Dissertationes Forestales 375

Large-scale estimation of boreal forest leaf area index with airborne laser scanning data

Shaohui Zhang

School of Forest Sciences
Faculty of Science, Forestry and Technology
University of Eastern Finland

Academic dissertation

To be presented, with the permission of the Faculty of Science, Forestry and Technology of the University of Eastern Finland, for public criticism in the auditorium F100 of the University of Eastern Finland, Yliopistokatu 7, Joensuu, on the 8th of August 2025, at 12 o' clock noon.

Title of dissertation: Large-scale estimation of boreal forest leaf area index with airborne laser scanning data
Author: Shaohui Zhang

Dissertationes Forestales 375
https://doi.org/10.14214/df.375

Thesis Supervisors:
Associate Professor Lauri Korhonen
School of Forest Sciences, University of Eastern Finland, Finland

Professor Matti Maltamo
School of Forest Sciences, University of Eastern Finland, Finland

Pre-examiners:
Dr. Janne Heiskanen
Department of Geosciences and Geography, University of Helsinki, Finland

Dr. Francesco Chianucci
Research Centre for Forestry and Wood, Italy

Opponent:
Research Professor Svein Solberg
Norwegian Institute of Bioeconomy Research, Norway

ISSN 1795-7389 (online)
ISBN 978-951-651-840-7 (pdf)

Publishers:
Finnish Society of Forest Science
Faculty of Agriculture and Forestry of the University of Helsinki
School of Forest Sciences of the University of Eastern Finland

Editorial Office:
Finnish Society of Forest Science
Viikinkaari 6, FI-00790 Helsinki, Finland
https://www.dissertationesforestales.fi

Zhang, S. (2025) Large-scale estimation of boreal forest leaf area index with airborne laser scanning data. Dissertationes Forestales 375. 39p. https://doi.org/10.14214/df.375

ABSTRACT

Leaf area index (LAI), defined as half of the two-sided leaf area per unit horizontal ground surface area, is an essential variable that describes forest canopy structure. It is a key input in various biosphere-atmosphere models and an important indicator of biodiversity. Temporally and spatially accurate large-area LAI maps are highly needed, but measuring LAI in the field is labour-intensive and time-consuming, especially over large areas. The aim of this thesis was to investigate the feasibility of estimating LAI at large scales using multiple airborne laser scanning (ALS) datasets. Various ALS-derived penetration indices were first compared with field-measured gap fractions at near-vertical angles. The all-echo penetration index (API) showed the least bias among the indices, making it a suitable input following the semi-physical modelling approach. I also explored the utility of ALS polar metrics following the empirical modelling approach, which were found useful and led to improved model accuracy. Furthermore, the performance of both empirical and semi-physical modelling approaches for LAI estimation was assessed at both regional and nationwide scales. While empirical LAI models achieved slightly higher accuracy, the semi-physical model demonstrated better potential for transferability across regions. The nationwide LAI model accuracy could be further improved by incorporating local plots into model calibration. Finally, we proposed a gamified framework for LAI data collection. It may become a valuable data source for validation and calibration of nationwide LAI models.

Keywords: plant area index, LiDAR, model transferability, forest canopy, citizen science, gamification

ACKNOWLEDGEMENTS

The thesis began quietly in 2021, when the world was sighing with relief as the shadow of you-know-what receded. I was wrapping up my MSc thesis in Nancy, France when that email from my supervisors arrived, very unexpectedly. So began the journey that would span the next four years.

It has been an adventure, full of unforgettable moments I could never have anticipated. Sometimes it was smooth, sometimes stormy. In the meantime, I have skied across frozen lakes on sunlit winter days and played tennis late into midsummer nights when the sun refused to set. I have worked alone in the quiet hum of the office long after dark, stepping out into silence and dark, tired, apprehensive, but always a little more resilient.

There were moments of joy and moments of doubts, especially while wrestling with words and ideas that refused to come, or with Reviewer No. 2 who certainly knew how to press my buttons.

I am deeply grateful to my supervisors, Dr. Lauri Korhonen and Prof. Matti Maltamo, who walked with me along this winding path with patient guidance, steady encouragement and unwavering support even when I lost sight of the goal myself.

To my friends and colleague, thank you for the coffees and laughter during lunch and coffee breaks. Your companionship never let me forget I was not alone. To my co-authors, thank you for your constructive feedback and shared commitment throughout the research process.

To my family, thank you for your love and faith in me. You are my light and my anchor.

I also wish to thank the UNITE flagship, the LUMETO doctoral programme as well as the Finnish Society of Forest Science for the support that made this thesis possible.

This work has taken four years and encompassed several existential crises. I hereby declare it is finished and impress it with my own seal.

Ylläs, Finland.

31st March 2025

Shaohui Zhang

LIST OF ORIGINAL ARTICLES

This thesis is based on data presented in the following articles, referred to by Roman Numerals I-III.

I **Zhang S**, Korhonen L, Lang M, Pisek J, Díaz GM, Korpela I, Xia Z, Haapala H, Maltamo M (2024) Comparison of semi-physical and empirical models in the estimation of boreal forest leaf area index and clumping with airborne laser scanning data. *IEEE Transactions on Geoscience and Remote Sensing* 62: 1-12, Article no. 5701212. https://doi.org/10.1109/TGRS.2024.3353410

II **Zhang S**, Korhonen L, Korpela I, Packalen P, Maltamo M (2025) Nationwide airborne laser scanning based models for leaf area index in Finland. Manuscript.

III **Zhang S**, Korhonen L, Nummenmaa T, Bianchi S, Maltamo M (2024) How to implement the data collection of leaf area index by means of citizen science and forest gamification? *Silva Fennica* 58(5): article id 24044 https://doi.org/10.14214/sf.24044

Study I: Zhang participated in field data collection, processed and analysed the data, and wrote the manuscript.

Study II: Zhang was responsible for data curation, model construction and validation, and writing the manuscript.

Study III: Zhang contributed to the conceptualisation of the research idea, data simulation and analysis, original draft preparation, and wrote the manuscript.

OTHER PUBLICATIONS

The following publications are not discussed in this thesis; however, they are included here as they are related to my doctoral research or involve my co-authorship.

I **Zhang S**, Korhonen L, Nummenmaa T, Bianchi S, Maltamo M (2025, April) Note on how to implement the data collection of leaf area index by means of citizen science and forest gamification [Work-in-progress note]. GamiFIN 2025 Conference, Ylläs, Finland.

II **Zhang S**, Haapala H, Korpela I, Maltamo M, Korhonen L (2024, September) Nationwide airborne laser scanning based models for leaf area index mapping in Finland [Paper presentation]. ForestSAT 2024 Conference, Rotorua, New Zealand.

III Kaha M, Lang M, **Zhang S**, Pisek J (2023) Note on the compatibility of ICOS, NEON, and TERN sampling designs, different camera setups for effective plant area index estimation with digital hemispherical photography. *Forestry Studies* 79(1): 21-36. https://doi.org/10.2478/fsmu-2023-0010

IV **Zhang S**, Korhonen L, Korpela I, Maltamo M (2023, September) The utility of airborne laser scanning polar metrics in the estimation of leaf area index [Poster]. SilviLaser 2023 Conference, London, the UK.

V **Zhang S**, Korhonen L, Lang M, Pisek J, Díaz GM, Korpela I, Maltamo M (2022, August) The effect of digital hemispherical photograph binarization methods on the estimation of plant area index with LiDAR data [Paper presentation]. ForestSAT 2022 Conference, Berlin, Germany.

VI **Zhang S**, Vega C, Deleuze C, Durrieu S, Barbillon P, Bouriaud O, Renaud JP (2022) Modelling forest volume with small area estimation of forest inventory using GEDI footprints as auxiliary information. *International Journal of Applied Earth Observation and Geoinformation* 114: 103072. https://doi.org/10.1016/j.jag.2022.103072

TABLE OF CONTENTS

ABBREVIATIONS

ABA	Area-based approach
ALS	Airborne laser scanning
BA	Basal area
CS	Citizen science
CSs	Citizen scientists
DBH	Diameter at breast height (1.3 m)
DCP	Digital cover photography
DCPs	Digital cover photographs
DHP	Digital hemispherical photography
DHPs	Digital hemispherical photographs
EV	Exposure value
EVI	Enhanced vegetation index
FG	Forest gamification
FIs	Forest inventories
GEDI	Global Ecosystem Dynamics Investigation
GF	Gap fraction
GNSS	Global navigation satellite system
Hdom	Dominant tree height
LAI	Leaf area index
LAI_e	Effective leaf area index
LAI_c	Clumping-corrected leaf area index
LiDAR	Light detection and ranging
NDVI	Normalised difference vegetation index
PI	Penetration index
PIs	Penetration indices
PAI	Plant area index
RMSE	Root mean square error
SES	Structuring element size
sd	Standard deviation
VIs	Vegetation indices

1. INTRODUCTION

1.1 Background

The primary objectives of forest inventories (FIs) have traditionally focused on assessing commercially valuable forest attributes, such as timber volume and stand productivity. These conventional FIs have played a vital role in ensuring timber supply, supporting forest management and informing economic decision-making. However, they often largely overlook the broader ecological and social functions of forests, which are increasingly recognised as integral to modern societies. From an ecological perspective, the forest canopy is a three-dimensional subsystem that plays a critical role in ecosystem functioning. It is considered the most species-rich environment within forests, earning it the title of the "last biotic frontier" (Levin 2013). Among various parameters (e.g., canopy cover and canopy closure) used to describe canopy structure, leaf area index (LAI) is one of the most important ones.

LAI is here defined as half of the two-sided leaf area per unit horizontal ground surface area (Chen and Black 1992). It quantifies the amount of foliage present in the canopy, where the vital mass and energy exchanges occur between the biosphere and the atmosphere. Due to its central role in these interactions, LAI has been recognised by the Global Climate Observing System (GCOS) as one of the key variables in global biosphere-atmosphere models, influencing processes such as photosynthesis, evapotranspiration and carbon cycles (Norman and Jarvis 1974; Broge and Leblanc 2001; Baldocchi et al. 2002; Ryu et al. 2011; GCOS 2022). In forestry, LAI serves as a versatile input for various forest-related applications, including growth and yield modelling, surface albedo estimation, forest disturbance assessment and soil nutrient cycling analysis (Pierce et al. 1994; Heiskanen et al. 2012; Härkönen et al. 2013; Hardwick et al. 2015; Wang et al. 2016). Furthermore, LAI plays a fundamental role in biodiversity studies, as it shapes ecosystem structure and functions (Skidmore et al. 2015).

Therefore, temporally and spatially accurate LAI data are of great importance for forest ecology and environmental research. Consequently, high-resolution LAI maps at national and global scales are highly desired.

1.2 LAI estimation methods

1.2.1 In situ measurements

There are two methods of measuring LAI in the field: direct and indirect methods (Breda 2003). Direct methods involve measuring the leaf area of litterfall or destructively sampled leaves, which produces the 'true' LAI (Asner et al. 1998). However, such methods are time-consuming and labour-intensive when applied to large areas (Weiss et al. 2004). LAI can be indirectly estimated based on the allometric relationships with other forest attributes such as basal area and canopy cover. Different allometric models, however, can result in significantly different LAI estimates (Fang et al. 2019).

So far, the majority of in situ LAI measurements has been obtained indirectly by optical instruments, based on the logarithmic relationship between LAI and gap fraction according to the Beer-Lambert law (Eq. 1) (Ross 1981).

$$T(\theta) = e^{\frac{-LAI\ \Omega(\theta)G(\theta)}{\cos(\theta)}} \tag{1}$$

where T(θ) denotes canopy gap fraction at the viewing angle θ, **Ω**(θ) is the canopy clumping index, and G(θ) is the leaf projection function that is the projection coefficient function of unit foliage area on a plane perpendicular to the viewing direction θ.

Devices designed for LAI measurements include the LAI-2200C Plant Canopy Analyser and its predecessor, the LAI-2000 (LI-COR Environmental 2025), as well as digital cameras equipped with standard or hemispherical lenses (Macfarlane et al. 2007; Díaz and Lencinas 2018).

