Comparison of Semi-Physical and Empirical Models in the Estimation of Boreal Forest Leaf Area Index and Clumping With Airborne Laser Scanning Data

Shaohui Zhang, *Graduate Student Member, IEEE*, Lauri Korhonen, Mait Lang, Jan Pisek, Gastón M. Díaz, Ilkka Korpela, Zhongyu Xia, Hanna Haapala, and Matti Maltamo

***Abstract*— Leaf area index (LAI) is an important forest canopy variable that is related to various biophysical processes of forest ecosystems. Airborne laser scanning (ALS) has shown promise in modeling and mapping LAI using different types of ALS metrics. The most common ways of modeling LAI with ALS data are multivariate empirical models and the semi-physical model shape derived from the Beer–Lambert law of radiation attenuation. We tested the utility of ALS-based empirical and semi-physical models in the estimations of effective LAI (LAI_e), canopy clumping index (Ω_E), and clumping-corrected LAI at three boreal forest sites in Finland. In semi-physical models, the all echo penetration index (API) showed consistently the best performance in predicting LAI_e. It is, therefore, a robust and potentially the most transferable predictor using this model shape. Empirical models overall yielded slightly better model fits compared to the semi-physical models, yet they are also more prone to overfitting. In addition, empirical models had constantly lower accuracies when predicting LAI than LAI_e. We also tested the utility of ALS-based multi-angular canopy gap fraction metrics that were derived from polar transformed ALS point clouds. Images derived from polar transformed point clouds can be analyzed similarly to digital hemispherical photographs (DHPs) to obtain canopy gap fractions. The results showed that polar metrics derived from polar transformed ALS data can provide supporting information to empirical models in the estimation of LAI_e, LAI, and especially Ω_E. In particular, a combination of ALS penetration indices and polar metrics yielded positive results in Ω_E estimation.**

***Index Terms*— Airborne laser scanning (ALS), canopy clumping, forest canopy, leaf area index (LAI), light detection and ranging (LiDAR).**

I. Introduction

LEAF area index (LAI), here defined as half of the total leaf area per unit horizontal ground surface area [1], [2], is a key parameter that describes canopy properties of forest ecosystems. It measures the amount of leaf material present in the canopy, making it a suitable input for modeling biosphere-atmosphere mass and energy exchanges, such as photosynthesis and transpiration [3]. The Global Climate Observing System (GCOS) has identified LAI as one of the variables that are crucial in global biosphere-atmosphere models. LAI also contributes to two essential biodiversity variable classes: ecosystem structure and ecosystem function [4]. In forestry, LAI is a key input of process-based growth models [5], a main driver of forest albedo modeling [6], and an indicator of defoliation when monitoring forest health [7].

Effective LAI (LAI_e) is the LAI value derived from indirectly measured gap fractions assuming that the canopy only consists of foliage elements that are opaque and randomly distributed according to the Beer–Lambert law. In reality, however, foliage rarely has a random distribution due to clumping at different levels, let alone that canopies consist of a diversity of canopy elements, not only foliage. Previous research has attempted to classify clumping at shoot, branch, crown, and landscape levels, or simply at between-crown and within-crown levels [8], [9]. The clumping index (Ω), which is defined as the ratio of LAI_e to LAI, is used to quantify the clumping effect [10], [11]. It denotes the degree of non-randomness of foliage in an observed canopy. When the foliage has a random spatial distribution, $\Omega = 1$. When canopy elements are clumped in a way that more canopy gaps are observed than if they were distributed randomly, $\Omega < 1$. On the contrary, $\Omega > 1$ implies that foliage elements are regularly distributed and less canopy gaps are observed than if they were distributed randomly. Accounting for the clumping

Manuscript received 10 July 2023; revised 6 October 2023 and 30 November 2023; accepted 7 January 2024. Date of publication 12 January 2024; date of current version 30 January 2024. The work of Shaohui Zhang, Lauri Korhonen, and Matti Maltamo was supported by the Academy of Finland Flagship Program Forest-Human-Machine Interplay-Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) under Grant 337127 and Grant 357906. The work of Lauri Korhonen, Zhongyu Xia, Hanna Haapala, and Matti Maltamo was supported by the Academy of Finland under Grant 332707. The work of Jan Pisek was supported by the Estonian Research Council under Grant PRG1405 and by Ministry of Education and Research, Centre of Excellence for Sustainable Land Use (TK232). *(Corresponding author: Shaohui Zhang.)*

Shaohui Zhang, Lauri Korhonen, Zhongyu Xia, Hanna Haapala, and Matti Maltamo are with the School of Forest Sciences, University of Eastern Finland, 80101 Joensuu, Finland (e-mail: shaohui.zhang@uef.fi).

Mait Lang is with the Institute of Forestry and Engineering, Estonian University of Life Sciences, 51014 Tartu, Estonia.

Jan Pisek is with the Tartu Observatory, University of Tartu, Tõravere, Tartu 61602, Estonia.

Gastón M. Díaz is with the National Scientific and Technical Research Council—Argentina (CONICET), Department of Geomatics, Patagonian Andes Forest Research and Extension Centre (CIEFAP), Esquel 9200, Argentina.

Ilkka Korpela is with the Department of Forest Sciences, University of Helsinki, 00014 Helsinki, Finland.

Digital Object Identifier 10.1109/TGRS.2024.3353410

effect is therefore crucial for accurate LAI estimation. Failure to do so would lead to an underestimation of LAI ranging from 30% to 70% [12], [13], [14]. Furthermore, simultaneous large-area mapping of LAI_e, LAI, and Ω remains particularly challenging due to their temporal and spatial variability [15], [16]. However, their importance demands that more effort be paid to obtaining reliable, spatially continuous maps of these variables.

Remote sensing technology offers a great opportunity for collecting forest information over large coverages of land area in a cost-effective way. In many countries, airborne laser scanning (ALS) is routinely used in the estimation of forest attributes, such as basal area and timber volume [17], [18], [19]. ALS has also been successfully applied for the estimation of LAI [20], [21].

Different kinds of light detection and ranging (LiDAR) systems, such as full-waveform (FW) and discrete return (DR) sensors, have been deployed on airborne platforms for the retrieval of forest attributes. Armston et al. [22] found that airborne FW LiDAR sensors produced more accurate results than DR sensors in the estimation of canopy gap fraction. Nevertheless, commercial DR scanners remain the most popular LiDAR systems in topographic and forest surveys, and the availability of DR data is therefore considerably better when large-area applications are considered.

The most common way of estimating LAI with ALS data is to build empirical models between field-measured LAI and ALS predictors. It means that field data are required to calibrate and validate the models, which imposes a challenge as field measurements of LAI are often unavailable. The majority of field data used for model construction and validation are from indirect optical measurements, such as digital hemispherical photographs (DHPs) [23], [24], [25], digital cover photographs (DCPs) [26] or the LAI-2000 Plant Canopy Analyser [27], [28]. The LAI derived from canopy gap fractions obtained from DHP, DCP or LAI-2000 data (indirect measurements) is the LAI_e, which does not account for woody materials and clumping effect. However, it can be converted to "true" LAI if leaf angle distribution, woody areas, and Ω are accounted for [29] and [30]. Thus, the estimation of Ω from ALS data is also important from the perspective of obtaining reliable LAI maps.

Over the years, many studies have assessed which types of metrics derived from DR LiDAR data should be used to predict LAI. For example, [31] found that in an empirical framework, height-related predictor variables were often useful as predictors of LAI. Intensity-based variables, which can be calculated for example as the ratio of the intensities from ground echoes and the total intensity, have also been applied to predict LAI [32]. However, in practice, the necessity of calibrating LiDAR intensity for variation in scan range and scanner settings may present extra difficulties [33]. Thus, the use of only geometric variables is in many cases more reliable.

In fully empirical models, the model predictors are selected from a pool of LiDAR metrics describing the point cloud obtained from the canopy. However, there is also a model shape for ALS-based LAI prediction as (1), which is derived from the Beer–Lambert law that is used to obtain LAI estimates from in situ measurements of multiangular gap fractions. In this semi-physical model, the near-vertical gap fraction is approximated by a canopy penetration index T computed from ALS data [20], [21], [34]

$$\mathrm{LAI}_e = -\beta * \ln(T) \tag{1}$$

where β is a coefficient estimated by regression analysis. If T is an unbiased estimate of near-vertical gap fraction, β can be interpreted as an estimate of leaf orientation within the canopy, which must be known to convert gap fraction in a given view direction to LAI [35]. Even if T is biased, the model will work if β is re-estimated by regression analysis based on LAI estimated in the field. However, if unbiased estimates of T can be obtained, the model can be applied in all forests where β is assumed to be similar. Thus, it is important to study how different canopy penetration indices compare with field-measured estimates of T. Yet, no single LiDAR penetration index has been uniformly agreed to have a superior accuracy as an estimator of gap fraction, as the penetration index is dependent on the type of LiDAR sensor and acquisition settings, such as flying heights and scanning angles [36].

