



Robust processing of airborne laser scans to plant area density profiles

Johan Arnqvist¹, Julia Freier², and Ebba Dellwik³

¹Johan Arnqvist, Uppsala University, department of Earth Sciences, Uppsala, Sweden

²Fraunhofer Institute for Energy Economics and Energy System Technology, Kassel, Germany

³Technical University of Denmark, Roskilde, Denmark

Correspondence: Johan Arnqvist (johan.arnqvist@geo.uu.se)

Abstract. We present a new algorithm for the estimation of plant area density (PAD) profiles and plant area index (PAI) for forested areas based on data from airborne lidar.

The new element in the algorithm is to scale and average returned lidar intensities for each lidar pulse, whereas other methods either do not use the intensity information at all, only use average intensity values or do not scale the intensity information, which can cause problems for heterogeneous vegetation. We compare the performance of the new and three
5 previously published algorithms over two contrasting types of forest: a boreal coniferous forest with a relatively open structure and a dense beech forest. For the beech forest site, both summer (full leaf) and winter (bare trees) scans are analyzed, thereby testing the algorithm over a wide spectrum of PAIs.

Whereas all tested algorithms give qualitatively similar results, absolute differences are large (up to 400 % for the average
10 PAI at one site). A comparison with ground-based estimates shows that the new algorithm performs well for the tested sites, and further and more importantly - it never produces clearly dubious results. Specific weak points for estimation of PAD from airborne lidar data are addressed; the influence of ground reflections and the effect of small-scale heterogeneity, and we show how the effect of these points is minimized using the new algorithm. We further show that low-resolution gridding of PAD will lead to a negative bias in the resulting estimate according to Jensen's inequality for concave functions, and that the severity of
15 this bias is method-dependent. As a result, PAI magnitude as well as heterogeneity scales should be carefully considered when setting the resolution for PAD gridding of airborne lidar scans.

1 Introduction

Plant or leaf area is a key parameter for quantification of air-vegetation exchange of momentum, latent and sensible heat as well as carbon dioxide. Its vertical distribution, the plant (or leaf) area density (PAD), describes the distribution of plant elements
20 from ground to canopy top. The PAD profile is a key parameter in high-accuracy atmospheric models for the estimation of fluxes between the atmosphere and canopies (Williams et al., 1996; Sogachev et al., 2002; Patton et al., 2016; Smallman et al., 2013). PAD profiles have also been introduced into wind modelling of heterogeneous forests in the context of understanding how local inhomogeneities affect the wind field above the crowns (Boudreault et al., 2015; Ivanell et al., 2018). Since wind turbines are very tall, the wind approaching the turbine is affected by surface conditions from far away. Therefore, large areas



of consistent high-quality PAD profiles are attractive as model input. Likewise, atmosphere-biosphere interactions for weather and climate modelling require PAD information from large areas. With the advent of airborne lidar scans (ALS), such a product is possible to achieve.

ALS scans, made with a density of 5-10 laser shots over each square meter of ground are now performed routinely on a country level scale. Besides information on where the laser beam from the airplane was reflected in x , y and z coordinates down to centimeter precision, each point is associated with several other properties. In this paper, we attempt to use more of this information compared to previously published methods with the aim of producing a more consistent and less-biased estimate of PAD.

The previously established methods for calculating the PAD profile, and hence also the PAI, from ALS scans, were based on approximations of the Beer-Lambert law, first introduced in the field of vegetation density estimation by Monsi and Saeki in 1953 (see Monsi and Saeki (2005) for a translated version). Below, we recall the Beer-Lambert law and its adaptation for vegetative canopies, and then describe three reference models, published by Hopkinson and Chasmer (2009), Boudreault et al. (2015) and Almeida et al. (2019). We then introduce the new method. The rest of the paper follows standard formats.

2 Background

2.1 Beer-Lambert law and its adaptation for vegetative canopies

The Beer-Lambert law estimates the effective canopy density in form of PAD as a function of height z , by means of light penetration.

$$Q(z) = Q_0 e^{-\frac{\mu}{\cos\theta} \int_z^{h_c} PAD(z) dz}, \quad (1)$$

where Q is the amount of radiation that penetrates the canopy to height, Q_0 is the incoming radiation, μ is the extinction coefficient and θ is the angle of the incoming radiation. The value of the extinction coefficient depends on the distribution of the angles of reflecting elements. Values have been reported in the literature between 0.28 to 0.58 for common types of trees, see Bréda (2003) for a detailed investigation.

In contrast to in-situ radiation measurements, measurements from ALS are based on the *reflected radiation* R . The crucial assumption is that the amount of reflections from a vertical section of forest is dependent on the reflected radiation from that section, $R = \alpha Q$, where α is the albedo of the canopy elements to the infrared light in the lidar pulse;

$$\int_0^z R(z) dz = R_0 e^{-\frac{\mu}{\cos\theta} \int_z^{h_c} PAD(z) dz}, \quad (2)$$

where R is the reflected radiation at height z and $R_0 = \alpha Q_0$, is the total reflected radiation. Integration on the right hand side all the way from the forest floor gives the PAI.

In practice, the lidar data are discrete, and hence the forest is split into a number of vertical layers $1, 2, \dots, k, \dots, k_{top} - 1, k_{top}$, where 1 is the layer closest to the ground and k_{top} is the highest layer containing plant area density. The effective PAD



(assuming constant extinction coefficient, which implies locally homogeneous distribution of canopy elements) between layer z_k and z_{k+1} can be found by isolation in Eq. 2:

$$\overline{PAD}\Delta z = -\frac{\cos\theta_l}{\mu} \ln\left(\frac{\sum_{i=1}^k R_i}{\sum_{i=1}^{k+1} R_i}\right) \quad (3)$$

where the overbar indicates average, $\Delta z = z_{k+1} - z_k$ and θ_l is the zenith angle of the lidar beam.

5 2.2 Reference methods

As previously mentioned, the foundation of each method is the way Q , the light absorption in a volume $\Delta A\Delta z$, where ΔA is the area of the volume as seen from above, is found from the lidar data. Specifically, what is used as R in Eq. 3. Two main frameworks of the Beer-Lambert law have been applied to discrete small-footprint airborne lidar; one based on measured *intensities* of the reflected pulses (Hopkinson and Chasmer, 2009) and one based on the probability of detecting a reflection (Solberg et al., 2006; Morsdorf et al., 2006; Richardson et al., 2009; Boudreault et al., 2015; Almeida et al., 2019).

Including the option of whether or not to count higher order returns in the probability of detecting a return, the previously published methods can be split into three:

- **Intensity ratio, IR.** Assuming the intercepted radiation is related to the measured intensity, I , of the laser pulse from a volume $\Delta A\Delta z$ (Hopkinson and Chasmer, 2009).

$$\frac{\sum_{i=1}^k R_i}{\sum_{i=1}^{k+1} R_i} = \frac{\sum_{i=1}^k I_i}{\sum_{i=1}^{k+1} I_i} \quad (4)$$

- **First returns ratio, FR.** Assuming the intercepted radiation is related to the number of first returns, r_1 , from a volume $\Delta A\Delta z$ (Solberg et al., 2006; Morsdorf et al., 2006; Richardson et al., 2009; Boudreault et al., 2015)

$$\frac{\sum_{i=1}^k R_i}{\sum_{i=1}^{k+1} R_i} = \frac{\sum_{i=1}^k r_{1i}}{\sum_{i=1}^{k+1} r_{1i}} \quad (5)$$

- **All returns ratio, AR.** Assuming the intercepted radiation is related to the total number of returns, r , from a volume $\Delta A\Delta z$ (Almeida et al., 2019).

$$\frac{\sum_{i=1}^k R_i}{\sum_{i=1}^{k+1} R_i} = \frac{\sum_{i=1}^k r_i}{\sum_{i=1}^{k+1} r_i} \quad (6)$$

The FR approach has been used by Solberg et al. (2006); Morsdorf et al. (2006); Richardson et al. (2009), but then only differentiating between ground returns and total returns, giving information of the PAI, but not PAD. In Boudreault et al. (2015) and Almeida et al. (2019), the method was used to determine PAD.