Eq. 1 can be simplified by integrating the entire hemispherical gap fraction measurements without the prior knowledge of G(θ) (Eq. 2) (Miller 1967). Assuming that the foliage is randomly distributed in the canopy (**Ω** = 1), the LAI derived this way is more precisely called effective LAI (LAI_e).

$$LAI_e = -2\int_0^{\frac{2}{\pi}} \ln\big(T(\theta)\big)\cos(\theta)\sin(\theta)\,d\theta \tag{2}$$

When data are only available at a series of discrete viewing angles, Eq. 2 is approximated by the sum (Eq. 3):

$$LAI_e = -2\sum\nolimits_{i=1}^{n} \ln(\bar{T}_i)\cos(\theta_i)\,w_i \tag{3}$$

Another method of calculating LAI_e is by using the gap fraction at the so-called ‘hinge angle’ (57° from the zenith), which relies on the advantage that the G(θ) function at this angle remains constant as 0.5 (Eq. 4) (Wilson 1963; Zhao et al. 2019):

$$LAI_{HA} = -\frac{\cos(\theta)}{G(\theta)}\ln\big(T(\theta)\big) = -1.089 \times \ln\,(T(\theta)) \tag{4}$$

1.2.2 Remote sensing

Remote sensing offers a viable alternative for efficiently collecting LAI information over large geographical areas. Over the past decades, extensive research has focused on estimating LAI using remotely sensed data (Breda 2003; Chen 2018; Fang et al. 2019; Yan et al. 2019). Remotely sensed LAI products are primarily derived from passive optical sensors and active Light Detection and Ranging (LiDAR) instruments across multiple platforms (Chen 1996; Zhao et al. 2011; Fernandes et al. 2014; Li et al. 2017).

Using data acquired from passive optical sensors, LAI is typically estimated by establishing statistical relationships with canopy reflectance or vegetation indices (VIs) (le Maire et al. 2004; Houborg et al. 2009; Chen 2018). While the near-infrared band has been widely used for LAI estimation, many studies recommend incorporating multiple spectral bands to mitigate atmospheric effects and background noise (Cohen et al. 2003; Kobayashi et al. 2007). Consequently, multi-band VI methods have gained popularity, with commonly used indices including the normalised difference vegetation index (NDVI) and the enhanced vegetation index (EVI) (Huete et al. 2002; Wang et al. 2005; Bajocco et al. 2024). Although optical imagery enables large-scale LAI estimation, also at global level through satellite missions such as Sentinel-2, it is often limited by coarse spatial resolution and susceptibility to atmospheric effects and background noise. Additionally, LAI products from passive optical sensors are often constrained by saturation effects in dense canopies (Gower et al. 1999).

Light Detection and Ranging (LiDAR) is a technique that uses laser light to measure distances and general detailed 3D representations of the targets (forests in this case). Estimating LAI using active LiDAR data often relies on gap fraction analysis, from which LAI is subsequently derived (Eq. 1). LiDAR systems operate across multiple platforms, including spaceborne, airborne (mounted on aircrafts or drones) and terrestrial systems. Spaceborne LiDAR instruments, such as the Global Ecosystem Dynamics Investigation (GEDI), provide a great opportunity of collecting LAI sampling data (rather than continuous wall-to-wall coverage) at the global scale (Dubayah et al. 2020). However, GEDI operates only between 51.6° N and S latitude, limiting its applicability in high-altitude boreal forests in countries such as Finland. While terrestrial and drone-based LiDAR systems offer highly detailed 3D canopy information, their limited spatial coverage makes them impractical for large-area mapping at nationwide level (Zhu et al. 2018; Tian et al. 2025).

Among the available remote sensing platforms, airborne laser scanning (ALS), a LiDAR system mounted on an aircraft with a scanning mechanism, has been widely adopted for large-scale forest data collection. Recent implementations have taken place in many countries, including the Netherlands (Kissling et al. 2023), Norway (Astrup et al. 2019), Sweden (Nilsson et al. 2017), Denmark (Magnussen et al. 2018) and Canada (Wulder et al. 2017). ALS has also proven successful in mapping LAI at local (i.e., site level) and regional levels (Solberg 2010; Korhonen et al. 2011). It offers a cost-effective balance between spatial coverage and resolution, making it a well-suited tool for nationwide LAI mapping.

Using ALS data, LAI is typically estimated through statistical regression models (Lim et al. 2003; Riaño et al. 2004; Korhonen and Morsdorf 2014; Heiskanen et al. 2015). Empirical modelling is one of the most common approaches, where parameters are estimated using methods such as ordinary least squares, among other algorithms (Fang et al. 2019). ALS-derived metrics used in these models can be categorised into three main types: height-based, density-based, and penetration indices (PIs), which are derived from different echo types or intensity values. While empirical models are straightforward to implement and often yield high accuracy, their applicability outside the calibration domain is often limited (Richardson et al. 2009). As an alternative, the semi-physical model estimates LAI by establishing relationships with ALS-derived PI alone, which is simpler and more general (Solberg et al. 2009). Thus, the semi-physical model potentially has a better transferability, which is a desired merit when it comes to nationwide LAI mapping.

Given the country-specific objectives of nationwide ALS campaigns, variations in sensor specifications and acquisition parameters are expected. Previous research has extensively examined the influence of flying altitude and speed, scanning angle (off nadir), pulse density, beam divergence, footprint size and pulse repetition frequency (Næsset et al. 2004; Hopkinson 2007; Næsset 2009; Ørka et al. 2010; Bater et al. 2011). Notably, Næsset (2002) found that these ALS parameters did not significantly affect forest attribute predictions in the area-based approach (ABA). However, for LAI prediction, one of the most critical ALS parameters is arguably the scan angle. Previous studies have shown that enlarging the scan angle reduces the probability of receiving ground echoes while increasing the likelihood of vegetation echoes (Disney et al. 2010; Montaghi 2013). As a result, models that rely on ALS PIs may be especially sensitive to variations in scan angle.

1.3 Research gaps

Despite substantial technological and methodological advances in large-scale LAI estimation, notable gaps remain. While current optical LAI products are valuable for large-scale applications, they often fall short in providing accurate LAI predictions at high spatial resolution. Meanwhile, national ALS campaigns represent an underutilised opportunity for LAI mapping, offering both high-resolution data quality and extensive geographical coverage. However, to realise this potential, robust models that are transferable across regions are needed, and their performance remain to be evaluated.

A fundamental requirement for developing such large-scale ALS-based LAI models is the availability of reliable in situ LAI reference data. Yet, collecting in situ LAI data remains challenging, as it is not routinely measured by conventional FIs and requires specialised instruments as well as favourable weather conditions. Consequently, in situ LAI data remain scarce, especially across large areas.

To address this challenge, citizen science has emerged as a promising approach. By involving the public in structured data collection efforts, citizen science programmes can significantly expand the spatial distribution of LAI reference datasets. This approach offers a practical solution to the current data shortage and can play a crucial role in supporting the development, calibration and validation of large-scale ALS-based LAI models.

1.4 Objectives

The primary objective of this thesis is to evaluate the feasibility of developing nationwide LAI models using multiple ALS datasets following both empirical and semi-physical modelling approaches. Specifically, the study investigates whether incorporating ‘novel’ polar ALS metrics can improve LAI estimation, assesses the performance of different LAI modelling approaches at both regional and nationwide scales, examines the effectiveness of calibrating nationwide models using local sample plots, and proposes a gamified framework for innovative in situ LAI data collection. This thesis comprises three sub-studies:

- **Study I** examines the utility of ALS-derived polar metrics in empirical models. It assesses model accuracy by comparing different modelling approaches using a subset of the study areas at the regional level.
- **Study II** extends the first study by introducing nationwide models involving all study areas. It discusses the advantages and limitations of empirical and semi-physical modelling approaches and assesses the effectiveness of calibrating nationwide LAI models using a limited sample of local plots.
- **Study III** addresses the challenge of in situ LAI data collection that can be potentially used for validating nationwide models. It explores the potential of integrating citizen science and forest gamification as an innovative approach to collecting LAI data and assesses its feasibility using simulated data.

Through these studies, this research aims to advance the development of ALS-based large-scale LAI estimation while addressing its key challenges, recommending modelling techniques, and proposing data collection strategies (Figure 1).

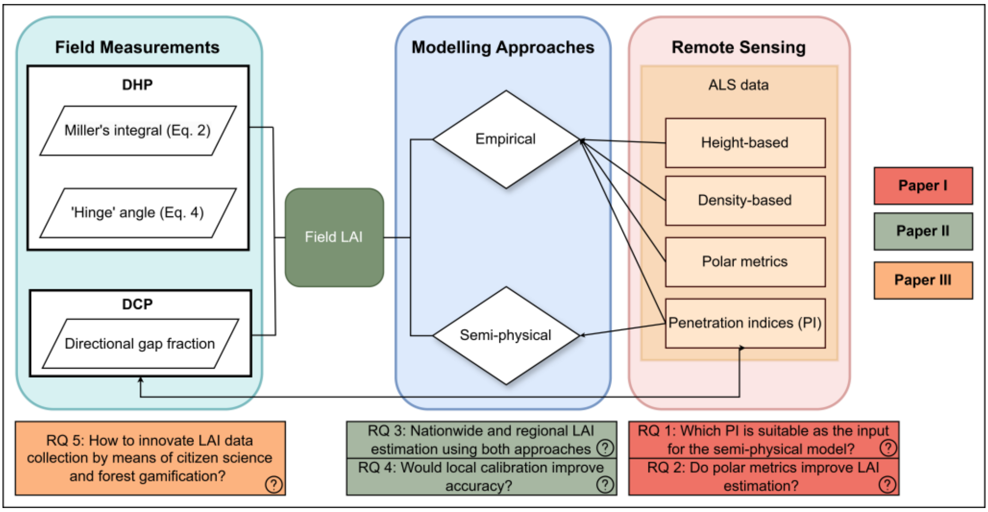

Figure 1. Thesis framework and interconnections of the sub-studies. Abbreviations: DHP (digital hemispherical photography), DCP (digital cover photography) and ALS (airborne laser scanning).

2. MATERIALS AND METHODS

2.1 Study sites

A total of 253 field plots from nine study sites across Finland were included in this study (Figure 2). The main species consisted of Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* (L.) Karst.) and birches (*Betula spp* L.). The plots were selected in a way that would cover a wide range of forest structures. The plot centres were recorded using a Trimble Geo 7 GNSS. Common forest inventory attributes, such as diameter at breast height (DBH), dominant tree height (Hdom) and basal area (BA), were measured at plot level (Table 1).

2.2 Data acquisitions and processing

Given the use of different data types and multiple datasets in the sub-studies, we provide Figure 3 for clarity. While each type of data followed specific processing methods, the datasets obtained from different sites were processed using the standardised procedure.

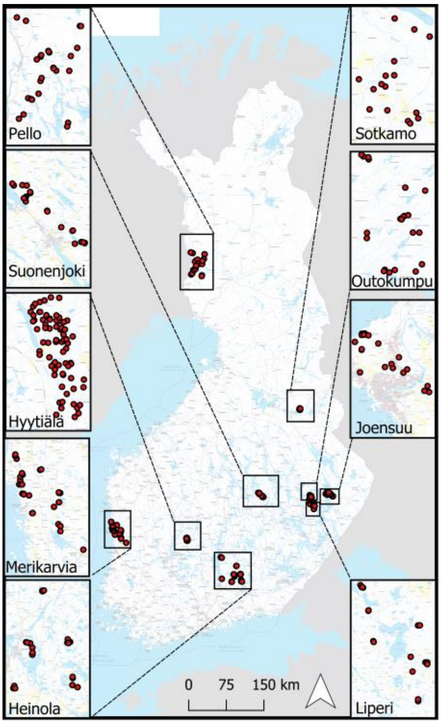

Figure 2. Locations of the study sites and their subordinate plots in Finland

Table 1. Forest attributes of LAI plots across different study sites in Finland

Study sites	n	Dominant height (m)				Basal area (m^2ha^{-1})			
		min	mean	max	sd	min	mean	max	sd
Hyytiälä (2011)	73	2.2	16.8	34.3	7.0	0.5	22.9	51.3	10.7
Suonenjoki (2014)	20	4.0	15.7	26.9	6.7	4.0	18.0	34.0	8.5
Liperi (2016)	20	4.2	16.2	32.6	7.2	1.0	18.1	44.0	11.4
Outokumpu (2021)	20	6.0	17.0	26.3	6.2	2.5	19.7	44.0	12.8
Sotkamo (2021)	16	6.3	15.4	22.9	5.5	2.0	17.5	34.0	10.1
Heinola (2022)	30	2.6	13.8	26.9	7.5	1.0	14.8	34.0	8.2
Merikarvia (2022)	30	3.8	16.9	26.6	7.1	3.0	27.0	83.0	18.5
Pello (2022)	27	1.7	12.5	19.3	4.8	2.0	19.9	50.0	9.8
Joensuu (2023)	17	4.5	18.1	38.5	8.6	5.0	18.71	33.0	7.7

Note: The names of individual study sites and their measurement years in brackets. Abbreviations: n (the number of plots); min (minimum); max (maximum) and sd (standard deviation). The mean values represent the basal area median trees of the dominant tree species.