The semi-physical model shape is simple and robust and has been shown to yield accurate predictions in different biomes [34], [35]. Thus, it is suitable for large-scale mapping of LAI_e. However, fully empirical models could perform even better, since increasing the number of predictors in a model usually leads to more accurate results, given that overfitting is avoided [37]. For example, LiDAR metrics computed from polar transformed ALS data, such as polar grid fractions, may contain canopy structural information that could help to improve the estimates of LAI and clumping [38], [39]. Polar transformed ALS point clouds can be rasterized to create figures that resemble in situ hemispherical figures and then processed to obtain polar metrics that describe canopy penetration in non-vertical directions. However, the utility and performance of polar metrics in the modeling of LAI for large areas still need further investigation.

The first aim of this study is to directly compare several ALS-based canopy penetration indices against field-measured estimates of near-vertical canopy gap fraction and assess if the least biased indices also perform the best as predictors in semi-physical LAI models. Next, we compare empirical and semi-physical modeling techniques in the estimation of LAI_e, LAI, and Ω with ALS data at three boreal forest sites using discrete-return ALS data. With empirical models, we specifically test the utility of polar-transformed metrics in addition to the commonly used LiDAR canopy height and density distribution metrics as well as canopy penetration indices.

II. Methodology

A. Study Sites

Three study sites including a total of 123 field plots in Southern and Eastern Finland were used in this study (Fig. 1). The plots at all three sites covered a wide range of tree species compositions and forest structures (Table I), but overall, the

TABLE I
FOREST INVENTORY ATTRIBUTES OF THE THREE STUDY SITES

Study sites	Number of plots	Height (m)				Basal area (m^2)			
		Min.	Mean	Max.	SD	Min.	Mean	Max.	SD
Heinola	30	2.6	13.8	26.9	7.5	1.0	14.8	34	8.2
Hyytiälä	73	2.2	16.8	34.3	7.0	0.5	22.9	51.3	10.7
Outokumpu	20	6.0	17.0	26.3	6.2	2.5	19.7	44	12.8

* Note: Abbreviation Min., Max., SD refer to the minimum, maximum and standard deviation respectively.

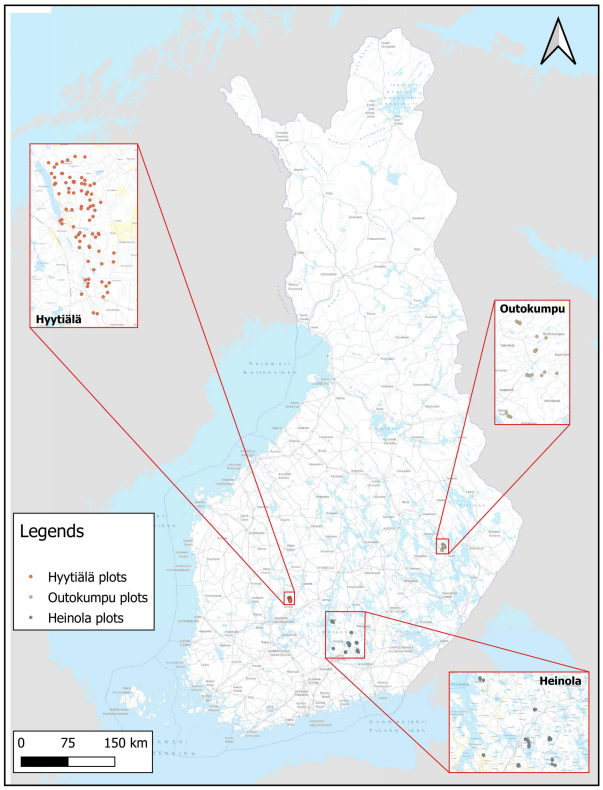

Fig. 1. Locations of the three LAI modeling study sites in Finland and their subordinate plots. The map contains data from the National Land Survey of Finland (2022) under the creative commons attribution 4.0 international license CC BY 4.0.

sites were similar to each other. The most common tree species were Scots pine (Pinus sylvestris L.), Norway spruce (Picea abies L. Karst), and birches (Betula spp L.). The forest structures ranged from sparsely wooded homogeneous pine stands growing on boglands to clumped heterogeneous old growth stands growing on fertile soils.

B. Digital Hemispherical and Cover Photographs

The Hyytiälä site was measured in 2011, and the DHP was acquired using a Nikon Coolpix 8800 camera and an FC-E9 fisheye converter. At the Outokumpu and Heinola sites that were measured in 2021–2022, a Canon EOS2000 camera and a Sigma 4.5 mm fisheye lens were used. The image acquisition schemes also varied at the different sites. At the Hyytiälä site, 12 hemispherical images were taken at plot level, whereas at the Outokumpu and Heinola sites, only five images were collected per plot. All measurements were made under overcast sky or near sunset to avoid direct sunlight. The cameras were fixed to a tripod at approximately 1.3 m height above ground and leveled using a two-axis bubble level. The

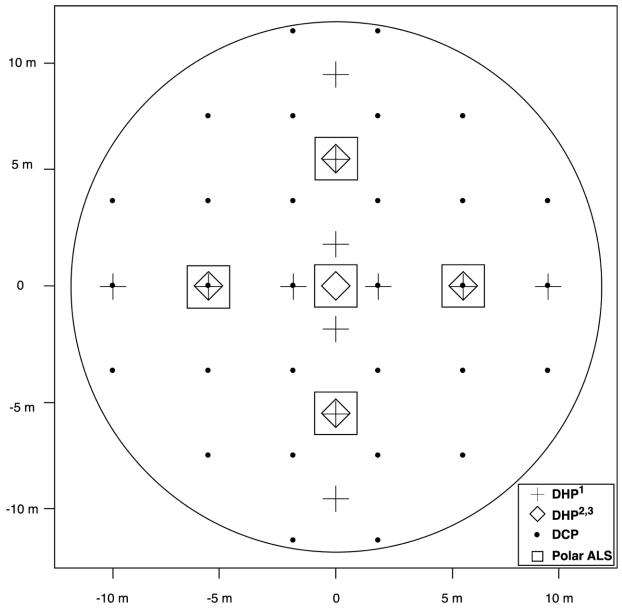

Fig. 2. DHPs and DCPs observation spots at plot level. In case the observation spots were close to tree stems, both types of photographs were taken at least 1 m away. DHP^1 refers to observations in site 1 (Hyytiälä) and $DHP^{2,3}$ observations in site 2 (Outokumpu) and site 3 (Heinola). Polar ALS refers to reference points at which polar metrics were calculated.

lens was then pointed upward with its focus set to infinity (Nikon) or using autofocus (Canon). Auto exposure bracketing function was used, with the base exposure set to −2 EV at +/− 1 EV stop, which resulted in a series of images at three consecutive EVs: −3, −2 and −1, and the one with the best exposure was manually picked for further image processing. The shooting mode was set to aperture priority and aperture was kept at $f/8$ to decrease the vignetting effect [25]. The DHP was saved in raw image format.

DCP was acquired at the same time as DHP. In each plot, a total of 30 DCP were acquired using an Olympus μ700 (Hyytiälä) or Canon SX200 IS (Outokumpu and Heinola) camera (Fig. 2). By default, the camera was set to aperture priority mode and automatic exposure was decreased by 1–2 stops to avoid overexposure. In very dense canopies this was not enough, and manual exposure had to be used instead. In such conditions, the aperture and shutter speed were adjusted so that the background sky was not overexposed. The images were taken by pointing the camera in the skywards direction and saved in JPEG format.

1) Processing of Hemispherical Photographs: The hemispherical photographs were processed in hemispherical project manager (HSP), a software that implements the LinearRatio method [23] for a single camera ($LinearRatio_{cs}$), as instructed in [40]. We chose this software as our preliminary results showed that the binarized images obtained by HSP had a slightly higher consistency than those obtained by the commonly used thresholding methods [41]. The software applies a linear conversion to digital raw files, based on the assumption that the digital numbers of raw DHP files are linearly related to the incident radiance. It then reconstructs an above-canopy reference image based on DHP taken below the canopy by

using sky radiance sampled from sufficiently large canopy gaps. The software first converts the raw DHP to 16-bit simple portable gray maps (PGM format) with the help of dcraw software (version 9.28). Only original blue pixels were extracted in the process because at this spectrum they have the highest contrast between the sky and the canopy. We used the switches of dcraw: $-d$ (document mode, no color, and interpretation), $-W$ (do not automatically brighten the image), $-g$ 1 1 (linear 16-bit custom gamma curve). We input camera-specific parameters to correct lens projection distortion and adopted the default value of 1.0 to offset the camera vignetting effect.

Next, sky pixel sampling markers were manually placed to create an above-canopy reference image, assuming that the canopy gaps are large enough (3 × 3 pixels). To do this for each hemispherical image, a mixed method combining a simple inverse distance weighted interpolation and a sky radiance model fit to the sky pixel samples was employed [25]. Two nearest sky samples were used to interpolate each pixel in the reconstructed above-canopy reference image, and the searching distance was set to 200–300 pixels, depending on the sky condition. For validating the reference image, the sky pixel values in canopy gaps had to have a light transmission P(gap_sky) close to 1. Finally, binarized gap fraction images were exported using automatic thresholds that yielded the same gap fraction as in the ratio images.