From the ALS, the PAD can now be found using Eq. 3 in combination with either of Eq. 4-5.



2.3 Shortcomings of existing models

A problem using only first returns is that for dense vegetation, the attenuation of the laser beam increases to a problematic level, since the probability of having a first return from the ground becomes too low. The distance between points where the laser beam penetrates the vegetation uninterrupted all the way to the ground can become so large that for common grid size values there are many grid boxes without first returns from the ground. This is particularly true for forest types with flat wide crowns in closed canopies, like beech forests. Counting the higher order returns and the first returns equally can be problematic if the amount of backscatter from each return is not taken into account, since the actual backscatter depends on the area of the scattering object. This was also noted by Almeida et al. (2019). To facilitate use of higher order returns Hopkinson and Chasmer (2009) presented a method where instead of the number of returns, the intensity of the backscatter was summed (Eq. 4), which enables a consistent use also of the higher order returns. Furthermore, Hopkinson and Chasmer (2009) reasoned that for higher order returns, the integration in Eq. 1 should be carried out both on the downward and the upward (reflected) path of the beam, accounting for losses both on the downward and upward path of the laser beam.

Hopkinson and Chasmer (2009) showed improvements using the intensity information compared to using only the first returns (based on correlation with gap fractions from hemispheric photos) and concluded that an intensity based method is less error prone due to the implicit information about the size of the return object that is carried in the intensity value. A problem using the Hopkinson and Chasmer (2009) method is that the intensity of the ground level can influence the PAI estimate to a large degree, since the ground level may have a different value of α . Wagner et al. (2008) for example, reported on difference in reflective properties between non-vegetation and vegetation. This is particularly problematic for less dense forests where many of the returns come from the ground. Ideally, the vegetation density estimate should not be sensitive to the ground type.

The Beer-Lambert law was developed for continuously absorbed light (such as incoming solar radiation), but can be shown to also apply for discrete light absorption (such as infinitely thin laser beams either passing or being obstructed by canopy elements) (Monsi and Saeki, 2005). However, in practice, ALS data lies somewhere between those two cases. With a beam size of around 0.1 to 1 m diameter the ALS beam frequently get partially obstructed, complicating the application of Beer-Lambert law. When laser beams have a large cross section compared to the size of the scattering elements, the FR method becomes problematic, since the probability of reflection becomes less correlated with object size. On the other hand, for very thin laser beams, the probability of uninterrupted penetration to the ground level becomes higher, making IR sensitive to differences in reflectivity of the ground and canopy. Thus, a method that combines the respective strengths of the IR and FR methods has potential to be less sensitive to parameters that affect beam width, such as flight altitude and instrument type, and thereby be more suited for PAI and PAD estimates from ALS.

In this work, we present a novel method which aims to combine the benefits of the Hopkinson and Chasmer (2009) method and the Boudreault et al. (2015) method by using the intensity information in a way that removes the influence of the intensity value from the first returns at ground level.



3 Methods

3.1 New method

In this section, a new method to determine PAD from ALS is described. The method will be referred to as SR, Scaled Ratio. An early version of the SR algorithm was analysed and compared to the FR and the IR method in Freier (2017). The method uses standard outputs in the .las format. Table 1 shows a list of the variables used, and also lists the variables required in the IR, FR and AR methods described above.

Table 1. Variables used to calculate the PAD from ALS with three different methods. The variables are usually available with data in the .las format

	Scaled return ratio (SR)	Intensity ratio (IR)	First return ratio (FR)	All returns ratio (AR)
Position, X, Y, Z	×	×	×	×
Intensity, I	×	×		
Return number, i_r	×		×	
Number of Returns (per pulse), N_R	×			
Classification (ground identification)	×	×	×	×
Scanning angle, θ_l	×	×	×	×

The idea of the method is to scale each return intensity by the total return intensity of the pulse the return belongs to. Recently, Vincent et al. (2017) presented a method which uses both the number of returns and the intensity information, similar to the method presented here. However, the method proposed by Vincent et al. (2017) use the intensity information in an *average* sense while the method proposed in this paper uses individual intensity data for each pulse. If the individual intensities are not available in the data set, or the data set is not ordered in consecutive returns, using the average data might be the only option. For a discussion on the implications of using average intensities as scales, please see Sec. 5.2.

By scaling the returns by their individual share of pulse reflected intensity, the method will reduce to the FR method, if there are only first returns in a data set. For the more common case with several recorded reflections per pulse, each return intensity is scaled with the total reflected intensity of that pulse:

$$r_s = I_{i_r} / \sum_{i_r=1}^{N_r} I_{i_r}. \quad (7)$$

After the intensity scaling the PAD is calculated in the same way as with the other methods; (1) by estimating the ratio of incoming to outgoing radiation following

$$\frac{\sum_{i=1}^k R_i}{\sum_{i=1}^{k+1} R_i} = \frac{\sum_{i=1}^k r_{s_i}}{\sum_{i=1}^{k+1} r_{s_i}}, \quad (8)$$

and (2) by calculating PAD using Eq. 3. In this way, for all pulses with only one return ($N_r = 1$), the weight is 1, whereas the weight for all other pulses are scaled by their respective backscatter intensity. In summary, for the special case of a data set



with only first returns, the SR method reduces to the FR method and for the special case of only one pulse, the SR method reduces to the IR method.

When applying Eq. 3 we have opted to carry out the integration only on the downward path of the beam for all methods. This is contrary to Hopkinson and Chasmer (2009) where attenuation was also accounted for on the upward path. The reason for only using one-way attenuation is that the scattering does not come from continuous attenuation, but rather distinct reflection on objects. That, in combination with the fact that changes in θ_l (position of the airplane) is negligible under the short time it takes for the light pulse to return to the sensor, leads to the conclusion that if the path was free of obstacles on the way down, it should also be free on the way up.

The SR method requires that the backscatter intensity of each reflection can be connected to the total backscatter intensity of the pulse that the point belongs to. In other words, the data set with the point cloud must be structured in a way that the points are ordered, so that all reflections of a pulse are traceable. In order to check if this is true, the data sets were scanned and it was required that all points with $i_r = N_r$ along with $N_r - 1$ preceding points satisfy $\prod_{i_r=1}^{N_r} i_r = N_r!$. For the data sets that has been part of this study, the number of points satisfying the above criterion is typically more than 99 % of the data set.

In the comparison with reference methods we will follow Solberg et al. (2006); Morsdorf et al. (2006); Richardson et al. (2009); Boudreault et al. (2015) in assuming a spherical distribution of the reflecting surfaces of vegetation, corresponding to a value of $\mu = 0.5$. That is equivalent of approximating that the average forest density is the same looking from above, from the side, or from any other direction of the forest. While choosing a value of μ fitting for the vegetation of interest improves the PAI estimates (Yan et al., 2019), the focus of this study is the application of Beer-Lambert law itself, and any change in μ would be common for the investigated methods, and therefor the value $\mu = 0.5$ has been kept throughout the study. To speed up the calculation, instead of using pulse specific θ_l , we have used the average absolute angle within the grid box.

3.2 Sites

The methods have been tested on three different data sets from two different sites. One is a dense deciduous (beech) forest and the other one is a medium dense coniferous (mainly pine) forest. The data sets were chosen based on that they cover two different and commonly occurring types of forests as well as the availability of both ALS data and ground based PAI estimates.

25 **Dense beech forest: Tromnaes**

This approximately 80 year old forest of predominantly European beech (*Fagus sylvatica* L.) is located near the coast on the island Falster in Denmark. During the leaf-on summer period the canopy is very dense, whereas the winter time tree structure is relatively open as seen in Fig. 1. In order to accurately estimate the canopy height and density for the purpose of wind flow modelling, the forest was scanned during the summer time in 2013. The site was scanned again in a national survey in 2014 during the leaf-off winter period. In connection with a field campaign of wind observations upwind and downwind of the forest edge, the forest density was further measured using two LAI-2000 PCAs (LI-COR, Lincoln, NE, USA). The PAI was found to be approximately 1 and $6 \text{ m}^2 \text{ m}^{-2}$ during leaf-off and leaf-on conditions, respectively. These measurements were performed using two instruments; one was placed well within the forest and the other, which was used for reference measurement, was



placed well outside the forest. More information on the site can be found in Dellwik et al. (2014) and technical data on scanning height, density and instrumentation is stated in Table 1.

