{\sigma_{\delta_e}^2}\frac{\frac{1}{2}2\frac{\sigma_{\delta_e}^2}{\overline{\delta_e}}}{\sqrt{1-2\frac{\sigma_{\delta_e}^2}{\overline{\delta_e}}\widehat{\Lambda}}}\right)^2 = \frac{\sigma_{\widehat{\Lambda}_2}^2}{1-2\frac{\sigma_{\delta_e}^2}{\overline{\delta_e}}\widehat{\Lambda}} \approx \sigma_{\widehat{\Lambda}_2}^2\left(1 + 2\frac{\sigma_{\delta_e}^2}{\overline{\delta_e}}\widehat{\Lambda} + 4\left(\frac{\sigma_{\delta_e}^2}{\overline{\delta_e}}\widehat{\Lambda}\right)^2\right) \tag{S2-12}$$

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Proof of lemma (Eq. S2-2)

With $\mathrm{g(x)} = \log(x)$, the integral formulation of (S6-1) leads to:

$$\frac{1}{\mathrm{S}}\iint_{\mathrm{s}\in\mathrm{S}} \log\left((1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}\right)\mathrm{dS} \approx \log\left(\frac{1}{\mathrm{S}}\iint_{\mathrm{s}\in\mathrm{S}}(1-\lambda_1\delta(\mathrm{s}))^{\frac{\lambda}{\lambda_1}}\mathrm{dS}\right) - \frac{1}{2}\frac{var\left((1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}\right)}{\overline{(1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}}^2} \quad \text{(S2-13)}$$

The left member is:

$$\frac{1}{\mathrm{S}}\iint_{\mathrm{s}\in\mathrm{S}} \log\left((1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}\right)\mathrm{dS} = \frac{\lambda}{\lambda_1}\overline{\log(1-\lambda_1\delta)} = -\overline{\delta_e}\lambda \quad \text{(S2-14)}$$

With $\mathrm{g(x)} = \exp(\lambda x)$, $var\left((1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}\right) = var\left(g\left(\frac{log(1-\lambda_1\delta(\mathrm{s}))}{\lambda_1}\right)\right)$

According to (S6-2),

$$var\left((1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}\right) \approx g'\left(\frac{\overline{log\left(1-\lambda_1\delta(\mathrm{s})\right)}}{\lambda_1}\right)^2 var\left(\frac{log\left(1-\lambda_1\delta(\mathrm{s})\right)}{\lambda_1}\right)$$

$$= \lambda^2 exp\left(\lambda\frac{\overline{log\left(1-\lambda_1\delta(\mathrm{s})\right)}}{\lambda_1}\right)^2 var(\delta_e) \quad \text{(S2-15)}$$

Since at the first order, $\mathrm{g}(\overline{\mathrm{x}}) = \overline{\mathrm{g(x)}}$,

we can write, $exp\left(\lambda\frac{\overline{log(1-\lambda_1\delta(\mathrm{s}))}}{\lambda_1}\right) \approx \overline{exp\left(\lambda\frac{log(1-\lambda_1\delta(\mathrm{s}))}{\lambda_1}\right)} = \overline{(1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}}$, and then:

$$\frac{1}{2}\frac{var\left((1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}\right)}{\overline{(1-\lambda_1\delta(\mathrm{s}))^{\lambda/\lambda_1}}^2} \approx \frac{1}{2}\lambda^2 var(\delta_e) \quad \text{(S2-16)}$$

Combining this with the above results leads to:

$$\log\left(\frac{1}{\mathrm{S}}\iint_{\mathrm{s}\in\mathrm{S}}(1-\lambda_1\delta(\mathrm{s}))^{\frac{\lambda}{\lambda_1}}\mathrm{dS}\right) \approx -\overline{\delta_e}\lambda + \frac{1}{2}\sigma_{\delta_e}^2\lambda^2 \quad \text{(S2-17)}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Supplementary S3. Point and variance estimators based on MLE

S3.1. Log likelihood and MLE of the attenuation coefficient

Let use denote $\{z_j\}_{j=1,N}$ the N free paths with respective path lengths $\{\delta_j\}_{j=1,N}$.

