https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Open Access Earth System Science Data Discussions

Multi-temporal high-resolution data products of ecosystem structure derived from country-wide airborne laser scanning surveys of the Netherlands

Yifang Shi* & W. Daniel Kissling

University of Amsterdam, Institute for Biodiversity and Ecosystem Dynamics (IBED), P.O. Box 94240, 1090 GE Amsterdam, The Netherlands

Correspondence to: Yifang Shi (y.shi@uva.nl)

Short summary

We present a new set of multi-temporal LiDAR metrics of ecosystem structure derived from four national ALS surveys of the Netherlands (AHN1–AHN4), capturing vegetation height, cover, and structural variability over the past two decades (1998–2022). Around 70 TB point clouds have been processed to read-to-use raster layers at 10 m resolution (~ 59 GB), enabling a wide use and uptake of ecosystem structure information in biodiversity and habitat monitoring, ecosystem and carbon dynamic modelling.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Abstract

Recent years have seen a rapid surge in the use of Light Detection and Ranging (LiDAR) technology for characterizing the structure of ecosystems. Even though repeated airborne laser scanning (ALS) surveys are increasingly available across several European countries, only few studies have so far derived data products of ecosystem structure at a national scale, possibly due to a lack of free and open-source tools and the computational challenges involved in handling the large volumes of data. Nevertheless, high-resolution data products of ecosystem structure generated from multi-temporal country-wide ALS datasets are urgently needed if we are to integrate such information into biodiversity and ecosystem science. By employing a recently developed, open source, high-throughput workflow (named "Laserfarm"), we processed around 70 TB of raw point clouds collected from four national ALS surveys of the Netherlands (AHN1–AHN4, 1996–2022). This resulted in ~ 59 GB raster layers in GeoTIFF format as ready-to-use multi-temporal data products of ecosystem structure at a national extent. For each AHN dataset, we generated 25 LiDAR-derived vegetation metrics at 10 m spatial resolution, representing vegetation height, vegetation cover, and vegetation structural variability. The data enable an in-depth understanding of ecosystem structure at fine resolution across the Netherlands and provide opportunities for exploring ecosystem structural dynamics over time. To illustrate the utility of these data products, we present ecological use cases that monitor forest structural change and analyse vegetation structure differences across various Natura 2000 habitat types, including dunes, marshes, grasslands, shrublands, and woodlands. The provided data products and the employed workflow can facilitate a wide use and uptake of ecosystem structure information in biodiversity and carbon modelling, conservation science, and ecosystem management. The full data products are publicly available on Zenodo (https://doi.org/10.5281/zenodo.13940846) (Shi and Kissling 2024).

1 Introduction

Monitoring ecosystem structure is essential for sustainable forest management (Lindenmayer et al., 2000), species distribution research (Jetz et al., 2019; Kissling et al., 2018), dynamic ecosystem modelling (Kucharik et al., 2000), biodiversity monitoring (Noss, 1990), and the conservation and restoration of terrestrial ecosystems (Ruiz-Jaén and Aide, 2005). As one of the Essential Biodiversity Variables (EBVs) classes (Pereira et al., 2013), ecosystem structure provides detailed insights into both the vertical and horizontal profiles of ecosystems, facilitating a deeper understanding of the relationship between vegetation structure and animal ecology (Davies and Asner, 2014) as well as carbon and biomass dynamics (Zhao et al., 2018; Dalponte et al., 2019). However, until a decade ago, the collection of vegetation structure data was difficult and labour intensive, especially over large spatial scales (Davies and Asner, 2014). Although previous studies have explored the use of passive remote sensing technologies, such as high-resolution satellite imagery and aerial photographs, alongside field

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

measurements to obtain structural information (e.g. Wolter et al., 2009; Lamonaca et al., 2008), these applications have largely been confined to plot or local scales with limited scalability and uncertain transferability between different regions.

Over the past few decades, the advent of airborne laser scanning has enabled the direct measurement of ecosystem structural properties such as high-resolution topographic variation and accurate estimation of vegetation height, cover, and canopy structure (Lefsky et al., 2002). The LiDAR technology used in ALS surveys generates discrete returns (point clouds) and/or full-waveform signals by emitting laser pulses from the sensor towards the target objects (e.g. ground, trees, and buildings, etc), recording the distance between the sensor and the objects ("X", "Y", "Z" coordinates), the amount of energy returned to the sensor ("Intensity"), the type of the object ("Classification"), the sequence of returns generated from one pulse ("Return number" and "Number of returns"), the time of the pulse emitted ("GPS time"), and so on. Advances in sensor systems and techniques also allow many countries to carry out ALS campaigns over national or regional extents, producing fine-scale ecosystem measurements across broad spatial extents (Kissling et al., 2022; Assmann et al., 2022). ALS surveys often generate massive amounts of data (e.g. point clouds with a multi-terabyte data volume) which contain ecosystem structural information that is essential for ecological and biodiversity research (Kissling et al., 2022; Koma et al., 2021b; Bakx et al., 2019). Although tools and software for processing large amounts of LiDAR data are increasingly available (Roussel et al., 2020; Isenburg, 2017; Meijer et al., 2020; Kissling et al., 2022), significant challenges remain, including the need for specialist expertise, extensive data storage, and substantial computational power (Assmann et al., 2022). Ultimately, ecologists, foresters, biodiversity researchers and land managers require raster layers with structural information that can be readily integrated into analytical workflows using software that they are familiar with (e.g. GIS, R, Python). Such raster layers, e.g. LiDAR-derived vegetation metrics, are often generated by statistically aggregating the 3D point cloud information within spatial units such as voxels or 2D raster cells (Meijer et al., 2020; Kissling et al., 2022). These LiDAR-derived vegetation metrics typically capture three key dimensions of ecosystem structure: vegetation height (e.g. maximum vegetation height, vegetation height at a certain percentile), vegetation cover (e.g. the density of vegetation at a given height layer), and vegetation structural variability (e.g. the vertical or horizontal distribution and variability of vegetation within a spatial unit) (Kissling et al., 2023; Bakx et al., 2019). Providing high-resolution (~ 10 m) ready-to-use LiDAR metrics and making them accessible for the public is, therefore, critical for monitoring Essential Biodiversity Variables (EBVs) (Valbuena et al., 2020), modelling species distributions (De Vries et al., 2021; Koma et al., 2021b; Zellweger et al., 2013), and estimating species diversity (Moeslund et al., 2019; Zellweger et al., 2017; Aguirre-Gutiérrez et al., 2017) at a regional or national scale.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Open Access Earth System Science Data Discussions

Ecosystem structure is a three-dimensional phenomenon with horizontal and vertical components that change over time (Zenner and Hibbs, 2000). The increasing frequency of ALS data acquisition offers a unique opportunity to monitor ecological changes and ecosystem dynamics at fine spatial and temporal scales. Several countries have been conducting repeated (sub-)national ALS surveys to obtain fine-scale information on topography and forest ecosystems (Nilsson et al., 2017). For example, the Dutch national ALS programme (AHN, *Actueel Hoogtebestand Nederland*, https://www.ahn.nl/) has been collecting country-wide LiDAR data since 1996, providing four complete ALS datasets (AHN1–AHN4) with an ongoing fifth survey (AHN5), conducted at intervals of 3 to 5 years. In Spain, under the PNOA-LiDAR project, two national ALS campaigns have taken place during 2008–2015 (LiDAR 1st coverage) and during 2015–2021 (LiDAR 2nd coverage), while the third acquisition (LiDAR 3rd coverage) has started in 2023 and is planned to finish in 2025 (http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do?Serie=LIDAR, last access: 19 October 2024). While the primary goal of many ALS campaigns is to produce terrain models, such as Digital Terrain Models (DTMs) or Digital Surface Models (DSMs), the multi-temporal LiDAR datasets also capture detailed 3D characteristics on vegetation structure over time, providing valuable information for evaluating changes in biomass (Cao et al., 2016; Feng et al., 2024), forest structure (Mccarley et al., 2017; Riofrío et al., 2022; Vepakomma et al., 2011), and forest carbon stocks (Dalponte et al., 2019; Zhao et al., 2018). Furthermore, these datasets are increasingly being integrated with other remote sensing data, such as satellite imageries from Landsat, Sentinel-2, and synthetic aperture radar (SAR), to assess forest changes caused by disturbances like wildfires (Li et al., 2023; Feng et al., 2024) and to model aboveground biomass (Musthafa and Singh, 2022). However, despite the growing availability of multi-temporal ALS datasets, there is a noticeable lack of publicly available data products, i.e. LiDAR-derived vegetation metrics, from national ALS surveys.

Several challenges are posed in generating accurate and standardized data products from multi-temporal ALS data (Valbuena et al., 2020). Over the past three decades, advances in LiDAR sensors and associated technologies have led to improvements in point density, classification accuracy, and additional attributes provided in each point (Riofrío et al., 2022). However, these advancements also introduce complexities in data harmonization. In addition to the challenges associated with processing large datasets and high computational costs (Meijer et al., 2020), discrepancies in sensor technology and flight configurations across different ALS surveys can hinder the generation of consistent data products (Lin et al., 2022). For instance, the first Dutch national ALS campaign (AHN1, 1996–2003) had an average point density ranging from 1 point per 16 square meters to 1 point per square meter, with no detailed point classification available. By contrast, in the fourth campaign (AHN4, 2020–2022), the point density has improved to 20–30 points per square meter, with detailed classification code provided for each point following the ASPRS standard (Asprs, 2019). These technological variations inevitably result in data products with varying quality and accuracy, introducing uncertainties in their usability (Tompalski et al.,

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

2021; Hopkinson et al., 2008). To understand ecosystem dynamics accurately, changes detected from multi-temporal ALS datasets should reflect actual ecological changes in the target of interest rather than differences in data acquisition or quality (Riofrío et al., 2022). Identifying the limitations and providing usage notes of derived data products are important for users to interpret the data products correctly and apply them optimally in their analyses.

Here, we present a new set of multi-temporal data products of ecosystem structure derived from four national ALS surveys of the Netherlands (AHN1–AHN4). The data products, with a spatial resolution of 10 m, include four sets of 25 LiDAR-derived vegetation metrics representing ecosystem height, vegetation cover, and structural variability, aimed at supporting a wide range of ecological applications. In this paper, we (1) describe the ALS data collection from AHN1–AHN4 and the employed "Laserfarm" workflow to generate the data products, (2) present the detailed characteristics of the generated multi-temporal data products (i.e. LiDAR-derived vegetation metrics as GeoTIFF raster layers) and their known limitations and corresponding usage notes, (3) demonstrate two use cases for using the generated data products in ecological applications, and (4) discuss the potential use and recommendations for utilizing these data products in future research. To facilitate open science, we make the data products, employed workflow, Python script, and related documentation publicly available. We anticipate that this will not only allow the upscaling of ecological and biodiversity research but also benefit a broad range of scientists and decision-makers who are interested in using ecosystem structure information for environmental monitoring and management.

2 Raw data and processing workflow

2.1 Geography and ecology of the Netherlands

The Netherlands is situated in Northwest Europe (52°22′N, 4°53′E), covering a total land area of 33893 km^2. It has mostly flat coastal lowlands and reclaimed land (polders) with an average elevation of approximately 30 meters above sea level. The primary ecosystems in the Netherlands include agricultural land, dunes and beaches, forests, wetlands, grasslands, and other (semi)natural environments (Hein et al., 2020). The Netherlands has a temperate maritime climate with continental influence, resulting in an average annual precipitation of 854.7 mm and a mean temperature of 10.5 °C.

2.2 Four Dutch national ALS campaigns

The initial purpose of the AHN programme was to monitor and manage water systems in the Netherlands. It is a collaboration between 26 regional water boards, provinces and Rijkswaterstaat (the executive directorate general for public works and water management of the Dutch government) with the aim of producing accurate digital elevation models of the Netherlands. To minimize the impact of foliage on ground detection during the laser scanning, the AHN data acquisition is performed in the winter period,

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

from December to April. The first generation of AHN (AHN1) was conducted during 1996–2003, with a point density of 1 point per 1–16 square meters, which largely depended on the viability of the technology and the date of acquisition (Swart, 2010). Due to errors in the AHN1 data (e.g. inaccuracies in the inertial navigation system, misalignment of overlapping scanning strips, and the presence of artifacts), the data quality of AHN1 is rather poor, especially for areas covered by vegetation (Brand et al., 2003). To support both water and dike management, the second generation of AHN (AHN2) was started in 2007, with improved specifications such as a higher point density (on average 6–10 pts m^{-2}) and a higher planimetric/vertical accuracy (5–15 cm). It also required some raster data (i.e. DTMs and DSMs) to be delivered with grid cell sizes of 0.5 m and 5 m. With the main aim of obtaining terrain surface information, both AHN1 and AHN2 datasets were delivered in two separate parts: point clouds representing the terrain ("gefilterde puntenwolk") and point clouds representing non-ground points, i.e. trees, buildings, bridges and other objects ("uitgefilterde puntenwolk").