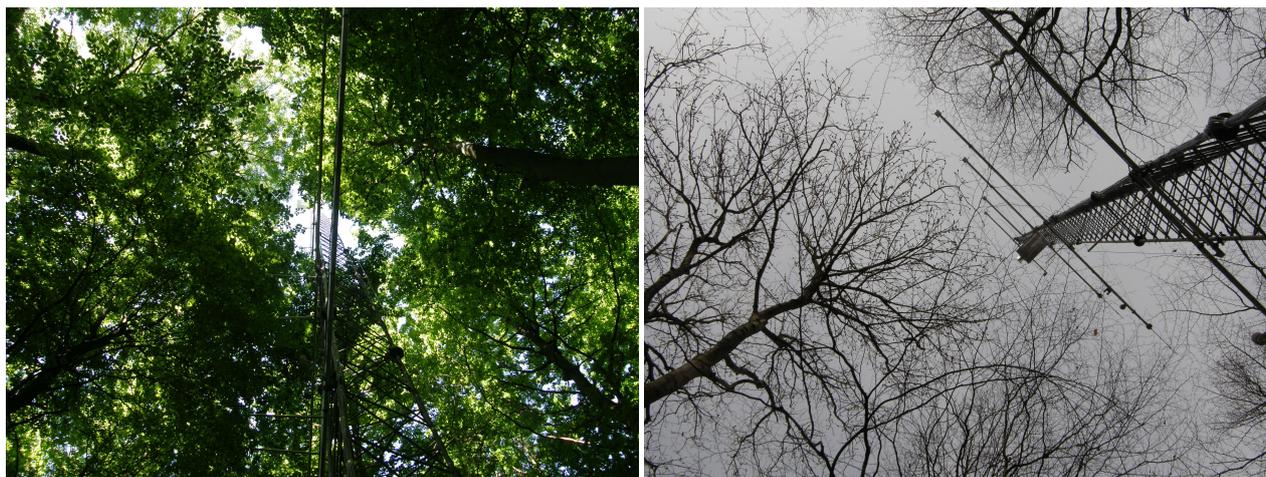


Figure 1. Photos taken at the position of the beech forest mast at the Falster site (Dellwik et al., 2014) illustrating the large difference between summer (left) and winter (right) canopies.

From the point cloud of the two scans, we focus our analysis on a $300\text{ m} \times 300\text{ m}$ tile which has the center coordinate at 54.7638°N , 12.0396°E . The data sets will be referred to as *Falster winter* and *Falster summer*.

5 Norunda

The Norunda site is a research site under the ICOS, Integrated Carbon Observation System, infrastructure, located in the south east Sweden (60.0833°N , 17.4833°E). The forest cover is predominately Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*). A minor fraction (15 %) of deciduous trees, mainly birch (*Betula* sp.) is also present. The forest at the site is marked by its many clearings, characteristic of Swedish forest management. As part of the ICOS project, the forest density at the site has been analyzed with digital hemispheric photometry according to the ICOS protocol for ancillary vegetation measurements Gielen et al. (2018). The site has also been scanned with airborne laser as part of a national scan over Sweden Lantmäteriet (2016). The scan was made in November 2010 and the coverage is generally more than 0.5 ground reflections per m^2 , but sometimes as low as around 5 ground reflections per $10\text{ m} \times 10\text{ m}$ square in the most dense parts of the forest. The data set will be referred to as *Norunda*.

Ground based estimates of the forest were made in November 2018. Despite the difference of eight years between the ground based measurements and the airborne scan, the two data sets should be comparable, since the age of the particular forest stand is more than hundred years and thus the growth rate is very low.



4 Results

In order to compare the new routine with the three reference methods described above, PAD, PAI, ground height and forest height were calculated for the three datasets at a resolution of $10\text{ m} \times 10\text{ m}$. To begin with, an overview of the PAD estimates for the three data sets is shown, highlighting their characteristics for different conditions, then a detailed case study is performed for one $10\text{ m} \times 10\text{ m}$ grid cell and finally a grid size sensitivity study is presented.

4.1 Maps based on ALS methods

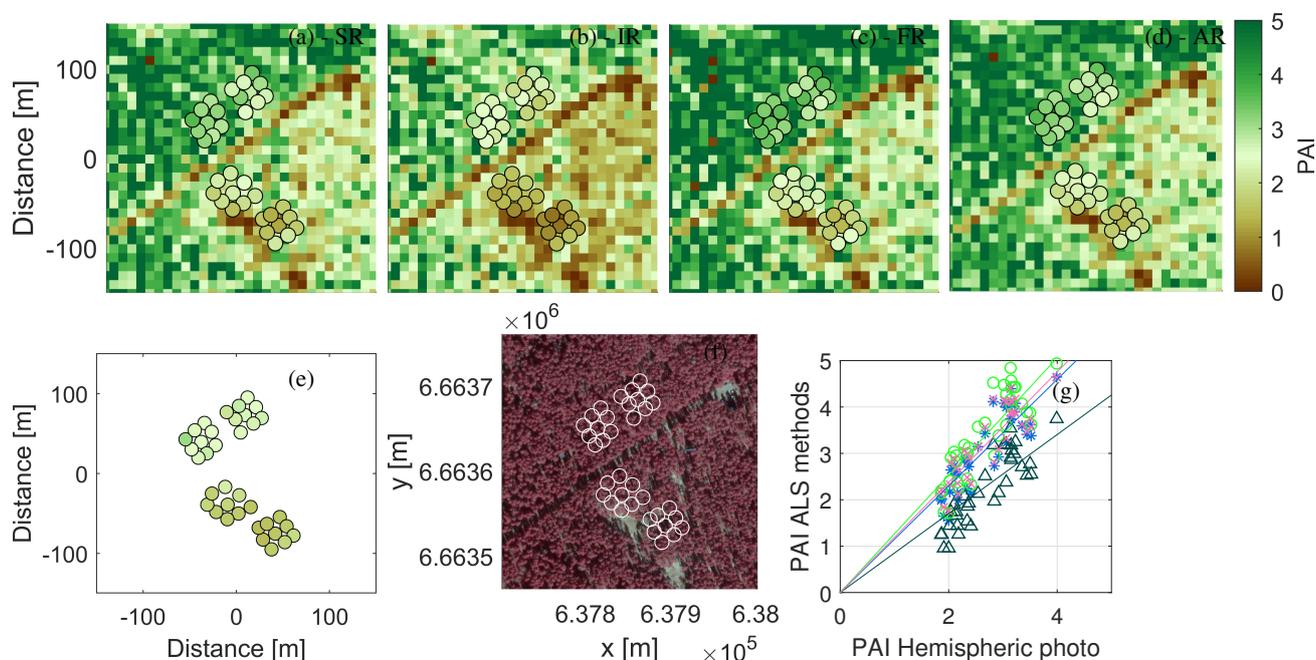


Figure 2. PAI estimates from Norunda by various methods. *a)*, *b)*, *c)* and *d)* show PAI in a $10\text{ m} \times 10\text{ m}$ square grid based on the methods SR, IR and FR and AR respectively. Circles in *a)* to *d)* displays estimates based on points within 14 m radius from the centre of the circle. *e)* shows estimates from hemispheric photo, *f)* shows an aerial photo over the site and *g)* a point to point comparison for the locations of the photometry estimates between hemispheric photo estimates and ALS methods, where SR is represented by blue stars, IR is represented by dark green triangles, FR is represented by bright green circles and AR is represented by pink crosses.

Maps of the calculated PAI can be seen in Fig. 2, Fig. 3 and 4 for the Norunda and Falster site respectively. It is immediately clear that forest characterization by ALS has a big advantage compared to ground based estimates due to the full horizontal coverage. In Fig. 2, details of a road and clearings shown in the flight photo are clearly visible also in the PAI estimates. In terms of PAI magnitudes, FR gives the highest values for all three data sets. SR and AR are in the average very similar, apart from the Falster winter scan, where AR gives higher values in the low PAD regions. In Norunda, also the variation of PAI is similar in SR and AR, but for the Falster scans the PAI distribution of AR is narrower than that of SR. SR and FR gives similar



estimates in the low PAI regions of Fig. 2 and 3, but FR shows significantly higher estimates in high PAI areas, where instead SR and IR are more similar.