From Eq. (17), the likelihood of Z is:

$$\mathcal{L}(\lambda; z_1, z, \dots, z_N) = \prod_{j=1}^{N} f_Z(z_j; \delta_j) = \prod_{z_j<\delta_j} \lambda(1-\lambda_1 z_j)^{\lambda/\lambda_1 - 1} \prod_{z_j=\delta_j} (1-\lambda_1 z_j)^{\lambda/\lambda_1} \quad \text{(S3-1)}$$

The ML estimator is the value $\tilde{\lambda}$ that cancels the first derivative of $\mathcal{L}$ (Kay, 1993, chapter 7). Deriving the logarithm of the likelihood and equating to zero provides

$$\frac{d\log\mathcal{L}}{d\lambda} = \frac{N_i}{\lambda} + \sum_{j=1}^{N} \frac{\log(1-\lambda_1 z_j)}{\lambda_1} = 0 \quad \text{(S3-2)}$$

Hence, with $z_{ej} = -\frac{\log(1-\lambda_1 z_j)}{\lambda_1}$

$$MLE_\lambda = \frac{\lambda_1 N_i}{\sum_{j=1}^{N} \log(1-\lambda_1 z_j)} = \frac{I}{\overline{z_e}} \quad \text{(S3-3)}$$

S3.2. Bias correction terms for the MLE

The bias correction for the MLE is derived from (S6-6) with $f(x,y) = \frac{x}{y}$, since the MLE is $\frac{I}{\overline{z_e}}$. The three terms corresponding to bias corrections are, assuming that beams are independent:

$$-\frac{1}{2}\sigma_I^2 \frac{\partial^2 f}{\partial x^2}(I, \overline{z_e}) = -\frac{1}{2}\sigma_I^2 \times 0 = 0 \quad \text{(S3-4)}$$

$$-\frac{1}{2}\sigma_{\overline{z_e}}^2 \frac{\partial^2 f}{\partial y^2}(I, \overline{z_e}) = -\frac{1}{2}\sigma_{\overline{z_e}}^2 \frac{2I}{\overline{z_e}^3} = -\frac{I}{N\overline{z_e}^3} var(z_e) \quad \text{(S3-5)}$$

$$-\sigma_{I,\overline{z_e}} \frac{\partial^2 f}{\partial x \partial y}(I, \overline{z_e}) = \sigma_{I,\overline{z_e}} \frac{1}{\overline{z_e}^2} = \frac{1}{N\overline{z_e}^2} covar\left(\mathbf{1}_{z_j<\delta_j}, z_e\right) \quad \text{(S3-6)}$$

Combining S3-4, S3-5, and S3-6 leads to:

$$\tilde{\Lambda} = \frac{I}{\overline{z_e}} - \frac{I}{N\overline{z_e}^3} var(z_e) + \frac{1}{N\overline{z_e}^2} covar\left(\mathbf{1}_{z_j<\delta_j}, z_e\right) \quad \text{(S3-7)}$$

Although practically computable for a distribution of $\{z_j\}$ and $\{\delta_j\}$, the variance of the estimator can not be analytically derived, so that it is not possible to estimate the variance and thus the confidence interval for the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

estimation. The next subsection is dedicated to the development of analytical estimator for $var(z_e)$ and $covar\left(\mathbf{1}_{z_j<\delta_j}, z_e\right)$.

S3.3. Estimates for $var(z_e)$ and $covar\left(\mathbf{1}_{z_j<\delta_j}, z_e\right)$

From variance formulae:

$$var(z_e) = E[{z_e}^2] - E[z_e]^2 \tag{S3-8}$$

From the probability distribution (Eq. (17)) and definition of z_e, the expectation of z_e is:

$$E[z_e] = \frac{1}{S}\int_S \left(\int_0^{\delta(s)} -\frac{\log(1-\lambda_1 z)}{\lambda_1}\lambda(1-\lambda_1 z)^{\lambda/\lambda_1 - 1}dz + -\frac{\log(1-\lambda_1\delta(s))}{\lambda_1}(1-\lambda_1\delta(s))^{\lambda/\lambda_1}\right)ds \tag{S3-9}$$

Integrating by parts the integral leads to:

$$\begin{aligned} E[z_e] &= \frac{1}{S}\int_S \left(\left[\frac{\log(1-\lambda_1 z)}{\lambda_1}(1-\lambda_1 z)^{\frac{\lambda}{\lambda_1}}\right]_0^{\delta(s)}\right. \\ &\qquad \left. - \int_0^{\delta(s)} -\frac{1}{1-\lambda_1 z}(1-\lambda_1 z)^{\frac{\lambda}{\lambda_1}}dz - \frac{\log(1-\lambda_1\delta(s))}{\lambda_1}(1-\lambda_1\delta(s))^{\frac{\lambda}{\lambda_1}}\right)ds \\ &= \frac{1}{S}\int_S \left(\int_0^{\delta(s)} (1-\lambda_1 z)^{\frac{\lambda}{\lambda_1}-1}dz\right)ds = \frac{P(Z<\delta)}{\lambda} = \frac{E\left[\mathbf{1}_{z_j<\delta_j}\right]}{\lambda} \end{aligned} \tag{S3-10}$$