Benefitting from the advances in LiDAR sensors and related technologies, the third generation of AHN (AHN3) provided not only a higher density of point clouds, but also more information stored for each point, such as point classification code, intensity values, number of returns, and so on (Table 1). Even though both AHN2 and AHN3 were collected within a 6-year cycle (2007–2012 for AHN2, and 2014–2019 for AHN3), the actual time difference between AHN2 and AHN3 varies between 4–10 years depending on the area of interest (Fig. 1). For the latest completed AHN (AHN4), the survey was conducted between 2020 and 2022 (3-year cycle), making the country-wide dataset more quickly available for the whole Netherlands. All four AHN datasets were provided in LAZ format (i.e. version 1.2 for AHN1–AHN3, and version 1.4 for AHN4), under the local Dutch coordinate system "RD_new" (EPSG: 28992, NAP:5709). The datasets from AHN1 to AHN4 show an increase in data volume and improved classification as well as additional attributes stored for each point (Table 1). An ongoing fifth ALS survey (AHN5) has started in 2023 (the first part of the data is available, see https://www.ahn.nl/heel-westelijk-nederland-gereed) and the data acquisition will be completed in 2025.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

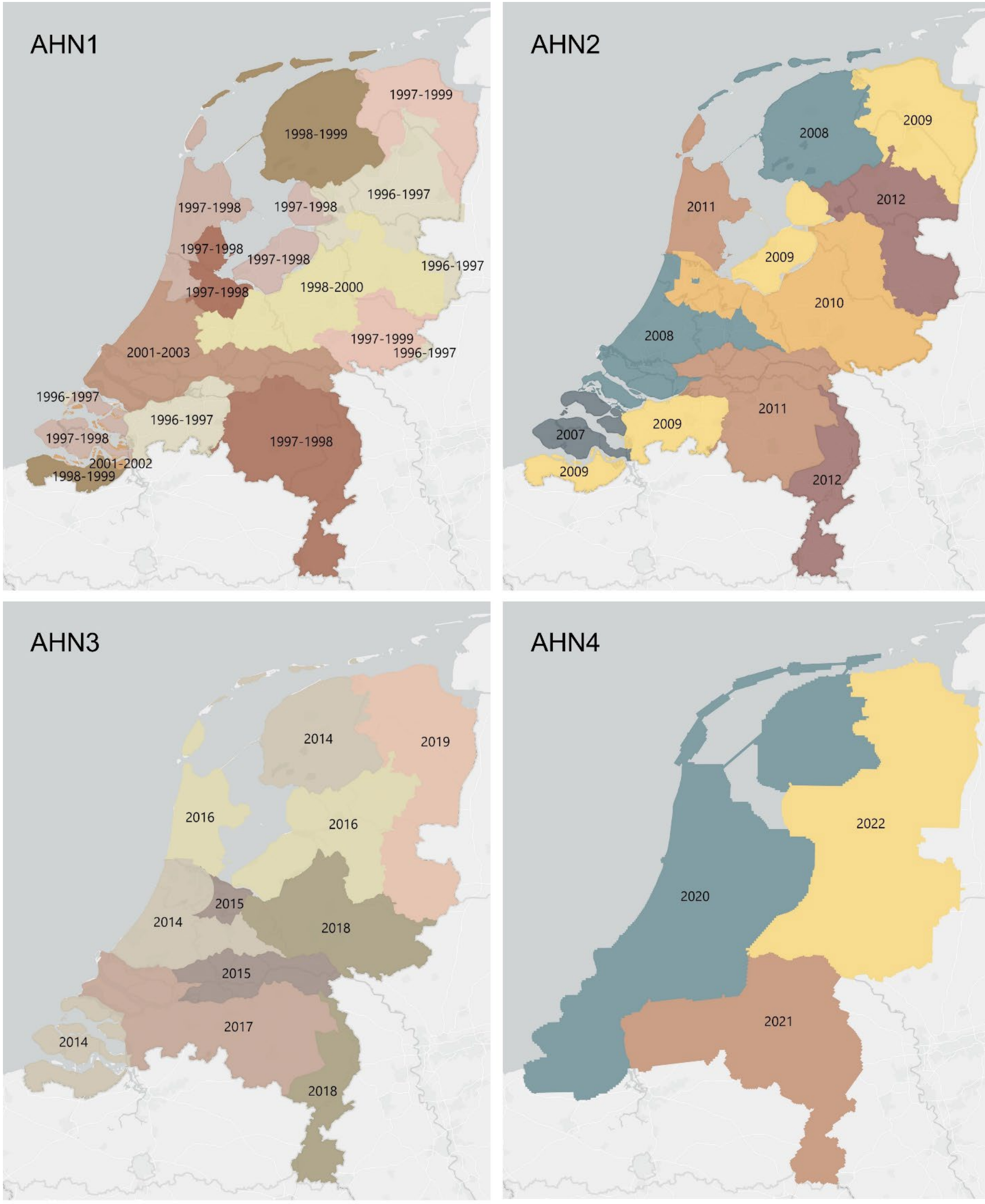


Fig.1 Data acquisition times for AHN1–AHN4. Different colours indicate the different years of data collection for each dataset.

Table 1. Summary of raw point cloud characteristics collected by different AHN surveys (AHN1–AHN4). Some flight configurations are not available, for instance, the type of sensor, the flight height, flight speed, and the scan angle, especially for the AHN1 dataset. NAP: Normal Amsterdam Level.

Data characteristic	AHN1	AHN2	AHN3	AHN4
Acquisition year	1996–2003	2007–2012	2014–2019	2020–2022

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Acquisition season	Leaf-off	Leaf-off	Leaf-off	Leaf-off
Horizontal projection	RD_new	RD_new	RD_new	RD_new
Vertical projection	NAP	NAP	NAP	NAP
Point density (pts m^{-2})	0.05–1	6–15	10–20	20–30
Scan angle (°)	-	± 30	± 35	± 35
Overlapping rate	-	20–35%	20–35%	20–35%
Point cloud format	Laz (1.2)	Laz (1.2)	Laz (1.2)	Laz (1.4)
Vertical accuracy (cm)	5–35	5–15	5–15	5–10
Number of files	2720	60185	1367	1381
Data volume (compressed)	33.1 GB	986.7 GB	2564.8 GB	6408.6GB
Attributes in each point	X, Y, Z	X, Y, Z	X, Y, Z, intensity, return number, number of returns, classification, scan angle, point ID, GPS time	X, Y, Z, intensity, return number, number of returns, classification, scan angle, point ID, GPS time, amplitude, reflectance, deviation
Classification	uitgefilterd (0) gefilterd (0)	uitgefilterd (0) gefilterd (0)	unclassified (1) ground (2) building (6) water (9) reserved (26)	unclassified (1) ground (2) building (6) water (9) powerline (14) reserved (26)
Available additional layers	-	DSM, DTM	DSM, DTM	DSM, DTM

2.3 Processing workflow

We employed the high-throughput workflow "Laserfarm" (https://laserfarm.readthedocs.io/en/latest/) to process the multi-temporal AHN datasets. Laserfarm is an open-source workflow designed for processing large amount of LiDAR point cloud data into geospatial data products of ecosystem structure (Kissling et al., 2022). It consists of four main modules: (1) re-tiling, where the original LAZ files (covering 5 km × 6.5 km per tile) are re-tiled into 1 km × 1 km LAZ files for an efficient, scalable and distributed processing; (2) normalization, where the height (z value) of the lowest point within a 1 m × 1 m grid cell is subtracted from each point in the cell, so that the influence of terrain on the height of above-ground points is removed from subsequent processing; (3) feature extraction, where user-defined features (e.g. LiDAR metrics such as the 95th percentile of vegetation height and the skewness of vegetation height) are calculated at 10

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Open Access Earth System Science Data Discussions

meter resolution using points within an infinite square cell (Meijer et al., 2020); and (4) rasterization, where the extracted feature files (.PLY files) are merged and exported as single-band GeoTIFF raster files. Note that in all four AHN datasets, vegetation points are not classified separately based on the ASPRS standard. Instead, they are assigned a classification value 0 (“uitgefilterd”) in AHN1 and AHN2, and a value 1 (“unclassified”) in AHN3 and AHN4. These classification values were used as vegetation class during the feature extraction. We chose the Laserfarm workflow to process the four country-wide AHN datasets because (1) it enables the efficient, scalable and distributed processing of multi-terabyte LiDAR point clouds at a national scale, (2) it is a free and open-source tool implemented in Python and available as Jupyter Notebooks, and (3) it allows the automated generation of consistent and reproducible geospatial data products of ecosystems structure from different ALS data.

Due to the different characteristics of each AHN dataset (Table 1), several pre-processing steps were implemented before executing the main modules of the Laserfarm workflow (Fig. 2). In particular, for the AHN1 and AHN2 datasets, the step “Reclassification” was carried out before re-tiling, as both datasets only have “gefilterd” (ground) and “uitgefilterd” (non-ground) files provided and the raw classification value was set to 0 (never classified) for all points. We therefore reassigned a classification value “2” to the ground points (“gefilterd”) and a classification value “0” to the non-ground points (“uitgefilterd”). These classification values were later used for the normalization and feature extraction. For the AHN4 dataset, the volume of a single original LAZ file varies from 0.3 MB to 16.5 GB, with an average size of 4.6 GB per file (Table 2). Since handling such volumes is challenging for many computing infrastructures (due to their CPUs and random-access memory, RAM), we applied a “Splitting” step before the re-tiling (Fig. 2), with a maximum data volume of ~ 500 MB being used for splitting the original tiles into smaller ones.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Open Access Earth System Science Data Discussions

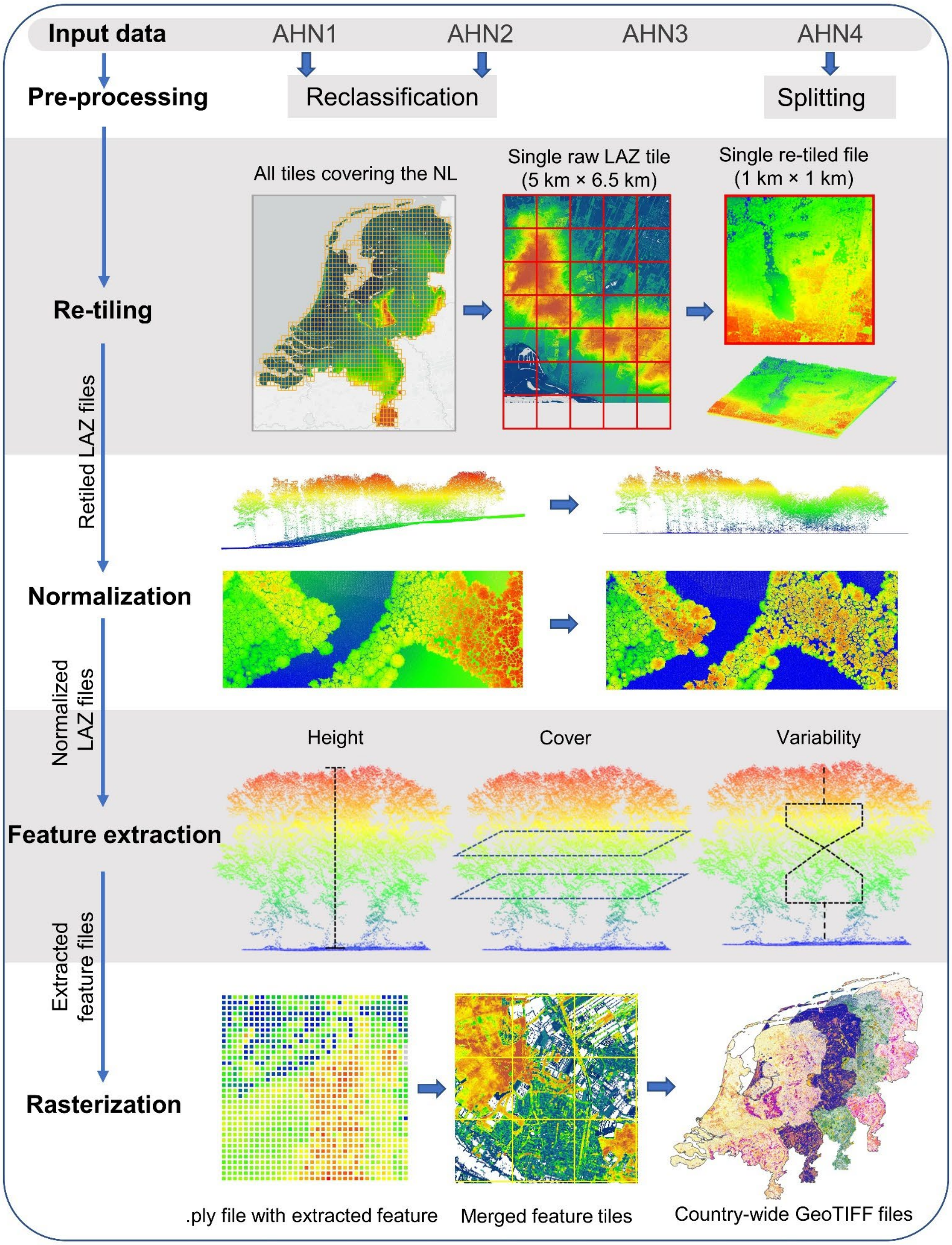


Fig. 2 Overview of the processing workflow employed for four country-wide AHN datasets of the Netherlands (AHN1–AHN4). The pre-processing step "reclassification" was only conducted for the AHN1 and AHN2 datasets, where ground points were reassigned a classification value "2". The "splitting" step was added to split the large LAZ files from AHN4 into smaller ones before re-tiling. Re-tiling, normalization, feature extraction and rasterization are four main modules of the Laserfarm

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

workflow, which have been applied for all four AHN datasets to generate country-wide LiDAR-derived vegetation metrics. The input data were raw LAZ files with different point density, and the output data were 25 single-band GeoTIFF raster layers at 10 meter resolution for each AHN dataset.

2.4 IT infrastructure and computational cost

All four AHN datasets were processed on the IT infrastructure services provide by SURF, the Dutch national facility for information and communication technology (https://www.surf.nl/). Specifically, we used the dCache platform for data storage (https://www.surf.nl/en/services/dcache) and the HPC Cloud (https://www.surf.nl/en/services/hpc-cloud) or Spider platform (https://www.surf.nl/en/services/high-performance-data-processing) for high-performance data processing. The data processing platforms have fast access to the data storage while enabling scalable and flexible processing of multi-terabytes datasets on distributed resources. We first downloaded the raw AHN1–AHN4 LiDAR point clouds from the PDOK webservices (https://www.pdok.nl/introductie/-/article/actueel-hoogtebestand-nederland-ahn) to the dCache data storage using a customized python script (https://github.com/ShiYifang/AHN/tree/main/AHN_downloading). We then ran the Laserfarm workflow for processing the AHN1–AHN3 datasets on the HPC Cloud, where we set up a cluster of 11 VMs, each VM with 2 cores, 32 GB or 64 GB RAM, and 256 GB local HDD. Due to migration of the computing resources by SURF (from HPC Cloud to Spider), we processed the AHN4 dataset with the Laserfarm workflow on Spider, where a number of flexible and customisable workers with assigned CPU cores were defined based on the computing requirement for each workflow step. We used 2–10 workers, each with 2–4 cores and 16–32 GB RAM for splitting, re-tiling, normalization, and feature extraction, and 2 workers, each with 12 cores and 94 GB RAM for the rasterization step. All input data (i.e. raw LAZ files), intermediate results (e.g. re-tiled LAZ files, normalized LAZ files, featured PLY tiles), and final output (i.e. GeoTIFF raster layers) were automatically stored (and/or retrieved for the next step) on the dCache data storage.

The computing time for each AHN dataset varies based on the input data volume, the required processing steps (Table 2), and the settings of the employed infrastructure. The increase in data volumes from AHN1 to AHN4 resulted in a strong increase of the processing time (Table 2). In total, it required 57.6 days (wall-time) to process the multi-temporal AHN datasets (AHN1–AHN4). The AHN1 (data volume of 33.1 GB) only took a wall-time of 4.8 days to complete whereas the AHN4 (data volume of 6408.6 GB) took a total wall-time of 26.8 days. It is worth noting that the actual computing time of the process might be longer than the wall-time estimates, e.g. due to processing errors, worker failures, and system maintenance.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

262 **Table 2.** Overview of the number of input files, the total volume and the average volume per file for each
263 processing step, and the total processing wall-time for each AHN dataset. Note that the total wall-time
264 was estimated based on different infrastructure settings for processing the AHN1–AHN3 (HPC Cloud)
265 and AHN4 (Spider) datasets.