In addition to the fixed $10\text{ m} \times 10\text{ m}$ square grid, PAI was also calculated for the positions of the ground based measurements, within a diameter equal to the tree height. Those estimates are plotted in circles on top of the $10\text{ m} \times 10\text{ m}$ grid. Scatter plots of the PAI estimates from ALS methods and ground based estimates are included in Fig. 2 and 3. It is important to notice that the ground based measurements are all taken in relatively low PAI locations. For these locations, the SR, FR and AR estimates are of similar magnitudes, while the IR gives lower values. In comparison with the ground based estimates the SR, FR and AR are in better agreement than IR.

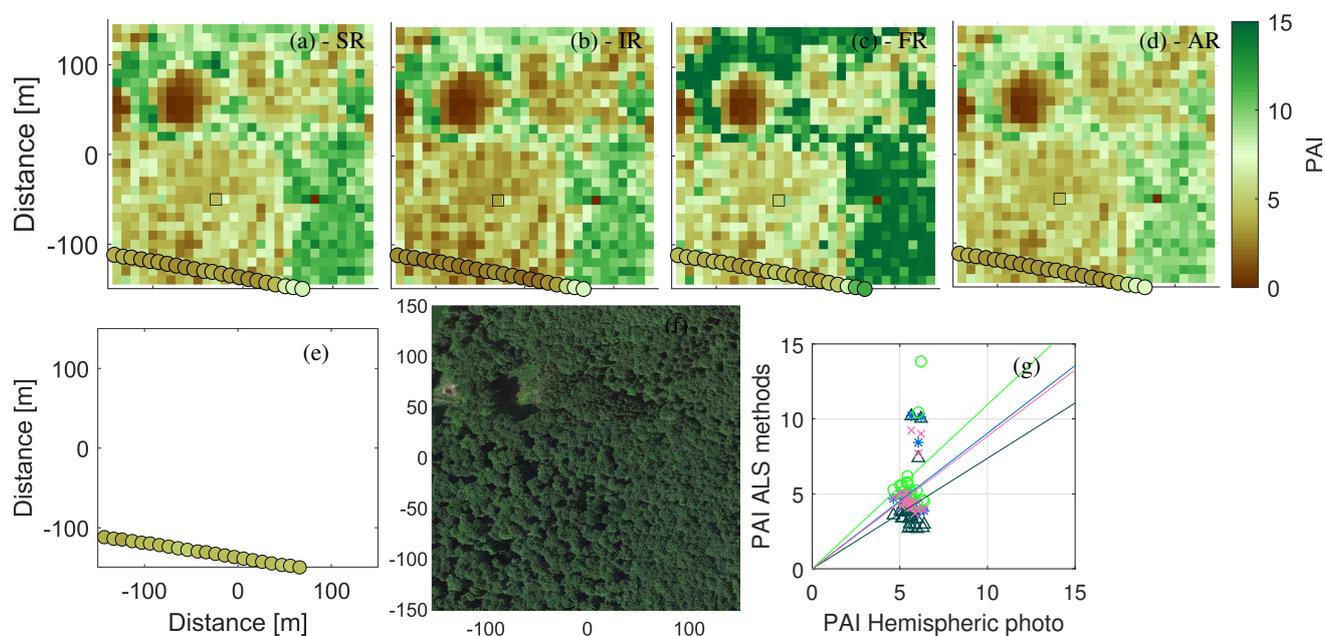


Figure 3. PAI estimates from Falster summer scan by various methods. *a)*, *b)*, *c)* and *d)* show PAI in a $10\text{ m} \times 10\text{ m}$ square grid based on the methods SR, IR and FR and AR respectively. Circles in *a)* to *d)* displays estimates based on points within 16 m radius from the centre of the circle, and the black square marks the edges for the data used in Section 4.2. *e)* shows estimates from PCA, *f)* shows an aerial photo over the site and *g)* a point to point comparison between the PCA and ALS methods, where SR is represented by blue stars, IR is represented by dark green triangles, FR is represented by bright green circles and AR is represented by pink crosses.

To further illustrate the comparison with ground based methods, in Fig. 2 and Fig. 3, the point to point comparisons of PAI are shown. While the correlation is low for both sites, the slope of a linear fit is reasonably close to 1 for the SR, FR and AR methods, while IR methods shows a systematically lower PAI than the ground based methods. It is later shown in Section 4.3, that the PAI estimates form ALS are sensitive to grid size which somewhat limits the value of a direct comparison between ALS based methods and ground based methods, since the area representation of ground based methods is difficult to properly define.



While for Norunda and Falster summer the four methods show similar magnitude of the PAI estimates, that is not the case in Falster winter, as shown in Fig. 4, where there is a striking difference between the intensity methods on the one hand and probabilistic FR and AR methods on the other. The overall PAI is more than three times higher with FR than with SR. The reason for this is examined in detail in the next section, but as can be seen from Fig. 1, even though the PAD of the upper canopy is low in winter compared to summer, there are relatively few gaps in the forest without any branches at all, effectively limiting the number of first order returns reaching the ground.

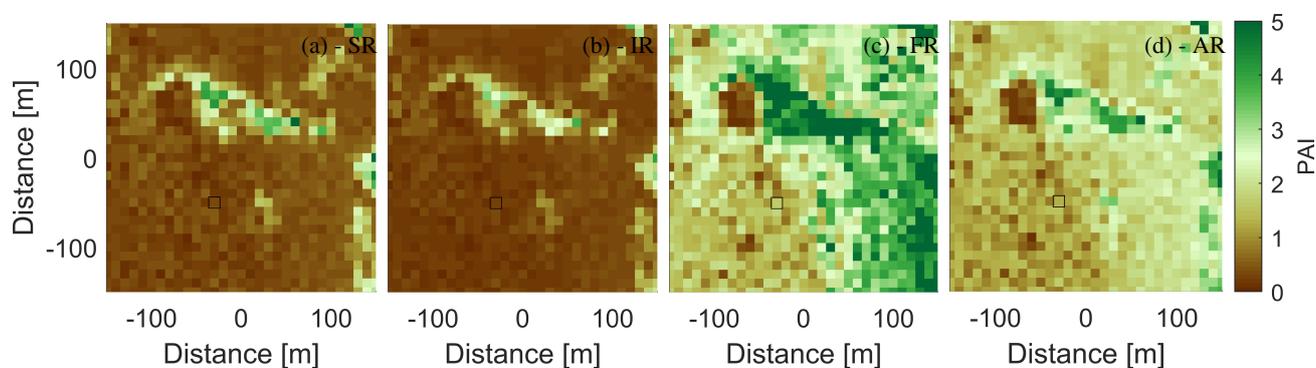


Figure 4. PAI estimates from Falster winter scan by various methods. *a)*, *b)*, *c)* and *d)* show PAI in a $10\text{ m} \times 10\text{ m}$ square grid based on the methods SR, IR, FR and AR respectively. The black square marks the edges for the data used in Section 4.2.

4.2 Single grid cell illustration

In order to illustrate how the methods work with the ALS data a grid cell encapsulating a single tree was selected, for which all returns are shown (Fig. 5). The location of the grid cell is highlighted in Fig. 3 and 4 with a black square. The size of the markers in the plot indicate the intensity of the return *as seen by the algorithm*.

There is a marked difference between the lidar returns in the summer and winter scan. In summer, the majority of the returns come from the upper part of the foliage, predominately by first and second order returns of relatively high backscatter intensity. In winter, the intensity of the returns from the upper canopy is much lower, and the only reflections with high intensity comes from the stem and larger branches, as well as the ground.

In the winter scan, the difference in scaling technique between the SR and IR methods is visible through the dominance of the ground reflections in IR. Since SR uses scaling of each pulse, the intensity of a higher order ground reflection is always smaller than that of a single first reflection, no matter the actual back scattered intensity. This can be an advantage, as some surfaces may very effectively block the pulse despite sending relatively low back scattered intensity.