This demonstrates that:

$$\lambda = \frac{E\left[\mathbf{1}_{z_j<\delta_j}\right]}{E[z_e]} = \frac{I_\infty}{\overline{z_e}^\infty} \tag{S3-11}$$

Similarly,

$$E[{z_e}^2] = \frac{1}{S}\int_S \left(\int_0^{\delta(s)} \left(\frac{\log(1-\lambda_1 z)}{\lambda_1}\right)^2 \lambda(1-\lambda_1 z)^{\lambda/\lambda_1 - 1}dz + \left(\frac{\log(1-\lambda_1\delta(s))}{\lambda_1}\right)^2 (1-\lambda_1\delta(s))^{\lambda/\lambda_1}\right)ds \tag{S3-12}$$

Integrating twice by parts the integral leads to:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

$$E[z_e{}^2] = \frac{1}{S}\int_S \left(\left[-\left(\frac{\log(1-\lambda_1 z)}{\lambda_1}\right)^2 (1-\lambda_1 z)^{\frac{\lambda}{\lambda_1}} \right]_0^{\delta(s)} - \int_0^{\delta(s)} \frac{2}{1-\lambda_1 z}\left(\frac{\log(1-\lambda_1 z)}{\lambda_1}\right)(1-\lambda_1 z)^{\frac{\lambda}{\lambda_1}} dz + \left(\frac{\log(1-\lambda_1 \delta(s))}{\lambda_1}\right)^2 (1 - \lambda_1 \delta(s))^{\frac{\lambda}{\lambda_1}} \right) ds = \frac{2}{\lambda}\frac{1}{S}\int_S \left(\int_0^{\delta(s)} -\frac{\log(1-\lambda_1 z)}{\lambda_1} \lambda (1-\lambda_1 z)^{\frac{\lambda}{\lambda_1}-1} dz \right) ds = \frac{2}{\lambda} E\left[\mathbf{1}_{z_j<\delta_j} z_e\right] \quad \text{(S3-13)}$$

Thus, $$var(z_e) = \frac{2}{\lambda} E\left[\mathbf{1}_{z_j<\delta_j} z_e\right] - E[z_e]^2 \quad \text{(S3-14)}$$

And $$covar\left(\mathbf{1}_{z_j<\delta_j}, z_e\right) = E\left[\mathbf{1}_{z_j<\delta_j} z_e\right] - E\left[\mathbf{1}_{z_j<\delta_j}\right] E[z_e] \quad \text{(S3-15)}$$

S3.4. Point estimator

Plugging S3-14 and S3-15 in, S3-7 leads to:

$$\tilde{\Lambda} = \frac{\mathrm{I}}{\bar{z}_e} - \frac{1}{N}\frac{\mathrm{I}}{\bar{z}_e{}^3}\left(\frac{2}{\lambda} E\left[\mathbf{1}_{z_j<\delta_j} z_e\right] - E[z_e]^2\right) + \frac{1}{N\bar{z}_e{}^2}\left(E\left[\mathbf{1}_{z_j<\delta_j} z_e\right] - E\left[\mathbf{1}_{z_j<\delta_j}\right] E[z_e]\right) \quad \text{(S3-16)}$$

And since $E\left[\mathbf{1}_{z_j<\delta_j}\right] \approx I$, $E[z_e] \approx \bar{z}_e$, $E\left[\mathbf{1}_{z_j<\delta_j} z_e\right] \approx \overline{\mathbf{1}_{z_j<\delta_j} z_e}$ and $\lambda \approx \frac{\mathrm{I}}{\bar{z}_e}$:

$$\tilde{\Lambda} = \frac{\mathrm{I}}{\bar{z}_e} - \frac{\overline{\mathbf{1}_{z_j<\delta_j} z_e}}{N\bar{z}_e{}^2} \quad \text{(S3-17)}$$