Data characteristic	AHN1	AHN2	AHN3	AHN4
Input for re-tiling	*(Reclassified)*	*(Reclassified)*		*(Splitted)*
Number of input files	2720	60185	1367	18797
Total volume	33.1 GB	986.7 GB	2564.8 GB	6408.6 GB
Average volume per file (mean ± SD)	12.20 ± 10.68 MB	16.40 ± 14.73 MB	1.75 ± 0.93 GB	4.60 ± 2.41 GB
Re-tiling				
Number of re-tiled files	37715	37627	37457	37990
Total volume	33.1 GB	986.7 GB	2564.8 GB	6408.6 GB
Average volume per file (mean ± SD)	0.83 ± 1.64 MB	26.90 ± 35.98 MB	0.07 ± 0.18 GB	0.17 ± 0.09 GB
Normalization				
Number of normalized files	37715	37627	37457	37990
Total volume	64.0 GB	3682.4 GB	6067.5 GB	9593.3 GB
Average volume per file (mean ± SD)	1.70 ± 2.13 MB	97.87 ± 59.23 MB	0.16 ± 0.09 GB	0.25 ± 0.13 GB
Feature extraction				
Number of featured files	37715 × 25	37627 × 25	37457 × 25	37990 × 25
Total volume	257.1 GB	282.5 GB	285.9 GB	212.5 GB
Average volume per file (mean ± SD)	0.29 ± 0.02 MB	0.30 ± 0.03 MB	0.33 ± 0.05 MB	0.23 ± 0.04 MB
Rasterization				
Number of rasterized files	25	25	25	25
Total volume	4.8 GB	19.4 GB	18.8 GB	15.6 GB
Average volume per file (mean ± SD)	202.1 ± 101.6 MB	774.5 ± 303.5 MB	759.8 ± 226.2 MB	625.5 ± 160.7 MB
Processing time				
Total processing wall-time (days)	4.8	11.7	14.3	26.8

266

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

3 Data products description

3.1 Overview of data products

The generated data products from each AHN campaign cover the whole Netherlands, ranging from 50.77 °N to 53.36 °N and from 3.57 °E to 7.11 °E. The data products are provided as 10 meter resolution GeoTIFF raster files (25 single-band raster layers for each AHN dataset) in the local Dutch coordinate system "RD_new" (EPSG: 28992, NAP:5709). The total volume of the four data products is approximately 58.6 GB. The pixel value is stored in 32 bit floating point precision. The data products are freely accessible via a permanent Zenodo repository (see Sect. 7).

3.2 LiDAR-derived vegetation metrics

In total, 25 LiDAR-derived vegetation metrics were generated from each AHN dataset, representing vegetation height, vegetation cover, and vegetation structure variability (Table 3). For vegetation height, we generated 7 LiDAR metrics (i.e. maximum, mean, median, 25th, 50th, 75th, 95th percentile of vegetation height) representing the height of vegetation at the canopy surface and for low, middle, and upper vegetation strata (Fig. 3a). We filtered out the points with a z value higher than 10000 m (outliers) during "Normalization" step of the Laserfarm workflow and used a square infinite cell (10 × 10 m) as the target volume to calculate the height metrics (see detailed description of target volumes in Meijer et al. (2020)). To ensure positive height values after normalization, we generally normalized the vegetation points based on the height of the lowest point within a 1 m × 1 m grid cell. For vegetation cover, we derived 11 LiDAR metrics consisting of one metric describing the openness of vegetation (i.e. pulse penetration ratio), one metric describing the density of upper vegetation layer (i.e. canopy cover), and 9 metrics quantifying vegetation density at different height layers (i.e. below 1 m, between 1–2 m, 2–3 m, 3–4 m, 4–5 m, 5–20 m, above 3 m, below 5 m, and above 20 m) (Fig. 3b). The height layers reflect the most relevant height strata to capture the vegetation distribution of major growth forms (e.g. grass, reed, shrubs and trees) (Morsdorf et al., 2010; Miura and Jones, 2010). Special attention was given to represent low vegetation strata (1–5 m) as they are essential for low-stature terrestrial ecosystems such as grasslands, shrublands or agricultural areas when monitoring animal habitats and species distributions (Koma et al., 2021a; Bakx et al., 2019). Note that the pulse penetration ratio is the only LiDAR metric (among the 25 metrics) that used ground points for the calculation. All other 24 metrics are only calculated with vegetation points (i.e. "unclassified" in AHN). For vegetation structural variability, we derived 7 LiDAR metrics representing the vertical variability of vegetation distribution within a cell (Fig. 3c), including the coefficient of variation, Shannon index, kurtosis, skewness, standard deviation, variance, and roughness (sigma) of vegetation height. The detailed description of how those metrics are calculated and their ecological relevance can be found in Table 3.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Open Access Earth System Science Data Discussions

Table 3. Twenty-five LiDAR-derived vegetation metrics capturing ecosystem structure in three key dimensions (vegetation height, vegetation cover and vegetation structural variability), together with their file names in the data products, the formulas for calculation, their descriptions and example of their ecological relevance. Each LiDAR metric is provided as a single-band GeoTIFF raster layer at 10 meter resolution, with the file name "ahn#_10m_xx", where # is the number of AHN campaign ("1–4") and xx is the name of the LiDAR metrics. For instance, "ahn4_10m_ perc_95_normalized_height" represents the 95th percentile of vegetation height derived from the AHN4 dataset. For the calculation formulas, N is the total number of normalized vegetation points within a cell, z_i represents all normalized z values in a cell, and $\bar{z}$ is the mean normalized z value in a cell.

LiDAR metric (abbreviation)	File name (ahn#_10m_xx)	Calculation formula	Description	Ecological relevance
Vegetation height				
Maximum vegetation height (Hmax)	max_normalized _height	z_{max}	Maximum of normalized z within a cell	Height of canopy surface, tree tops
Mean of vegetation height (Hmean)	mean_ normalized_heig ht	z_{mean}	Mean of normalized z within a cell	Average height of vegetation, mean tree height
Median of vegetation height (Hmedian)	median_ normalized_heig ht	z_{median}	Median of normalized z within a cell	Vegetation height, vertical distribution of vegetation
25th percentiles of vegetation height (Hp25)	perc_25_normali zed_height	$z_{25\ percentile}$	25th percentile of normalized z within a cell	Density of vegetation in the low stratum
50th percentiles of vegetation height (Hp50)	perc_50_normali zed_height	$z_{50\ percentile}$	50th percentile of normalized z within a cell. It corresponds to the Hmedian.	Average height and vertical distribution of vegetation
75th percentiles of vegetation height (Hp75)	perc_75_normali zed_height	$z_{75\ percentile}$	75th percentile of normalized z within a cell	Density of vegetation in the upper stratum
95th percentiles of vegetation height (Hp95)	perc_95_normali zed_height	$z_{95\ percentile}$	95th percentile of normalized z within a cell	Height of the vegetation canopy surface, avoiding the effect of outliers (compared to Hmax)
Vegetation cover				

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Pulse penetration ratio (PPR)	pulse_penetration_ratio	$\frac{N_{ground}}{N_{total}}$	Ratio of number of ground points to total number of points within a cell	Openness of vegetation, canopy fractional cover, laser penetration index
Canopy cover (Density_above_mean_z)	density_absolute_mean_normalized_height	$100 \times \sum [z_i > \bar{z}]/N$	Number of returns above mean height within a cell	Density of upper vegetation layer
Density of vegetation points below 1 m (BR_below_1)	band_ratio_normalized_height_1	$N_{z<1}/N_{total}$	Ratio of number of vegetation points below 1 m to the total number of vegetation points within a cell	Density of vegetation below 1 m
Density of vegetation points between 1–2 m (BR_1_2)	band_ratio_1_normalized_height_2	$N_{1<z<2}/N_{total}$	Ratio of number of vegetation points between 1–2 m to the total number of vegetation points within a cell	Density of vegetation in 1–2 m layer
Density of vegetation points between 2–3 m (BR_2_3)	band_ratio_2_normalized_height_3	$N_{2<z<3}/N_{total}$	Ratio of number of vegetation points between 2–3 m to the total number of vegetation points within a cell	Density of vegetation in 2–3 m layer
Density of vegetation points above 3 m (BR_above_3)	band_ratio_3_normalized_height	$N_{z>3}/N_{total}$	Ratio of number of vegetation points above 3 m to the total number of vegetation points within a cell	Density of vegetation in above 3 m layer
Density of vegetation points between 3–4 m (BR_3_4)	band_ratio_3_normalized_height_4	$N_{3<z<4}/N_{total}$	Ratio of number of vegetation points between 3–4 m to the total number of vegetation points within a cell	Density of vegetation in 3–4 m layer
Density of vegetation points between	band_ratio_4_normalized_height_5	$N_{4<z<5}/N_{total}$	Ratio of number of vegetation points between 4–5 m to the total number of	Density of vegetation in 4–5 m layer

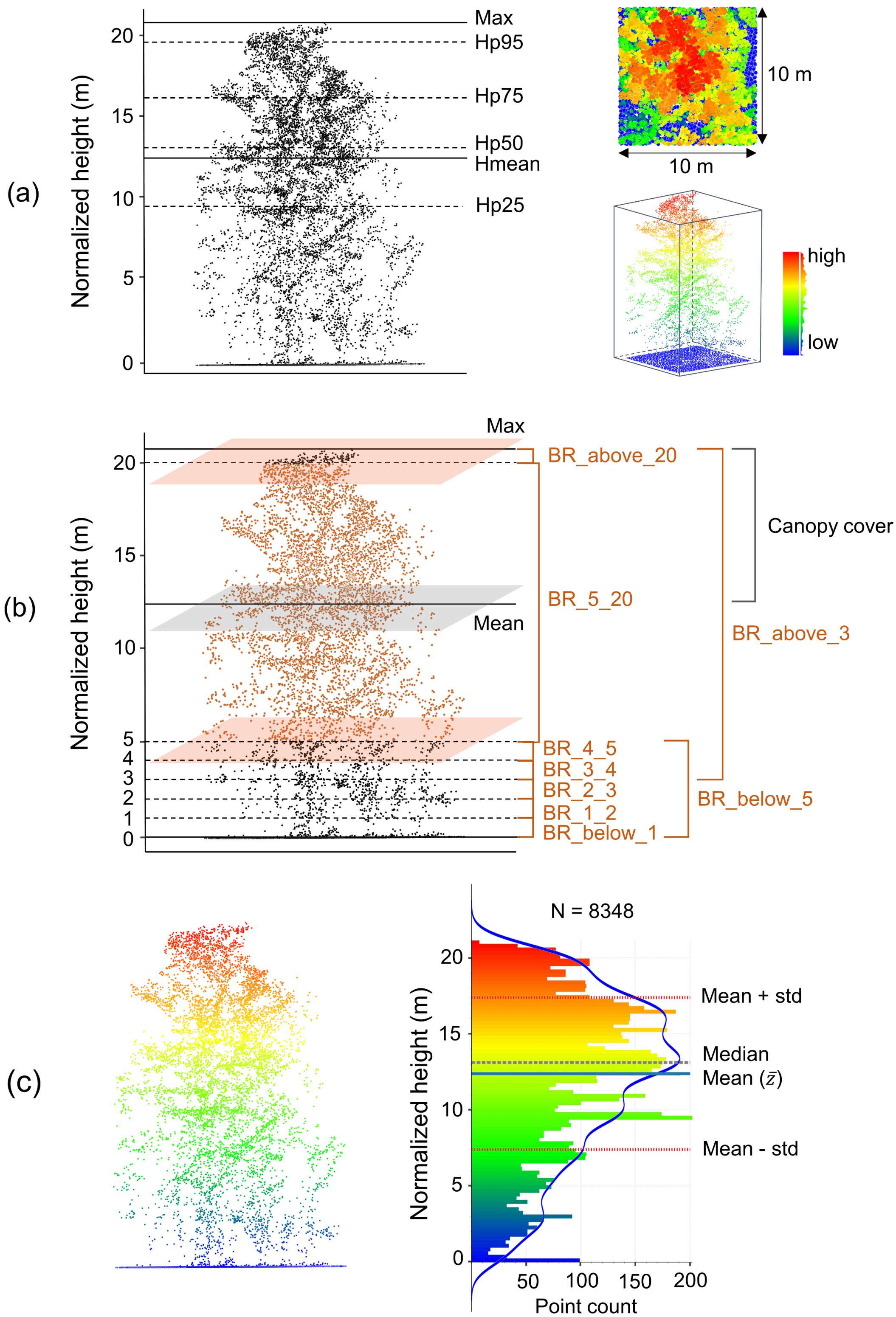
https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

4–5 m (BR_4_5)			vegetation points within a cell	
Density of vegetation points below 5 m (BR_below_5)	band_ratio_normalized_height_5	$N_{z<5}/N_{total}$	Ratio of number of vegetation points below 5 m to the total number of vegetation points within a cell	Density of vegetation below 5 m
Density of vegetation points between 5–20 m (BR_5_20)	band_ratio_5_normalized_height_20	$N_{5<z<20}/N_{total}$	Ratio of number of vegetation points between 5–20 m to the total number of vegetation points within a cell	Density of vegetation in 5–20 m layer
Density of vegetation points above 20 m (BR_above_20)	band_ratio_20_normalized_height	$N_{z>20}/N_{total}$	Ratio of number of vegetation points above 20 m to the total number of vegetation points within a cell	Density of vegetation in above 20 m layer
Vegetation structural variability				
Coefficient of variation of vegetation height (Coeff_var)	coeff_var_normalized_height	$\frac{1}{\bar{z}} \times \sqrt{\sum \frac{(z_i - \bar{z})^2}{N-1}}$	Coefficient of variation of normalized z within a cell	Vertical variability of vegetation distribution
Shannon index (Entropy_z)	entropy_normalized_height	$-\sum_i p_i \times log_2 p_i$ where $p_i = N_i / \sum_j N_j$, and N_i is the points in bin i.	The negative sum of the proportion of points within 0.5 m height layers multiplied with the logarithm of the proportion of points within 0.5 m height layers within a cell	Vertical complexity of vegetation, foliage height diversity
Kurtosis of vegetation height (Hkurt)	kurto_normalized_height	$\frac{1}{\sigma^4} \times \sum (z_i - \bar{z})^4 / N$ where σ is the standard deviation of the z value in a cell.	Kurtosis of normalized z within a cell	Vertical distribution of vegetation

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Roughness of vegetation (Sigma_z)	sigma_z	$\sqrt{\sum(R_i-\bar{R})^2/(N-1)}$ where R_i are the residual after plane fitting, and $\bar{R}$ the mean of residuals.	Standard deviation of the residuals of a locally fitted plane within a cylinder	Small-scale roughness and variability of vegetation
Skewness of vegetation height (Hskew)	skew_ normalized_height	$\frac{1}{\sigma^3}\times\sum(z_i-\bar{z})^3/N$	Skewness of normalized z within a cell	Vertical distribution of vegetation
Standard deviation of vegetation height (Hstd)	std_ normalized_height	$\sqrt{\sum\frac{(z_i-\bar{z})^2}{N-1}}$	Standard deviation of normalized z within a cell	Vertical variability of vegetation distribution
Variance of vegetation height (Hvar)	var_ normalized_height	$\sum\frac{(z_i-\bar{z})^2}{N-1}$	Variance of normalized z within a cell	Vertical variability of vegetation distribution

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024



313

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Fig 3. Examples of LiDAR metric generation in a 10 m × 10 m grid cell (the number of all points: N = 8348). (a) Metrics of vegetation height (mean, max, and percentiles of normalized height). (b) Vegetation cover metrics representing vegetation density within specific height layers. (c) Metrics of vegetation structural variability (e.g. standard deviation and variance of vegetation height are calculated based on mean height $\bar{z}$; kurtosis and skewness of vegetation height are calculated based on the standard deviation and mean height within a cell) (see detailed calculation formula in Table 3). The blue line in (c) represents a kernel density estimate (KDE) showing the shape of the points distribution. See abbreviation and calculation formula of all metrics in Table 3.