Gap fractions GF(θ) were then calculated from the binarized DHP following a ring-wise analysis. DHP was divided into six concentric rings (15° interval) similar to the well-known LAI-2000 plant canopy analyser. The weight of each ring was calculated as follows:

$$W_i = \frac{\sin\theta_i}{\sum_{n=1}^{n}\sin\theta_i} \tag{2}$$

where θ_i was the mean zenith angle of the ring (7°, 23°, 38°, 53°, 68° and 83°), and W_i the weight of the ring i. Note that the weight of the sixth ring was assigned to the fifth one, similar to the LAI-2000 method.

LAI_e was estimated from gap fractions using Miller [42]'s integral as follows:

$$\text{LAI}_e = -2\int_0^{\frac{\pi}{2}} \ln T(\theta)\cos(\theta)\sin\theta d\theta. \tag{3}$$

In practice, the above equation was approximated by the sum as follows:

$$\text{LAI}_e = -2\sum_{i=1}^{n} \ln(\overline{T_i})\cos(\theta_i)W_i \tag{4}$$

where $\overline{T_i}$ was the mean GF(θ) of each annulus ring from the DHPs collected at plot level. The effect of woody components (such as tree trunks and branches) was not removed, and for simplicity reasons, we used LAI_e to denote the effective plant area index.

Morphological closing and opening operations were employed to extract between- and within- crown gaps from the binarized DHPs [43]. This operation was controlled by a parameter called structuring element size, which was set to 10 for the images obtained by the Canon camera and 8 for the Nikon camera due to their different image resolutions. The structuring element size was determined by manually tuning it to the setting that yielded the closest approximation of the manually painted between-crown gaps at the plot level. The structuring element size was kept constant for all rings as its influence on the obtained gap fraction was negligible. The resultant mask was assumed to split the image into large between-crown gaps and a continuous canopy area with only small within-crown gaps.

Various methods have been proposed to correct the effect of canopy clumping, such as the LX [11], CC [10], CLX [44] and the recent LXG [45] methods. A detailed comparison can be found in [8]. Overall, all methods have consistent physical meanings with the Ω definition. We chose the CC method (5) because it intuitively corresponds to the method of morphological image analysis used in this study

$$\Omega_{\text{CC}}(\theta) = \frac{\ln[F_m(0,\theta)]}{\ln[F_{mr}(0,\theta)]}\frac{[1-F_{mr}(0,\theta)]}{[1-F_m(0,\theta)]} \tag{5}$$

where $F_m(0, \theta)$ denotes the measured mean canopy gap fraction at five annulus rings from images collected in the same plot, covering the range of 0°–75°. $F_{mr}(0, \theta)$ denotes the canopy gap fraction when the canopy has a random distribution of foliage, which was approximated by the mean within-crown gap fraction obtained by subtracting the mean between-crown gap fraction from the mean total gap fraction.

The element clumping index Ω_E that quantifies canopy foliage clumping at plot level was aggregated as an average of the directional $\Omega_{\text{CC}}(\theta_i)$ obtained from the five rings as follows:

$$\Omega_E = \frac{1}{n}\sum_{i=1}^{n}\Omega_{\text{CC}}(\theta_i). \tag{6}$$

An additional shoot-level clumping correction was introduced for stands with coniferous trees. Thus, the clumping-corrected LAI (neglecting plant woody materials) was calculated as follows:

$$\text{LAI} = \frac{\text{LAI}_e}{\Omega_E\cdot\Omega_s}P_c + \frac{\text{LAI}_e}{\Omega_E}(1-P_c) \tag{7}$$

where P_c is the proportion of the basal area of coniferous trees measured in the field, and Ω_s is the shoot silhouette area to total needle area ratio that was assumed to have a constant value of 0.56 [13], [46].

2) Processing of Cover Photographs: The DCP were processed with an in-house MATLAB script to obtain more representative estimates of the near-vertical canopy gap fraction GF(θ) [43]. The images were binarized to separate the sky and the background using the thresholding algorithm presented by [47]. Only the view angles 0°–15° from the zenith were included in the computation of GF(θ) to keep the view geometry near vertical.

C. LiDAR Data

Different LiDAR sensors and acquisition parameters were used at each site (Table II). Initial processing of the LiDAR data was done using LAStools (version 220310). Ground echoes were first classified using the lasground tool. Subsequently, all echoes were re-classified into four classes: single,

TABLE II
ALS SENSOR SPECIFICATIONS FOR THE LAI MODELING STUDY SITES

Sensor properties	Heinola	Hyytiälä	Outokumpu
Sensor name	Riegl VQ780i	Leica ALS60	Riegl VQ-1560 II
Date of Acquisition	June 14, 2021	August 2, 2011	12 June, 2020
Maximum scan angle, degrees	20	17	15
Mean flying altitude AGL* (m)	1265	760	2100
Mean pulse density (m^{-2})	6.87	9.39	5.58
Mean echo density (m^{-2})	11.87	11.92	8.86
Pulse repetition frequency (kHz)	100	118	134
Beam divergence (mrad)	0.25	0.22	0.25
Footprint diameter (cm)	50	17	50

* Note: above ground level

first of many, intermediate, and last of many. The heights of all echoes relative to the ground surface were calculated by subtracting their corresponding ground heights; thus, all echo heights were normalized. In addition, we filtered LiDAR echoes labeled with scan angles >15° to ensure the viewing angle was compatible with DCP. Lastly, various LiDAR variables were calculated at plot level using a radius of 20 m, following the area-based approach [48]. The cut-off height was set at 1.3 m, i.e., the same height at which DHP and DCP were taken.

LiDAR metrics used in this study included echo height percentiles, echo density percentiles ("bincentiles"), multiple penetration indices, and canopy variables from polar transformed LiDAR coordinates. Echo heights (p_*) and densities (b_*) were calculated using all echoes at 5% increments ($0\%, 5\%, \ldots, 95\%, and 100\%$). For example, p_5 means the echo height observed at the fifth percentile. In addition, means and standard deviations of echo heights were also calculated.

For penetration indices, we computed the all echo penetration index [API, (8)], first echo penetration index [FPI, (9)], last echo penetration index [LPI, (10)], and Solberg's penetration index [SPI, (11)] [21], [34], using 1.3 m elevation as the cut-off height to separate vegetation and ground echoes

$$\mathrm{API} = 1 - \frac{\sum \mathrm{All}_v}{\sum \mathrm{All}} \tag{8}$$

$$\mathrm{FPI} = 1 - \frac{\sum \mathrm{Single}_v + \sum \mathrm{First}_v}{\sum \mathrm{Single} + \sum \mathrm{First}} \tag{9}$$

$$\mathrm{LPI} = 1 - \frac{\sum \mathrm{Single}_v + \sum \mathrm{Last}_v}{\sum \mathrm{Single} + \sum \mathrm{Last}} \tag{10}$$

$$\mathrm{SPI} = \frac{\sum \mathrm{Single}_g + 0.5(\sum \mathrm{First}_g + \sum \mathrm{Last}_g)}{\sum \mathrm{Single} + 0.5(\sum \mathrm{First} + \sum \mathrm{Last})} \tag{11}$$

where All, Single, First, and Last denote echo types, and their subscripts indicate whether the echo hits vegetation (v) or ground (g).

In addition, we included another echo-weighted penetration index (EWI) which was directly derived from the echo numbers as follows [49]:

$$\mathrm{EWI} = \frac{N_{\mathrm{ground}}}{N_{\mathrm{ground}} + N_{\mathrm{vegetation}}} \tag{12}$$

where a weight was added for each echo as $(1/i)$ and i was the number of echoes of the given pulse. Hence, $N_{\mathrm{ground}} = g_1 + (1/2)g_2 + (1/3)g_3 + \cdots + (1/n)g_n$ and $V_{\mathrm{vegetation}} = v_1 + (1/2)v_2 + (1/3)v_3 + \cdots + (1/n)v_n$.

Furthermore, we tested the efficiency of canopy variables obtained from polar transformed echo coordinates at the plot level. Specifically, we converted Cartesian coordinates (X, Y, Z) of all echo types at plot level into azimuth (ϕ) and zenith (θ) polar angles after shifting original Cartesian coordinates to comply with the sampling scheme shown in Fig. 2. Instead of converting only the echoes inside the plots, we used an enlarged plot radius of 40 m to derive polar metrics as an attempt to represent the larger plot information captured by DHP due to the large field of view (FOV ≥ 180°) of the fisheye lens. After this conversion, it was possible to construct DHP-like images from ALS data [Fig. 3(a)]. Namely, all echoes were binned into a 2-D systematic grid defined by azimuth and zenith angles (ϕ, θ), and then the grid was rasterized to an image of 480 × 480 pixels that covered the entire hemisphere. We did several tests on the resolution of the raster image and chose this resolution as a compromise of spatial details and having sufficient echoes per pixel. The value of each pixel was initialized as the number of echoes (n) assigned to it. To calculate the fractional gap at the pixel level, the maximum number of echoes within the 0°–75° ($n\mathrm{max}$) was determined first. The fractional cover was calculated as $n/(n\mathrm{max}/2)$, with values > 1 truncated to 1; thus, the fractional gap was calculated as $1-n/(n\mathrm{max}/2)$. Hereafter this rasterization is referred to as grayscale polar image [Fig. 3(b)]. Furthermore, the greyscale image was binarized into canopy gaps (0) and vegetation (1), hereafter referred to as binarized polar image [Fig. 3(c)]. Finally, the binarized polar image was processed with morphological image processing operations using a structuring element size of 7 in the same way as DHP was processed [Fig. 3(d)]. Thus, we obtained estimates of angular gap fractions at five rings for each reference point, and their values were averaged to obtain the plot-level metrics. We used the symbols binarized-gaps* and greyscale-gaps* to denote gap fractions at 1–5 rings obtained from binarized polar images and greyscale polar images, as well as morphological-gaps* to denote between-crown gaps obtained from morphologically processed polar images respectively. For example, binarized-gaps1 denotes the gap fraction obtained at the first ring using the binarized polar images. Together, binarized-gaps*, greyscale-gaps* and morphological-gaps* are hereafter referred to as "polar metrics."