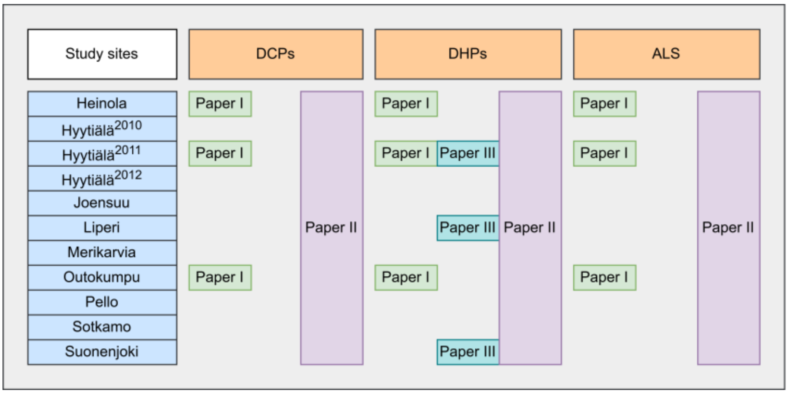

Figure 3. Illustration of datasets used in the three sub-studies

2.2.1 Digital hemispherical photographs

Digital hemispherical photography (DHP) using two sets of digital camera equipment was employed to measure in situ LAI as reference data. A Nikon Coolpix 8800 camera equipped with an FC-E9 fisheye converter was used at the Hyytiälä, Suonenjoki, Liperi and Sotkamo sites, while a Canon EOS2000 camera with a Sigma 4.5 mm fisheye lens was used at the Outokumpu, Merikarvia, Heinola, Pello and Joensuu sites. The image acquisition schemes differed slightly across the sites. Twelve images per plot were captured at Hyytiälä, Joensuu, Liperi and Suonenjoki, while five images were taken at Heinola, Merikarvia, Outokumpu, Pello and Sotkamo sites (Figure 4). All measurements were taken under diffuse sky conditions, either under uniform overcast skies or soon after sunset. The cameras were mounted on a tripod at a height of approximately 1.3 m above ground level, pointed upwards and levelled using a two-axis bubble level. They were set to aperture priority with the aperture fixed at *f*/8. Focus was set to infinity for the Nikon camera and autofocus for the Canon camera. Auto exposure bracketing function was enabled, with the base exposure value (EV) set at -2 and a bracketing range of ± 1, producing a sequence of three images at EVs of -3, -2 and -1. All DHPs were saved in raw image format, and the image with the optimal exposure was manually selected for further processing.

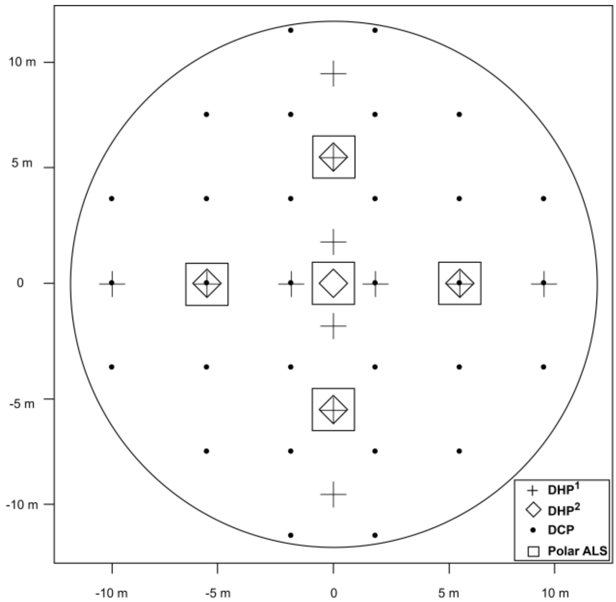

Figure 4. Measurement schemes of digital hemispherical photography (DHP) and digital cover photography (DCP) at plot level. DHP[1] refers to twelve DHP measurement scheme at Hyytiälä, Joensuu, Liperi and Suonenjoki sites, and DHP[2] refers to five DHP measurement scheme at Heinola, Merikarvia, Outokumpu, Pello and Sotkamo sites. Polar ALS means the reference spots where polar ALS metrics were computed.

The acquired DHPs were later processed with the image processing software Hemispherical Project Manager which implements the LinearRatio method (Cescatti 2007). This method considers the camera's linear response to light for a single camera operated below the canopy (Lang et al. 2010). The software first extracts the original blue pixels with the help of dcraw (version 9.28), using the following switches: -d (document mode, no colour and interpretation), -W (do not automatically brighten the image), - g 1 1 (linear 16-bit custom gamma curve). Camera-specific parameters were used to correct lens projection distortion, and the default value of 1.0 was used to minimise the effects of vignetting. By fitting unobscured sky pixel sampled from canopy gaps and interpolated pixel values nearby to a standard overcast sky radiance model (ISO/CIE), the above-canopy reference images were reconstructed (Lang et al. 2017). Finally, binarized DHPs were exported using automatic thresholds that yielded the same gap fraction as in the ratio images. Gap fraction was subsequently derived from the binarized DHPs using a ring-wise analysis, in which each image was divided into six concentric rings at 15° interval, following the same design of the widely used LAI-2200C plant canopy analyser (LI-COR Environmental 2025). The weight assigned to each ring was calculated using Eq. 5:

$$W_i = \sin\theta_i \Big/ \sum_{j=1}^{n} \sin\theta_i \tag{5}$$

where θ_i represents the mean zenith angle of the ring (7°, 23°, 38°, 53°, 68° and 83°), and W_i is the corresponding weight for ring i. Note that the weight for the sixth ring was reassigned to the fifth ring.

Morphological operations were performed on the DHPs to extract between- and within-crown gaps (Korhonen and Heikkinen 2009). The operations were controlled by a parameter known as the structuring element size (SES), which was set to 8 for DHPs taken with the Nikon Camera and 10 for those from the Canon camera due to their difference in image resolution. The SES values were selected to best match manually estimated between-crown gaps at plot level. It was kept consistent across all five concentric rings, as the impact on the derived T(θ) was negligible. The resulting mask was assumed to effectively separate the binarized DHPs into large between-crown gaps and a continuous canopy layer containing small within-canopy gaps.

Based on the T(θ) obtained at plot level, LAI_e can be calculated using Eq. 3. Various methods of canopy clumping correction have been proposed in the literature, such as the LX (Lang and Xiang 1986), CC (Chen and Cihlar 1995), CLX (Leblanc et al. 2005) and LXG methods (Chianucci et al. 2019). A detailed review on the clumping correction methods can be found in (Fang 2021). Overall, all methods align with the physical meanings of the definition of canopy clumping **Ω**. We selected the CC method (Eq. 6), as it intuitively corresponds to morphological image analysis described above:

$$\Omega_{CC}(\theta) = \frac{\ln[F_m(0,\theta)]}{\ln[F_{mr}(0,\theta)]}\frac{[1 - F_{mr}(0,\theta)]}{[1 - F_m(0,\theta)]} \tag{6}$$

where $F_m(0,\theta)$ represents the mean canopy gap fraction at each concentric ring measured at plot level. $F_{mr}(0,\theta)$ refers to the gap fraction when the canopy has a random foliage distribution, estimated by subtracting the mean between-crown gap fraction from the mean total gap fraction.

The element clumping index Ω_E, which quantifies the plot-level foliage clumping, was calculated as the mean of the directional $\Omega_{CC}(\theta)$ obtained from the five concentric rings, as defined in Eq. 7:

$$\Omega_E = \frac{1}{n}\sum_{i=1}^{n}\Omega_{CC}(\theta_i) \tag{7}$$

For plots dominated by coniferous trees, an additional correction was made to account for clumping at the shoot level. The clumping-corrected LAI (LAI_c) was then derived from using Eq. 8:

$$LAI_c = \frac{LAI_e}{\Omega_E \times \Omega_s} \times P_c + \frac{LAI_e}{\Omega_E} \times (1 - P_c) \tag{8}$$

where P_c denotes the field-measured proportion of coniferous trees basal area, and Ω_s is the ratio of shoot silhouette area to total needle area, which was assumed to be a fixed value of 0.56 (Stenberg 1996; Stenberg et al. 2003). Throughout this thesis, the term LAI refers to both LAI_e and LAI_c, unless otherwise specified.

2.2.2 Digital cover photographs

Digital cover photographs (DCPs) were acquired alongside DHPs, with 30 images collected per plot (Figure 4), using an Olympus µ700 at Hyytiälä site or Canon SX200 IS at other sites. The cameras operated in aperture priority mode, with automatic exposure values decreased by 1–2 stops. For plots with dense forest canopies, aperture and shutter speed settings were adjusted to prevent overexposure. The DCPs were then taken with the cameras pointed upwards and saved in JPEG format.

The DCPs were processed using a custom MATLAB script to obtain gap fraction T(θ) estimates around zenith. The images were converted to binary format to distinguish between sky and background using the thresholding method proposed by Nobis and Hunziker (2005). The JPEG format was considered sufficient for this thresholding method, as it is relatively robust to compression artifacts. To maintain a near-vertical geometry, only the view angles 0°–15° from the zenith were used in calculating T(θ).

2.2.3 Airborne laser scanning data

Discrete-return airborne laser scanning (ALS) data were acquired during leaf-on seasons of the same year as field measurements. Note that the Hyytiälä site was also scanned prior to (2011) and after (2013) the field measurements. Thus, the Hyytiälä plots were included in the nationwide modelling three times to better consider sensor effects. We assumed that the forest condition over the one-year period remained the same, as none of the plots appeared as an outlier in the models. Details of the sensor specifications and scanning parameters of the ALS campaigns were listed in Table 2.

Although the ALS data came from different surveys, the data processing had the same procedure. First, all echoes were classified into four types: single, first of many, middle, and last of many, based on their echo number and the number of echoes per pulse. Ground echoes were classified using the Triangular Irregular Network (TIN) method and Digital terrain models (DTMs) were constructed from the ground echoes. Next, the echo heights above the ground were normalised by subtracting their DTM values from the recorded heights. As the scan angle effects were not within the scope of this study, we only kept echoes having the scan angles ≤ 15 degrees to minimise the sensor effects. Finally, ALS metrics were calculated at plot level using a radius of 20 m, following the area-based approach (ABA) (Næsset 2002; Bouvier et al. 2015).

The main ALS metrics included in this study were height and density-based metrics, ALS polar metrics and various ALS penetration indices (PIs). Height (h*) and density (d*) percentiles were calculated at 5% increments (i.e., 5%, 10%, ..., 95%). The minimum, maximum, mean and standard deviation values were also calculated.