3.3 Auxiliary data

Since the point density of AHN datasets changes across space and time, we also provide a raster layer of point density (using all point classes) for each AHN dataset (four in total) (Fig. 4). The AHN1 has a much lower point density (average less than 0.5 pts m^{-2}) throughout the whole country than other AHN datasets due to sensor limitations back in 1996. AHN2 and AHN3 have a similar point density (on average 10–20 pts m^{-2}), while AHN4 has the highest point density (25–30 pts m^{-2}). Especially for the AHN2–AHN4 datasets, distinct patterns (patches, lines, edges) can be observed in different parts of the Netherlands. They are partially due to the influence of the water surface (yellow areas in AHN2, AHN3, and AHN4, Fig. 4), but also related to flight lines and operational configurations (e.g. flying altitude and flight speed) during the campaign.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

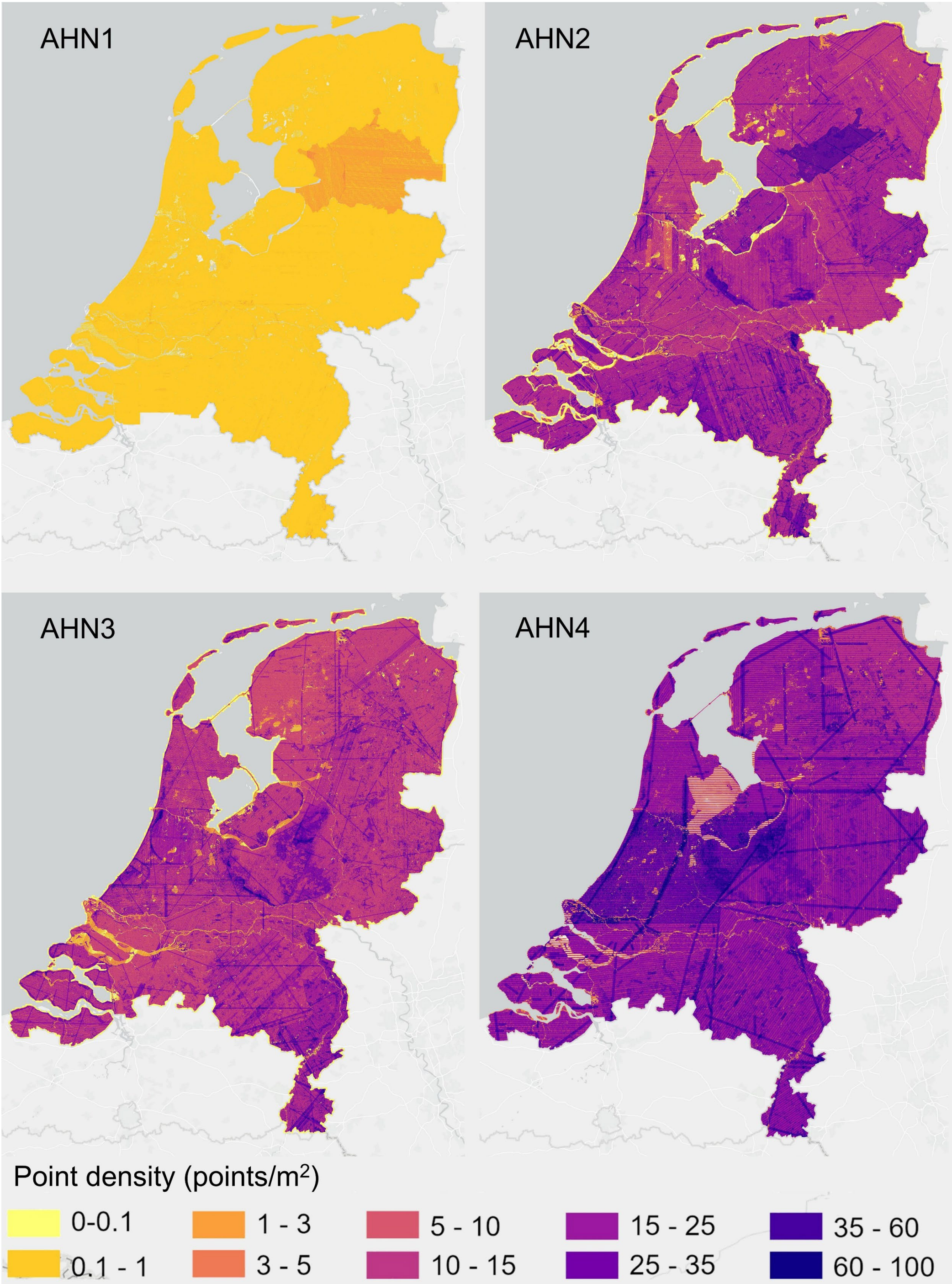


Fig. 4 Point density of AHN1–AHN4 ALS campaigns across the Netherlands. The total number of points was used for calculating the density of points at 10 meter spatial resolution. The four point density layers are made available in the data repository as auxiliary data together with the derived LiDAR metrics (see Sect. 7).

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

3.4 Limitations and usage notes

3.4.1 Classification related errors and masks

In the pre-classification of the raw AHN point clouds, there is no "vegetation" class provided based on the ASPRS standard (i.e. class 3: low vegetation, class 4: medium vegetation, or class 5: high vegetation). Instead, the vegetation points in the raw AHN1 and AHN2 datasets are included in the non-ground class ("uitgefilterd", classification value of 0), whereas they belong to the class "unclassified" (classification value 1) in the AHN3 and AHN4 datasets (Table 1). This can introduce errors and biases when using the "uitgefilterd" or "unclassified" class for calculating ecosystem structure properties because points belonging to human infrastructures can still be included in these classes. Particularly, buildings and bridges are included (together with other objects other than ground) in the class "uitgefilterd" in the AHN1 and AHN2 datasets, while they are classified separately (buildings in class 6: "buildings", and bridges in class 26: "reserved") in the AHN3 and AHN4 dataset — eliminating the errors caused by buildings and bridges in the final data products of the AHN3 and AHN4. Powerlines are not separated from the "uitgefilterd" class in the AHN1 and AHN2 datasets, and included in the class "unclassified" in the AHN3 dataset, but they are classified separately in the AHN4 dataset as class 14: "powerline". Yet, other human objects and infrastructures (e.g. cars, fences, and transmission towers) are not separated in any of the four AHN datasets and thus included in the non-ground class ("uitgefilterd") of the AHN1 and AHN2 datasets and in the class "unclassified" in the AHN3 and AHN4 datasets, introducing some errors and biases in the final data products. There are also points appearing on water surfaces (e.g. reflected by boats and birds) which are included in the class "uitgefilterd" or "unclassified", causing inaccuracies in the final products. In a previous study (Kissling et al., 2023), the accuracy of the 25 LiDAR metrics generated from the AHN3 dataset was assessed, particularly in relation to the error caused by using the class "unclassified" for calculating ecosystem structure properties. The results showed that the overall accuracy of the generated LiDAR metrics was high (0.90 ± 0.04, n = 25 LiDAR metrics, tested in 100 randomly selected plots throughout the Netherlands, with 10 m × 10 m size per plot), ranging from 0.87–1. It is worth noting that the impact of those errors on the 25 LiDAR metrics varies, for instance, a stronger bias (i.e. the difference between the generated LiDAR metrics and the ground truth) can be observed in height metrics describing the top canopy layer (i.e. Hmax and Hp95) than in other height metrics or in metrics of vegetation cover in the low strata (i.e. BR_below_1 and BR_below_5) (Kissling et al., 2023).

To minimize the inaccuracies of the data products caused by human infrastructures and water surfaces, we provide mask layers of water areas, roads, and buildings for both the AHN3 and AHN4 data products based on the Dutch cadaster data (TOP10NL) from 2018 (corresponding to AHN3) and 2021 (corresponding to AHN4) (https://www.kadaster.nl/zakelijk/producten/geo-informatie/topnl, last access 19 October 2024). TOP10NL is part of the Basic Topography Registry (BRT) which provides the standard topographic base files for the whole Netherlands. Like the LiDAR metrics, the masks are calculated at 10

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

m resolution with the RD_new / EPSG 28992 projection coordinate system and provided as raster layers in GeoTIFF format. In the masks, water surfaces, buildings and roads were merged into one class with a pixel value assigned to 1 and the rest with a pixel value of 0 (Fig. 5). Since the historical versions of TOP10NL data are not available for AHN1 (1996–2003) and AHN2 (2007–2012), we can only provide the masks for the AHN3 and AHN4 datasets (see Sect. 7 for data availability). However, despite the potential changes in buildings and roads over time, it is still possible to apply the generated masks to all four AHN data products, for instance, to minimize errors and to have comparable areas of interest.

Since powerlines are not classified separately for AHN1–AHN3 datasets and thus included in the calculation, it may cause abnormal values of vegetation structure, especially for vegetation height and vegetation cover above 20 m (Shi and Kissling, 2023). However, points belonging to powerlines are classified separately in AHN4 (Table 1), which provides a way to minimize errors caused by powerlines in the data products generated from AHN1–AHN3. We therefore extracted all powerline points from the AHN4 raw point cloud and generated a mask (at 10 m resolution) where pixels containing powerlines are assigned a value 1 and the rest as NoData (Fig. 5). Since the transmission towers are not classified separately in all four AHN datasets, the mask only covers the powerlines but not the transmission towers. Users can apply the powerline mask generated from AHN4 to the data products from AHN1–AHN3 and consequently improve the comparability of the LiDAR metrics across time. Note that the powerline infrastructure may also change over time, and the classification of powerlines from the AHN4 may not be fully representative for powerline distributions in earlier time periods.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Open Access Earth System Science Data Discussions

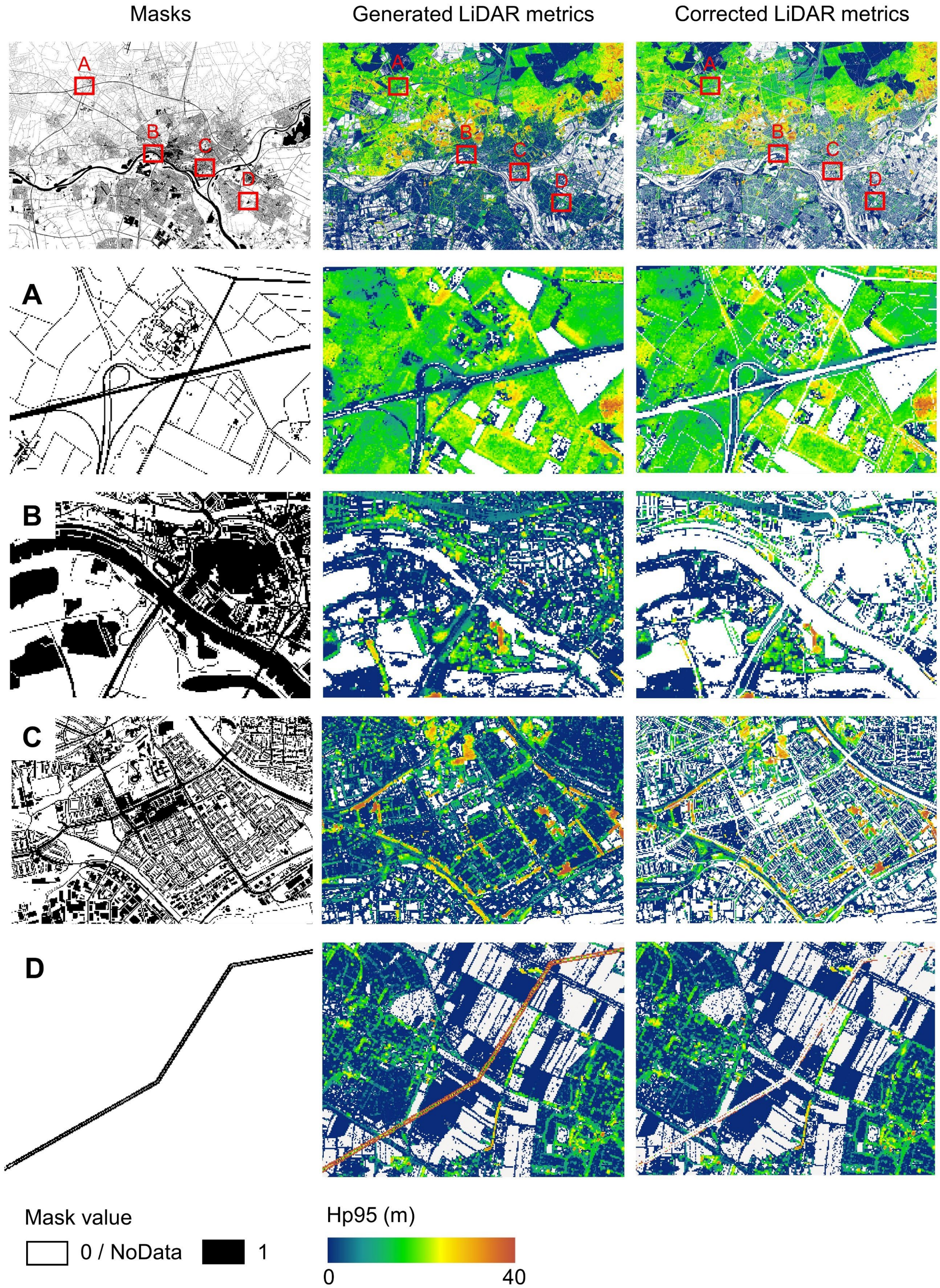


Fig. 5 Examples of masking roads, water surfaces, and buildings from the 2018 Dutch cadaster data (areas A, B, and C) and powerlines generated from the AHN4 (area D). Illustrated is the rasterized mask (first column), the generated vegetation height metric (i.e. Hp95) from AHN3 (second column), and the

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

corrected LiDAR metric using the masks (third column). Four subareas show the inaccuracies in the originally generated LiDAR metric and the removal effect of using the mask for roads (area A), water (area B), buildings (area C), and powerlines (area D). A mask value of 1 represents the pixels with roads, water surfaces, buildings, and powerlines, while value 0 or NoData represents the rest. The masks and the LiDAR metrics are at 10 × 10 m resolution. Hp95 = 95th percentile of vegetation height.