D. Model Construction and Validation

We first directly compared the LiDAR penetration indices with DCP-derived near-vertical GF(θ) without using any models. The comparison was based on computing the root mean square error (RMSE) and bias in the following equations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{13}$$

$$\mathrm{Bias} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{14}$$

where $\hat{y}_i$ is the value of the given penetration index, y_i is the observed value from ground DCP measurements, and n is the number of plots.

Next, we predicted LAI_e using the selected LiDAR penetration indices and the semi-physical model form (1). The

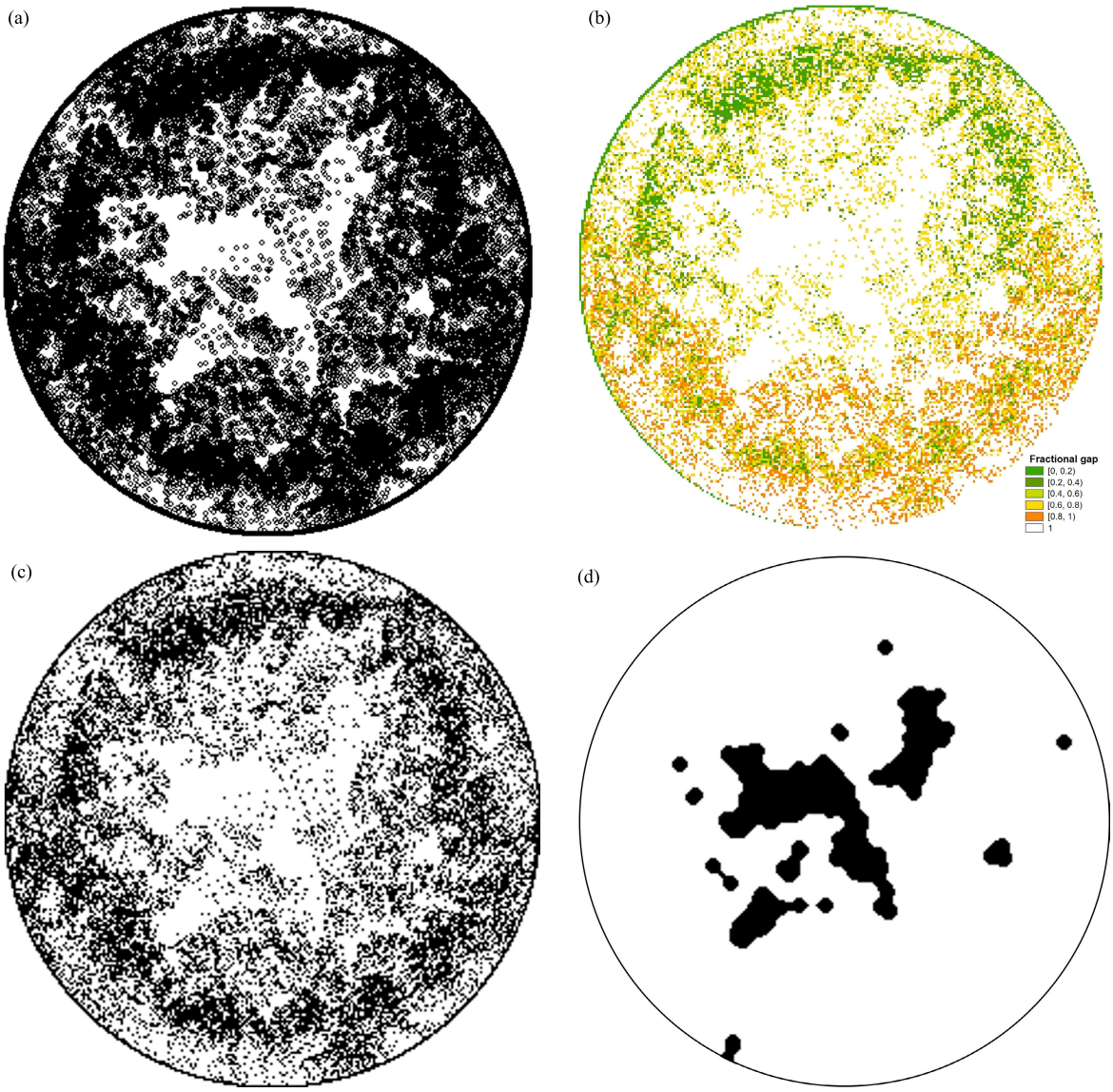

Fig. 3. (a) DHP-like image constructed using polar transformed ALS coordinates at one observation spot. (b) Greyscale polar image displaying fractional gaps. (c) Binarized polar image. (d) Between-crown canopy gaps after morphological operations (large gaps are shown in black).

coefficient β, when it is assumed as a leaf orientation correction parameter, should equal two when the entire hemisphere is considered (3) or the leaf angle distribution is random [42]. The latter is not necessarily the case in boreal forests, so β must be estimated empirically by using field measured LAI_e and regression analysis.

Furthermore, we constructed empirical models for LAI_e, LAI, and Ω_E using ordinary least squares (OLSs) with up to three LiDAR-based predictors. The predictors were selected using an exhaustive search of the whole LiDAR-based variable pool, including penetration indices and their log transformations, echo height percentiles, and echo density percentiles using all echoes, and polar metrics. Adding variables into the multivariate model not only increased the model fit but also the risk of overfitting. Therefore, we applied the following rules in selecting the predictors for the empirical models.

1) All predictors had to be statistically significant, $p < 0.05$.
2) We used models with up to three predictors with variance inflation factor (VIF) < 5 in most cases.
3) In the case of models having the same RMSE, we chose the one with the lower Akaike information criterion (AIC).

The models were validated following the leave one out cross validation (LOOCV) approach and the three sites were modeled separately. In the result section, we report the relative RMSE [RMSE%, (15)] as well as R^2 and mean absolute error [MAE, (16)] values to measure the model accuracy

$$\mathrm{RMSE\%} = \frac{\mathrm{RMSE} \cdot 100\%}{\bar{y}_i} \tag{15}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{16}$$

where y_i is the observed value from ground measurements, $\hat{y}_i$ is the value of the predicted counterpart, $\bar{y}_i$ is the mean observed value of all plots and n is the number of plots used to construct the model.

TABLE III

COMPARISON OF DCP-MEASURED NEAR-VERTICAL GAP FRACTION WITH LIDAR PENETRATION INDICES

LiDAR index	Heinola		Hyytiälä		Outokumpu	
	RMSE	Bias	RMSE	Bias	RMSE	Bias
FPI	0.18	0.15	0.19	0.18	0.12	0.10
SPI	**0.08**	**-0.02**	0.10	0.08	0.08	-0.07
LPI	0.23	-0.18	**0.08**	**-0.02**	0.26	-0.25
API	0.10	0.07	0.10	0.08	**0.05**	**-0.02**
EWI	0.13	0.11	0.14	0.13	0.08	0.07

Note: bold type indicates better performance.

III. RESULTS

A. Direct Comparison of DCP-Derived GF(θ) With LiDAR Data

We first compared the GF(θ) obtained by DCP with different LiDAR penetration indices to find a stable LiDAR penetration index T as the input to the semi-physical model shape (Table III). Overall, LiDAR penetration indices could represent field measured GF(θ) with varying performances in the different sites. The FPI and EWI underestimated GF(θ) with a positive bias, while the LPI overestimated GF(θ) at all three sites. As expected, the FPI leads to underestimation as it is not sensitive to detect small gaps in the canopy. On the contrary, the LPI, which is based on last echoes and single echoes, tends to overestimate GF(θ). SPI and API generally had smaller biases but were inconsistent across the sites. Both indices overestimated GF(θ) at the Outokumpu site and underestimated at the Hyytiälä site. Yet, SPI overestimated at the Heinola site while API showed the opposite.