Polar metrics were derived by converting the Cartesian coordinates (X, Y, Z) of all ALS echoes into polar coordinates defined by azimuth (φ) and zenith (θ) angles. The location spots for which polar metrics were computed followed the same layout as DHP image acquisition (Figure 4). To account for the broader area captured by DHP due to its wide field of view (FOV > 180°), an extended plot radius (40 m) was used for calculating polar metrics. After coordinate transformation, DHP-like images were constructed from ALS data by binning all ALS echoes into a systematic grid defined by polar angles (φ, θ) before being rasterized to an image of 480 × 480 pixels covering the full hemisphere (Figure 5A). We tested various image resolutions and selected this one as a balance of spatial detail and echo density per pixel. Initially, each pixel's value reflected the number of echoes (n) it contained. The maximum echo count within the 0°–75° zenith range ($nmax$) was used to normalise fractional cover as $n/(nmax/2)$, with values > 1 truncated to 1. The corresponding gap fraction was calculated as $1 - n/(nmax/2)$. Hereafter this rasterization is referred to as greyscale polar image (Figure 5B). This greyscale image was then binarized to separate canopy (1) and gaps (0), hereafter referred to as binarized polar image (Figure 5C). Finally, the binarized polar image underwent morphological operations using a SES of 7, similar to the processing method applied to the real DHPs (Figure 5D). Gap fractions were calculated following the same ring-wise analysis at each location spot, and their means were used to obtain the plot-level ALS polar metrics. The terms binarized-gaps* and greyscale-gaps* refer to gap fractions at 1–5 rings derived from the binarized and greyscale polar images respectively, and the term morphological-gaps* denotes between-crown gaps derived from using morphologically processed polar images. For instance, greyscale-gaps2 refers to the gap fraction of the second ring derived from the greyscale polar images. Collectively, these three terms are referred to as “polar ALS metrics”. The main motivation of computing this novel set of polar ALS metrics was the possibility of calculating between- and within- crown gap fractions in a similar manner as with real DHPs that are naturally displayed in polar angles. As a result, they were expected to improve the accuracy of LAI estimation.

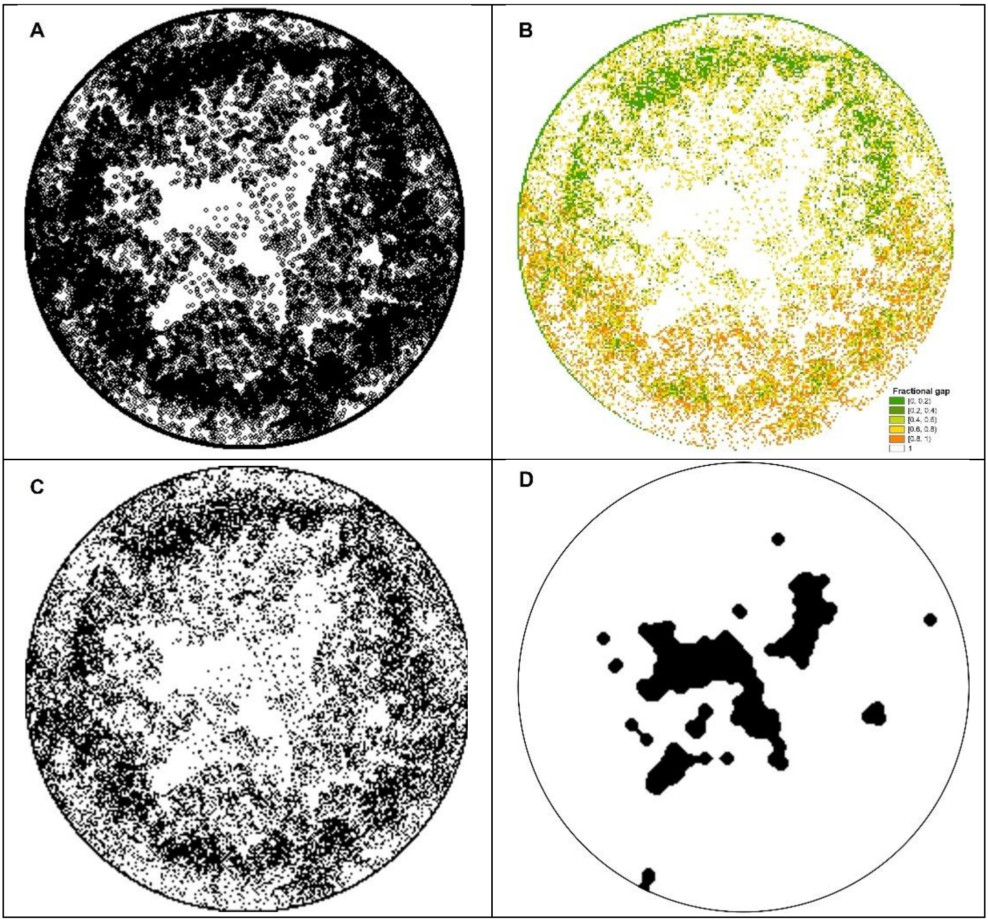

Figure 5. A) DHP-like image from polar transformed ALS data. B) Greyscale polar image coloured by fractional gaps. C) Binarized polar image. D) Between-crown canopy gaps extracted through morphological operations, with gaps displayed in black.

Table 2. ALS sensor specifications for the leaf area index modelling study sites

Sensor Properties	Hyytiälä[2010]	Hyytiälä[2011]	Hyytiälä[2012]	Suonenjoki	Liperi	Outokumpu	Sotkamo	Heinola	Merikarvia	Pello	Joensuu
Models	Leica ALS60	Leica ALS60	Leica ALS60	Leica ALS70-HA	Optech Titan[b]	Riegl VQ-1560 II	Riegl VQ-1560 II	Riegl VQ-780i	Riegl VQ-780i	Riegl VQ-780II	Leica ALS70-HA
Dates	19 Jul 2010	2 Aug 2011	5 Jul 2012	4 Sep 2014	2–10 Jul 2016	12 Jun 2020	25 Jun – 6 Jul, 2021	14 Jun 2021	6–9 Jun 2021	3-6 Jul 2021	15–16 Jun 2020
Max scan angle, °	30	17	15	30	22	15	20	20	20	20	22
Flying altitude AGL[a] (m)	1180	760	2000	2000	900	2100	2100	1265	1265	1292	2561
Strip overlap (%)	60	55	45	20	55	40	33	40	21	24	31
Pulse density (m^{-2})	14.4	9.4	10.5	0.8	18.4	5.6	5.8	6.8	6.5	7.4	0.7
Pulse repetition frequency (kHz)	173	118	59	140	250	134	134	100	100	120	106
Beam divergence (mrad)	0.26	0.22	0.22	0.22	0.35	0.25	0.25	0.25	0.25	0.25	0.25
Footprint diameter (cm)	30	17	44	44	32	52	52	31	31	32	64

[a] above ground level

[b] Only the 1064 nm band was use

Various ALS PIs were used in this study, including the all-echo penetration index (API, Eq. 9), first-echo penetration index (FPI, Eq. 10), last-echo penetration index (LPI, Eq. 11) and Solberg's penetration index (SPI, Eq. 12). Another echo-weighted penetration index (EWI, Eq. 13) based on echo numbers was also included. We intentionally omitted intensity-based ALS PIs (Hopkinson and Chasmer 2009; Armston et al. 2013), because intensities can be substantially different for different sensors, and normalizing them between the areas was beyond our scope.

$$API = 1 - \frac{\sum All_v}{\sum All} \tag{9}$$

$$FPI = 1 - \frac{\sum Single_v + \sum First_v}{\sum Single + \sum First} \tag{10}$$

$$LPI = 1 - \frac{\sum Single_v + \sum Last_v}{\sum Single + \sum Last} \tag{11}$$

$$SPI = 1 - \frac{\sum Single_v + 0.5 * (\sum First_v + \sum Last_v)}{\sum Single + 0.5 * (\sum First + \sum Last)} \tag{12}$$

$$EWI = 1 - \frac{N_v}{N_v + N_g} \tag{13}$$

where *All*, *Single*, *First* and *Last* denote echo types and their subscripts denote that the echo hits vegetation (*v*) or ground (*g*). For EWI, a weight was added to each echo as $(\frac{1}{i})$ and i was the number of echoes of the given pulse. Hence, $N_v = v_1 + \frac{1}{2}v_2 + \frac{1}{3}v_3 + \cdots + \frac{1}{n}v_n$ and $N_g = g_1 + \frac{1}{2}g_2 + \frac{1}{3}g_3 + \cdots + \frac{1}{n}g_n$.

2.3 Model construction and validation

2.3.1 Direct comparison

We directly compared ALS-based PIs with DCP-derived gap fraction at near-vertical angles of 0°–15° without modelling. We evaluated how well the ALS PIs can represent near-vertical gap fraction based on RMSE (Eq. 14) and bias (Eq. 15).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{14}$$

$$Bias = \sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)}{n} \tag{15}$$

2.3.2 Empirical modelling approach

We constructed empirical LAI models using ordinary least squares. We decided to use empirical models with two ALS-based predictors to make the models as general as they possibly can. The predictors were selected using an exhaustive search of all different variable combinations. We also verified that the selected predictors of the models were statistically significant.

2.3.3 Semi-physical modelling approach

The semi-physical LAI model shape (Eq. 16), which requires only one model input T, is simpler and more general.

$$LAI = -\beta * \ln(T) \tag{16}$$

where β is the coefficient to be estimated and T denotes the ALS PI through the canopy. The value of β is influenced by the foliage angle distribution and the technical properties of the ALS sensors. If T accurately represents the near-vertical gap fraction (i.e., is unbiased), β can serve as an estimate of the foliage angle distribution of the canopy. If T is biased, the model remains functional by re-estimating the value through regression analysis using field-measured LAI. In cases where ALS-derived T is unbiased, the model can be applied across forest types assuming β values are similar. Thus, the semi-physical model has a great potential for model transferability, which is a desired merit when it comes to large-scale LAI mapping.

2.3.4 Model validation

All models were validated using leaving-one-out cross-validation (LOOCV). Specifically, the nationwide models were cross validated by leaving each study site out at a time. The regional models were validated by leaving one plot out at a time. All ALS-based LAI models were cross validated and compared by their relative root mean square error (RMSE%, Eq. 17) and their relative mean absolute error (MAE%, Eq. 18).

$$RMSE\% = \frac{100\% \times RMSE}{\bar{y}_i} \tag{17}$$

$$MAE\% = 100\% \times \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} / \bar{y}_i \tag{18}$$

2.4 Nationwide model calibration using local samples

In Paper II, we tested to what extent local calibration of the nationwide models would improve the accuracy of estimation. The local calibration was performed by using mixed-effects models. The same predictors used in the basic nationwide models were used as fixed effects, and the study site was included as the random effect (i.e. grouping variable). In sites with multiple ALS acquisitions (Hyytiälä), each scan was considered as a separate site. Specifically, the calibration started by 1) sequencing one study site as calibration dataset and the remaining ten sites as training data; 2) building linear mixed-effects models with the training dataset; 3) randomly selecting 20% of the calibration dataset to estimate the random effect with the best linear unbiased predictor (BLUP) estimator; 4) predicting LAI_e and LAI_c with fixed and random effects. Local calibration was iterated 1000 times and the mean RMSE% was reported for each study site. The BLUP estimator of random effect was calculated as Eq. 19:

$$\tilde{b} = DZ'_k(Z_k DZ'_k + R_k)^{-1}(y_k - X_k\beta) \tag{19}$$

where D is the variance-covariance matrix of random effects, R is the variance-covariance matrix of residuals, Z and X contain the same ALS metrics, and y information of the field metrics of the randomly sampled calibration plots, β is the fixed effects of the fitted model (Mehtätalo and Lappi 2020). In our case, the residual variance-covariance matrix was approximated by the residual variance.

2.5 Simulating directional photography for gamified LAI data collection

While we attempted to build nationwide LAI models using ALS data, validating these large-scale models with local in situ LAI measurements remains essential. In Paper III, we explored the potential of using citizen science (CS) and forest gamification (FG) for LAI_e data collection through a simulation study. The question under investigation was simplified to determine how many locations within a plot should be used for taking directional photographs, and how many images should be taken at each location.

The simulation was based on sub-sampling the DHP datasets from the Suonenjoki, Hyytiälä^{2011} and Liperi sites. Thus, there were twelve potential locations where directional photographs could be simulated. Different simulation scenarios were designed, each varying in the number of images to be taken at different locations.

The simulation began by randomly selecting the centre coordinates (φ, θ) of the simulated directional photographs. The azimuth angle (φ) was randomly drawn from 0° to 360°, reflecting the flexibility of citizen scientists to capture canopy images from any direction in a forest gamification setting. The zenith angle (θ) was fixed at 57°, as calculating LAI_e at this angle does not need to consider the full hemisphere. To emulate the field of view of modern smartphone cameras, which generally have a horizontal viewing angle of approximately 60° and a vertical viewing angle of approximately 50°, the simulated image extent was adjusted accordingly: horizontally from φ ± 30° and vertically within to ± 7.5° of the hinge angle (i.e., 49.5° to 64.5°). This angle-defined bounding box was applied to real DHPs collected in the field to extract gap fractions, which were subsequently used to calculate the LAI_e (Figure 6).