3.4.2 Strip issues

Several strip patterns occur in the data products from AHN2 (Fig. 6). This strip issue specifically affects the pulse penetration ratio layer (representing vegetation openness), where both ground points ("ground" class) and vegetation points ("unclassified" class) were used for the metric calculation. A possible reason could be that the scan angle of the laser scanner used for point cloud acquisition was rather wide, and that the scanner thus has received more laser pulses from the areas located at the edges of the flight lines. Those overlapping areas (edges of the flight lines) often have a doubled point density, which also contributes to the strip patterns in the calculation of the LiDAR metrics using ground points (e.g. pulse penetration ratio). This issue only occurs in an area in the centre of the Netherlands (Fig. 6). Other LiDAR-derived vegetation metrics representing vegetation height, cover, and structural variability do not seem to be influenced by this strip issue. This strip issue was not observed in other AHN data products.

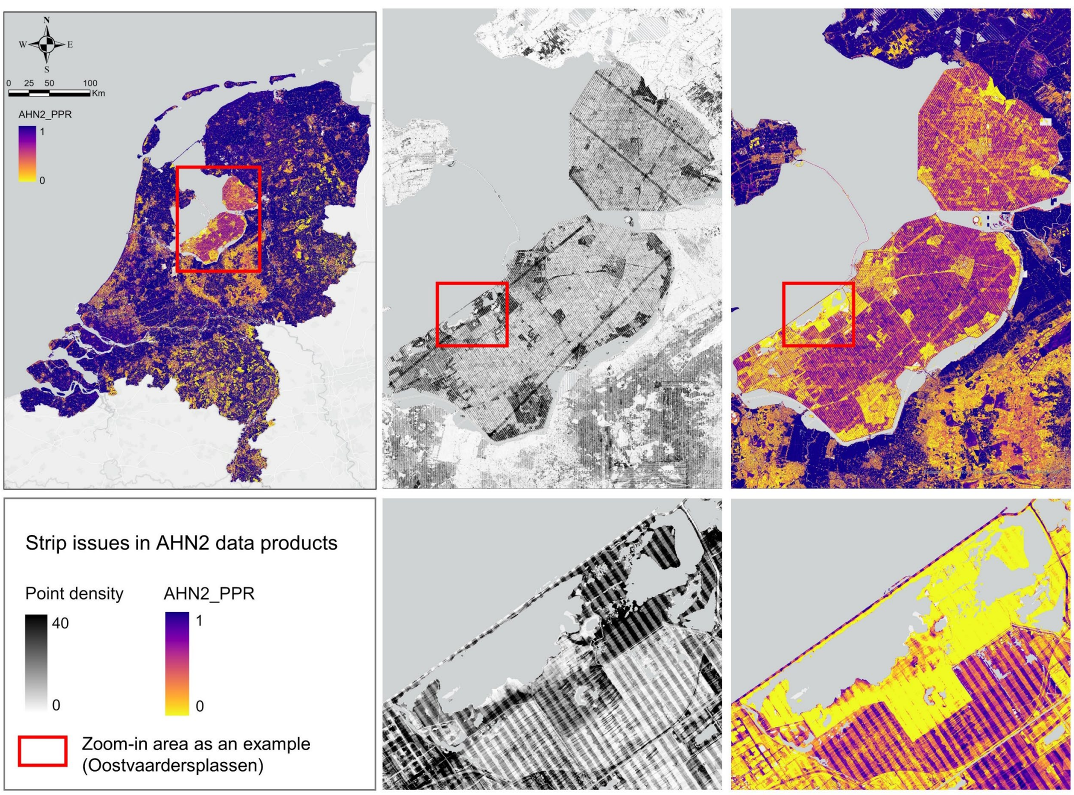


Fig. 6 Strip issues in the AHN2 dataset. The point density (black and white, including all points) and the pulse penetration ratio (colour, representing vegetation openness) show similar strip patterns.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

3.4.3 Abnormal values

A few pixels with abnormal values still exist in the final products. For instance, several pixels in the Hp95 layer have a value higher than 100 m, which cannot represent the upper canopy of vegetation since the tallest tree in the Netherlands (a Douglas Fir, *Pseudotsuga menziesii*, i.e. a tall and fast-growing conifer native to western North America which was planted between 1860 and 1870 in Apeldoorn, the Netherlands) has been measured to be ~50 meter tall. More generally, most measurements of the tall trees in the Netherlands range between 20–45 m. Hence, abnormal values of vegetation height (e.g. > 50 m) most likely reflect the occurrence of human infrastructures that are not included in the AHN1 and AHN2 class "uitgefilterd" or not sufficiently captured in the AHN3 and AHN4 classes "building" and "reserved", e.g. aerial and radio masts (up to 350 m tall), tall industrial and meteorological towers and chimneys (50–200 m), cranes (50–130 m), elements of bridges (e.g. pylons and steel cables up to 140 m tall), wind turbines (up to 260 m) and powerlines (up to 80 m). Flying objects, such as birds and planes, can also be captured in the datasets, resulting in abnormal height values in the data products. We recommend filtering out those abnormal values before using the data products for further analysis, e.g. by removing grid cells with Hp95 > 50 m.

Although the Netherlands has rather flat terrain, it is worth noting that the normalization method implemented in the Laserfarm workflow may introduce inaccuracies in normalized vegetation height values, especially if steep terrain occurs within a grid cell (Kissling et al., 2022). When applying the same workflow for other country or regions, abnormal values may occur in the areas with drastic topographic changes (e.g. cliffs, mountainous area). Users may consider using a different normalization method, for instance, normalizing non-ground points by subtracting the derived DTM from all points, or by interpolating the elevation of non-ground points using the exact position of ground points beneath (Roussel et al., 2020). Some studies also have suggested to use raw point clouds (e.g. the un-normalized DSM) to preserve the geometry of tree tops or plant area index profile in high slope areas (Khosravipour et al., 2015; Liu et al., 2017).

4 Demonstration of ecological use cases

4.1 Monitoring forest structural change across time using multi-temporal ALS data

As a use case, we demonstrate here how the multi-temporal data products generated from the Dutch ALS surveys can capture forest structural change over the past two decades (2000–2023). We included the ongoing ALS campaign (AHN5) since the data were made available for the sample area (central location coordinates: 52.3250517°N, 5.7409230°E) at the time when the analysis was conducted. This provided a longer time series for detecting forest change. The sample area (in a forest area north of the national park De Hoge Veluwe) has experienced a clear forest cut in 2011 (between AHN2 and AHN3 surveys), with further forest loss and some regenerations captured by AHN4, while the latest AHN5 showed a forest

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

regrowth in the middle-low vegetation strata (< 10 m) compared to AHN4 (Fig. 7). The histograms derived from point clouds from AHN1–AHN5 show the distribution of points shifting from tall vegetation (above 20 m, AHN1–AHN3) to low vegetation (below 10 m, AHN4 and AHN5). Due to the very low point density of the AHN1 data, detailed information on vegetation structure in the year 2000 is lacking. However, the histogram from AHN1 implies a similar pattern of canopy height as that from AHN2 (Fig. 7). Google Earth imageries obtained on the closest dates available from each AHN survey also provide a good reference for the forest change events, except for the time of AHN1.

Six selected LiDAR-derived vegetation metrics derived from AHN1–AHN5 at 10 m resolution effectively capture the changes in vegetation structure over time (Fig. 8). The 95th percentile of vegetation height (Hp95) and mean vegetation height (Hmean) highlight reductions in forest canopy height due to cutting in 2011 (between AHN2 and AHN3) and in 2019 (between AHN3 and AHN4). The pulse penetration ratio (PPR) reveals shifts in vegetation openness, with openness peaking in AHN4, while the density of vegetation points at 2–3 m (BR_2_3) indicates regrowth in the understory, particularly in AHN4 and AHN5 (after 2021). The Shannon index (entropy_z) reflects the vertical distribution of vegetation points (i.e. evenness), with AHN2 showing the highest value due to a more even point distribution of the canopy foliage before the canopy was cut. AHN3 shows the widest Shannon index range, capturing both high canopy trees and new re-growth. The standard deviation (i.e. vertical variability) of vegetation height (Hstd) shows a similar pattern as seen in Hp95.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

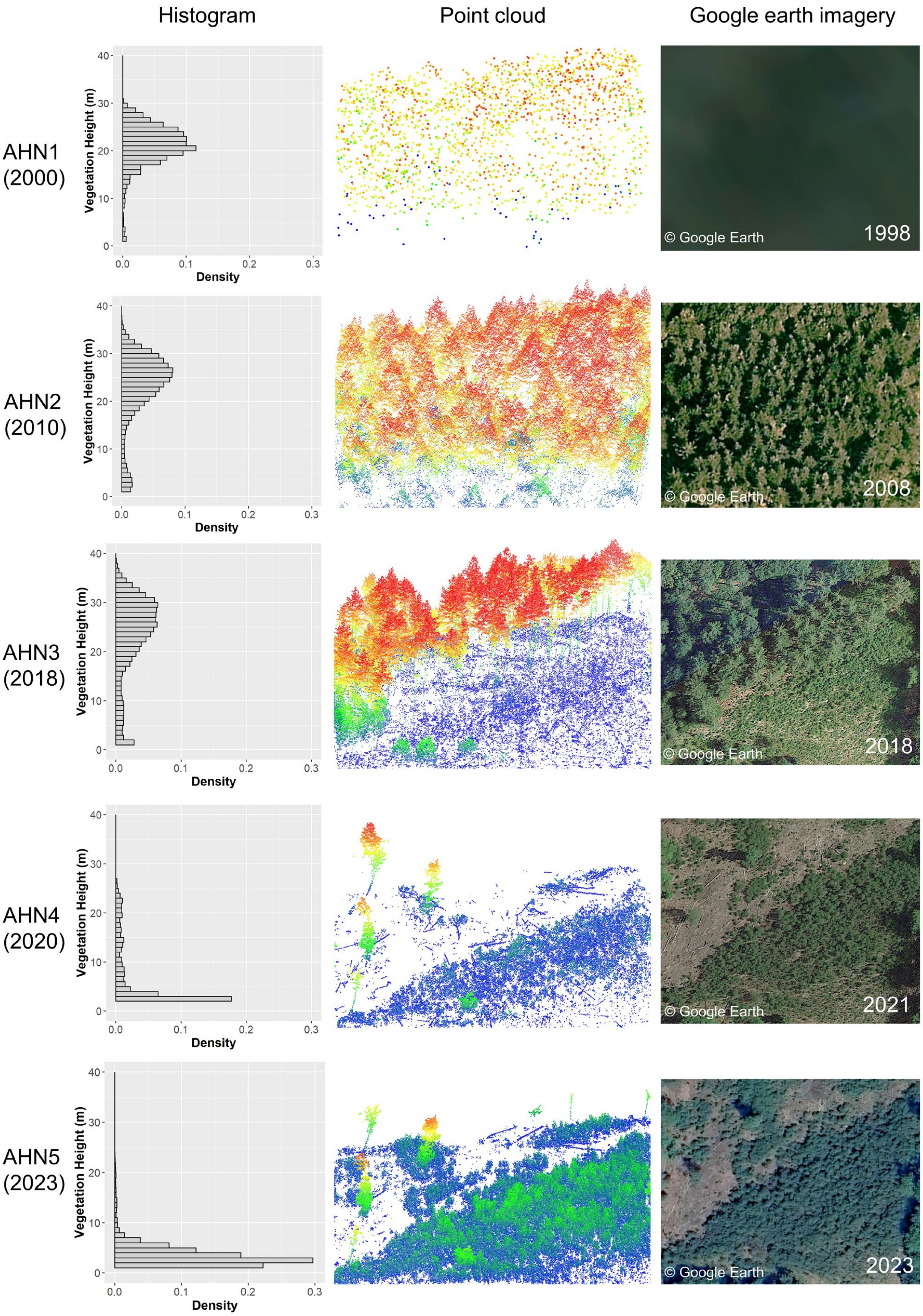


Fig. 7 Forest structural change in a sample plot (100 m × 100 m) between 1998–2023 captured by the multi-temporal AHN datasets (AHN1–AHN5). The histograms were generated from each AHN point cloud, showing the distribution of the normalized vegetation height within the plot. The point clouds were

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Open Access Earth System Science Data Discussions

coloured by height (blue indicates lower vegetation height and red indicates higher vegetation height). AHN1 has a rather poor point density, but shows a histogram of vegetation height that is similar to AHN2. The forest cut can be observed from the point clouds of AHN3 and AHN4 compared to AHN2, with forest regrowth occurring in AHN5. Google Earth imageries from the example area show the changes of the forest. Note that the dates of the Google Earth imageries do not correspond exactly to the dates of the airborne laser scanning surveys, but to the closest dates available. Map data: © Google Earth.