Overall, FPI and LPI yielded relatively larger RMSE in all three sites. The LPI gave the largest RMSE in Heinola (0.23) and Outokumpu (0.26) sites, yet the RMSE reduced considerably in Hyytiälä (0.08), making it the least biased index at Hyytiälä site. One possible reason could be the LiDAR sensor that registered more last echoes and thus provided a smaller LPI at this site. The EWI outperformed FPI and LPI, albeit not reaching the level of precision achieved by SPI and API. The SPI, which added weight on the first and last echoes, had the smallest bias in Heinola. The API provided the smallest bias in the Outokumpu site.

B. Effective LAI

1) Semi-Physical Model: Although all LiDAR indices can be used as T in the semi-physical model to predict LAI_e, here we only report the performances of the API and SPI indices. These indices had less bias and a more stable correlation with the GF(θ) (Table III), and also produced a better fit with DHP-based LAI_e across the different sites than the other indices.

Both indices that predicted LAI_e as a linear function of the negative logarithm had an almost 1:1 relationship with the field measured LAI_e obtained by DHP (Fig. 4). As expected, the estimated coefficient β took a value around two in both cases, which varied due to different LiDAR sensor properties, flight parameters, or foliage angle distributions by site. It ranged from 2.08 to 2.63 when modeling with the API index and from 2.01 to 2.87 with the SPI index. The API (RMSE% = 9.0%–18.2%) yielded more accurate results than the SPI (RMSE% = 17.5%–27.0%) in all datasets. The best fit was achieved in the Outokumpu site (RMSE% = 9.0%, MAE = 0.16, R^2 = 0.97) with the coefficient value β reaching 2.63, indicating an erectophile foliage angle distribution by the lower contact frequency in the zenith direction than around the horizon. In Heinola, the accuracy was slightly lower (RMSE = 16.9%, MAE = 0.23, R^2 = 0.92). In Hyytiälä, the accuracy was even lower (RMSE% = 18.2%, MAE = 0.31, R^2 = 0.75), which was particularly influenced by one plot having a large LAI_e value. In the overestimated plot (LAI_e difference: 1.44), the sun still illuminated the tree crowns, which apparently led to the overestimation of gap fractions and consequently the underestimation of LAI_e.

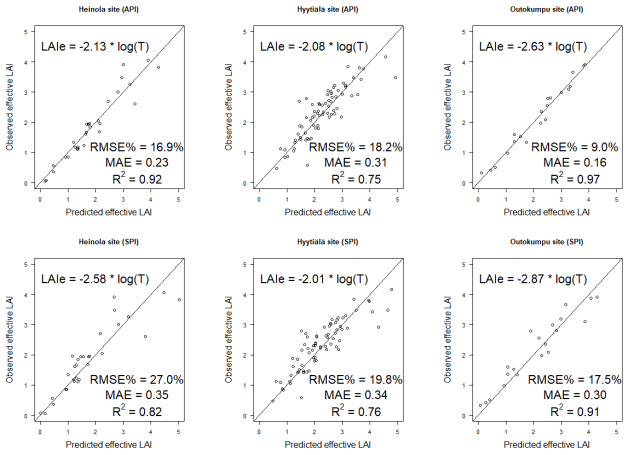

Fig. 4. Scatterplot of the field measured LAI_e and LiDAR predicted LAI_e after LOOCV in three study sites. T refers to the (top row) API and (bottom row) SPI indices.

TABLE IV

SELECTED LIDAR-BASED METRICS USING OLSS

Area	Response variable	Predictor variable(s)	AIC	VIF
Heinola	LAI_e	log(API)	16.6	N.A.
		greyscale-gaps5, log(API)	14.9	3.9, 3.9
		p_5, log(FPI), log(API)	8.6	1.2, 9.3, 9.8
	sqrt(LAI)[a]	log(FPI), log(API)	-3.4	9.0, 9.0
	Ω_E	FPI, greyscale-gaps5	-93.4	2.5, 2.5
Hyytiälä	LAI_e	SPI	72.8	N.A.
		log(LPI), morphological-gaps5	45.8	1.2, 1.2
		b_60, morphological-gaps4, log(LPI)	32.4	1.4, 1.8, 1.4
	log(LAI)[b]	LPI, morphological-gaps5	-18.4	1.4, 1.4
	Ω_E	FPI, morphological-gaps5	-290.2	1.6, 1.6
Outokumpu	LAI_e	log(API)	-4.3	N.A.
		b_95, log(API)	-4.0	1.1, 1.1
		p_10, b_95, log(API)	-7.5	1.8, 1.6, 1.2
	LAI	log(SPI), log(API)	50.3	23.4, 23.4
	Ω_E	FPI, binarized-gaps5	-50.6	2.6, 2.6

Note: [a] square root transformation and [b] log transformation was applied to the response variable.

2) Empirical Models With 1–3 Predictors: We predicted LAI_e using empirical regression models with up to three LiDAR-based predictors. The results showed that a variety of LiDAR-based predictors could be used in empirical models to estimate LAI_e (Table IV). LiDAR penetration indices, especially API, were often selected as model predictors. Predictors derived from polar transformed point clouds also appeared at all sites, highlighting their potential importance in empirical models. Predictors based on LiDAR height and density percentiles had relatively a lower rate of selection.

The results also showed that the model accuracy could be improved by feeding additional predictors, which was

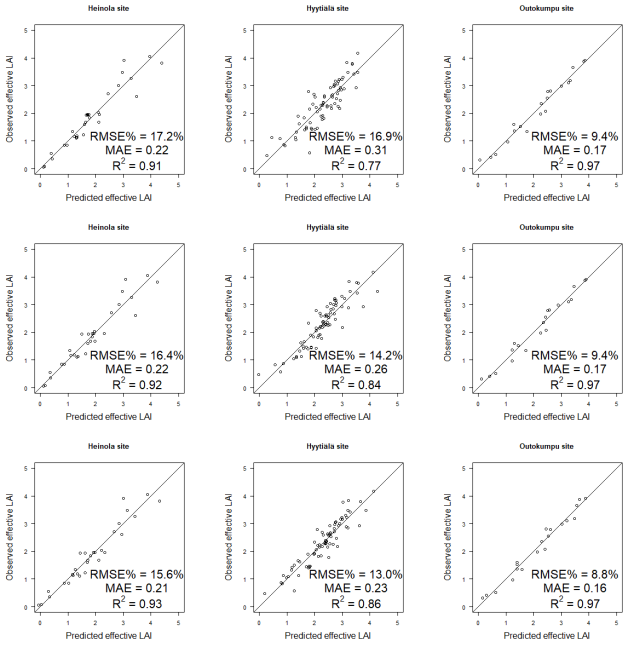

Fig. 5. Scatter plots of field measured LAI_e and LiDAR predicted LAI_e using empirical models of (top row) one variable, (middle row) two variables, and (bottom row) three variables.

indicated by decreased RMSE% and MAE as well as increased R^2 for all study sites (Fig. 5). The model fit was improved in Hyytiälä (RMSE% = 16.9%, 14.2% and 13.0%) when the respective models had one, two and three predictors, and the model improvements were relatively smaller in Heinola (RMSE% = 17.2%, 16.4% and 15.6%) and Outokumpu (RMSE% = 9.4%, 9.4% and 8.8%). However, concerns arose when having additional model predictors, as every additional predictor not only decreased the model's RMSE%, but also brought the risk of overfitting. In our case, models having more than three predictors always accompanied by high VIF that indicated strong multicollinearity. Therefore, it seemed that multivariate models with $\leq$3 predictors remained a safe choice.

C. Canopy Clumping Ω_E and LAI With Empirical Models

As previous results suggested that the risk of overfitting should be concerned with multivariate models, we used only two predictors to estimate LAI and Ω_E to avoid such issues. Fig. 6 llustrates the comparison of LiDAR predicted LAI and Ω_E with their field measured counterparts obtained by DHP. Regarding LAI, the models' RMSE% and MAEs were similar (RMSE% = 18.7%–19.7%, MAE = 0.14–0.64) in three sites while the R^2 was slightly lower in Hyytiälä (R^2 = 0.60) compared with Heinola (R^2 = 0.79) and Outokumpu (R^2 = 0.78). Similar results were observed for Ω_E across the sites. In addition, the accuracies of LAI models were considerably lower when compared with the LAI_e models of 1–3 predictors (RMSE% = 8.8%–17.2%, MAE = 0.16–0.31). Nevertheless, the results showed that LiDAR-based metrics were successful in modeling LAI and Ω_E.