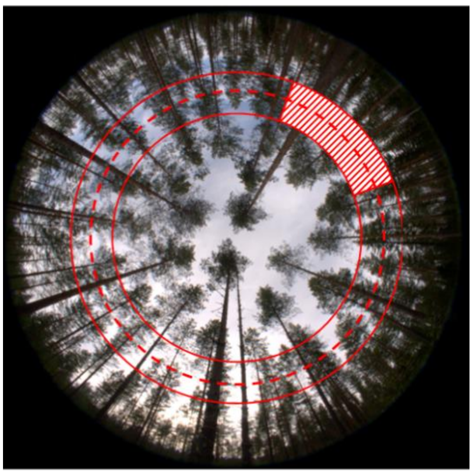

Figure 6. Illustration of a hemispherical photograph with the dash line representing the hinge angle (θ = 57°). The bounding box covering a 60° azimuth range and ± 7.5° around the hinge angle was used to simulate smartphone directional photograph.

Assuming the gap fraction extracted from the bounding box could represent the gap fraction at the full hinge angle, the LAI_e at the hinge angle (LAI_{HA}) was calculated for each simulation using Eq. 4. At each plot, this simulation was repeated 100 times. The performance of each simulation scenario was evaluated based on the mean RMSE% ($\overline{RMSE\%}$, Eq. 20) and mean standard deviation ($\overline{SD}$, Eq. 21).

$$\overline{RMSE\%} = \frac{\sum_{i=1}^{m} RMSE\,\%}{m} \tag{20}$$

$$\overline{SD} = \frac{\sum_{i=1}^{m} \sqrt{\frac{\sum_{i=1}^{n}(LAI_{HA} - \overline{LAI_{HA}})^2}{n}}}{m} \tag{21}$$

where LAI_{HA} was derived from the simulated images using the truncated gap fraction in each simulation, $\overline{LAI_{HA}}$ was the mean of LAI_{HA}, m was the total number of plots (127), and n was the number of simulations per scenario (100).

3. RESULTS AND DISCUSSION

3.1 DCP-measured vs. ALS-derived near vertical gap fraction (sub-studies I and II)

We directly compared gap fraction obtained from DCPs and ALS PIs at near-vertical angles (0°–15°). In Paper I, the comparison was done using data from three study sites, including Heinola, Hyytiälä[2011] and Outokumpu, while all study sites were included in Paper II. The combined results are shown in Table 3.

Overall, all ALS PIs were able to represent DCP-measured near-vertical gap fraction with varying performance. The FPI underestimated gap fraction with a positive (0.13) across sites. Conversely, the LPI overestimated gap fraction with a negative mean bias (-0.22). Both SPI and EWI had considerably smaller biases but were inconsistent across sites, with SPI (mean bias: -0.04) in general overestimating gap fraction and EWI (mean bias: 0.05) underestimating gap fraction. The API had the smallest mean bias (0.01) among all ALS PIs and was relatively stable across sites (ranging from -0.06 to 0.07).

The purpose of this direct comparison was to identify the most suitable ALS PI for the semi-physical model, which is only valid when the input T is strongly correlated with field-measured near-vertical gap fraction. Although ALS (top-down) and DCP (bottom-up) have inherently different viewing directions, both measurement devices capture comparable gap fractions since ALS beams travel similarly to light. To ensure the accuracy of the comparison, we only included ALS echoes with scan angles ≤ 15° to match the view angles of ALS and DCP.

Table 3. Direct comparison of DCP-measured near-vertical gap fraction with ALS PIs

Sites	API		SPI		FPI		LPI		EWI	
	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
Heinola	0.10	-0.03	0.19	-0.13	0.10	0.04	0.35	-0.30	0.10	-0.01
Hyytiälä[2010]	0.11	-0.06	0.11	-0.06	0.09	0.05	0.20	-0.17	0.20	0.12
Hyytiälä[2011]	0.11	-0.05	0.10	-0.05	0.09	0.05	0.19	-0.15	0.08	-0.01
Hyytiälä[2012]	0.13	-0.05	0.12	-0.05	0.11	0.05	0.21	-0.15	0.10	-0.01
Joensuu	0.13	0.02	0.12	-0.01	0.23	0.22	0.30	-0.24	0.14	0.10
Liperi	0.09	0.06	0.05	-0.01	0.15	0.14	0.17	-0.16	0.11	0.10
Merikarvia	0.10	0.03	0.11	-0.05	0.16	0.14	0.29	-0.25	0.11	0.08
Outokumpu	0.05	-0.02	0.08	-0.07	0.12	0.10	0.26	-0.25	0.46	-0.16
Pello	0.11	0.07	0.07	-0.03	0.17	0.16	0.25	-0.21	0.12	0.11
Sotkamo	0.07	0.05	0.04	-0.02	0.15	0.14	0.20	-0.18	0.10	0.09
Suonenjoki	0.13	0.07	0.12	0.01	0.37	0.35	0.38	-0.33	0.20	0.18
Mean	**0.10**	**0.01**	**0.10**	**-0.04**	**0.16**	**0.13**	**0.25**	**-0.22**	**0.16**	**0.05**

Our findings aligned with previous research. Studies have also found that the FPI is not sensitive to detect small within-crown gaps due to the relatively larger size of ALS footprints, which leads to underestimated gap fractions (Lovell et al. 2003; Morsdorf et al. 2006). Conversely, the LPI tends to overestimate gap fractions (Korhonen et al. 2011). While the SPI and EWI can represent near-vertical gap fractions with some biases, the API is robust across sites and thus is suitable for large-scale estimation. Therefore, among all the ALS PIs, the API may best serve as a proxy of gap fraction and is a suitable input for the semi-physical model.

3.2 Utility of polar metrics (Sub-study I)

We first tested the utility of polar metrics in training empirical LAI_c and Ω_E models using data from the Hyytiälä[2011] site. Results showed that using polar metrics alone as predictors achieved good accuracy, with RMSE% values of 24.9% and 9.0% for LAI_c and Ω_E respectively. Incorporating polar metrics with other types of ALS metrics resulted in higher accuracy (20.6% for LAI_c and 4.3% for Ω_E), while excluding polar metrics from the predictors yielded similar RMSE% values of 21.8% and 4.1% respectively. The test results therefore suggested that polar metrics may store canopy information that is not captured by commonly used ALS metrics (e.g., height- and density-based metrics), which may enhance LAI_c and Ω_E estimations.

Next, we extended the assessment of polar metrics in estimating LAI_e, LAI_c and Ω_E in three study sites, including Heinola, Hyytiälä[2011] and Outokumpu (Paper I). The results showed that polar metrics were frequently selected as predictors in regional models following the empirical modelling approach (Table 4). Notably, combining polar metrics with ALS PIs obtained reliable Ω_E estimation (RMSE%: 4.2%–8.5%). This provided new insights into generating high-resolution Ω_E maps using ALS data.

Table 4. Selected ALS metrics in the empirical models

Sites	Variables	Predictors	RMSE%
Heinola	LAI_e	log(API)	17.2%
		greyscale-gaps5, log(API)	16.4%
		h5, log(FPI), log(API)	15.6%
	$sqrt(LAI_c)^a$	log(FPI), log(API)	19.7%
	Ω_E	FPI, greyscale-gaps5	7.1%
$Hyytiälä^{2011}$	LAI_e	log(API)	16.9%
		log(LPI), morphological-gaps5	14.2%
		d60, morphological-gaps4, log(LPI)	13.0%
	$log(LAI_c)^b$	LPI, morphological-gaps5	18.9%
	Ω_E	FPI, morphological-gaps5	4.2%
Outokumpu	LAI_e	log(API)	9.4%
		d95, log(API)	9.4%
		h10, d95, log(API)	8.8%
	LAI_c	SPI, log(API)	18.7%
	Ω_E	FPI, binarized-gaps5	8.5%

[a] square root transformation and [b] log transformation was applied to the variables.

Unlike the angle-based polar grid suggested by Vaughn et al (2013), we calculated polar metrics from pixel-based raster images. We computed a 2D planar systematic grid, allowing the derived DHP-like polar images to undergo the same morphological operations as real DHPs. The resolution of the image was optimised by maintaining the spatial details and having enough ALS echoes pixelwise.

In addition to satisfactory accuracy, polar metrics may offer other benefits. Each polar image corresponds to a specific sample spot at plot level, accounting for the local canopy variation. In contrast, common ALS metrics only describe a fixed area. Furthermore, integrating polar metrics with other ALS metrics may help mitigate LAI saturation effect (Luo et al. 2018). More discussion of ALS polar metrics can be found in Paper I.

3.3 Regional LAI models (Sub-studies I and II)

Building on the regional LAI models from the three study sites examined in Paper I, we incorporated six additional study sites across Finland in Paper II to construct nationwide models. Although polar metrics proved to be useful, we did not include them this time as we wanted to make the LAI models as general as possible with common ALS metrics. For the semi-physical model, we used API as the input T in the semi-physical model as it demonstrated more stability across sites.

The results showed that the empirical and semi-physical LAI_e models achieved comparable accuracy at regional level (Table 5). The smallest RMSE% was observed at Outokumpu site (empirical: 9.4%; semi-physical: 9.0%) while the lowest accuracy was observed at Suonenjoki sites (empirical: 21.8%; semi-physical: 24.6%). While semi-physical

LAI_e models showed slightly better accuracy at the Heinola and Outokumpu sites, empirical LAI_e models considerably outperformed at the Hyytiälä, Joensuu, Liperi, Pello, Sotkamo and Suonenjoki sites. In Merikarvia, no difference was found between the two modelling approaches.

For empirical LAI_e models, the commonly selected predictors consisted of various ALS PIs and their logarithmic transformations as well as height- and density- based metrics at upper or lower thresholds (e.g., h95, d5). For semi-physical LAI_e models, the β parameter ranged from 2.09 (Liperi) to 2.63 (Outokumpu), with a mean of 2.25.

Table 5. Regional empirical and semi-physical LAI_e models and their accuracies

Regions	Regional models	RMSE%	MAE%
Heinola	$LAI_e = -0.132 - 2.647 \times \ln(API) + 0.278 \times \ln(FPI)$	17.6%	12.6%
	$LAI_e = -2.13 \times \ln(API)$	16.9%	12.4%
	Difference:	**0.7%**	**0.2%**
Hyytiälä[2010]	$LAI_e = 15.215 + 14.923 \times min - 5.142 \times \mathrm{API}$	15.6%	12.0%
	$LAI_e = -2.15 \times \ln(API)$	17.6%	13.4%
	Difference:	**-2.0%**	**-1.4%**
Hyytiälä[2011]	$LAI_e = 4.943 - 0.037 \times d20 - 5.072 \times LPI$	13.3%	10.6%
	$LAI_e = -2.12 \times \ln(API)$	16.2%	12.5%
	Difference:	**-2.9%**	**-1.9%**
Hyytiälä[2012]	$LAI_e = -0.261 + 0.055 \times h95 - 0.991 \times \ln(FPI)$	14.8%	12.1%
	$LAI_e = -2.12 \times \ln(API)$	18.0%	13.8%
	Difference:	**-3.2%**	**-1.7%**
Joensuu	$LAI_e = -0.214 + 0.119 \times h05 - 2.111 \times \ln(SPI)$	19.6%	15.8%
	$LAI_e = -2.51 \times \ln(API)$	24.2%	20.0%
	Difference:	**-4.6%**	**-4.2%**
Liperi	$LAI_e = -0.111 + 0.061 \times h05 - 1.854 \times \ln(API)$	11.3%	8.5%
	$LAI_e = -2.09 \times \ln(API)$	13.7%	10.7%
	Difference:	**-2.4%**	**-2.2%**
Merikarvia	$LAI_e = 0.402 - 0.061 \times d05 - 2.667 \times \ln(SPI)$	20.5%	15.3%
	$LAI_e = -2.42 \times \ln(API)$	20.5%	15.5%
	Difference:	**0%**	**-0.2%**
Outokumpu	$LAI_e = -2.644 + 0.029 \times d95 - 2.659 \times \ln(API)$	9.4%	7.9%
	$LAI_e = -2.63 \times \ln(API)$	9.0%	7.3%
	Difference:	**0.4%**	**0.6%**
Pello	$LAI_e = 1.871 - 1.971 \times EWI - 0.604 \times \ln(FPI)$	10.5%	8.4%
	$LAI_e = -2.25 \times \ln(API)$	14.5%	12.3%
	Difference:	**-4.0%**	**-3.9%**
Sotkamo	$LAI_e = -0.379 - 7.062 \times \ln(API) + 3.796 \times \ln(EWI)$	14.2%	13.3%
	$LAI_e = -2.14 \times \ln(API)$	19.0%	15.9%
	Difference:	**-4.8%**	**-2.6%**
Suonenjoki	$LAI_e = -2.571 + 1.302 \times min - 2.136 \times \ln(EWI)$	21.8%	16.8%
	$LAI_e = -2.19 \times \ln(API)$	24.6%	19.5%
	Difference:	**-2.8%**	**-2.7%**