S3.5. Variance estimator

The variance for the MLE is derived from (S6-4) with $f(x,y) = \frac{x}{y}$, since the MLE is $\frac{\mathrm{I}}{\bar{z}_e}$. The three terms corresponding to bias corrections are:

$$\sigma_I^2 \left(\frac{\partial f}{\partial x}\right)^2 (I, \bar{z}_e) = \frac{\sigma_I^2}{\bar{z}_e{}^2} = \frac{I(1-I)}{N\bar{z}_e{}^2} = \frac{I}{N\bar{z}_e{}^2} - \frac{I^2}{N\bar{z}_e{}^2} \quad \text{(S3-18)}$$

$$\sigma_{\bar{z}_e}^2 \left(\frac{\partial f}{\partial y}\right)^2 (I, \bar{z}_e) = \frac{var(z_e)}{N}\frac{\mathrm{I}^2}{\bar{z}_e{}^4} = \frac{1}{N}\left(\frac{2}{\lambda}\overline{\mathbf{1}_{z_j<\delta_j} z_e} - \bar{z}_e{}^2\right)\frac{\mathrm{I}^2}{\bar{z}_e{}^4} = \frac{2}{\lambda}\frac{\mathrm{I}^2}{N\bar{z}_e{}^4}\overline{\mathbf{1}_{z_j<\delta_j} z_e} - \frac{\mathrm{I}^2}{N\bar{z}_e{}^2} \quad \text{(S3-19)}$$

$$2\sigma_{I,\bar{z}_e}\left(\frac{\partial f}{\partial x}\right)(I,\bar{z}_e)\left(\frac{\partial f}{\partial y}\right)(I,\bar{z}_e) = 2\frac{\left(\overline{\mathbf{1}_{z_j<\delta_j}z_e} - I\bar{z}_e\right)}{N}\left(\frac{1}{\bar{z}_e}\right)\left(-\frac{\mathrm{I}}{\bar{z}_e{}^2}\right)$$
$$= -\overline{\mathbf{1}_{z_j<\delta_j}z_e}\frac{2I}{N\bar{z}_e{}^3} + \frac{2I^2}{N\bar{z}_e{}^2} \quad \text{(S3-20)}$$

Since $\lambda \approx \frac{\mathrm{I}}{\bar{z}_e}$, summing S3-18 to S3-20 leads to:

$$var\left(\frac{I}{\bar{z}_e}\right) \approx \frac{I}{N\bar{z}_e{}^2} \quad \text{(S3-21)}$$

(S3-14) can be rewritten:

$$\tilde{\Lambda} = \frac{\mathrm{I}}{\bar{z}_e}\left(1 - \frac{\overline{\mathbf{1}_{z_j<\delta_j}z_e}}{NI\bar{z}_e}\right) = \mathrm{MLE}_\lambda\left(1 - \frac{\overline{\mathbf{1}_{z_j<\delta_j}z_e}}{NI\bar{z}_e}\right) \quad \text{(S3-22)}$$

MLE_λ is corrected by a factor of which the variance is assumed to be small when compared to the variance of the standard MLE, so that, we can write (when I>0):

$$\sigma_{\tilde{\Lambda}}^2 = \frac{I}{N\bar{z}_e{}^2}\left(1 - \frac{\overline{\mathbf{1}_{z_j<\delta_j}z_e}}{NI\bar{z}_e}\right)^2 \quad \text{(S3-23)}$$

This formulation does not account for the asymptotic variability of attenuation coefficient estimators, caused by the variability of element positions in vegetation samples. However, this asymptotic variability can be estimated from Eq. (31):

$$\lim_{N\to\infty}\sigma_{\tilde{\Lambda}}^2 = \frac{\sigma_{\mathrm{I}\infty}^2(\mathrm{I},\lambda_1\bar{\delta})}{\bar{\delta}_e{}^2(1-I)^2} \quad \text{(S3-24)}$$

And thus:

$$\sigma_{\tilde{\Lambda}}^2 = \frac{I}{N\bar{z}_e{}^2}\left(1 - \frac{\overline{\mathbf{1}_{z_j<\delta_j}z_e}}{NI\bar{z}_e}\right)^2 + \frac{\sigma_{I\infty}^2(\mathrm{I_b},\lambda_1\bar{\delta})}{\bar{\delta}_e^2(1-\mathrm{I_b})^2} \quad \text{(S3-25)}$$

With $\mathrm{I_b}$ defined as deal with case I=1 as in Supplementary S2:

$$\mathrm{I_b} = \min\left(\mathrm{I}, 1 - \frac{1}{2N+2}\right) \quad \text{(S3-26)}$$

S3.6. Cramer Rao bound

The Fisher information (Kay, 1993, Chapter 3) measures the amount of information that is carried about the attenuation coefficient, by the set of distances travelled by beams within a voxel $\{x_i\}_{i\le N}$. It is defined as:

$$\mathrm{I}_F(\lambda) = \mathrm{E}\left[\left(\frac{\mathrm{dlog}\mathcal{L}}{\mathrm{d}\lambda}\right)^2\right] = -\mathrm{E}\left[\frac{\mathrm{d}^2\mathrm{log}\mathcal{L}}{\mathrm{d}\lambda^2}\right] \quad \text{(S3-27)}$$