We observed that LiDAR penetration indices as well as polar metrics were often selected as empirical model predictors. Notably, the empirical models always took a combination

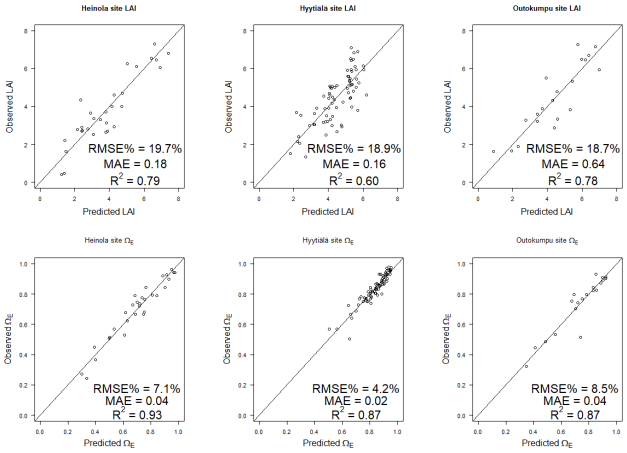

Fig. 6. Scatterplot of DHP measured LAI and canopy element clumping index (Ω_E) with their LiDAR predicted counterparts after LOOCV. Selected predictors can be found in Table IV.

of penetration indices and polar metrics as predictors in the estimation of Ω_E. These results again highlighted the potential of using polar metrics to describe forest canopy properties (e.g., LAI and Ω_E) following an empirical modeling approach. It may be that polar metrics from the fifth ring (60°–75°) provided particularly useful information given their selected frequency (Table IV). It is possible that polar metrics, derived by using an extended radius of 40 m at the plot level, provided useful information to the models as they also captured canopy information located outside the plot boundary. This information was captured by DHP due to its large FOV but not by the LiDAR penetration indices that were calculated using the radius of 20 m. It seemed that including both canopy penetration indices and polar metrics in empirical models provided the most reliable estimates for both LAI and Ω_E. On the other hand, the commonly used ALS variables related to height and density did not appear in these models.

D. Comparison of Model Accuracies for the Various Response Variables

Table V displays a comparison of the model accuracies at the various sites based on the RMSE%. The direct comparison between DCP-derived GF(θ) with the API index had the RMSE% ranging from 11.1% to 29.5%. The Ω_E was predicted most accurately with RMSE% ranging from 4.2% to 8.5%, and the clumping corrected LAI was by far the most difficult to predict accurately (RMSE% = 18.7%–19.7%). The differences in RMSE% between the semi-physical and empirical models for LAI_e were small at Heinola and Outokumpu, but the 5.0 percent decrease in Hyytiälä suggested that empirical models may have added value in sites with a large variety of forest structures.

IV. Discussion

A. Comparison of the Two Modeling Approaches in the Estimation of LAI_e

Both semi-physical models and empirical models showed satisfactory performance in estimating LAI_e when calibrated against field data. Overall, the best result of predicting LAI_e

TABLE V
SUMMARY OF THE MODEL ACCURACIES BASED ON RMSE% AT THE THREE STUDY SITES

Response variables	Heinola	Hyytiälä	Outokumpu
GF(θ), directly from API	21.5	29.5	11.1
LAI_e, semi-physical	16.9	18.2	9.0
LAI_e, empirical	15.6	13.0	8.8
LAI, empirical	19.7	18.9	18.7
Ω_E, empirical	7.1	4.2	8.5

with the semi-physical model shape was achieved with the API index; and with empirical models, the best results were achieved when three LiDAR-derived predictors were included. The empirical models were slightly more accurate than the semi-physical models, as their RMSE% were on average 2% smaller. Although the RMSE% values were overall lower across sites when using the empirical multivariate models with 2–3 predictors than the semi-physical models, empirical multivariate models may be less robust in real-world prediction scenarios than the semi-physical models, whose form is simple and built upon a solid theoretical basis.

Solberg et al. [21] suggested that the semi-physical modeling approach is only valid on the condition that a strong relationship exists between the LiDAR penetration index and the gap fractions GF(θ). In our case, both the SPI and API indices were found to be highly correlated with GF(θ) measured by DCP. The API index was the least biased at the Outokumpu site (Bias = -0.02) and also achieved the best accuracy (RMSE% = 9.0%). However, the opposite was observed with the SPI index, as it appeared almost unbiased at the Heinola site (Bias = -0.02) but resulted in the weakest model fit (RMSE% = 27.0%). Therefore, the least biased LiDAR penetration index does not necessarily result in the best performance when included as a predictor in the semi-physical model shape.

In direct comparison with DCP-measured gap fractions, the performances of different penetration indices depended on the LiDAR scan and acquisition settings, such as scan angle and footprint size. The impact of ALS pulse density on the computation of penetration indices should be small because our data had a sufficient density to form a reliable height distribution for each plot [50]. We normalized the variations in the ALS scan angle by setting it identical to the view angle of DCP (15° off zenith). The effects of footprint size were difficult to assess as they also depended on pulse power that was unknown for all data sets. However, large footprint sizes typically result in more echoes from both the canopy and the ground. In addition, the Leica scanner had a 17 cm footprint size could digitize fewer echoes per pulse (four) than the Riegl scanners with 50 cm footprints (six). Regardless of these differences, Fig. 4 showed that API always had less scatter than SPI and consequently it constantly outperformed SPI in all sites. Thus, API appears to be a robust proxy of GF(θ), even though it may be more prone to bias from sensor effects than other indices, as it is computed using all echo types. This is in line with the findings of [51], who also stated that the indices computed using all echoes produced better results.

Furthermore, the β retrievals were expected to take a value around two, given that the foliage angle distribution is spherical and that the used LiDAR penetration index equals the GF(θ). Previous studies showcased that in boreal forests the β estimates based on the SPI index ranged from 2.3 to 2.7 [21], [32], [34], which was a rather narrow spread and raised hopes that the use of SPI index with $\beta \approx 2.5$ might provide reasonably accurate estimates of boreal forest LAI_e even if field calibration data were not available. In our case, the β values based on the SPI index were 2.58 and 2.87 when the GF(θ) were overestimated (Heinola and Outokumpu, respectively) and 2.01 when the GF(θ) was underestimated (Hyytiälä). Therefore, sensor effects had a considerable effect on the SPI-based β values, and it does not seem feasible to assign a specific value to the β when estimating LAI_e based on the SPI index in south-eastern Finland. With the API index, the β values were closer to each other: 2.13, 2.08, and 2.63, which may indicate that it is more robust than the SPI index. However, the resultant β is still sensor-dependent, and further research is needed to determine if transferable semi-physical LAI_e models are feasible based on the API index.

Although estimating LAI_e using empirical models with three predictors yielded better accuracy than the use of semi-physical models, empirical models are more complex and thus more prone to errors in prediction and less transferable. Transferability is a critical issue in LAI estimation, because in situ LAI data are rarely available for large-area mapping. In general, model transferability can be poor when an existing model is applied with different LiDAR systems and forest structures [52]. For a limited area such as south-eastern Finland, it could still be possible to construct large-area models for LAI_e or LAI by combining data from multiple LiDAR projects, which has already been done for forest attributes such as biomass [53]. Model calibration with small and easily obtainable sets of field reference data may also be a feasible solution [54].

B. Utility of Polar Metrics in Empirical Models

We showcased that polar transformed ALS point clouds can be used to construct DHP-like images, which can be processed in a similar manner as real DHPs and consequently produce polar metrics. The results showed that polar metrics were frequently selected in the empirical models, especially in the estimation of Ω_E (Table IV). Currently, canopy clumping is commonly mapped from the Moderate Resolution Imaging Spectroradiometer (MODIS) at 500 m resolution [55]; however, mapping clumping with a finer resolution is highly desired in the field of remote sensing. Our results showed that using the combination of penetration indices and polar metrics in empirical models yielded satisfactory model fits (RMSE% = 4.3%–8.5%, $R^2 = 0.86$–0.93), which may provide an insight into constructing finer Ω_E maps using LAI field data.

The use of polar metrics also has multiple advantages. First, polar metrics, which are derived from DHP-like images constructed using ALS echoes, provided useful information to support LAI_e, LAI, and Ω_E empirical models. The main difference in polar metrics is that each image is specific to

a given sample point, whereas commonly used height and density percentiles apply to a fixed area. To account for the local variation between sampling spots, we constructed five polar transformed images per plot following the same sampling design as DHP and used their averaged values as inputs of multivariate models. Vaughn et al. [39] suggested that polar grids consisting of azimuth and zenith intervals could produce stable LAI estimates. In the current study, the polar transformed variables were pixel-based instead of angle based, because the application of morphological image processing operations would have been difficult if the image elements were defined by zenith and azimuth angles. Our approach, however, enabled the ring-wise analysis based on the five annulus rings, which is similar to standard DHP image processing.

Alexander et al. [38] investigated the influence of different radii on the computation of ALS polar metrics. Their study revealed that employing a 50-m plot radius yielded the best correlation between angular canopy cover and understory light condition. Considering that trees in Finland are on average smaller than their study site in Denmark, we used the plot radius of 40 m to derive polar metrics. These metrics were also proven effective in empirical models. In some rare cases, such as seed tree stands characterized by tall trees but a small stem density, this decision may potentially introduce a bias, as visible trees further than 40 m may be excluded from consideration. Nevertheless, in the majority of forest environments, this bias is insignificant, as trees positioned at greater distances are already obscured from the view by trees closer to the viewing point. Using the plot radius of 40 m also provides a computational challenge of deriving high-resolution LAI and Ω_E maps with for example 20 m spatial resolution because the predictor values for each pixel should also include ALS echoes outside of the pixel. A possible workaround is to compute multiple low-resolution LAI and Ω_E maps with slightly different pixel locations and merge them to form a high-resolution raster of polar metrics.