Although the LAI_c model accuracies were slightly lower compared to the regional LAI_e model, the regional LAI_c models overall yielded satisfactory results (Table 6). Empirical LAI_c models had RMSE% values ranging from 15.5% (Liperi) to 27.1% (Suonenjoki) while semi-physical models had RMSE% values ranging 23.6% (Merikarvia) to 32.1% (Sotkamo). The empirical models consistently outperformed their semi-physical counterparts across all sites, with varying accuracy differences. The largest accuracy difference was observed at the Liperi site, where the empirical model outperformed the semi-physical model by 14.1%; while only a minor gain of 2.2% was obtained at the Suonenjoki site. The selected LAI_c model predictors also included various ALS PIs and their logarithmic transformations as well as height- and density- based metrics at upper or lower thresholds (e.g., d95, min). For semi-physical LAI_e models, the β parameter ranged from 3.94 (Hyytiälä^{2012}) to 5.01 (Outokumpu), with the mean of 4.24.

Various ALS-based predictors were used in different regional empirical LAI models and the value of the β parameter also varied in regional semi-physical LAI models. This likely originated from different ALS acquisition settings, highlighting the specific sensor effects on both modelling approaches. On one hand, the empirical modelling approach in general yielded higher accuracy. On the other hand, the semi-physical modelling approach is simple and robust. It requires only one model input and yielded a comparable accuracy.

3.4 Nationwide LAI models (Sub-study II)

We attempted to build nationwide LAI models using plots from all study sites to reach nationwide representation. In a leave-one-site out cross validation, the nationwide LAI models following both the empirical and semi-physical modelling approaches overall achieved comparable accuracies (Table 7).

The empirical approach outperformed the semi-physical approach in predicting both LAI_e (RMSE% by 6.4%) and LAI_c (RMSE% by 9.3%). Both modelling approaches employed the API and its logarithmic transformation as model predictors, indicating the importance of this ALS PI in predicting LAI.

For the semi-physical LAI models, the β values were 2.19 for LAI_e and 4.11 for LAI_c, which was similar to the regional means of 2.25 (LAI_e) and 4.24 (LAI_c). The β values in the API-coupled semi-physical LAI model are influenced by both canopy foliage distribution and ALS sensor properties. So far, no study has investigated the value range for LAI_c. For LAI_e, Solberg et al (2009) suggested that in theory the value is expected to take 2 if the foliage distribution is spherical and the API is unbiased against gap fraction. In our case, the β value for LAI_e proved to be relatively stable. This suggests that although forest conditions and canopy structures vary among different regions across Finland, the foliage angle distribution is relatively consistent in boreal forests. The estimated β value of 2.2 in our nationwide model suggests that API generally overestimates vertical gap fraction. It also indicates an erectophile foliage angle distribution with lower contact frequency in zenith direction than in horizontal direction, which is often the case in boreal forests where trees have long and narrow crowns.

Table 6. Regional empirical and semi-physical LAI_c models and their accuracies

Regions	Regional models	RMSE%	MAE%
Heinola	$LAI_c = 0.472 - 8.149 \times \ln(EWI) + 3.771 \times \ln(FPI)$	23.8%	17.4%
	$LAI_c = -4.07 \times \ln(API)$	30.2%	22.2%
	Difference:	**-6.4%**	**-4.8%**
Hyytiälä[2010]	$LAI_c = 50.918 + 44.476 \times min - 6.876 \times API$	19.6%	15.8%
	$LAI_c = -4.01 \times \ln(API)$	27.1%	21.5%
	Difference:	**-7.5%**	**-5.7%**
Hyytiälä[2011]	$LAI_c = 5.366 + 0.067 \times max - 5.167 \times LPI$	20.0%	15.5%
	$LAI_c = -4.00 \times \ln(API)$	26.7%	20.3%
	Difference:	**-6.7%**	**-4.8%**
Hyytiälä[2012]	$LAI_c = 3.452 + 0.112 \times h95 - 4.195 \times FPI$	20.2%	16.0%
	$LAI_c = -3.94 \times \ln(API)$	28.6%	21.8%
	Difference:	**-8.4%**	**-5.8%**
Joensuu	$LAI_c = 77.647 - 0.781 \times d95 - 2.702 \times \ln(SPI)$	20.7%	15.5%
	$LAI_c = -4.41 \times \ln(API)$	27.7%	22.0%
	Difference:	**-7.0%**	**-6.5%**
Liperi	$LAI_c = 25.527 - 0.279 \times d60 - 6.924 \times LPI$	15.5%	12.0%
	$LAI_c = -3.97 \times \ln(API)$	29.6%	23.8%
	Difference:	**-14.1%**	**-11.8%**
Merikarvia	$LAI_c = 5.356 - 4.936 \times API - 0.757 \times \ln(FPI)$	17.4%	12.8%
	$LAI_c = -4.50 \times \ln(API)$	23.6%	18.8%
	Difference:	**-6.2%**	**-6.0%**
Outokumpu	$LAI_c = 0.815 - 9.970 \times \ln(API) + 6.314 \times \ln(SPI)$	18.7%	14.3%
	$LAI_c = -5.01 \times \ln(API)$	25.7%	22.4%
	Difference:	**-7.0%**	**-8.1%**
Pello	$LAI_c = 7.165 + 12.027 \times FPI - 18.585 \times \ln(EWI)$	19.7%	16.8%
	$LAI_c = -4.46 \times \ln(API)$	26.0%	21.8%
	Difference:	**-6.3%**	**-5.0%**
Sotkamo	$LAI_c = 43.089 - 42.723 \times LPI + 24.118 \times \ln(LPI)$	23.9%	19.7%
	$LAI_c = -4.50 \times \ln(API)$	32.1%	20.3%
	Difference:	**-8.2%**	**-0.6%**
Suonenjoki	$LAI_c = -7.718 + 4.895 \times min - 4.989 \times \ln(API)$	27.1%	21.1%
	$LAI_c = -4.19 \times \ln(API)$	29.3%	23.9%
	Difference:	**-2.2%**	**-2.8%**

Table 7. Nationwide empirical and semi-physical LAI models and their accuracies

Variables	Approaches	Nationwide models	RMSE%	MAE%
LAI_e	Empirical	$LAI_e = 0.9052 - 1.1158 \times FPI - 1.6086 \times \ln(API)$	23.8%	17.4%
	Semi-physical	$LAI_e = -2.19 \times \ln(API)$	30.2%	22.2%
		Difference:	**-6.4%**	**-4.8%**
LAI_c	Empirical	$LAI_c = 6.0342 + 0.0450 \times h95 - 6.2456 \times API$	18.9%	14.1%
	Semi-physical	$LAI_c = -4.11 \times \ln(API)$	28.2%	21.3%
		Difference:	**-9.3%**	**-7.2%**

3.5 Calibrating nationwide models with local plots (Sub-study II)

In Paper II, we also tested how much the accuracy of nationwide LAI models would improve when calibrated using a small sample of local plots. In general, local calibration improved nationwide model performance, although the improvement was modest and not consistent across sites.

Table 8 shows that the calibrated nationwide models had improved performance over the cross-validated nationwide models for both LAI_e and LAI_c. Following the empirical modelling approach, the calibrated nationwide models had slight improvements, with the decreased mean RMSE% by 0.6% for both LAI_e and LAI_c. The semi-physical LAI_e model also showed a mean RMSE% improvement of 0.6% with local calibration, whereas no overall improvement was observed for the semi-physical LAI_c model. Regional LAI models in general had the best accuracy when compared to the nationwide and calibrated nationwide models.

Table 8. Comparison of nationwide, calibrated nationwide and regional LAI models

Approaches	Sites	LAI_e (RMSE%)			LAI_c (RMSE%)		
		Nationwide	**Calibrated**	**Regional**	**Nationwide**	**Calibrated**	**Regional**
Empirical	Heinola	17.2	17.2	17.6	25.9	25.8	23.8
	Hyytiälä^{2010}	16.4	16.4	15.6	19.4	19.4	19.6
	Hyytiälä^{2011}	15.1	15.1	13.3	20.0	20.1	20.0
	Hyytiälä^{2012}	17.1	17.1	14.8	20.9	20.9	20.2
	Joensuu	23.0	22.3	19.6	22.9	22.7	20.7
	Liperi	12.8	12.9	11.3	20.0	20.0	15.5
	Merikarvia	23.3	22.4	20.5	22.6	22.4	17.4
	Outokumpu	20.5	16.7	9.4	25.9	25.9	18.7
	Pello	12.0	12.0	10.5	20.0	20.0	19.7
	Sotkamo	17.5	16.7	14.2	26.9	22.6	23.9
	Suonenjoki	28.9	28.3	21.8	29.1	29.2	27.1
	Mean (sd)	**18.5 (4.8)**	**17.9 (4.5)**	**15.3 (4.0)**	**23.1 (3.2)**	**22.6 (3.1)**	**20.6 (3.1)**
Semi-physical	Heinola	16.2	16.5	16.9	28.3	30.4	30.2
	Hyytiälä^{2010}	17.5	17.2	17.6	26.8	27.9	27.1
	Hyytiälä^{2011}	16.5	17.1	16.2	26.6	27.8	26.7
	Hyytiälä^{2012}	18.2	18.8	18.0	28.3	29.3	28.6
	Joensuu	26.7	24.5	24.2	26.2	26.0	27.7
	Liperi	13.7	15.2	13.7	27.2	30.3	29.6
	Merikarvia	23.2	21.5	20.5	24.8	22.1	23.6
	Outokumpu	21.2	14.4	9.0	31.9	23.9	25.7
	Pello	14.2	14.2	14.5	26.2	24.5	26.0
	Sotkamo	17.5	18.8	19.0	35.7	38.7	32.1
	Suonenjoki	23.6	23.6	24.6	27.7	29.1	29.3
	Mean (sd)	**18.9 (4.0)**	**18.3 (3.4)**	**17.7 (4.3)**	**28.2 (2.9)**	**28.2 (4.2)**	**27.9 (2.3)**

Note: **Nationwide**: accuracies of nationwide models applied to respective regions using leave-one-site-out cross-validation. **Calibrated**: accuracies of nationwide models calibrated at respective regions with local sample plots. **Regional**: accuracies of regional models using leave-one-plot-out cross-validation. sd: standard deviation.

Overall, calibration with local plots proved effective. While improvements were not consistent across all sites, calibration led to an average increase in accuracy compared to the cross-validated nationwide models. This suggests that the nationwide LAI models are robust enough to be applied directly to new regions. Using only 20% of local plots for calibration achieved comparable or slightly improved accuracy than the cross-validated nationwide models. However, the results of local calibration may vary depending on the selected sample plots. In sites like Sotkamo and Joensuu, as few as four plots were sufficient for calibrating the nationwide models. Given the time-consuming nature of in situ LAI data collection, this finding has important implications for optimising field data collection. Another approach to calibration is to incorporate additional predictors from external data sources or to use sample plots that represent the study site (Kotivuori et al. 2018).