Since the log likelihood function of Z (S3-1) is twice differentiable, the Fisher information can be expressed as:

$$\mathrm{I}_F(\lambda) = \frac{\mathrm{E}[\mathrm{N_i}]}{\lambda^2} = \frac{\mathrm{NI}_\infty(\lambda)}{\lambda^2} \quad \text{(S3-28)}$$

The Cramer-Rao bound is the inverse of the Fisher Information:

$$\mathrm{CRB}_\lambda = \frac{\lambda^2}{\mathrm{NI}_\infty(\lambda)} \tag{S3-28}$$

In the case of a spherical voxel, the Cramer-Rao bound can be analytically computed (see section S4.2).

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Supplementary S4. Technical derivation in a spherical voxel

S4.1. Distribution of optical depths in a spherical voxel

In this subsection, we present the derivation leading to the PDF of optical depths $\{L_j\}$ in a spherical voxel, required for the numerical simulation described in subsection 4.3.

If r is the distance between the beam and the sphere center, the path length of the beam within the sphere is:

$$L_j = \lambda\delta_j = \lambda 2\sqrt{R^2 - r^2} \tag{S4-1}$$

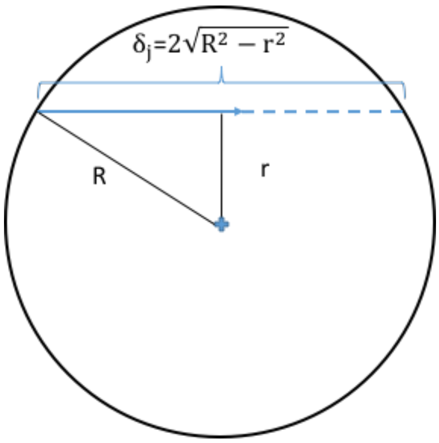


Fig. S4-1. Illustration of the numerical simulation of a TLS beam crossing a spherical voxel with radius R and path length δ_j.

Assuming a constant beam density within the sphere, the beam density with length L_j is:

$$P\left(\lambda\delta_j(r) \leq L_j < \lambda\delta_j(r + dr)\right) = \frac{2\pi r dr}{\pi R^2} = \frac{2rdr}{R^2} \tag{S4-2}$$

With $u = \frac{r}{R}$ between 0 and 1 and $h(u) = \sqrt{1 - u^2}$, it becomes:

$$P\left(\lambda 2Rh(u) \leq L_j < \lambda 2Rh(u + du)\right) = 2udu \tag{S4-3}$$

Or equivalently, using (44):

$$P\left(L_j(u) \leq L_j < L_j(u + du)\right) = 2udu \tag{S4-4}$$

with $L_j(u) = \frac{3}{2}L\sqrt{1 - u^2}$

S4.2. Cramer-Rao bound for dimensionless spherical voxels

As defined in Supplementary S5.2, the Cramer-Rao bound for a spherical voxel is given by:

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

$$\mathrm{CRB_L} = \frac{L^2}{\mathrm{NE(I)}} \tag{S4-5}$$

For the Beer-Lambert law, the expectation of the relative density index E(I) can be expressed as a function of the optical depth Li:

$$\mathrm{E[I]} = 1 - \int_0^1 e^{-L_i(u)} 2u\,du \tag{S4-6}$$

With $y = \frac{3}{2}L\sqrt{1-u^2}$, $dy = \frac{3}{2}L\frac{1}{\sqrt{1-u^2}}\frac{1}{2}(-2u\,du)$ so that $u\,du = \frac{4}{9L^2}y\,dy$ and:

$$\int_0^1 e^{-L_i(r)} 2u\,du = -\frac{8}{9L^2}\int_{\frac{3}{2}L}^{0} e^{-y} y\,dy = \frac{8}{9L^2}\int_0^{\frac{3}{2}L} e^{-y} y\,dy \tag{S4-7}$$

Integrating by parts:

$$\int_0^1 e^{-L_i(r)} 2u\,du = \frac{8}{9L^2}\left(\left[-ye^{-y}\right]_0^{\frac{3}{2}L} - \int_0^{\frac{3}{2}L} -e^{-y}dy\right) = \frac{8}{9L^2}\left(1 - e^{-\frac{3}{2}L} - \frac{3}{2}Le^{-\frac{3}{2}L}\right) \tag{S4-8}$$

Which leads to:

$$\mathrm{E[I]} = 1 - \frac{8}{9L^2}\left(1 - e^{-\frac{3}{2}L} - \frac{3}{2}Le^{-\frac{3}{2}L}\right) \tag{S4-9}$$