As polar metrics are, in essence, derived from ALS point cloud data, they are prone to LAI saturation and might not work as stand-alone predictors. With empirical regression models, this means that the predicted LAI values do not increase after the predictors exceed certain values, usually at the upper extents [56]. However, commonly used penetration indices can always be included as auxiliary predictors, as we did in the current study. For example, [57] managed to delay the saturation effect to some extent with a combination of different types of LiDAR metrics. However, LAI saturation remains to be a complex issue as it depends on many conditions, such as site type and LiDAR systems, which also links back to the issue of model transferability.

V. Conclusion

We conclude that ALS data can provide many kinds of metrics that are suitable for modeling LAI_e, Ω_E, and LAI with ALS data when field data are available for model calibration. Both empirical and semi-physical modeling approaches are effective in predicting LAI_e over a variety of forest conditions at multiple geographical locations in Finland. With the semi-physical modeling approach, the input LiDAR penetration index is expected to be reasonably unbiased against the vertical gap fraction. However, the most unbiased penetration index did not necessarily produce the best model fit in our case. Both the SPI and API indices were suitable candidates to be included; yet, the API index provided the strongest correlation with field measured LAI_e. The API index therefore offers the most potential for model transferability. When following the empirical modeling approach, the models for clumping-corrected LAI were considerably less accurate than the models for LAI_e. In addition, we demonstrated that LiDAR echoes can produce polar metrics that can facilitate the estimations of LAI_e, LAI and Ω_E. These polar metrics can provide empirical models with additional information on canopy structure, and they were frequently selected as model predictors. Especially the canopy clumping coefficient Ω_E was modeled with great accuracy using polar metrics and canopy penetration indices.

Acknowledgment

The authors would like to express their sincere gratitude to the four anonymous reviewers who provided constructive comments on improving the quality of this article.

References

[1] J. M. Chen and T. A. Black, "Defining leaf area index for non-flat leaves," *Plant, Cell Environ.*, vol. 15, no. 4, pp. 421–429, May 1992.

[2] F. Baret et al., *Global Leaf Area Index Product Validation Good Practices*. Cambridge, MA, USA: Academia, 2014.

[3] Y. Ryu et al., "Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales," *Global Biogeochem. Cycles*, vol. 25, no. 4, 2011.

[4] A. K. Skidmore et al., "Priority list of biodiversity metrics to observe from space," *Nature Ecol. Evol.*, vol. 5, no. 7, pp. 896–906, May 2021.

[5] S. Härkönen et al., "Predicting forest growth based on airborne light detection and ranging data, climate data, and a simplified process-based model," *Can. J. Forest Res.*, vol. 43, no. 999, pp. 364–375, 2013.

[6] T. Manninen et al., "Airborne measurements of surface albedo and leaf area index of snow-covered boreal forest," *J. Geophys. Res., Atmos.*, vol. 127, no. 1, 2022, Art. no. e2021JD035376.

[7] K. H. Hanssen and S. Solberg, "Assessment of defoliation during a pine sawfly outbreak: Calibration of airborne laser scanning data with hemispherical photography," *Forest Ecol. Manage.*, vol. 250, nos. 1–2, pp. 9–16, Oct. 2007.

[8] H. Fang, "Canopy clumping index (CI): A review of methods, characteristics, and applications," *Agricult. Forest Meteorol.*, vol. 303, Jun. 2021, Art. no. 108374.

[9] W. Woodgate et al., "Validating canopy clumping retrieval methods using hemispherical photography in a simulated eucalypt forest," *Agricult. Forest Meteorol.*, vol. 247, pp. 181–193, Dec. 2017.

[10] J. M. Chen and J. Cihlar, "Quantifying the effect of canopy architecture on optical measurements of leaf area index using two gap size analysis methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 777–787, May 1995.

[11] A. Lang and X. Yueqin, "Estimation of leaf area index from transmission of direct sunlight in discontinuous canopies," *Agricult. Forest Meteorol.*, vol. 37, no. 3, pp. 229–243, 1986.

[12] H. Jiang et al., "Clumping effects in leaf area index retrieval from large-footprint full-waveform LiDAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4406220.

[13] P. Stenberg, "Correcting LAI-2000 estimates for the clumping of needles in shoots of conifers," *Agricult. Forest Meteorol.*, vol. 79, nos. 1–2, pp. 1–8, Mar. 1996.

[14] M. Weiss, F. Baret, G. Smith, I. Jonckheere, and P. Coppin, "Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling," *Agricult. Forest Meteorol.*, vol. 121, nos. 1–2, pp. 37–53, 2004.

[15] L. He et al., "Inter- and intra-annual variations of clumping index derived from the MODIS BRDF product," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 44, pp. 53–60, Feb. 2016.

[16] J. Pisek et al., "Intercomparison of clumping index estimates from POLDER, MODIS, and MISR satellite data over reference sites," *ISPRS J. Photogramm. Remote Sens.*, vol. 101, pp. 47–56, Mar. 2015.

[17] M. Maltamo, P. Packalen, and A. Kangas, "From comprehensive field inventories to remotely sensed wall-to-wall stand attribute data—A brief history of management inventories in the Nordic countries," *Can. J. Forest Res.*, vol. 51, no. 2, pp. 257–266, 2021.

[18] E. Næsset, "Area-based inventory in Norway–from innovation to an operational reality," in *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*. Dordrecht, The Netherlands: Springer, 2013, pp. 215–240. [Online]. Available: https://link.springer.com/chapter/10.1007/978-94-017-8663-8_11

[19] P. Tompalski, J. C. White, N. C. Coops, and M. A. Wulder, "Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data," *Remote Sens. Environ.*, vol. 227, pp. 110–124, Jun. 2019.

[20] F. Morsdorf, B. Kötz, E. Meier, K. I. Itten, and B. Allgöwer, "Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction," *Remote Sens. Environ.*, vol. 104, no. 1, pp. 50–61, Sep. 2006.

[21] S. Solberg et al., "Mapping LAI in a Norway spruce forest using airborne laser scanning," *Remote Sens. Environ.*, vol. 113, no. 11, pp. 2317–2327, Nov. 2009.

[22] J. Armston et al., "Direct retrieval of canopy gap probability using airborne waveform LiDAR," *Remote Sens. Environ.*, vol. 134, pp. 24–38, Jul. 2013.

[23] A. Cescatti, "Indirect estimates of canopy gap fraction based on the linear conversion of hemispherical photographs: Methodology and comparison with standard thresholding techniques," *Agricult. Forest Meteorol.*, vol. 143, nos. 1–2, pp. 1–12, 2007.

[24] G. M. Díaz and J. D. Lencinas, "Model-based local thresholding for canopy hemispherical photography," *Can. J. Forest Res.*, vol. 48, no. 10, pp. 1204–1216, Oct. 2018.

[25] M. Lang, A. Kuusk, M. Mõttus, M. Rautiainen, and T. Nilson, "Canopy gap fraction estimation from digital hemispherical images using sky radiance models and a linear conversion method," *Agricult. Forest Meteorol.*, vol. 150, no. 1, pp. 20–29, Jan. 2010.

[26] C. Macfarlane et al., "Estimation of leaf area index in eucalypt forest using digital photography," *Agricult. Forest Meteorol.*, vol. 143, nos. 3–4, pp. 176–188, 2007.

[27] A. Kuusk, "Specular reflection in the signal of LAI-2000 plant canopy analyzer," *Agricult. Forest Meteorol.*, vol. 221, pp. 242–247, May 2016.

[28] Y. Ryu et al., "On the correct estimation of effective leaf area index: Does it reveal information on clumping effects?" *Agricult. Forest Meteorol.*, vol. 150, no. 3, pp. 463–472, Mar. 2010.

[29] H. Fang, F. Baret, S. Plummer, and G. Schaepman-Strub, "An overview of global leaf area index (LAI): Methods, products, validation, and applications," *Rev. Geophys.*, vol. 57, no. 3, pp. 739–799, Sep. 2019.

[30] G. Yan et al., "Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives," *Agricult. Forest Meteorol.*, vol. 265, pp. 390–411, Feb. 2019.

[31] J. Jensen, K. Humes, L. Vierling, and A. Hudak, "Discrete return LiDAR-based prediction of leaf area index in two conifer forests," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3947–3957, Oct. 2008.

[32] S. Solberg, "Mapping gap fraction, LAI and defoliation using various ALS penetration variables," *Int. J. Remote Sens.*, vol. 31, no. 5, pp. 1227–1244, Mar. 2010.

[33] I. Korpela, H. O. Ørka, J. Hyyppä, V. Heikkinen, and T. Tokola, "Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 4, pp. 369–379, Jul. 2010.

[34] L. Korhonen, I. Korpela, J. Heiskanen, and M. Maltamo, "Airborne discrete-return LiDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index," *Remote Sens. Environ.*, vol. 115, no. 4, pp. 1065–1080, Apr. 2011.

[35] J. J. Richardson, L. M. Moskal, and S.-H. Kim, "Modeling approaches to estimate effective leaf area index from aerial discrete-return LiDAR," *Agricult. Forest Meteorol.*, vol. 149, nos. 6–7, pp. 1152–1160, Jun. 2009.