3.6 Gamified directional photography data for validating LAI nationwide models (Sub-study III)

The fundamental basis for estimating LAI_e used in Study III involves inferring the complete gap fraction at the hinge angle using a truncated gap fraction, as the projection function $G(\theta)$ remains constant at 0.5 at this angle. Our simulations showed that twenty directional canopy photographs at the hinge angle (Scenario 18) yielded LAI_e estimates comparable to those derived from twelve DHPs at plot level, with an RMSE% of 10.2% (Figure 7). However, additional images may be required in forests with dense or heterogeneous canopy structures.

Alternatively, taking four images at a single location with 90° azimuth intervals (Scenario 4) produced adequate results (with reduced $\overline{RMSE\%}$ by 4.3%). The optimal data collection scheme involves balancing the desired accuracy with the practical feasibility of image acquisition in the field.

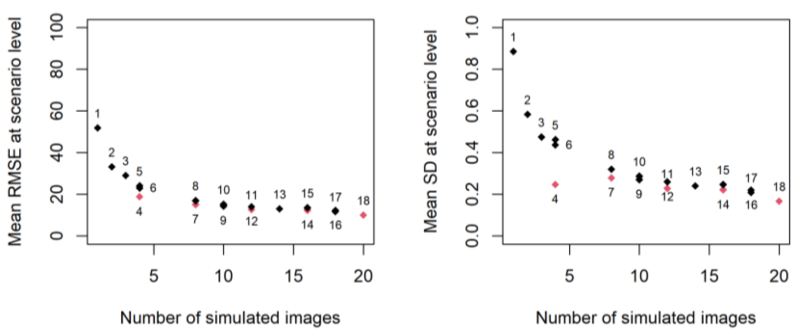

Figure 7. Non-linear relationship between the numbers of simulated images and both the $\overline{\mathrm{RMSE\%}}$ and $\overline{\mathrm{SD}}$ across scenarios. Scenario IDs were shown, with red colour indicating those where images were simulated 90° azimuth intervals

Drawing on previous CS project experience, we observed that simply asking participants to follow scientific protocols without adequate engagement or clear guidance often led to suboptimal data quality. In response, we propose a gamified data collection framework where citizen scientists capture canopy photographs guided by an interactive on-screen element — such as a virtual bird appearing between forest canopies. Birds are natural forest inhabitants capable of moving freely across the hemisphere, making them an intuitive guide to help participants locate the hinge angle. In this framework, the simulated bird appears at specific azimuth (φ) and zenith (θ) angles, with azimuth 0° aligned to the magnetic north and zenith angle fixed at 57° off vertical to ensure the correct image orientation. Participants take a directional image each time they spot the bird in the canopy. After taking the initial image, three additional images are taken at the same location by following with the virtual bird's movements at 90° azimuth intervals. To further improve accuracy, citizen scientists can be encouraged to take images from four additional locations within the plot boundaries and repeat the same process. However, this introduces a trade-off between improving the accuracy and the potential risk of the task becoming overly burdensome for the participants.

4. CONCLUSIONS AND FUTURE PERSPECTIVES

This thesis investigated the feasibility of various modelling approaches and multiple discrete-return ALS datasets for large-area mapping of LAI at both regional and nationwide scales. It concluded that both empirical and semi-physical modelling approaches can achieve satisfactory accuracy. While empirical models yielded slightly higher accuracy, semi-physical models offered greater robustness and transferability across varying conditions. Each modelling approach has distinct benefits and limitations, and the determination of appropriate modelling approach should be guided by the specific objectives of future projects.

Despite these strengths, ALS-based LAI estimation faces several key challenges. First, LAI estimation using remotely sensed data remains constraint by the saturation effect. In the context of ALS, saturation refers to situations where ALS pulses are unable to sufficiently penetrate through the forest canopy, resulting in reduced ground echoes and consequently biased LAI estimates, especially in structurally complex forests. Additionally, it may be argued that there exists an inherent discrepancy between ALS and DHP. ALS, with its top-down perspective, tends to capture more information on foliage, whereas DHP, with its bottom-up perspective, is more sensitive to other canopy elements such as branches and trunks.

As an attempt to address these challenges, this thesis introduced novel polar ALS metrics, based on the suggestion that incorporating various types of ALS metrics could mitigate the saturation effect. This improvement may be attributed to the increased information derived from different types of ALS metrics. With empirical models, the incorporation of polar ALS metrics notably improved model performance, particularly in the estimation of Ω_E. Following the semi-physical modelling approach, the ALS penetration index API was the least biased against vertical gap fraction, making it a reliable input for semi-physical models. Both ALS- and DHP- derived LAI estimates were obtained through gap fraction analysis, albeit over different angular ranges (i.e., near-vertical for ALS and hemispherical for DHP). Since gap fraction is independent of specific canopy elements, this helps address the measurement mismatch that stems from their different viewing geometries.

Looking forward, a key recommendation of this thesis is the simultaneous modelling of LAI_e, LAI_c and Ω_E. Each of these parameters describe distinct biophysical functions of the canopy and has specific applications in forest ecosystem modelling. While empirical models allow the simultaneous estimation of all three parameters, the semi-physical approach is limited to LAI_e and LAI_c. This gives empirical models a practical advantage in applications where a comprehensive suite of canopy parameters is required.

The modelling approaches examined in this thesis showcased the feasibility of mapping LAI at nationwide level. However, these maps are inherently static, as they are based on data collected during a single acquisition period, typically corresponding to peak growing season. Consequently, they are unable to capture temporal dynamics or provide temporarily continuous LAI estimates. Although the National Land Survey of Finland acquires ALS data on a six-year cycle, this frequency is insufficient to support multi-temporal LAI estimation. The integration of satellite-based remote sensing data, such as from the Sentinel-2 mission, may provide a viable solution for achieving time-series LAI estimation and near-real-time updates.

Calibrating nationwide models using a small sample of local plots generally resulted in improved model accuracy. Although the improvement was modest, it indicated that the models are robust and relatively transferable across different regions. In practical terms, calibration with as few as four field plots can improve model performance. This would significantly reduce the labour and costs associated with extensive field data collection.

A final and important consideration concerns the validation of large-area LAI products. While satellite-based LAI products could have been used for intercomparison, doing so would treat other remotely sensed datasets as ground truth and consequently ignore the uncertainties inherent in those products. Field-based validation remains the most reliable method. Therefore, this thesis strongly advocates the use of in situ LAI measurements as the primary means of validating LAI products derived from remote sensing data. To address the limited availability of in situ reference data, a novel gamified framework was proposed to engage citizen scientists in LAI data collection. Data gathered through this framework could be instrumental in validating the nationwide LAI models as well as other LAI products. A key future step involves the development of a mobile application to operationalise the proposed gamified approach. This application will engage users in structured and game-like tasks as they collect directional photographic data for LAI estimation. Future work should focus on testing the application in real-world conditions to evaluate both the quality of the collected data and its suitability for validating large-scale ALS-based models.

REFERENCES

Armston J, Disney M, Lewis P, Scarth P, Phinn S, Lucas R, Bunting P, Goodwin N (2013) Direct retrieval of canopy gap probability using airborne waveform lidar. Remote Sens Environ 134: 24–38. https://doi.org/10.1016/j.rse.2013.02.021

Asner GP, Braswell BH, Schimel DS, Wessman CA (1998). Ecological Research Needs from Multiangle Remote Sensing Data. Remote Sens Environ 63(2): 155–165. https://doi.org/10.1016/S0034-4257(97)00139-9

Bajocco S, Ferrara C, Savian F, Ginaldi F, Puletti N, Crecco L, Bregaglio S, Chianucci F (2024) LAIr: an R package to estimate LAI from Normalized Difference Vegetation Index. SoftwareX 27: 101776. https://doi.org/10.1016/j.softx.2024.101776

Baldocchi DD, Wilson KB, Gu L (2002) How the environment, canopy structure and canopy physiological functioning influence carbon, water and energy fluxes of a temperate broad-leaved deciduous forest--an assessment with the biophysical model CANOAK. Tree Physiol 22(15–16): 1065–1077. https://doi.org/10.1093/treephys/22.15-16.1065

Bater CW, Wulder MA, Coops NC, Nelson RF, Hilker T, Nasset E (2011) Stability of Sample-Based Scanning-LiDAR-Derived Vegetation Metrics for Forest Monitoring. IEEE T Geosci Remote 49(6): 2385–2392. https://doi.org/10.1109/TGRS.2010.2099232

Bouvier M, Durrieu S, Fournier RA, Renaud JP (2015) Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. Remote Sens Environ 156: 322–334. https://doi.org/10.1016/j.rse.2014.10.004

Breda NJJ (2003) Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. J Exp Bot 54(392): 2403–2417. https://doi.org/10.1093/jxb/erg263

Broge NH, Leblanc E (2001) Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sens Environ 76(2): 156–172. https://doi.org/10.1016/S0034-4257(00)00197-8

Cescatti A (2007) Indirect estimates of canopy gap fraction based on the linear conversion of hemispherical photographs. Agr Forest Meteorol 143(1–2): 1–12. https://doi.org/10.1016/j.agrformet.2006.04.009

Chen JM (1996) Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. Agr Forest Meteorol 80(2–4): 135–163. https://doi.org/10.1016/0168-1923(95)02291-0

Chen JM (2018) Remote Sensing of Leaf Area Index and Clumping Index. In Comprehensive Remote Sens: 53–77. Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.10540-8

Chen JM, Black TA (1992) Defining leaf area index for non-flat leaves. Plant Cell Enviro 15(4): 421–429. https://doi.org/10.1111/j.1365-3040.1992.tb00992.x

Chen JM, Cihlar J (1995) Quantifying the effect of canopy architecture on optical measurements of leaf area index using two gap size analysis methods. IEEE T Geosci Remote 33(3): 777–787. https://doi.org/10.1109/36.387593

Chianucci F, Zou J, Leng P, Zhuang Y, Ferrara C (2019) A new method to estimate clumping index integrating gap fraction averaging with the analysis of gap size distribution. Can J Remote Sens 49(5): 471–479. https://doi.org/10.1139/cjfr-2018-0213

Cohen WB, Maiersperger TK, Gower ST, Turner DP (2003) An improved strategy for regression of biophysical variables and Landsat ETM+ data. Remote Sens Environ 84(4): 561–571. https://doi.org/10.1016/S0034-4257(02)00173-6

Díaz GM, Lencinas JD (2018) Model-based local thresholding for canopy hemispherical photography. Can J Remote Sens 48(10): 1204–1216. https://doi.org/10.1139/cjfr-2018-0006

Dubayah R, Blair JB, Goetz S, Fatoyinbo L, Hansen M, Healey S, Hofton M, Hurtt G, Kellner J, Luthcke S, Armston J, Tang H, Duncanson L, Hancock S, Jantz P, Marselis S, Patterson PL, Qi W, Silva C (2020) The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Sci Remote Sens 1: 100002. https://doi.org/10.1016/j.srs.2020.100002

Fang H (2021) Canopy clumping index (CI): A review of methods, characteristics, and applications. Agr Forest Meteorol 303: 108374. https://doi.org/10.1016/j.agrformet.2021.108374

Fang H, Baret F, Plummer S, Schaepman-Strub G (2019) An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. Rev Geophys 57(3): 739–799. https://doi.org/10.1029/2018RG000608

Fernandes R, Plummer S, Nightingale J, Baret F, Camacho F, Fang H, Garrigues S, Gobron N, Lang M, Lacaze R, LeBlanc S, Meroni M, Martinez B, Nilson T, Pinty B, Pisek J, Sonnentag O, Verger A, Welles J, … Widlowski JL (2014) *Global Leaf Area Index Product Validation Good Practices. Version 2.0.* In: Schaepman-Strub G, Román M, Nickeson J (eds.) Good Practices for Satellite-Derived Land Product Validation Land Product Validation Subgroup (WGCV/CEOS), pp. 76

GCOS (2022) The 2022 GCOS Implementation Plan. https://gcos.wmo.int/site/global-climate-observing-system-gcos

Gower ST, Kucharik CJ, Norman JM (1999) Direct and Indirect Estimation of Leaf Area Index, fAPAR, and Net Primary Production of Terrestrial Ecosystems. Remote Sens Environ 70(1): 29–51. https://doi.org/10.1016/S0034-4257(99)00056-5

Hardwick SR, Toumi R, Pfeifer M, Turner EC, Nilus R, Ewers RM (2015) The relationship between leaf area index and microclimate in tropical forest and oil palm plantation: Forest disturbance drives changes in microclimate. Agr Forest Meteorol 201: 187–195. https://doi.org/10.1016/j.agrformet.2014.11.010