Thus

$$\mathrm{CRB_L} = \frac{L^2}{N\left(1 - \frac{8}{9L^2}\left(1 - e^{-\frac{3}{2}L} - \frac{3}{2}Le^{-\frac{3}{2}L}\right)\right)} \tag{S4-10}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

Supplementary S5: Dimensionless quantities used in numerical simulations

S5.1. Finite element simulations (Described in section 4.2):

The dimensionless quantities of interest for these numerical simulations are:

$$\mathbf{1}_e = \frac{\delta_e}{\delta} = -\frac{\log(1-\lambda_1\delta)}{\lambda_1\delta} = -\frac{\log(1-L_1)}{L_1} \quad \text{(S5-1)}$$

$$\hat{I} = \hat{\lambda}\delta = \begin{cases} -\log(1-I) & \text{when } I < 1 \\ +\infty & \text{when } I = 1 \end{cases} \quad \text{(S5-2)}$$

$$\hat{L} = \hat{\Lambda}\delta = \begin{cases} -\frac{1}{1_e}\left(\log(1-I) + \frac{I}{2N(1-I)}\right) & \text{when } I < 1 \\ \frac{1}{1_e}\log(2N+2) & \text{when } I = 1 \end{cases} \quad \text{(S5-3)}$$

$$\sigma_{\hat{L}}^2 = \sigma_{\hat{\Lambda}}^2\delta^2 = \begin{cases} \frac{I}{1_e{}^2(1-I)}\left(\frac{1}{N} + h_\infty(I, L_1)\right)\left(1 - \frac{1}{2N(1-I)}\right)^2 & \text{when } I < 1 \\ \frac{2N-1}{1_e{}^2}\left(\frac{1}{N} + h_\infty\left(\frac{1}{2N+2}, L_1\right)\right) & \text{when } I = 1 \end{cases} \quad \text{(S5-4)}$$

$$\overline{y_e} = \lambda\overline{z_e} = -\lambda\frac{\overline{\log(1-\lambda_1 z_j)}}{\lambda_1} = -\frac{L}{L_1}\,\overline{\log\left(1 - \frac{L_1}{L}y_j\right)} \quad \text{(S5-5)}$$

$$\overline{\mathbf{1}_{y<L}y_e} = \lambda\overline{\mathbf{1}_{z<\delta}y_e} = -\lambda\frac{\overline{\mathbf{1}_{z<\delta}\log(1-\lambda_1 z_j)}}{\lambda_1} = -\frac{L}{L_1}\,\overline{\mathbf{1}_{y<L}\log\left(1 - \frac{L_1}{L}y_j\right)} \quad \text{(S5-6)}$$

$$\tilde{I} = \tilde{\lambda}\delta = \frac{L}{\bar{y}}I \quad \text{(S5-7)}$$

$$\tilde{L} = \tilde{\Lambda}\delta = \frac{L}{\overline{y_e}}\left(I - \frac{\overline{\mathbf{1}_{y<L}y_e}}{N\overline{y_e}}\right) \quad \text{(S5-8)}$$

$$\sigma_{\tilde{L}}^2 = \sigma_{\tilde{\Lambda}}^2\delta^2 = \frac{L^2 I}{N\overline{y_e}^2}\left(1 - \frac{\overline{\mathbf{1}_{y<L}y_e}}{NI\overline{y_e}}\right)^2 + \frac{\sigma_{I\infty}^2(I_b, L_1)}{1_e{}^2(1-I_b)^2} \quad \text{(S5-9)}$$

$$CRB_L = \frac{L^2}{NE(I)} = \frac{L^2}{N\left(1-(1-L_1)^{L/L_1}\right)} \quad \text{(S5-10)}$$

S5.2. Unequal path length simulations (Described in section 4.3)

The ratio of the volume to cross section of the voxel $\frac{V}{S}$ is $\bar{\delta}^\infty$, the limit of $\bar{\delta}$ when N tends to infinity, since a constant surface density of beams is assumed. Thus, the asymptotic optical depth is:

$$L = \lambda\frac{V}{S} = \lambda\bar{\delta}^\infty \quad \text{(S5-11)}$$

With

$$\overline{L_e} = \lambda\overline{\delta_e} = \lambda\bar{\delta} \quad \text{(S5-12)}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at Remote Sensing of Environment 2018, 215, 342-370; doi:10.1016/j.rse.2018.06.024

and

$$\sigma_L^2 = \sigma_\delta^2 \lambda^2 \tag{S5-13}$$

The dimensionless quantities of interest are:

$$1_e = \frac{\overline{\mathrm{L}_e}}{\mathrm{L}} \tag{S5-14}$$

$$\hat{\mathrm{l}} = \hat{\lambda}\delta = \begin{cases} -\dfrac{1}{1_e}\log(1-\mathrm{I}) & when\ I < 1 \\ +\infty & when\ I = 1 \end{cases} \tag{S5-15}$$

$$\hat{\mathrm{L}} = \hat{\Lambda}\delta = \begin{cases} -\dfrac{1}{1_e}\left(\log(1-\mathrm{I}) + \dfrac{\mathrm{I}}{2\mathrm{N}(1-\mathrm{I})}\right) & when\ I < 1 \\ \dfrac{1}{1_e}\log(2\mathrm{N}+2) & when\ I = 1 \end{cases} \tag{S5-16}$$

$$\hat{\mathrm{L}}_2 = \hat{\Lambda}_2 \frac{\mathrm{V}}{\mathrm{S}} = \frac{\mathrm{V}}{\mathrm{S}} \frac{\bar{\delta}}{\sigma_\delta^2}\left(1 - \sqrt{1 - 2\frac{\sigma_\delta^2}{\bar{\delta}}\hat{\Lambda}}\right) = \frac{\mathrm{L}\overline{\mathrm{L}_e}}{\sigma_{\mathrm{L}}^2}\left(1 - \sqrt{1 - 2\frac{\sigma_{\mathrm{L}}^2}{\mathrm{L}\overline{\mathrm{L}_e}}\hat{\mathrm{L}}}\right) \tag{S5-17}$$

$$\sigma_{\hat{\mathrm{L}}}^2 = \sigma_{\hat{\Lambda}}^2 \left(\frac{\mathrm{V}}{\mathrm{S}}\right)^2 = \begin{cases} \dfrac{I}{{1_e}^2 N(1-I)}\left(1 - \dfrac{1}{2N(1-I)}\right)^2 & when\ I < 1 \\ \dfrac{2 - \frac{1}{N}}{{1_e}^2} & when\ I = 1 \end{cases} \tag{S5-18}$$

$$\sigma_{\hat{\mathrm{L}}_2}^2 = \sigma_{\hat{\Lambda}_2}^2 \left(\frac{\mathrm{V}}{\mathrm{S}}\right)^2 = \sigma_{\hat{\mathrm{L}}}^2 \left(1 + \frac{\sigma_{\mathrm{L}}^2}{\mathrm{L}\overline{\mathrm{L}_e}}\hat{\mathrm{L}}\right)^2 \tag{S5-19}$$

$$\mathrm{CRB_L} = \frac{\mathrm{L}^2}{\mathrm{N}\left(1 - \frac{8}{9\mathrm{L}^2}\left(1 - \mathrm{e}^{-\frac{3}{2}\mathrm{L}} - \frac{3}{2}\mathrm{L}\mathrm{e}^{-\frac{3}{2}\mathrm{L}}\right)\right)} \tag{S5-20}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2017 doi:10.20944/preprints201709.0131.v1

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

Supplementary S6. Empirical expectation and variance of the function f of random variable X or two random variables X and Y and method for bias correction.

Here, f is assumed continue and twice differentiable with continuous second derivative. Then the following second order approximations hold:

$$\overline{\mathrm{f(X)}} \approx \mathrm{f}(\overline{\mathrm{X}}) + \frac{1}{2}\sigma_X^2 f''(\bar{X}) \quad \text{(S6-1)}$$

$$\mathrm{Var}\big(\mathrm{f(X)}\big) \approx \big(\mathrm{f}'(\overline{\mathrm{X}})\big)^2 \sigma_X^2 \quad \text{(S6-2)}$$

$$\overline{\mathrm{f(X,Y)}} \approx \mathrm{f}(\overline{\mathrm{X}},\overline{\mathrm{Y}}) + \frac{1}{2}\sigma_X^2 \frac{\partial^2 f}{\partial x^2}(\overline{\mathrm{X}},\overline{\mathrm{Y}}) + \frac{1}{2}\sigma_Y^2 \frac{\partial^2 f}{\partial y^2}(\overline{\mathrm{X}},\overline{\mathrm{Y}}) + \sigma_{X,Y}\frac{\partial^2 f}{\partial x \partial y}(\overline{\mathrm{X}},\overline{\mathrm{Y}}) \quad \text{(S6-3)}$$

$$\mathrm{Var}\big(\mathrm{f(X,Y)}\big) \approx \left(\frac{\partial f}{\partial x}(\overline{\mathrm{X}},\overline{\mathrm{Y}})\right)^2 \sigma_X^2 + \left(\frac{\partial f}{\partial y}(\overline{\mathrm{X}},\overline{\mathrm{Y}})\right)^2 \sigma_Y^2 + 2\frac{\partial f}{\partial x}(\overline{\mathrm{X}},\overline{\mathrm{Y}})\frac{\partial f}{\partial y}(\overline{\mathrm{X}},\overline{\mathrm{Y}})\sigma_{X,Y} \quad \text{(S6-4)}$$