[36] E. Næsset, "Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 148–159, Jan. 2009.

[37] D. N. Cosenza et al., "Effects of numbers of observations and predictors for various model types on the performance of forest inventory with airborne laser scanning," *Can. J. Forest Res.*, vol. 52, no. 3, pp. 385–395, Mar. 2022.

[38] C. Alexander, J. E. Moeslund, P. K. Bøcher, L. Arge, and J.-C. Svenning, "Airborne laser scanner (LiDAR) proxies for understory light conditions," *Remote Sens. Environ.*, vol. 134, pp. 152–161, Jul. 2013.

[39] N. R. Vaughn, G. P. Asner, and C. P. Giardina, "Polar grid fraction as an estimator of montane tropical forest canopy structure using airborne LiDAR," *Int. J. Remote Sens.*, vol. 34, no. 21, pp. 7464–7473, Nov. 2013.

[40] M. Lang et al., "Digital photography for tracking the phenology of an evergreen conifer stand," *Agricult. Forest Meteorol.*, vol. 246, pp. 15–21, Nov. 2017.

[41] S. Zhang et al., "The effect of digital hemispherical photograph binarization methods on the estimation of plant area index with LiDAR data," in *Proc. ForestSAT Conf.*, Berlin, Germany, 2022.

[42] J. Miller, "A formula for average foliage density," *Austral. J. Botany*, vol. 15, no. 1, pp. 141–144, 1967.

[43] L. Korhonen and J. Heikkinen, "Automated analysis of in situ canopy images for the estimation of forest canopy cover," *Forest Sci.*, vol. 55, no. 4, pp. 323–334, 2009.

[44] S. G. Leblanc, J. M. Chen, R. Fernandes, D. W. Deering, and A. Conley, "Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests," *Agricult. Forest Meteorol.*, vol. 129, nos. 3–4, pp. 187–207, Apr. 2005.

[45] F. Chianucci, J. Zou, P. Leng, Y. Zhuang, and C. Ferrara, "A new method to estimate clumping index integrating gap fraction averaging with the analysis of gap size distribution," *Can. J. Forest Res.*, vol. 49, no. 5, pp. 471–479, May 2019.

[46] P. Stenberg, T. Nilson, H. Smolander, and P. Voipio, "Gap fraction based estimation of LAI in scots pine stands subjected to experimental removal of branches and stems," *Can. J. Remote Sens.*, vol. 29, no. 3, pp. 363–370, Jan. 2003.

[47] M. Nobis and U. Hunziker, "Automatic thresholding for hemispherical canopy-photographs based on edge detection," *Agricult. Forest Meteorol.*, vol. 128, nos. 3–4, pp. 243–250, Feb. 2005.

[48] E. Næsset, "Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 88–99, Apr. 2002.

[49] Y. Qu et al., "Direct estimation of forest leaf area index based on spectrally corrected airborne LiDAR pulse penetration ratio," *Remote Sens.*, vol. 12, no. 2, p. 217, Jan. 2020.

[50] T. Gobakken and E. Næsset, "Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data," *Can. J. Forest Res.*, vol. 38, no. 5, pp. 1095–1109, May 2008.

[51] M. Sumnall et al., "Assessing the transferability of statistical predictive models for leaf area index between two airborne discrete return LiDAR sensor designs within multiple intensely managed loblolly pine forest locations in the south-eastern USA," *Remote Sens. Environ.*, vol. 176, pp. 308–319, Apr. 2016.

[52] L. Tian, Y. Qu, and J. Qi, "Estimation of forest LAI using discrete airborne LiDAR: A review," *Remote Sens.*, vol. 13, no. 12, p. 2408, Jun. 2021.

[53] E. Kotivuori, L. Korhonen, and P. Packalen, "Nationwide airborne laser scanning based models for volume, biomass and dominant height in Finland," *Silva Fennica*, vol. 50, no. 4, pp. 1–28, 2016.

[54] E. Kotivuori, M. Maltamo, L. Korhonen, and P. Packalen, "Calibration of nationwide airborne laser scanning based stem volume models," *Remote Sens. Environ.*, vol. 210, pp. 179–192, Jun. 2018.

[55] L. He, J. M. Chen, J. Pisek, C. B. Schaaf, and A. H. Strahler, "Global clumping index map derived from the MODIS BRDF product," *Remote Sens. Environ.*, vol. 119, pp. 118–130, Apr. 2012.

[56] K. Zhao and S. Popescu, "LiDAR-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the Southern USA," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1628–1645, Aug. 2009.

[57] S. Luo et al., "Comparative performances of airborne LiDAR height and intensity data for leaf area index estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 300–310, Jan. 2018.

Shaohui Zhang (Graduate Student Member, IEEE) received the M.Sc. degree in environment and development from the University of Edinburgh, Edinburgh, U.K., in 2019, and the M.Sc. degree in European forestry from the University of Eastern Finland (UEF), Joensuu, Finland, and AgroParisTech, Paris, France, in 2021.

He is currently an early stage Researcher with the School of Forest Sciences, UEF. His interests in remote sensing include large-area forest attributes mapping using multi-source LiDAR data.

Lauri Korhonen received the D.Sc. degree in forestry from the University of Eastern Finland (UEF), Joensuu, Finland, in 2011.

He was a Senior Researcher of forest mensuration science with the School of Forest Sciences, UEF. His research interests include the measurements of forest canopy properties and forest inventory with airborne and spaceborne LiDAR data.

Mait Lang was born in Tartu, Estonia, in 1973. He received the Ph.D. degree in forestry from the Estonian University of Life Sciences, Tartu, in 2006.

Currently, he shares his workload between Tartu Observatory, University of Tartu, Tartu, and the Estonian University of Life Sciences, as an Associate Professor. His research interests are mainly related to the remotes sensing methods for estimating forest canopy properties and implementation of remote sensing in forest inventories.

Jan Pisek received the M.Sc. degree in geoinformatics and cartography from Masaryk University, Brno, Czech Republic, in 2004, and the M.Sc. degree in physical geography and the Ph.D. degree in physical geography and remote sensing from the University of Toronto, Toronto, ON, Canada, in 2005 and 2009, respectively.

He is currently an Associate Professor with the Tartu Observatory, University of Tartu, Tõravere, Estonia. His remote sensing interests include field- and space-based multi-spectral remote sensing, biophysical parameter, and vegetation structure mapping. His primary focus has been on exploiting the signal from multi-angle remote sensing data across multiple platforms.

Gastón M. Díaz is an Adjunct Researcher at the National Scientific and Technical Research Council—Argentina (CONICET), Santa Fe, Argentina, involved in a Forest Research Center in Andean Patagonia (CIEFAP). He is a Forest Engineer with the National University of La Plata, La Plata, Argentina, in 2008, and a Doctor of biology with the National University of Comahue, Neuquén, Argentina, in 2013. His research focus is short-range remote sensing of the vegetation, particularly canopy hemispherical photography, laser scanning, and drone-based mapping. His goal is to customize methods and equipment to fit the socioeconomic context of developing countries. He is the developer of the R package CAnopy IMage ANalysis (rcaiman).

Ilkka Korpela received the Ph.D. degree (Hons.) in forest inventory from the University of Helsinki, Helsinki, Finland, in 2004.

He has worked with optical remote sensing of trees. He has done research in directional reflectance of trees and in the use of waveform recording LiDAR in vegetation mapping.

Dr. Korpela was awarded the Decennial Yrjö Ilvessalo Prize in 2004 by the Finnish Society of Forest Research for his work in image matching and the Hansa Luftbild Award in 2008 by the German Society of Photogrammetry and Remote Sensing for a method that combines LiDAR and photogrammetry.

Zhongyu Xia received the M.Sc. degree in agriculture and forestry from the University of Eastern Finland, Kuopio, Finland, and AgroParisTech, Paris, France, in 2023 with the Erasmus M.Sc. in European Forestry degree Program. She is currently pursuing Ph.D. degree with the Swiss Federal Institute of Technology Zurich (ETH), Zürich, Switzerland.

Her remote sensing interests include forest management with optical remote sensing data, such as RGB images and LiDAR point clouds, primarily focusing on forest inventory, forest attributes estimation, and mapping based on airborne/terrestrial LiDAR data.

Hanna Haapala received the M.Sc. degree in agriculture and forestry from the University of Eastern Finland, Joensuu, Finland, in 2023. He is planning on to pursuing the Ph.D. degree.

Her research interests include forest canopy structure measurements using airborne and satellite-based light detection and ranging (LiDAR) data and forest composition estimation with satellite-based optical data.

Matti Maltamo was born in Jyväskylä, Finland, in 1965. He received the M.Sc., Lic.Sc., and D.Sc. degrees (Hons.) in forestry from the University of Joensuu, Joensuu, Finland, in 1988, 1992, and 1998, respectively.

He is a Professor of forest mensuration science with the Faculty of Science and Forestry, University of Eastern Finland, Joensuu. He has authored over 220 scientifically refereed articles. His research interests include different forestry applications of airborne laser scanning.

Dr. Maltamo is also the Editor-in-Chief of *Silva Fennica*.