The reason for the high PAI estimation by FR and AR in the case of Falster winter becomes apparent from Fig. 5. The ground returns are biased towards higher order returns, which are not included in FR. AR does on the other hand include the higher order returns, but neither FR or AR takes into account the small intensity in the returns from the upper canopy, ultimately leading to an overestimation of the PAD and PAI.

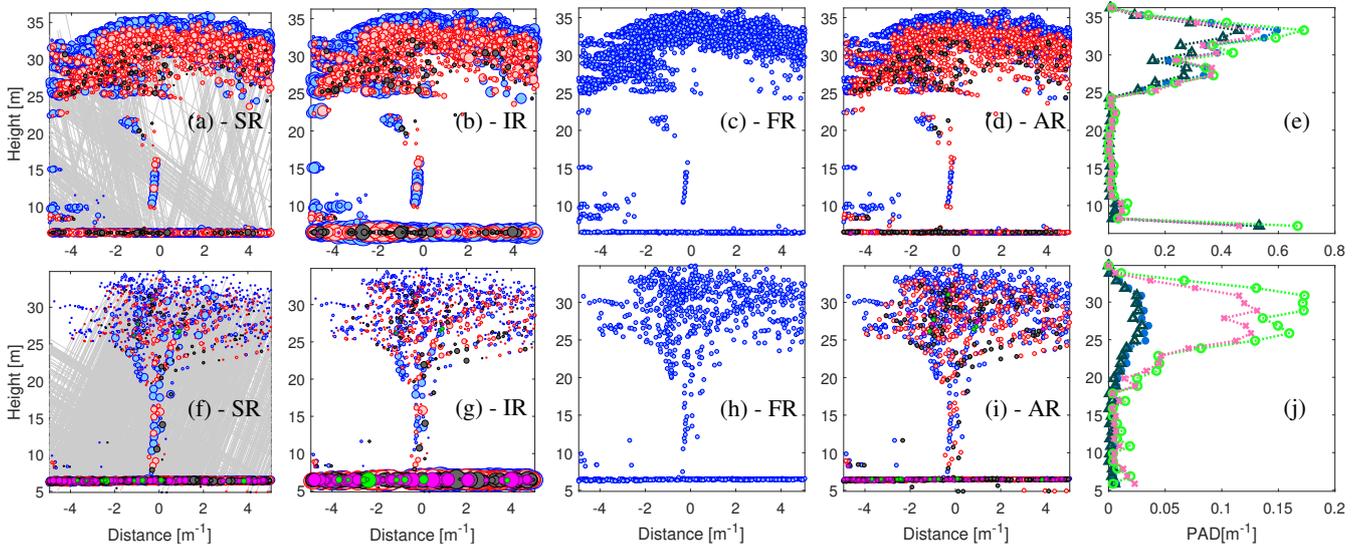


Figure 5. Backscatter reflections from within a $10\text{ m} \times 10\text{ m}$ grid box, taken from the Falster summer (upper plots) and winter (lower plots) scans. The color indicate return number: First returns - blue, second returns - red, third returns - black, fourth returns - purple and fifth returns - green. The size of the marker indicate weight of the return according to the four methods. Scaled return number, SR (*a* and *f*) is determined from Eq. 7, for IR (*b* and *g*) it is scaled by the medium intensity of the first order canopy returns in the grid box and for FR and AR (*c*, *d*, *h* and *i*), all returns are weighted the same. Note that the scale changes between the summer and winter. The grey lines indicate the path of the lidar beam between first and higher order returns. Thus, lines are only shown for beams which has more than one return. *e* and *j* shows the corresponding PAD for summer and winter scan respectively, SR is represented by blue stars, IR is represented by dark green triangles, FR is represented by bright green circles and AR is represented by pink crosses. The location of the grid cell can be seen in Fig. 3 and 4.

4.3 Sensitivity to grid size

Inherent in the Beer-Lambert law is the assumption of homogeneous spatial distribution of the scattering elements. The assumption is generally violated in forests, which have both vertically and horizontally inhomogeneous distribution of scattering elements. With vertical binning in thin planes, the effect of vertical inhomogeneity can be minimized, but heterogeneity in the horizontal is still problematic. The reason for this can be illustrated by studying a single layer PAI calculation for which the Beer-Lambert law simplifies to

$$\text{PAI} = -2 \ln \left(\frac{\sum_{\text{ground}} R_i}{\sum_{\text{all}} R_i} \right) \quad (9)$$

Calculating the PAI of an area by reducing the horizontal resolution so that there is only one single cell for the whole area is then equivalent to $-2 \ln \left(\overline{\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i} \right)$, where the vertical bar represents area average. Calculating the PAI for the whole area by taking the average of PAI from a higher resolution is instead $-2 \ln \left(\overline{\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i} \right)$. Because of the curvature of the logarithm function, Jensen's inequality says that $-2 \ln \left(\overline{\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i} \right) > -2 \ln \left(\overline{\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i} \right)$ which



means that sub grid size variations in the ratio of canopy to ground returns will lead to lower PAI estimates for larger grid size. A physical interpretation is that the Beer-Lambert law assumes that beneath a scattering element there is a certain probability that there exist other, shadowed, scattering elements. Smearing out the distribution of elements thus reduce the amount of shaded elements in the calculations. Some methods to correct for heterogeneity have been published, but of those, only the method of Hu et al. (2014) is applicable to ALS data (Yan et al., 2019).

The mapping of PAI from $\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i$ by Eq. 9 is shown in Fig. 6 in the x-y plane. Jensen's inequality is seen from the strong curvature of Eq. 9, which has the effect that an even distribution in $\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i$ becomes positively skewed, with the long tail on the high PAI side. Lowering the resolution means narrowing the distribution of $\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i$, thereby shortening the long tail more than the short tail in the PAI distribution, effectively lowering the PAI average.

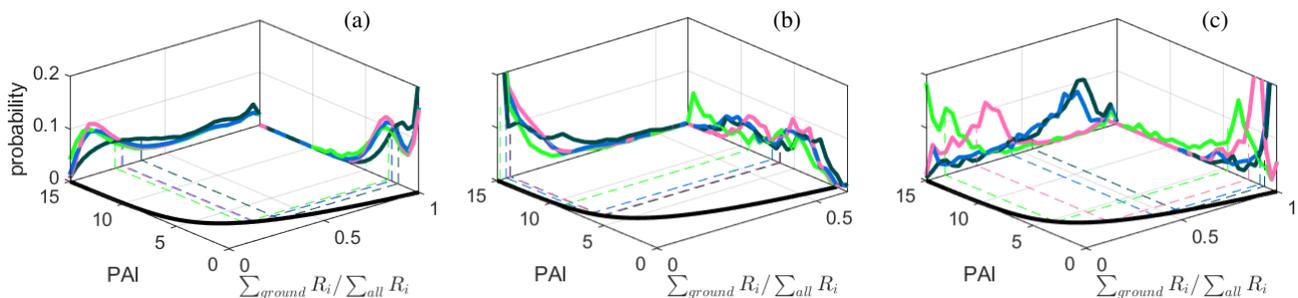


Figure 6. The mapping of PAI from the ratio of ground returns to total number of returns ($\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i$ at 10x10 m resolution) by Eq. 9 (black line in the x-y plane). On the vertical axis are shown the respective distributions of PAI (y-z plane) and $\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i$ (x-z plane), at 10x10 m resolution, for the four methods (SR -blue, IR - dark green, FR - light green and AR - pink). The dashed lines indicate the area average PAI from averaging PAI at 10x10 m resolution. For scans Norunda (a), Falster summer (b) and Falster winter (c).

The sensitivity of the routines to grid size was investigated by simply calculating PAD and PAI for the three data sets with different horizontal resolution, ranging from 10 m × 10 m to 100 m × 100 m. Before doing the PAI and PAD calculations the LAS data was flattened in the vertical direction by subtracting the 10 m × 10 m ground height in order to avoid effects coming from terrain elevation changes within the grid cell.