(S6-1) and (S6-3) are used to compute the bias correction for the estimators of $\mathrm{f}(\overline{\mathrm{X}})$ and $\mathrm{f}(\overline{\mathrm{X}},\overline{\mathrm{Y}})$, since

$$\overline{\mathrm{f(X)} - \frac{1}{2}\sigma_X^2 f''(X)} \approx \mathrm{f}(\overline{\mathrm{X}}) \quad \text{(S6-5)}$$

And

$$\overline{\mathrm{f(X,Y)} - \frac{1}{2}\sigma_X^2 \frac{\partial^2 f}{\partial x^2}(\mathrm{X,Y}) - \frac{1}{2}\sigma_Y^2 \frac{\partial^2 f}{\partial y^2}(\mathrm{X,Y}) - \sigma_{X,Y}\frac{\partial^2 f}{\partial x \partial y}(\mathrm{X,Y})} \approx \mathrm{f}(\overline{\mathrm{X}},\overline{\mathrm{Y}}) \quad \text{(S6-6)}$$

Proof:

The proof of (S6-1) is the following. For a set of N value x_j of the random variable X:

$$\mathrm{f}(\mathrm{x}_j) = \mathrm{f}\left(\bar{\mathrm{x}}_j + (\mathrm{x}_j - \bar{\mathrm{x}}_j)\right) \approx \mathrm{f}(\bar{\mathrm{x}}_j) + (\mathrm{x}_j - \bar{\mathrm{x}}_j) f'(\bar{\mathrm{x}}_j) + \frac{1}{2}(\mathrm{x}_j - \bar{\mathrm{x}}_j)^2 f''(\bar{\mathrm{x}}_j)$$

Summing over N and dividing by N leads to:

$$\frac{1}{N}\sum_{j=1}^{N}\mathrm{f}(\mathrm{x}_j) \approx \mathrm{f}(\bar{\mathrm{x}}_j) + \frac{f'(\bar{\mathrm{x}}_j)}{N}\sum_{j=1}^{N}(\mathrm{x}_j - \bar{\mathrm{x}}_j) + \frac{1}{2}\frac{f''(\bar{\mathrm{x}}_j)}{N}\sum_{j=1}^{N}(\mathrm{x}_j - \bar{\mathrm{x}}_j)^2$$

The second term is by definition equal to 0. Thus, with σ_X^2 the usual unbiased estimates for variance:

$$\overline{f(x_j)} \approx \mathrm{f}(\bar{\mathrm{x}}_j) + \frac{1}{2}\frac{N-1}{N}\sigma_X^2 f''(\bar{\mathrm{x}}_j) \approx \mathrm{f}(\bar{\mathrm{x}}_j) + \frac{1}{2}\sigma_X^2 f''(\bar{\mathrm{x}}_j)$$

Peer-reviewed version available at *Remote Sensing of Environment* **2018**, *215*, 342-370; doi:10.1016/j.rse.2018.06.024

The proof of (S6-2) is similar:

$$\left(\mathrm{f}(\mathrm{x}_j) - \mathrm{f}(\overline{\mathrm{x}}_j)\right)^2 \approx \left((\mathrm{x}_j - \overline{\mathrm{x}}_j) f'(\overline{\mathrm{x}}_j)\right)^2$$

Thus,

$$\frac{1}{N-1}\sum_{j=1}^{N}\left(\mathrm{f}(\mathrm{x}_j) - \mathrm{f}(\overline{\mathrm{x}}_j)\right)^2 \approx \frac{f'(\overline{\mathrm{x}}_j)^2}{N-1}\sum_{j=1}^{N}(\mathrm{x}_j - \overline{\mathrm{x}}_j)^2$$

So that,

$$\frac{1}{N-1}\sum_{j=1}^{N}\left(\mathrm{f}(\mathrm{x}_j) - \mathrm{f}(\overline{\mathrm{x}}_j)\right)^2 \approx f'(\overline{\mathrm{x}}_j)^2 \sigma_X^2$$

Similar derivations for a function of two variables lead to (S6-3) and (S6-4). Then (S6-5) and (S6-6) follow.