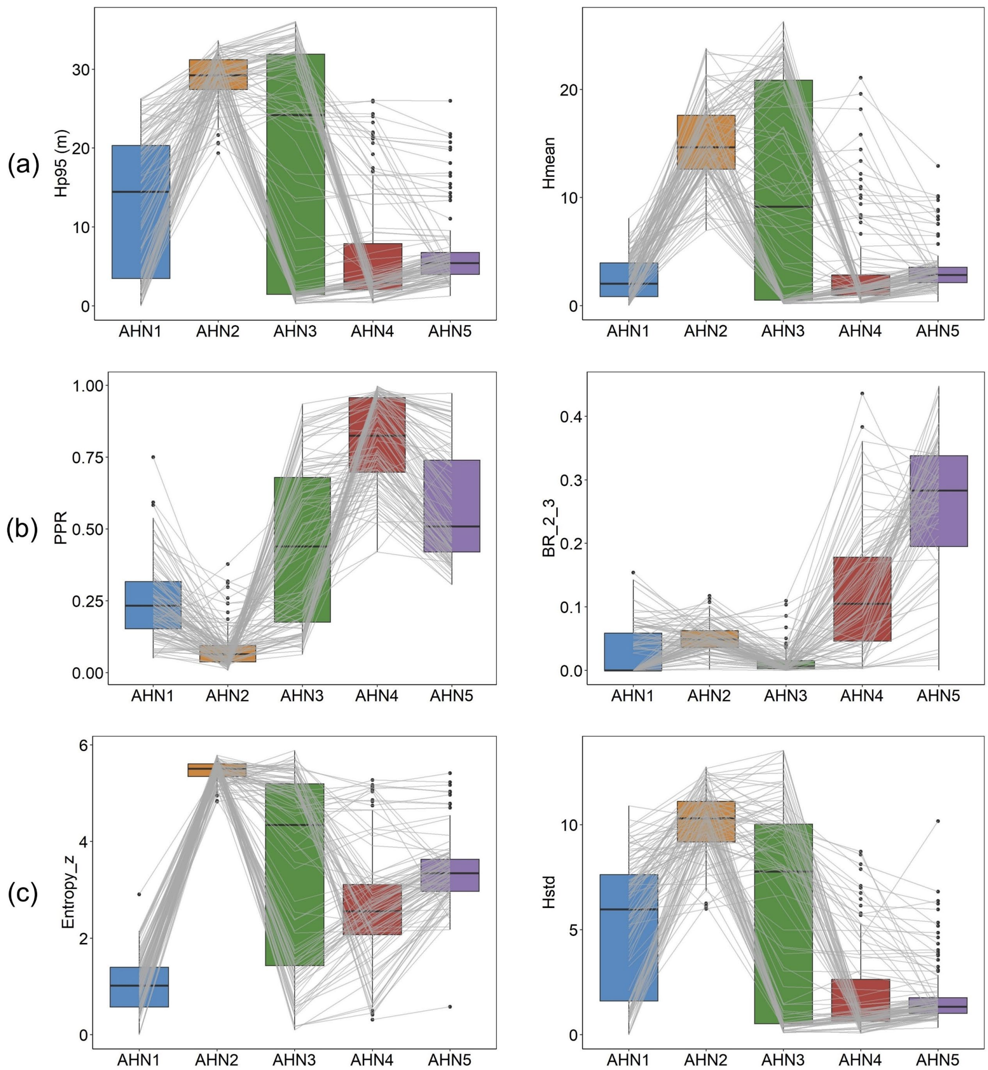


Fig. 8 Boxplots of LiDAR metrics derived from multi-temporal AHN datasets capturing the changes of the vegetation structure in a 100 m × 100 m sample area (compare Fig. 7). (a) The 95th percentile of vegetation height (Hp95) and the mean vegetation height (Hmean) representing vegetation height. (b) The pulse penetration ratio (PPR) and the density of vegetation points between 2–3 m (BR_2_3) representing vegetation cover. (c) The Shannon index (Entropy_z) and the standard deviation of vegetation height (Hstd) representing vegetation structural variability. Boxes show the median and interquartile range, with

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

whiskers extending to 1.5 times the interquartile range and outliers are plotted as dots. Each grey line represents a single pixel (10 m × 10 m) value changing from AHN1–AHN5, showing the influence of the events on vegetation within each pixel (e.g. forest cut and regrowth).

4.2 Comparison of vegetation structural difference within Natura 2000 sites

In a second use case, we analyse how vegetation structure varies spatially across different Natura 2000 habitat types in the Netherlands. Terrestrial habitats were categorized into five main classes: dunes, marshes, grasslands, shrublands, and woodlands, based on the dominant habitat type within each site (see details in Appendix A). For each habitat class, 100 random sample plots (10 m × 10 m, 500 plots in total) were selected where Hp95 is not NA (assuming vegetation occurring in the plots) (Figure A1). We used the data products from AHN4 for the analysis as they are the latest complete products for the whole Netherlands. Four LiDAR metrics were compared: the 95th percentile of vegetation height (Hp95), vegetation point density at 1–2 m (BR_1_2) and 4–5 m (BR_4_5), and the coefficient of variation in vegetation height (Coeff_var). Structural differences among the five habitat types were assessed using the non-parametric Kruskal-Wallis test by ranks (Kruskal and Wallis, 1952), which compares two or more independent groups of equal or different sample sizes without assuming a normal distribution of the residuals. Pairwise comparisons of the statistical significance were conducted among groups (i.e. habitat types) using the Wilcoxon rank-sum test (Wilcoxon et al., 1970).

The strongest structural differences among the five habitat types were observed in canopy height (Hp95) and vegetation density in the lower strata (BR_1_2), followed by vegetation vertical variability (Coeff_var) and vegetation density in the middle strata (BR_4_5) (Fig. 9). Canopy height (i.e. Hp95) of both woodlands and shrublands showed a statistically significant difference to all other habitat types, whereas grasslands, marshes and dunes did not differ in canopy height (Fig. 9a). The latter three habitat types showed a median canopy height of ~ 2.3 m, whereas it is around 9.9 m and 17.6 m for shrublands and woodlands, respectively. Vegetation density in the low vegetation stratum (between 1–2 m) also did not statistically differ between grasslands, marshes, and dunes (Fig. 9b). However, woodlands and shrublands with their more shaded understory and stronger light competition had much lower vegetation densities between 1–2 m than the three open habitat types (Fig. 9b). In the mid-layer (4–5 m), only the vegetation density of woodlands and marshes showed a statistically significant difference (Fig. 9c). The very low mid-layer density in woodlands may be due to the high canopy from trees limiting growth in the understory (e.g. shrubs), whereas shrubs and trees in marshes may generally have a lower canopy height than woodland trees, thus showing high vegetation density at 4–5 m. In terms of structural variability, grasslands and marshes have the highest median values of the coefficient of variation of vegetation height, showing significant differences to woodlands, shrublands and dunes (Fig. 9d). This probably reflects a high heterogeneity in vegetation structure in both grasslands and marshes, where a large variability from low to high vegetation is captured within the 10 m × 10 m plots. It is also the only metric among the four selected metrics where dunes showed statistically significant differences to grasslands and marshes.

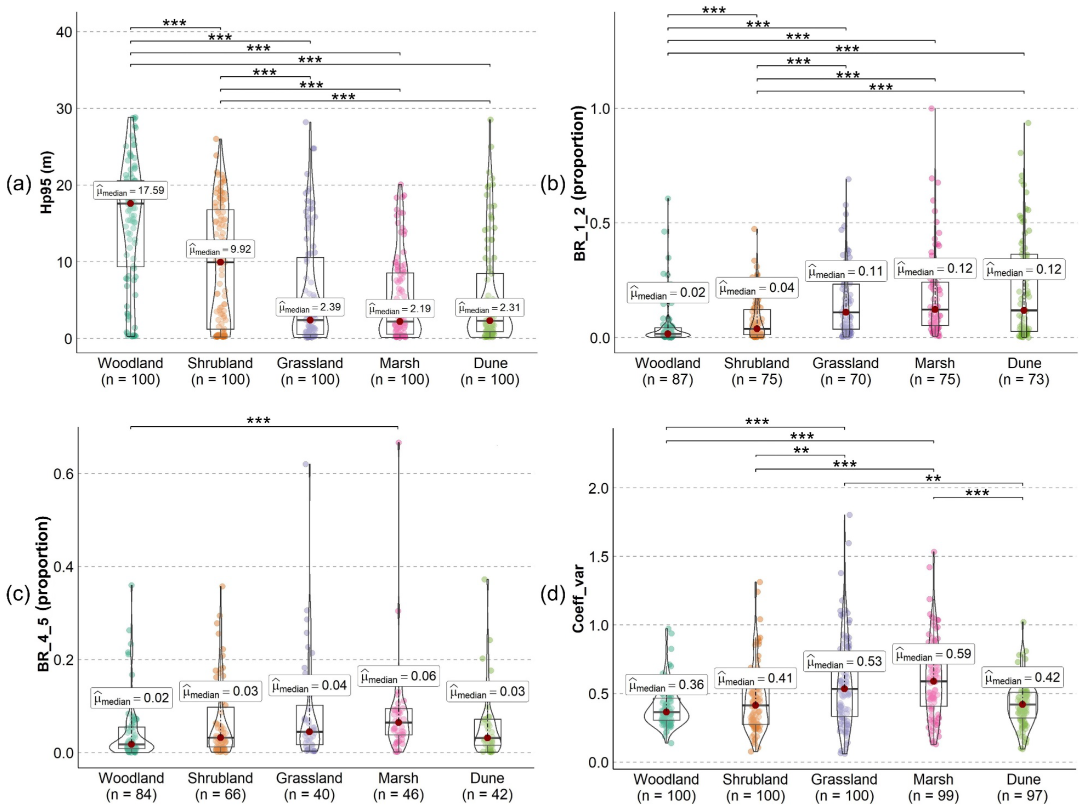


Fig. 9 Comparison of ecosystem structure between five Natura 2000 habitat types using four different LiDAR metrics of vegetation structure. (a) Canopy height (the 95th percentile of vegetation height, Hp95), (b) vegetation density at 1–2 m (BR_1_2), (c) vegetation density at 4–5 m (BR_4_5), and (d) structural variability of vegetation height (coefficient of variation in vegetation height, Coeff_var). The bars above the violin plot indicate whether there is a statistical significance between two compared habitat types. The pairwise comparisons of the statistical significance were conducted using the Wilcoxon rank-sum test after the non-parametric Kruskal-Wallis test by ranks. The significant level is marked as follows: *** ($p < 0.001$), ** ($p < 0.01$), and * ($p < 0.05$). Red dots indicate the median value ($\hat{\mu}_{median}$) of the LiDAR metrics measured for each habitat type. Note that not all sampled plots have vegetation points (from class "unclassified") between 1–2 m and between 4–5 m, therefore the total number of sample plots for the "BR_1_2" and "BR_4_5" analysis was < 100 for each habitat type (after removing NA value). The NA value also occurs for "Coeff_var" when there is only one point (from class "unclassified") in the sampled plot (see metric calculation in Table 3).

5 Discussion

We present a set of multi-temporal high-resolution data products of ecosystem structure derived from country-wide ALS surveys of the Netherlands (AHN1–AHN4), capturing vegetation structure dynamics over the past two decades (1998–2022). For each AHN dataset, we provide 25 LiDAR-derived vegetation

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Earth System Science Data
Open Access Discussions

metrics as GeoTIFF raster layers representing vegetation height, vegetation cover, and vegetation structural variability at 10 m resolution. In total, we processed ~ 70 TB (uncompressed) raw point clouds from four national ALS surveys into ~ 59 GB GeoTIFF raster layers as final data products. These data products hold great value for ecological and geospatial applications, including species distribution modelling, habitat characterization, and forest and biodiversity dynamics monitoring. The availability of these ready-to-use LiDAR metrics enables ecologists and researchers to integrate detailed ecosystem structural information from complex 3D point clouds into their studies without the burden of handling large ALS datasets and computational challenges. Additionally, the dataset serves as a valuable resource for detecting vegetation structural changes and analysing ecosystem dynamics using multi-temporal remote sensing techniques.

Several key aspects should be considered when utilizing the presented data products. First, many commonly used LiDAR-derived metrics, especially those related to vegetation height (e.g. maximum vegetation height, 95th percentile height, mean height), are often highly correlated (Kissling and Shi, 2023; Shi et al., 2018a). To gain a more comprehensive understanding of ecosystem structure, it is advisable to use a complementary set of LiDAR metrics that captures different dimensions of ecosystem structure, or to use dimensionality reduction methods (such as a principal component analysis) to avoid multi-collinearity (Kissling and Shi, 2023). For instance, using the coefficient of variation of vegetation height (Coeff_var) instead of the standard deviation (Hstd) as a metric of structural variability can avoid correlations with mean or canopy vegetation height (Hmean and Hp95) (Kissling and Shi, 2023). Second, vegetation cover in different height layers is a crucial component of forests and other ecosystems, influencing energy fluxes between the ecosystem and the atmosphere (Shugart et al., 2010; Toivonen et al., 2023). Unlike the cover metrics proposed by Moudrý et al. (2022), where herbaceous, shrub and tree layers were used to represent different vegetation strata, our metrics use fixed height intervals (e.g. 1–2 m, 2–3 m, 3–4 m, 4–5 m, 5–20 m, above 20 m) to ensure applicability across diverse ecosystems. Not all ecosystems share the same vegetation growth forms, making these height bin-defined metrics more ecosystem-agnostic. The cover metrics from different height layers can be used as predictors of animal species richness (Goetz et al., 2007), species distributions (Davies and Asner, 2014), and habitat characteristics (Vierling et al., 2008; Bakx et al., 2019). Third, LiDAR metrics related to vegetation structural variability (e.g. Hstd, Hskew, and Hkurt) are often influenced by various ecological and sensing methodology-related factors, making them potentially challenging to interpret (Assmann et al., 2022). However, metrics representing structural variability are valuable input for models assessing forest functional diversity and structural types, especially when combined with optical remote sensing (Kamoske et al., 2022; Zheng et al., 2021). Thus, careful selection of LiDAR metrics for specific applications is highly recommended. Terrain and surface descriptors such as DTMs and DSMs (or canopy height model as derivative) can be additionally considered because they are important for forest and habitat classifications (Shoot et al., 2021), quantifying soil moisture or wetness (Assmann et al., 2022),

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

and analysing species composition (Toivonen et al., 2023; Hill and Thomson, 2005). However, since the AHN programme has already provided DTM and DSM layers for the AHN2, AHN3, and AHN4 datasets at 0.5 m and 5 m resolutions in their repository, we did not reproduce these data products.

While multi-temporal ALS data offer valuable insights into fine-scale vegetation structural changes and ecosystem dynamics, there are also notable challenges, especially when performing change detection across point clouds with different characteristics, such as point density, scanning angle, and varying vertical and horizontal accuracy (White et al., 2016). Instead of performing change detection directly on point clouds (Xu et al., 2015; Kharroubi et al., 2022), many studies use rasterized LiDAR metrics for monitoring changes on vegetation structure. This is less computational intensive and better suited for areas with complex vegetation structure as it regularizes complex 3D point cloud information onto a 2D grid (Vastaranta et al., 2013; Choi et al., 2023). Several commonly used change detection methods can be applied to the multi-temporal data with rasterized LiDAR metrics. These include image differencing (i.e. subtracting the pixel values of one raster layer, such as Hp95 from AHN3, from the other, such as Hp95 from AHN4), threshold-based change detection (i.e. classifying the pixels as "changed" or "unchanged" based on a set threshold after image differencing), and post-classification comparison (i.e. comparing classified raster layers, such as maps of vegetation types based on derived LiDAR metrics, from different time periods) (Noordermeer et al., 2019; Dalponte et al., 2019). Those methods can be applied to the provided AHN data products, especially after masking water areas, roads, buildings, and powerlines. Change metrics derived from multi-temporal LiDAR data can also be combined with clustering methods to characterize areas of structural changes, such as modifications of forests by the eastern spruce budworm (Trotto et al., 2024). Together with the development of deep learning on change detection (Bai et al., 2023), more in-depth insights from the presented AHN datasets can be revealed, enabling accurate and comprehensive analysis of ecosystem dynamics. Given the consistent coordinate system used in the four AHN datasets (EPSG: 28992, NAP: 5709; see Table 1), additional georeferencing steps are unnecessary before conducting further analysis with the data products that we provide. The scan angle, overlapping rate, and vertical accuracy of AHN2–AHN4 are rather comparable (Table 1), potentially reducing errors related to systematic differences across time. However, the data products are generated from point clouds with different point density, which may introduce inconsistencies in capturing vegetation structure. Nevertheless, analyses of tree growth using multi-temporal LiDAR data with different point density in forests of Scotland implied that the accuracy does not decrease as long as the point density is exceeding 7 pts m^{-2} (Zhao et al., 2018). Several studies also indicated that the spatial distribution of the point cloud remains similar even if the point density varies and increasing point density does not increase area-based estimation accuracy (Hudak et al., 2012; Fekety et al., 2015; Cao et al., 2016). We therefore anticipate that the data products from AHN2, AHN3, and AHN4 are sufficiently comparable for reliable change detection. However, due to the low point density

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

and reduced accuracy, we do not recommend including the data products from AHN1 in multi-temporal analyses.