Table 2 summarizes the sensitivity to grid size showing that for a magnitude change in grid size, the change in PAI ranges from between a few percent to more than 30% depending on site and method. The AR method was found to be the least sensitive to grid size changes, followed by the SR method. The FR method was the most sensitive. The results indicate that IR and FR sees more heterogeneity than the AR and SR methods. This is also observed in Fig. 6, where the distributions of $\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i$ for IR is skewed towards high values (low PAI) and FR towards low values (high PAI). The observation in Section 4.1 that the AR method predicts less variations in PAI spatially, mainly by underprediction in dense areas relative to the other routines can also be seen in Fig. 6, particularly in Fig 6 (c), Falster winter, where the distribution of $\sum_{\text{ground}} R_i / \sum_{\text{all}} R_i$ is much less skewed for AR then for the other methods. As also noted in Section 4.1, the SR method behaves as the FR method in sparse regions and like the IR method in dense regions, thus avoiding overprediction in the most dense areas and



Table 2. Values of average PAI for the two sites, using different calculation methods and grid resolutions. *: A number of PAD estimates were not possible due to fully attenuated first returns before the ground level. The data was filled with PAD values of 0.1. Less than 1 % of the estimates were affected except for Falster summer at a resolution of 10 m × 10 m, where 4 % were affected.

Site and method	10 m × 10 m	20 m × 20 m	50 m × 50 m	100 m × 100 m
Norunda SR	2.56	2.42 (0.95)	2.35 (0.92)	2.29 (0.90)
Norunda IR	1.95	1.80 (0.92)	1.70 (0.87)	1.64 (0.84)
Norunda FR	2.87*	2.69 (0.94)	2.58 (0.90)	2.51 (0.87)
Norunda AR	2.60	2.51 (0.96)	2.45 (0.94)	2.41 (0.93)
Falster summer SR	7.09	6.58 (0.93)	5.94 (0.84)	5.68 (0.80)
Falster summer IR	6.34	5.74 (0.90)	5.03 (0.79)	4.75 (0.75)
Falster summer FR	9.23*	8.68* (0.94)	7.39 (0.80)	6.77 (0.73)
Falster summer AR	6.45	6.07 (0.94)	5.62 (0.87)	5.46 (0.85)
Falster winter SR	1.60	1.45 (0.91)	1.31 (0.82)	1.23 (0.77)
Falster winter IR	1.17	1.01 (0.86)	0.86 (0.74)	0.79 (0.68)
Falster winter FR	4.63*	4.13 (0.89)	3.76 (0.81)	3.48 (0.75)
Falster winter AR	2.66	2.59 (0.97)	2.50 (0.94)	2.48 (0.93)

underprediction in the least dense areas. As shown in Section 4.2 the weight of the first order ground returns is limited in the SR method, while for IR the reflectivity of the ground leads to lower PAI estimates, thereby widening the overall estimated PAI distribution.

Average PAD profiles for the three data sets under varying grid size is seen in Fig. 7. When split up into vertical profiles, the data in Table 2 shows some interesting differences between the methods. Overall, even though the profiles are rather similar apart from difference in magnitude, especially for Falster winter, there are also some differences in the shape of the profile. In Norunda and Falster summer, the SR and AR methods appear very similar from Table 1 and Fig. 2 (a), (d) and (g) as well as Falster summer (g). However, studying the PAD profiles in Norunda, the SR method shows a less bottom heavy profile, with a PAD maximum above 10 m height, in the canopy of most trees, whereas the AR method has the maximum PAD at around 5 m. For Falster summer, the difference between SR and AR is mainly in the dense regions (see, Fig. 3 (a) and (d)) which in the average profiles translates to a less pronounced upper canopy peak in AR. In the Falster winter case, the profiles reveal a difference between the intensity based SR and IR, with maximums at lower height and the FR and AR with maximums at higher heights. This is consistent with the case study in Sec. 4.2, where a lack of intensity information regards the upper canopy reflections with higher weight.

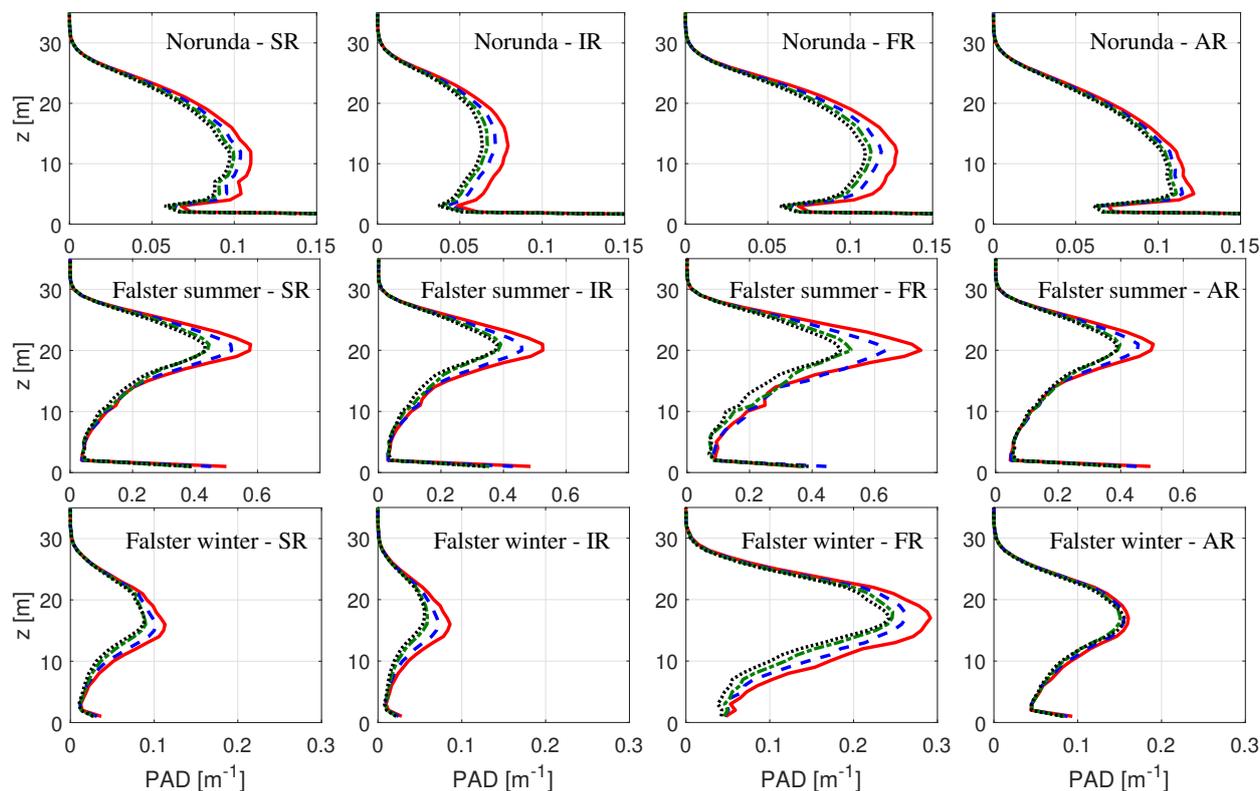


Figure 7. Average PAD profiles for the Norunda (upper), Falster summer (middle) and Falster winter (lower) sites. The averages are based on $10\text{ m} \times 10\text{ m}$ (red full line), $20\text{ m} \times 20\text{ m}$ (blue dashed line), $50\text{ m} \times 50\text{ m}$ (green dash dotted line) and $100\text{ m} \times 100\text{ m}$ (black dotted line) resolution. From methods SR, IR, FR and AR (left to right).

5 Discussion

5.1 Applicability of the results

The study includes both needle and broadleaf trees for in-leaf and leafless conditions, thereby covering important types of forests. However, there are still many types of canopies for which the routines have not been compared. Both theoretical considerations (Section 3.1) and the presented results indicate that the new method is less sensitive to extremes in forest density (very dense, very heterogeneous, or very sparse forests) than the other methods. Thus, there is reason to assume that the new method works well also in a wider range of conditions than tested in this study.