Härkönen S, Tokola T, Packalén P, Korhonen L, Mäkelä A (2013) Predicting forest growth based on airborne light detection and ranging data, climate data, and a simplified process-based model. Can J Forest Res 43(4): 364–375. https://doi.org/10.1139/cjfr-2012-0295

Heiskanen J, Korhonen L, Hietanen J, Pellikka PKE (2015) Use of airborne lidar for estimating canopy gap fraction and leaf area index of tropical montane forests. Int J Remote Sens 36(10): 2569–2583. https://doi.org/10.1080/01431161.2015.1041177

Heiskanen J, Rautiainen M, Stenberg P, Mõttus M, Vesanto V-H, Korhonen L, Majasalmi T (2012) Seasonal variation in MODIS LAI for a boreal forest area in Finland. Remote Sens Environ 126: 104–115. https://doi.org/10.1016/j.rse.2012.08.001

Hopkinson C (2007) The influence of flying altitude, beam divergence, and pulse repetition frequency on laser pulse return intensity and canopy frequency distribution. Can J Remote Sens 33(4): 312–324. https://doi.org/10.5589/m07-029

Hopkinson C, Chasmer L (2009) Testing LiDAR models of fractional cover across multiple forest ecozones. Remote Sens Environ 113(1): 275–288. https://doi.org/10.1016/j.rse.2008.09.012

Houborg R, Anderson M, Daughtry C (2009) Utility of an image-based canopy reflectance modeling tool for remote estimation of LAI and leaf chlorophyll content at the field scale. Remote Sens Environ 113(1): 259–274. https://doi.org/10.1016/j.rse.2008.09.014

Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ 83(1–2): 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Kobayashi H, Suzuki R, Kobayashi S (2007) Reflectance seasonality and its relation to the canopy leaf area index in an eastern Siberian larch forest: Multi-satellite data and radiative transfer analyses. Remote Sens Environ 106(2): 238–252. https://doi.org/10.1016/j.rse.2006.08.011

Korhonen L, Heikkinen J (2009) Automated Analysis of in Situ Canopy Images for the Estimation of Forest Canopy Cover. Forest Sci 55(4): 323–334. https://doi.org/10.1093/forestscience/55.4.323

Korhonen L, Korpela I, Heiskanen J, Maltamo M (2011) Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. Remote Sens Environ 115(4): 1065–1080. https://doi.org/10.1016/j.rse.2010.12.011

Korhonen L, Morsdorf F (2014) Estimation of Canopy Cover, Gap Fraction and Leaf Area Index with Airborne Laser Scanning. In: Maltamo, M., Næsset, E., Vauhkonen, J. (eds) Forestry Applications of Airborne Laser Scanning. Managing Forest Ecosystems, vol 27. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-8663-8_20

Kotivuori E, Maltamo M, Korhonen L, Packalen P (2018) Calibration of nationwide airborne laser scanning based stem volume models. Remote Sens Environ 210: 179–192. https://doi.org/10.1016/j.rse.2018.02.069

Lang A, Xiang Y (1986) Estimation of leaf area index from transmission of direct sunlight in discontinuous canopies. Agr Forest Meteorol 37(3): 229–243. https://doi.org/10.1016/0168-1923(86)90033-X

Lang M, Kuusk A, Mõttus M, Rautiainen M, Nilson T (2010) Canopy gap fraction estimation from digital hemispherical images using sky radiance models and a linear conversion method. Agr Forest Meteorol 150(1): 20–29. https://doi.org/10.1016/j.agrformet.2009.08.001

Lang M, Nilson T, Kuusk A, Pisek J, Korhonen L, Uri V (2017) Digital photography for tracking the phenology of an evergreen conifer stand. Agr Forest Meteorol 246: 15–21. https://doi.org/10.1016/j.agrformet.2017.05.021

le Maire G, François C, Dufrêne E (2004) Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. Remote Sens Environ 89(1): 1–28. https://doi.org/10.1016/j.rse.2003.09.004

Leblanc SG, Chen JM, Fernandes R, Deering DW, Conley A (2005) Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. Agr Forest Meteorol 129(3–4): 187–207. https://doi.org/10.1016/j.agrformet.2004.09.006

Levin SA (2013) Encyclopedia of biodiversity. Academic Press.

Li Y, Guo Q, Su Y, Tao S, Zhao K, Xu G (2017) Retrieving the gap fraction, element clumping index, and leaf area index of individual trees using single-scan data from a terrestrial laser scanner. ISPRS J Photogramm 130: 308–316. https://doi.org/10.1016/j.isprsjprs.2017.06.006

LI-COR Environmental (2025) LAI-2200C Plant Canopy Analyzer. https://www.Licor.Com/Products/Leaf-Area/LAI-2200C.

Lim K, Treitz P, Baldwin K, Morrison I, Green J (2003) Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. Can J Remote Sens 29(5): 658–678. https://doi.org/10.5589/m03-025

Lovell JL, Jupp DLB, Culvenor DS, Coops NC (2003) Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. Can J Remote Sens 29(5): 607–622. https://doi.org/10.5589/m03-026

Luo S, Chen JM, Wang C, Gonsamo A, Xi X, Lin Y, Qian M, Peng D, Nie S, Qin H (2018) Comparative Performances of Airborne LiDAR Height and Intensity Data for Leaf Area Index Estimation. IEEE J Sel Top Appl 11(1): 300–310. https://doi.org/10.1109/JSTARS.2017.2765890

Macfarlane C, Hoffman M, Eamus D, Kerp N, Higginson S, McMurtrie R, Adams M (2007) Estimation of leaf area index in eucalypt forest using digital photography. Agr Forest Meteorol 143(3–4): 176–188. https://doi.org/10.1016/j.agrformet.2006.10.013

Mehtätalo L, Lappi J (2020) Forest biometrics with examples in R. Chapman & Hall/CRC.

Miller J (1967) A formula for average foliage density. Aust J Bot 15(1): 141. https://doi.org/10.1071/BT9670141

Morsdorf F, Kötz B, Meier E, Itten KI, Allgöwer B (2006) Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. Remote Sens Environ 104(1): 50–61. https://doi.org/10.1016/j.rse.2006.04.019

Næsset E (2002) Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens Environ 80(1): 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5

Næsset E (2009) 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 113(1): 148–159. https://doi.org/10.1016/j.rse.2008.09.001

Næsset E, Gobakken T, Holmgren J, Hyyppä H, Hyyppä J, Maltamo M, Nilsson M, Olsson H, Persson Å, Söderman U (2004) Laser scanning of forest resources: the nordic experience. Scand J Forest Res 19(6): 482–499. https://doi.org/10.1080/02827580410019553

Nobis M, Hunziker U (2005) Automatic thresholding for hemispherical canopy-photographs based on edge detection. Agr Forest Meteorol 128(3–4): 243–250. https://doi.org/10.1016/j.agrformet.2004.10.002

Norman JM, Jarvis PG (1974) Photosynthesis in Sitka Spruce (Picea sitchensis (Bong.) Carr.). III. Measurements of Canopy Structure and Interception of Radiation. J Appl Ecol 11(1): 375. https://doi.org/10.2307/2402028

Ørka HO, Næsset E, Bollandsås OM (2010) Effects of different sensors and leaf-on and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. Remote Sens Environ *114*(7): 1445–1461. https://doi.org/10.1016/j.rse.2010.01.024

Pierce LL, Running SW, Walker J (1994) Regional-Scale Relationships of Leaf Area Index to Specific Leaf Area and Leaf Nitrogen Content. Ecol Appl 4(2): 313–321. https://doi.org/10.2307/1941936

Riaño D, Valladares F, Condés S, Chuvieco E (2004) Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. Agr Forest Meteorol 124(3–4): 269–275. https://doi.org/10.1016/j.agrformet.2004.02.005

Richardson JJ, Moskal LM, Kim, S-H (2009) Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR. Agr Forest Meteorol 149(6–7): 1152–1160. https://doi.org/10.1016/j.agrformet.2009.02.007

Ross J (1981) The radiation regime and architecture of plant stands. Dr. W. Junk Publishers.

Ryu Y, Baldocchi DD, Kobayashi H, van Ingen C, Li J, Black TA, Beringer J, van Gorsel E, Knohl A, Law BE, Roupsard O (2011) 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 Cy 25(4). https://doi.org/10.1029/2011GB004053

Skidmore AK, Pettorelli N, Coops NC, Geller GN, Hansen M, Lucas R, Mücher CA, O'Connor B, Paganini M, Pereira HM, Schaepman ME, Turner W, Wang T, Wegmann M (2015) Environmental science: Agree on biodiversity metrics to track from space. Nature 523(7561): 403–405. https://doi.org/10.1038/523403a

Solberg S (2010) Mapping gap fraction, LAI and defoliation using various ALS penetration variables. Int J Remote Sens 31(5): 1227–1244. https://doi.org/10.1080/01431160903380672

Solberg S, Brunner A, Hanssen KH, Lange H, Næsset E, Rautiainen M, Stenberg P (2009) Mapping LAI in a Norway spruce forest using airborne laser scanning. Remote Sens Environ 113(11): 2317–2327. https://doi.org/10.1016/j.rse.2009.06.010

Stenberg P (1996) Correcting LAI-2000 estimates for the clumping of needles in shoots of conifers. Agr Forest Meteorol 79(1–2): 1–8. https://doi.org/10.1016/0168-1923(95)02274-0

Stenberg P, Nilson T, Smolander H, Voipio P (2003) Gap fraction based estimation of LAI in Scots pine stands subjected to experimental removal of branches and stems. Can J Remote Sens 29(3): 363–370. https://doi.org/10.5589/m03-007

Tian X, Jia X, Da Y, Liu J, Ge W (2025) Evaluating the sensitivity of vegetation indices to leaf area index variability at individual tree level using multispectral drone acquisitions. Agr Forest Meteorol 364: 110441. https://doi.org/10.1016/j.agrformet.2025.110441

Vaughn NR, Asner GP, Giardina CP (2013) Polar grid fraction as an estimator of montane tropical forest canopy structure using airborne lidar. Int J Remote Sens 34(21): 7464–7473. https://doi.org/10.1080/2150704X.2013.820003

Wang Q, Adiku S, Tenhunen J, Granier A (2005) On the relationship of NDVI with leaf area index in a deciduous forest site. Remote Sens Environ 94(2): 244–255. https://doi.org/10.1016/j.rse.2004.10.006

Wang R, Chen JM, Pavlic G, Arain A (2016) Improving winter leaf area index estimation in coniferous forests and its significance in estimating the land surface albedo. ISPRS J Photogramm 119: 32–48. https://doi.org/10.1016/j.isprsjprs.2016.05.003

Weiss M, Baret F, Smith GJ, Jonckheere I, Coppin P (2004) Review of methods for in situ leaf area index (LAI) determination. Agr Forest Meteorol 121(1–2): 37–53. https://doi.org/10.1016/j.agrformet.2003.08.001

Wilson J (1963) Estimation of foliage denseness and foliage angle by inclined point quadrats. Aust J Bot 11(1): 95. https://doi.org/10.1071/BT9630095

Yan G, Hu R, Luo J, Weiss M, Jiang H, Mu X, Xie D, Zhang W (2019) Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. Agr Forest Meteorol 265: 390–411. https://doi.org/10.1016/j.agrformet.2018.11.033

Zhao K, Popescu S, Meng X, Pang Y, Agca M (2011) Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sens Environ 115(8): 1978–1996. https://doi.org/10.1016/j.rse.2011.04.001

Zhao K, Ryu Y, Hu T, Garcia M, Li Y, Liu Z, Londo A, Wang C (2019) How to better estimate leaf area index and leaf angle distribution from digital hemispherical photography? Switching to a binary nonlinear regression paradigm. Methods Ecol and Evol 10(11): 1864–1874. https://doi.org/10.1111/2041-210X.13273

Zhu X, Skidmore AK, Wang T, Liu J, Darvishzadeh R, Shi Y, Premier J, Heurich M (2018) Improving leaf area index (LAI) estimation by correcting for clumping and woody effects using terrestrial laser scanning. Agr Forest Meteorol 263: 276–286. https://doi.org/10.1016/j.agrformet.2018.08.026