All software and tools employed in the pipeline for producing the data products are free and open-source, ensuring a standardized yet flexible processing framework for country-wide ALS data and enabling reproducibility for future surveys. While existing ALS processing software such as OPALS (Pfeifer et al., 2014) and LAStools (http://lastools.org/) are not (fully) open-source, and others like FUSION (https://forsys.sefs.uw.edu/fusion/fusionlatest.html), CloudCompare (https://www.danielgm.net/cc/), and lidR (Roussel et al., 2020) lack horizontal scalability and do not provide reproducible end-to-end workflows for large ALS datasets, the employed “Laserfarm” workflow fills a niche by addressing these challenges. Laserfarm is a high-throughput, modular, and reproducible end-to-end workflow designed for efficiently extracting LiDAR metrics of ecosystem structure using distributed computing infrastructures (Kissling et al., 2022). With the workflow materials that we provide, users can implement additional pre-processing steps (e.g. splitting, reclassification) and customize required parameters based on the input ALS data and available computing resources. The demonstrated configurations of IT infrastructure, computational cost, and time efficiency for processing multi-temporal AHN datasets serve as a reference for users to estimate the processing requirements for future national or regional ALS datasets. It is worth noting that the normalization method implemented in the Laserfarm workflow subtracts the elevation of the lowest point within a given neighbourhood to remove the influence of the terrain. This approach was specifically chosen for its effectiveness in handling small ditches and canals that are common in the Dutch landscape, providing a straightforward way to generate positive height values after normalization. However, it may be less suited for capturing continuous normalized height values and fine-scale terrain variability in smaller grid cells (< 1 m) (Kissling et al., 2022).

The data products presented here also make a great contribution to multi-source data fusion in remote sensing and ecological research (Ghamisi et al., 2019). Through the two use cases in Sect. 4, we demonstrate the utility of these multi-temporal datasets for monitoring long-term forest dynamics and characterizing habitat types. These applications can be further extended to other studies, such as improving land cover classification accuracy, particularly for objects composed of similar materials (e.g. grasslands, shrubs, and trees). Moreover, the fusion of vegetation structural information from LiDAR, spectral data from optical remote sensing (e.g. high-resolution digital aerial photogrammetry, Landsat and Sentinel-2 imagery), climate data, and field measurements underscores the value of integrating complementary remote sensing data across diverse applications. These include wildlife habitat characterization (Boelman et al., 2016), tree species identification (Shi et al., 2018b), forest structure and carbon stock mapping (Li et al., 2024), as well as assessing disturbances and recovery of ecosystem process (Li et al., 2023). Additionally, combining ecosystem structure data from multiple LiDAR platforms, such as terrestrial, drone-based, airborne, and spaceborne LiDAR, could provide a more comprehensive understanding of ecosystem structure, spanning from understory to canopy level and across local plots to national or continental level.

6 Code availability

Jupyter Notebooks for processing AHN datasets: https://github.com/ShiYifang/AHN
Laserfarm workflow repository: https://github.com/eEcoLiDAR/Laserfarm
Laserchicken software repository: https://github.com/eEcoLiDAR/laserchicken
Code for downloading AHN dataset: https://github.com/ShiYifang/AHN/tree/main/AHN_downloading
Code for generating masks for AHN datasets: https://github.com/ShiYifang/AHN/tree/main/AHN_masks
Code for demonstration of ecological use cases: https://github.com/ShiYifang/AHN/tree/main/Use_case

7 Data availability

All data products from AHN1–AHN4 (25 GeoTIFF layers for each AHN dataset), three masks (two for roads, water surfaces, and buildings from both AHN3 and AHN4, and one for powerlines generated from AHN4), and four point density layers (for AHN1–AHN4) are available from a Zenodo repository (https://doi.org/10.5281/zenodo.13940846) (Shi and Kissling 2024). The data used for the demonstrated use cases are also provided in the same repository. A detailed description of the provided data can be found in the README file in the data repository.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

8 Conclusions

Ecosystem structure information derived from country-wide ALS data becomes increasingly needed for biodiversity science and ecosystem monitoring. The multi-temporal data products of ecosystem structure and the employed workflow presented here not only provide ready-to-use information for ecosystem monitoring and modelling within the Netherlands, but also enable reproducing desired data products from existing and upcoming large-scale ALS data beyond the Netherlands. We highlight the capability of multi-temporal ALS data products in capturing ecosystem structural dynamics across time and their usability in combination with other data sources. We also carefully evaluated the limitations and usability of generated data products and provided solutions or recommendations for future processing and usage. We envisage that the provided data products and the employed workflow will empower a wider use and uptake of ecosystem structure information in biodiversity and ecosystem science, land management, natural resource conservation, and policy support and decision making.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Appendix A

The source information about Natura 2000 sites was retrieved from the Europe Environment Agency (Natura 2000 (vector) - version 2021). The shapefile of the Natura 2000 sites and the attributes of each site that we used for the analysis were downloaded via https://sdi.eea.europa.eu/datashare/s/JWt9KJCFMrPQDc7/download. The information on the habitat class (from the table named "Natura2000_end2021_HABITATCLASS.csv") was used to group them into five habitat types (i.e. dunes, marshes, shrublands, grasslands, and woodlands). The table contains the following information: description of the habitat class, habitat code, site code, and percentage of habitat composition within the site.

We first selected all the Natura 2000 sites within the Netherlands (i.e. SITECODE starting with NL), then summarized the highest percentage of habitat class within each site and grouped them into six main habitat types: water, dunes, marshes, shrubland, grassland, and woodland. For water, we included marine areas, sea inlets (habitat code: N01), tidal rivers, estuaries, mud flats, sand flats, and lagoons (habitat code: N02), and inland water bodies (habitat code: N06). For dunes, we included costal sand dunes, sand beaches, and machair (habitat code: N04). For marsh, we included bogs, marshes, water fringed vegetation, and fens (habitat code: N07) and salt marshes, salt pastures, and salt steppes (habitat code: N03). For shrubland, we included heath, scrub, maquis and garrigue, and phygrana (habitat code: N08). For grassland, we included dry grassland, steppes (habitat code: N09), humid grassland, mesophile grassland (habitat code: N10), and improved grassland (habitat code: N14). For woodland, we included broadleaved deciduous woodland (habitat code: N16), coniferous woodland (habitat code: N17), evergreen woodland (habitat code: N18) and mixed woodland (habitat code: N19). For each Natura 2000 site, the habitat type with the highest composition percentage was chosen as the dominate habitat. In total, there were 197 Natura 2000 sites within the Netherlands, including 36 water sites, 25 dune sites, 23 marsh sites, 17 shrubland sites, 54 grassland sites, and 42 woodland sites. For our study, we excluded water sites for the vegetation structure analysis (remaining 161 sites in total). For each habitat type, we randomly selected 100 sample plots (10 m × 10 m for each plot, i.e. in total 500 plots) where Hp95 is not NA (assuming vegetation occurring in the plots) using the *sampleRandom()* function in R (Figure A1). The shapefile of the 500 sample plots across the Natura 2000 sites was then used to extract the pixel values of the LiDAR metrics for comparison.

The shapefile of the Natura 2000 sites within the Netherlands (with habitat class information in attributes), 100 sample plots for each habitat class, original and grouped habitat class information (.csv files), and the R processing script are provided in the data repository (see Sect.7).

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

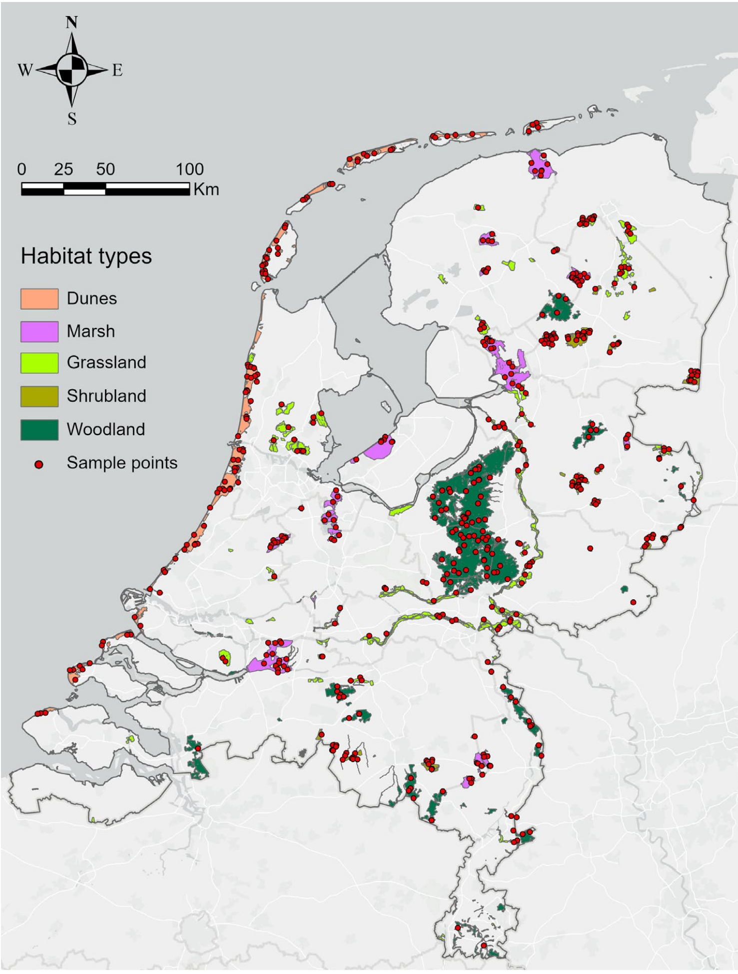


Figure A1. Natura 2000 sites and their habitat types in the Netherlands. The non-water habitat types were grouped into 5 classes (i.e. dunes, marshes, grasslands, shrublands, and woodlands) to conduct vegetation structure comparisons. For each class, we randomly sampled 100 plots (10 m × 10 m each) where Hp95 was not NA (assuming that vegetation occurs in the plots) for the analysis (n = 500 in total).

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Open Access Earth System Science Data Discussions

Author contributions

Yifang Shi: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **W. Daniel Kissling**: Conceptualization, Investigation, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

We acknowledge funding support from the European Commission (MAMBO project grant number 101060639) and the Netherlands eScience Center (grant number ASDI.2016.014). We thank Francesco Nattino and Meiert W. Grootes from the Netherlands eScience Center for leading the development of the Laserfarm workflow through the project 'eScience infrastructure for Ecological applications of LiDAR point clouds' (eEcoLiDAR) (Kissling et al., 2017). The development of the data products was also supported by LifeWatch ERIC (https://www.lifewatch.eu/), a European research infrastructure consortium with focus on biodiversity and ecosystem research. We acknowledge the computing resources provided by SURF, the Dutch national facility for information and communication technology (https://www.surf.nl/).

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

References

Aguirre-Gutiérrez, J., WallisDeVries, M. F., Marshall, L., van't Zelfde, M., Villalobos-Arámbula, A. R., Boekelo, B., Bartholomeus, H., Franzén, M., and Biesmeijer, J. C.: Butterflies show different functional and species diversity in relationship to vegetation structure and land use, Global Ecology and Biogeography, 26, 1126-1137, https://doi.org/10.1111/geb.12622, 2017.

ASPRS: LAS Specification 1.4-R15, 2019.

Assmann, J. J., Moeslund, J. E., Treier, U. A., and Normand, S.: EcoDes-DK15: high-resolution ecological descriptors of vegetation and terrain derived from Denmark's national airborne laser scanning data set, Earth Syst. Sci. Data, 14, 823-844, https://10.5194/essd-14-823-2022, 2022.

Bai, T., Wang, L., Yin, D., Sun, K., Chen, Y., Li, W., and Li, D.: Deep learning for change detection in remote sensing: a review, Geo-spatial Information Science, 26, 262-288, https://doi.org/10.1080/10095020.2022.2085633, 2023.

Bakx, T. R. M., Koma, Z., Seijmonsbergen, A. C., and Kissling, W. D.: Use and categorization of Light Detection and Ranging vegetation metrics in avian diversity and species distribution research, Diversity and Distributions, 25, 1045-1059, https://doi.org/10.1111/ddi.12915, 2019.

Boelman, N. T., Holbrook, J. D., Greaves, H. E., Krause, J. S., Chmura, H. E., Magney, T. S., Perez, J. H., Eitel, J. U. H., Gough, L., Vierling, K. T., Wingfield, J. C., and Vierling, L. A.: Airborne laser scanning and spectral remote sensing give a bird's eye perspective on arctic tundra breeding habitat at multiple spatial scales, Remote Sensing of Environment, 184, 337-349, https://doi.org/10.1016/j.rse.2016.07.012, 2016.

Brand, G., Crombaghs, M., Oude Elberink, S., Brügelmann, R., and de Min, E.: Precisiebeschrijving AHN 2002, Rijkswaterstaat AGI, 2003.

Cao, L., Coops, N. C., Innes, J. L., Sheppard, S. R. J., Fu, L., Ruan, H., and She, G.: Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data, Remote Sensing of Environment, 178, 158-171, https://doi.org/10.1016/j.rse.2016.03.012, 2016.

Choi, D. H., LaRue, E. A., Atkins, J. W., Foster, J. R., Matthes, J. H., Fahey, R. T., Thapa, B., Fei, S., and Hardiman, B. S.: Short-term effects of moderate severity disturbances on forest canopy structure, Journal of Ecology, 111, 1866-1881, https://doi.org/10.1111/1365-2745.14145, 2023.

Dalponte, M., Jucker, T., Liu, S., Frizzera, L., and Gianelle, D.: Characterizing forest carbon dynamics using multi-temporal lidar data, Remote Sensing of Environment, 224, 412-420, https://doi.org/10.1016/j.rse.2019.02.018, 2019.