5.2 Comparison with other ALS routines

As with forest types, the data sets presented in this study reflect only a subset of measurement devices and flight heights. It is clear that the FR method is sensitive to flight height as the beam width determines the probability of first returns penetrating



the lower canopy. In this regard the SR and IR methods are preferable. However, the IR method, and to some extent the SR method, could be sensitive to narrow beam widths, as then the reflectivity, as opposed to element size determines the backscatter intensity. In the case of very narrow beam widths, the SR method should be the preferable over the IR method since the probability of having higher order returns decreases with narrower beams, and in the case of only first order returns, the SR method is indifferent to the intensity information. If, as shown in Wagner et al. (2008), the reflective properties of the ground is higher than that of the canopy, the IR method will further underestimate the PAI.

As pointed out by Almeida et al. (2019), the AR method is dubious from a theoretical perspective as higher order returns only appear if there are first returns, thus making it questionable to regard them with the same weight. From a practical standpoint, the AR method seems reasonable, and could be a good alternative to the new method if intensity information is not available. It is important to notice though, that both FR and AR resulted in much higher PAI and PAD estimates for the leafless canopy, based on the too high weights of reflections from the upper canopy.

The method of Vincent et al. (2017) is similar to the method presented herein, and thus potentially carries many of the advantages of the SR method. However, their use of weights based on the *average* intensity per return number and number of returns, is a key difference from using the actual intensity fraction of the return within the shot. In their study, Vincent et al. (2017) found that the rank of return number and number of returns described 51 % of the variance in intensity, which indicates that using average intensity numbers omits roughly half of the variations that is captured by using pulse-specific intensity information. For the data sets presented in this study, the average intensity roughly splits evenly among the returns in the two full-leaf scans, but the leafless scan is very different, for which the average ground return consistently is more than twice as intense as any of the canopy returns, regardless of the number of returns. When we tested the difference between the SR method using individual and average weights the absolute relative difference in PAI estimates between the two was less than 5 % in the full leaf scans, but for the winter scan the difference was on average 25 %. Hence, if possible, the use of individual weights should be preferred, especially during the leafless state.

In cases of strong heterogeneity and focus on actual PAD, as opposed to effective PAD, the method presented in Detto et al. (2015) provides an alternative to the approximation Beer-Lambert law, and future research in that direction is interesting, especially with increasing access to computational power.

5.3 Comparison with ground based routines

Comparing the estimates of PAI and PAD from ALS based methods to those of ground based methods warrant a discussion. The size and shape of the ALS grid box needed for a fair comparison is difficult to anticipate. In this study we chose a cylindrical grid cell with diameter equal to the height of the forest. While it is clear that a cylindrical shape is preferable to a square, there is uncertainty in the optimal size of the cell. In addition to that, the timing of ALS and ground based measurements were not matching completely. Since there is further uncertainty connected to the degree of heterogeneity in the cells, the comparison with ground based methods should not be viewed as a strong argument for choosing one or the other of the ALS based techniques. Instead, the comparison with ground based methods illustrates that all four ALS based methods estimate the same order of magnitude of PAI as the more established ground based methods. It is furthermore clear from comparing the PAI



maps to the aerial photos that regions of dense and sparse forest, as well as minor clearings and roads are well identified by the ALS based routines, which for many applications vindicates using ALS based estimates despite uncertainty in the exact PAI magnitude.

5.4 Strategies for selecting the PAD and PAI resolution from ALS

- 5 How to optimally select resolution for PAI/PAD estimation by ALS is still unknown. It is clear that a coarser resolution leads to an underestimation in PAD if there is heterogeneity in the distribution of canopy elements (due to Jensen's inequality). Boudreault et al. (2015) presented a study for grid resolution using the FR algorithm in a spruce forest, varying resolution from $5\text{ m} \times 5\text{ m}$ to $30\text{ m} \times 30\text{ m}$. The findings in this study agree with the results of Boudreault et al. (2015) for the change between $10\text{ m} \times 10\text{ m}$ and $20\text{ m} \times 20\text{ m}$ resolution. As in Boudreault et al. (2015) we observe larger PAI changes for higher resolution.
- 10 It should also be noted that results shown in Fig. 6 demonstrate that grid size sensitivity is dependent on the magnitude of PAD/PAI, with dense forests having a more non-linear mapping from ALS to PAD/PAI than less dense, and that this dependence enhances grid size sensitivity. In addition to this forest density in combination with the scan density (number of returns per m^2) sets an upper limit to the possible resolution. It is evident from the presented study that the FR method is problematic in this regard, and a strategy to manage beam attenuation has to be adopted.
- 15 Furthermore, the optimal resolution also depends on the application. In wind modelling there is for example the optimal ratio of resolution of PAD field and flow field to consider, where the resolution of the flow field puts an upper limit to the resolution to the PAD field. Ivanell et al. (2018) concluded that for a predominantly forested area, where the drag force by the forest on the wind was modelled by means of PAD profiles, the value of the surface roughness below the canopy had very little impact on the flow above the forest. This implies that for wind modelling, uncertainty in the lower part of the canopy is acceptable,
- 20 thus enabling a higher PAD resolution. Further research in how to optimally choose grid resolution is needed.

6 Conclusion

We have presented a new method to calculate PAI and PAD from airborne laser scans, which uses all available information in the raw data. In order to evaluate the performance of the new method it was used together with three reference methods on sites with contrasting characteristics. The results show that the new routine shows no marked weakness with respect to sensitivity

25 to forest characteristics, grid resolution, beam attenuation and ratio of ground to canopy returns, whereas all three reference methods showed problems in either of those categories. In summary, the results indicate that the new algorithm is robust and would therefore perform well for a wider range of vegetation types than tested here. The results furthermore indicate that the new algorithm would be less sensitive to scanning characteristics such as flight height and scan density.

The new routine is applicable provided that each data point in the data set contains information on the number of returns in the pulse, the return number of the specific point and that the data is ordered such that consecutive points in the same pulse

30 follow each other in the data set. For the data sets that the authors have come across, this criterion has always been fulfilled. The results showed a dependence on the grid resolution of the derived PAI and PAD fields. The grid sensitivity was linked to



the violation of the homogeneity assumption in the Beer-Lambert law, which can also be explained by the Jensen's inequality for convex functions. The grid resolution sensitivity is varying between the methods (from 7 and 32 % reduction in PAI for a magnitude change in grid size), where the method by Almeida et al. (2019) showed the least sensitivity and the new method was second best.

5 Despite the identified issues with airborne lidars for the determination of plant area and plant density, the similarities between the four methods highlight the prospect of obtaining PAD and PAI with great spatial coverage, that are both precise and accurate, especially in regards to the uncertainty levels in established ground-based methods Yan et al. (2019). The use of airborne lidar scans to derive maps of PAD and PAI also looks like a promising way to greatly reduce the subjectivity, and likely uncertainty, of any atmospheric modelling that involves air-canopy interactions, such as drag, heat transfer and gas-exchange.

10 *Code availability.*

A Python (Python Software Foundation, <https://www.python.org/>) as well as MATLAB (Mathworks, Inc.) implementation of the algorithm that outputs PAD, PAI, forest height and ground elevation from laserdata in the .las format has been developed and is freely available at <https://github.com/johanarnqvist/ALS2PAD.git/>.

15 *Author contributions.* All authors contributed to study design, writing the manuscript, preparing data sets, analysis and discussion of the results. J. Arnqvist developed the new routine and had the main responsibility for writing, data analysis and figure preparation.

Competing interests. None of the authors have competing interests.

20 *Acknowledgements.* We acknowledge Irene Lehner for providing ground based measurements at the ICOS Sweden station Norunda.. Dr Ferhat Bingöl is acknowledged for the photos in Fig. 1 and Poul Therkild Soerensen is acknowledged for the reference measurements of PAI at the Falster site. Part of this work was conducted within STandUP for Wind, a part of the STandUP for Energy strategic research framework (Sweden). The Independent Research Fund Denmark, via the The Single Tree Experiment project [Grant No. 6111-00121B], and the European Commission, via the New European Wind Atlas Project, are also acknowledged for funding.



References

- Almeida, D. R. A. d., Stark, S. C., Shao, G., Schietti, J., Nelson, B. W., Silva, C. A., Gorgens, E. B., Valbuena, R., Papa, D. d. A., and Brancalion, P. H. S.: Optimizing the Remote Detection of Tropical Rainforest Structure with Airborne Lidar: Leaf Area Profile Sensitivity to Pulse Density and Spatial Sampling, *Remote Sensing*, 11, 92, <https://doi.org/10.3390/rs11010092>, 2019.
- 5 Boudreault, L. É., Bechmann, A., Tarvainen, L., Klemedtsson, L., Shendryk, I., and Dellwik, E.: A LiDAR method of canopy structure retrieval for wind modeling of heterogeneous forests, *Agricultural and Forest Meteorology*, 201, 86–97, <https://doi.org/10.1016/j.agrformet.2014.10.014>, 2015.
- Bréda, N. J. J.: Ground based measurements of leaf area index: a review of methods, instruments and current controversies, *Journal of Experimental Botany*, 54, 2403–2417, <https://doi.org/10.1093/jxb/erg263>, <https://dx.doi.org/10.1093/jxb/erg263>, 2003.
- 10 Dellwik, E., Bingöl, F., and Mann, J.: Flow distortion at a dense forest edge, *Quarterly Journal of the Royal Meteorological Society*, 140, 676–686, <https://doi.org/10.1002/qj.2155>, financial support: The Danish Council for Strategic research (Committee for Energy and Environment grant no. 2104-05-0076 and the Center for Computational Wind Turbine Aerodynamics and Atmospheric Turbulence grant no. 09-067216) and the Danish Agency for Science Technology and Innovation (grant no. 274-08-0380), 2014.
- Detto, M., Asner, G. P., Muller-Landau, H. C., and Sonnentag, O.: Spatial variability in tropical forest leaf area density from multireturn lidar and modeling, *Journal of Geophysical Research: Biogeosciences*, 120, 294–309, <https://doi.org/10.1002/2014JG002774>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JG002774>, 2015.
- 15 Freier, J.: Approaches to characterise forest structures for wind resource assessment using airborne laser scan data, Master's thesis, University of Kassel, Kassel, Germany, <http://publica.fraunhofer.de/dokumente/N-464768.html>, 2017.
- Gielen, B., Acosta, M., Altimir, N., Buchmann, N., Cescatti, A., Ceschia, E., Fleck, S., Hörtnagl, L., Klumpp, K., Kolari, P., Lohila, A., 20 Loustau, D., Marañón-Jimenez, S., Manise, T., Matteucci, G., Merbold, L., Metzger, C., Moureaux, C., Montagnani, L., Nilsson, M. B., Osborne, B., Papale, D., Pavelka, M., Saunders, M., Simioni, G., Soudani, K., Sonnentag, O., Tallec, T., Tuittila, E.-S., Peichl, M., Pokorny, R., Vincke, C., and Wohlfahrt, G.: Ancillary vegetation measurements at ICOS ecosystem stations, *International Agrophysics*, 32, 645–664, <https://doi.org/10.1515/intag-2017-0048>, <http://dx.doi.org/10.1515/intag-2017-0048>, 2018.
- Hopkinson, C. and Chasmer, L.: Remote Sensing of Environment Testing LiDAR models of fractional cover across multiple forest ecozones, 25 *Remote Sensing of Environment*, 113, 275–288, <https://doi.org/10.1016/j.rse.2008.09.012>, <http://dx.doi.org/10.1016/j.rse.2008.09.012>, 2009.
- Hu, R., Yan, G., Mu, X., and Luo, J.: Indirect measurement of leaf area index on the basis of path length distribution, *Remote Sensing of Environment*, 155, 239 – 247, <https://doi.org/https://doi.org/10.1016/j.rse.2014.08.032>, <http://www.sciencedirect.com/science/article/pii/S0034425714003393>, 2014.
- 30 Ivanell, S., Arnqvist, J., Avila, M., Cavar, D., Aurelio Chavez-Arroyo, R., Olivares-Espinosa, H., Peralta, C., Adib, J., and Witha, B.: Micro-scale model comparison (benchmark) at the moderately complex forested site Ryningsnäs, *Wind Energy Science*, 3, 929–946, 2018.
- Lantmäteriet: Tech. rep., Lantmäteriet, Sweden, https://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/hojddata/laserdata_nh_v2.6.pdf, 2016.
- Monsi, M. and Saeki, T.: On the factor light in plant communities and its importance for matter production, *ANNALS OF BOTANY*, 95, 35 549–567, <https://doi.org/10.1093/aob/mci052>, 2005.
- Morsdorf, F., Kötz, B., Meier, E., Itten, K. I., and Allgöwer, B.: Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction, 104, 50–61, <https://doi.org/10.1016/j.rse.2006.04.019>, 2006.



- Patton, E. G., Sullivan, P. P., Shaw, R. H., Finnigan, J. J., and Weil, J. C.: Atmospheric Stability Influences on Coupled Boundary Layer and Canopy Turbulence, *Journal of the Atmospheric Sciences*, 73, 1621–1647, <https://doi.org/10.1175/JAS-D-15-0068.1>, <https://doi.org/10.1175/JAS-D-15-0068.1>, 2016.
- Richardson, J. J., Moskal, L. M., and Kim, S.-H.: Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR, *Agricultural and Forest Meteorology*, 149, 1152 – 1160, <https://doi.org/https://doi.org/10.1016/j.agrformet.2009.02.007>, <http://www.sciencedirect.com/science/article/pii/S0168192309000409>, 2009.
- Smallman, T. L., Moncrieff, J. B., and Williams, M.: WRFv3.2-SPAv2: development and validation of a coupled ecosystem-atmosphere model, scaling from surface fluxes of CO₂ and energy to atmospheric profiles, *Geoscientific Model Development*, 6, 1079–1093, <https://doi.org/10.5194/gmd-6-1079-2013>, <https://www.geosci-model-dev.net/6/1079/2013/>, 2013.
- 5 Sogachev, A., Menzhulin, G. V., Heimann, M., and Lloyd, J.: A simple three-dimensional canopy – planetary boundary layer simulation model for scalar concentrations and fluxes, *Tellus B*, 54, 784–819, <https://doi.org/10.1034/j.1600-0889.2002.201353.x>, <https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0889.2002.201353.x>, 2002.
- Solberg, S., Næsset, E., Hanssen, K. H., and Christiansen, E.: Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning, *Remote Sensing of Environment*, 102, 364–376, <https://doi.org/10.1016/j.rse.2006.03.001>, <http://linkinghub.elsevier.com/retrieve/pii/S0034425706000964>, 2006.
- 15 Vincent, G., Antin, C., Laurans, M., Heurtebize, J., Durrieu, S., Lavalley, C., and Dautzat, J.: Mapping plant area index of tropical evergreen forest by airborne laser scanning. A cross-validation study using LAI2200 optical sensor, *Remote Sensing of Environment*, 198, 254 – 266, <https://doi.org/https://doi.org/10.1016/j.rse.2017.05.034>, <http://www.sciencedirect.com/science/article/pii/S003442571730233X>, 2017.
- Wagner, W., Hollaus, M., Briese, C., and Ducic, V.: 3D vegetation mapping using small-footprint full-waveform airborne laser scanners, *International Journal of Remote Sensing*, 29, 1433–1452, <https://doi.org/10.1080/01431160701736398>, <https://doi.org/10.1080/01431160701736398>, 2008.
- 20 Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Wofsy, S. C., Shaver, G. R., Melillo, J. M., Munger, J. W., Fan, S.-M., and Nadelhoffer, K. J.: Modelling the soil-plant-atmosphere continuum in a Quercus-Acer stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties, *Plant, Cell & Environment*, 19, 911–927, <https://doi.org/10.1111/j.1365-3040.1996.tb00456.x>, <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3040.1996.tb00456.x>, 1996.
- 25 Yan, G., Hu, R., Luo, J., Weiss, M., Jiang, H., Mu, X., Xie, D., and Zhang, W.: Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives, *Agricultural and Forest Meteorology*, 265, 390 – 411, <https://doi.org/https://doi.org/10.1016/j.agrformet.2018.11.033>, <http://www.sciencedirect.com/science/article/pii/S0168192318303873>, 2019.