Davies, A. B. and Asner, G. P.: Advances in animal ecology from 3D-LiDAR ecosystem mapping, Trends in Ecology & Evolution, 29, 681-691, https://doi.org/10.1016/j.tree.2014.10.005, 2014.

de Vries, J. P. R., Koma, Z., WallisDeVries, M. F., Kissling, W. D., and Tingley, R.: Identifying fine-scale habitat preferences of threatened butterflies using airborne laser scanning, Diversity and Distributions, 27, 1251-1264, https://doi.org/10.1111/ddi.13272, 2021.

Fekety, P. A., Falkowski, M. J., and Hudak, A. T.: Temporal transferability of LiDAR-based imputation of forest inventory attributes, Canadian Journal of Forest Research, 45, 422-435, https://doi.org/10.1139/cjfr-2014-0405, 2015.

Feng, T., Duncanson, L., Hancock, S., Montesano, P., Skakun, S., Wulder, M. A., White, J. C., Minor, D., and Loboda, T.: Characterizing Fire-Induced Forest Structure and Aboveground Biomass Changes in Boreal Forests Using Multitemporal Lidar and Landsat, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 10108-10125, https://doi.org/10.1109/JSTARS.2024.3400218, 2024.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., and Gloaguen, R.: Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art, IEEE Geoscience and Remote Sensing Magazine, 7, 6-39, https://doi.org/10.1109/MGRS.2018.2890023, 2019.

Goetz, S., Steinberg, D., Dubayah, R., and Blair, B.: Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA, Remote Sensing of Environment, 108, 254-263, https://doi.org/10.1016/j.rse.2006.11.016, 2007.

Hein, L., Remme, R. P., Schenau, S., Bogaart, P. W., Lof, M. E., and Horlings, E.: Ecosystem accounting in the Netherlands, Ecosystem Services, 44, 101118, https://doi.org/10.1016/j.ecoser.2020.101118, 2020.

Hill, R. A. and Thomson, A. G.: Mapping woodland species composition and structure using airborne spectral and LiDAR data, International Journal of Remote Sensing, 26, 3763-3779, https://doi.org/10.1080/01431160500114706, 2005.

Hopkinson, C., Chasmer, L., and Hall, R. J.: The uncertainty in conifer plantation growth prediction from multi-temporal lidar datasets, Remote Sensing of Environment, 112, 1168-1180, https://doi.org/10.1016/j.rse.2007.07.020, 2008.

Hudak, A. T., Strand, E. K., Vierling, L. A., Byrne, J. C., Eitel, J. U. H., Martinuzzi, S., and Falkowski, M. J.: Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys, Remote Sensing of Environment, 123, 25-40, https://doi.org/10.1016/j.rse.2012.02.023, 2012.

Isenburg, M.: LAStools-Efficient LiDAR Processing Software (version 141017, academic), obtained from http://rapidlasso.com, 2017.

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., and Turak, E.: Essential biodiversity variables for mapping and monitoring species populations, Nature Ecology & Evolution, 3, 539-551, https://doi.org/10.1038/s41559-019-0826-1, 2019.

Kamoske, A. G., Dahlin, K. M., Read, Q. D., Record, S., Stark, S. C., Serbin, S. P., and Zarnetske, P. L.: Towards mapping biodiversity from above: Can fusing lidar and hyperspectral remote sensing predict taxonomic, functional, and phylogenetic tree diversity in temperate forests?, Global Ecology and Biogeography, 31, 1440-1460, https://doi.org/10.1111/geb.13516, 2022.

Kharroubi, A., Poux, F., Ballouch, Z., Hajji, R., and Billen, R.: Three Dimensional Change Detection Using Point Clouds: A Review, Geomatics, 2, 457-485, https://doi.org/10.3390/geomatics2040025, 2022.

Khosravipour, A., Skidmore, A. K., Wang, T., Isenburg, M., and Khoshelham, K.: Effect of slope on treetop detection using a LiDAR Canopy Height Model, ISPRS Journal of Photogrammetry and Remote Sensing, 104, 44-52, https://doi.org/10.1016/j.isprsjprs.2015.02.013, 2015.

Kissling, W. D. and Shi, Y.: Which metrics derived from airborne laser scanning are essential to measure the vertical profile of ecosystems?, Diversity and Distributions, 29, 1315-1320, https://doi.org/10.1111/ddi.13760, 2023.

Kissling, W. D., Seijmonsbergen, A. C., Foppen, R. P. B., and Bouten, W.: eEcoLiDAR, eScience infrastructure for ecological applications of LiDAR point clouds: reconstructing the 3D ecosystem structure for animals at regional to continental scales, Research Ideas and Outcomes, 3, e14939, https://doi.org/10.3897/rio.3.e14939, 2017.

Kissling, W. D., Shi, Y., Koma, Z., Meijer, C., Ku, O., Nattino, F., Seijmonsbergen, A. C., and Grootes, M. W.: Laserfarm – A high-throughput workflow for generating geospatial data products of ecosystem structure from airborne laser scanning point clouds, Ecological Informatics, 72, 101836, https://doi.org/10.1016/j.ecoinf.2022.101836, 2022.

Kissling, W. D., Shi, Y., Koma, Z., Meijer, C., Ku, O., Nattino, F., Seijmonsbergen, A. C., and Grootes, M. W.: Country-wide data of ecosystem structure from the third Dutch airborne laser scanning survey, Data in Brief, 46, 108798, https://doi.org/10.1016/j.dib.2022.108798, 2023.

Kissling, W. D., Walls, R., Bowser, A., Jones, M. O., Kattge, J., Agosti, D., Amengual, J., Basset, A., van Bodegom, P. M., Cornelissen, J. H. C., Denny, E. G., Deudero, S., Egloff, W., Elmendorf, S. C., Alonso

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

García, E., Jones, K. D., Jones, O. R., Lavorel, S., Lear, D., Navarro, L. M., Pawar, S., Pirzl, R., Rüger, N., Sal, S., Salguero-Gómez, R., Schigel, D., Schulz, K.-S., Skidmore, A., and Guralnick, R. P.: Towards global data products of Essential Biodiversity Variables on species traits, Nature Ecology & Evolution, 2, 1531-1540, https://doi.org/10.1038/s41559-018-0667-3, 2018.

Koma, Z., Seijmonsbergen, A. C., and Kissling, W. D.: Classifying wetland-related land cover types and habitats using fine-scale lidar metrics derived from country-wide Airborne Laser Scanning, Remote Sensing in Ecology and Conservation, 7, 80-96, https://doi.org/10.1002/rse2.170, 2021a.

Koma, Z., Grootes, M. W., Meijer, C. W., Nattino, F., Seijmonsbergen, A. C., Sierdsema, H., Foppen, R., and Kissling, W. D.: Niche separation of wetland birds revealed from airborne laser scanning, Ecography, 44, 907-918, https://doi.org/10.1111/ecog.05371, 2021b.

Kruskal, W. H. and Wallis, W. A.: Use of ranks in one-criterion variance analysis, Journal of the American statistical Association, 47, 583-621, https://www.jstor.org/stable/2280779, 1952.

Kucharik, C. J., Foley, J. A., Delire, C., Fisher, V. A., Coe, M. T., Lenters, J. D., Young-Molling, C., Ramankutty, N., Norman, J. M., and Gower, S. T.: Testing the performance of a dynamic global ecosystem model: water balance, carbon balance, and vegetation structure, Global Biogeochemical Cycles, 14, 795-825, https://doi.org/10.1029/1999GB001138, 2000.

Lamonaca, A., Corona, P., and Barbati, A.: Exploring forest structural complexity by multi-scale segmentation of VHR imagery, Remote Sensing of Environment, 112, 2839-2849, https://doi.org/10.1016/j.rse.2008.01.017, 2008.

Lefsky, M. A., Cohen, W. B., Parker, G. G., and Harding, D. J.: Lidar Remote Sensing for Ecosystem Studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists, BioScience, 52, 19-30, https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2, 2002.

Li, H., Hiroshima, T., Li, X., Hayashi, M., and Kato, T.: High-resolution mapping of forest structure and carbon stock using multi-source remote sensing data in Japan, Remote Sensing of Environment, 312, 114322, https://doi.org/10.1016/j.rse.2024.114322, 2024.

Li, Q., Bonebrake, T. C., Michalski, J. R., Kit Wong, F. K., and Fung, T.: Combining multi-temporal airborne LiDAR and Sentinel-2 multispectral data for assessment of disturbances and recovery of mangrove forests, Estuarine, Coastal and Shelf Science, 291, 108444, https://doi.org/10.1016/j.ecss.2023.108444, 2023.

Lin, Y.-C., Shao, J., Shin, S.-Y., Saka, Z., Joseph, M., Manish, R., Fei, S., and Habib, A.: Comparative Analysis of Multi-Platform, Multi-Resolution, Multi-Temporal LiDAR Data for Forest Inventory, Remote Sensing, 14, 649, https://doi.org/10.3390/rs14030649, 2022.

Lindenmayer, D. B., Margules, C. R., and Botkin, D. B.: Indicators of Biodiversity for Ecologically Sustainable Forest Management, Conservation Biology, 14, 941-950, https://doi.org/10.1046/j.1523-1739.2000.98533.x, 2000.

Liu, J., Skidmore, A. K., Heurich, M., and Wang, T.: Significant effect of topographic normalization of airborne LiDAR data on the retrieval of plant area index profile in mountainous forests, ISPRS Journal of Photogrammetry and Remote Sensing, 132, 77-87, https://doi.org/10.1016/j.isprsjprs.2017.08.005, 2017.

McCarley, T. R., Kolden, C. A., Vaillant, N. M., Hudak, A. T., Smith, A. M. S., Wing, B. M., Kellogg, B. S., and Kreitler, J.: Multi-temporal LiDAR and Landsat quantification of fire-induced changes to forest structure, Remote Sensing of Environment, 191, 419-432, https://doi.org/10.1016/j.rse.2016.12.022, 2017.

Meijer, C., Grootes, M. W., Koma, Z., Dzigan, Y., Gonçalves, R., Andela, B., van den Oord, G., Ranguelova, E., Renaud, N., and Kissling, W. D.: Laserchicken—A tool for distributed feature calculation from massive LiDAR point cloud datasets, SoftwareX, 12, 100626, https://doi.org/10.1016/j.softx.2020.100626, 2020.

Miura, N. and Jones, S. D.: Characterizing forest ecological structure using pulse types and heights of airborne laser scanning, Remote Sensing of Environment, 114, 1069-1076, https://doi.org/10.1016/j.rse.2009.12.017, 2010.

Moeslund, J. E., Zlinszky, A., Ejrnæs, R., Brunbjerg, A. K., Bøcher, P. K., Svenning, J.-C., and Normand, S.: Light detection and ranging explains diversity of plants, fungi, lichens, and bryophytes across multiple habitats and large geographic extent, Ecological Applications, 29, e01907, https://doi.org/10.1002/eap.1907, 2019.

Morsdorf, F., Mårell, A., Koetz, B., Cassagne, N., Pimont, F., Rigolot, E., and Allgöwer, B.: Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning, Remote Sensing of Environment, 114, 1403-1415, https://doi.org/10.1016/j.rse.2010.01.023, 2010.

Moudrý, V., Cord, A. F., Gábor, L., Laurin, G. V., Barták, V., Gdulová, K., Malavasi, M., Rocchini, D., Stereńczak, K., Prošek, J., Klápště, P., and Wild, J.: Vegetation structure derived from airborne laser scanning to assess species distribution and habitat suitability: The way forward, Diversity and Distributions, 29, 39-50, https://doi.org/10.1111/ddi.13644, 2022.

Musthafa, M. and Singh, G.: Improving forest above-ground biomass retrieval using multi-sensor L-and C-Band SAR data and multi-temporal spaceborne LiDAR data, Frontiers in Forests and Global Change, 5, 822704, https://doi.org/10.3389/ffgc.2022.822704, 2022.

Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J., Egberth, M., Larsson, S., Nilsson, L., Eriksson, J., and Olsson, H.: A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory, Remote Sensing of Environment, 194, 447-454, https://doi.org/10.1016/j.rse.2016.10.022, 2017.

Noordermeer, L., Økseter, R., Ørka, H. O., Gobakken, T., Næsset, E., and Bollandsås, O. M.: Classifications of Forest Change by Using Bitemporal Airborne Laser Scanner Data, Remote Sensing, 11, 2145, https://doi.org/10.3390/rs11182145, 2019.

Noss, R. F.: Indicators for Monitoring Biodiversity: A Hierarchical Approach, Conservation Biology, 4, 355-364, https://doi.org/10.1111/j.1523-1739.1990.tb00309.x, 1990.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D. S., McGeoch, M. A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J. P. W., Stuart, S. N., Turak, E., Walpole, M., and Wegmann, M.: Essential Biodiversity Variables, Science, 339, 277-278, https://doi.org/10.1126/science.1229931, 2013.

Pfeifer, N., Mandlburger, G., Otepka, J., and Karel, W.: OPALS – A framework for Airborne Laser Scanning data analysis, Computers, Environment and Urban Systems, 45, 125-136, https://doi.org/10.1016/j.compenvurbsys.2013.11.002, 2014.

Riofrío, J., White, J. C., Tompalski, P., Coops, N. C., and Wulder, M. A.: Harmonizing multi-temporal airborne laser scanning point clouds to derive periodic annual height increments in temperate mixedwood forests, Canadian Journal of Forest Research, 52, 1334-1352, https://doi.org/10.1139/cjfr-2022-0055, 2022.

Roussel, J.-R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R. H., Meador, A. S., Bourdon, J.-F., de Boissieu, F., and Achim, A.: lidR: An R package for analysis of Airborne Laser Scanning (ALS) data, Remote Sensing of Environment, 251, 112061, https://10.1016/j.rse.2020.112061, 2020.

Ruiz-Jaén, M. C. and Aide, T. M.: Vegetation structure, species diversity, and ecosystem processes as measures of restoration success, Forest Ecology and Management, 218, 159-173, https://doi.org/10.1016/j.foreco.2005.07.008, 2005.

Shi, Y., & Kissling, W. D.: Multi-temporal high-resolution data products of ecosystem structure derived from country-wide airborne laser scanning surveys of the Netherlands [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13940846, 2024

Shi, Y. and Kissling, W. D.: Performance, effectiveness and computational efficiency of powerline extraction methods for quantifying ecosystem structure from light detection and ranging, GIScience & Remote Sensing, 60, 2260637, https://doi.org/10.1080/15481603.2023.2260637, 2023.

Shi, Y., Wang, T., Skidmore, A. K., and Heurich, M.: Important LiDAR metrics for discriminating forest tree species in Central Europe, ISPRS Journal of Photogrammetry and Remote Sensing, 137, 163-174, https://doi.org/10.1016/j.isprsjprs.2018.02.002, 2018a.

Shi, Y., Skidmore, A. K., Wang, T., Holzwarth, S., Heiden, U., Pinnel, N., Zhu, X., and Heurich, M.: Tree species classification using plant functional traits from LiDAR and hyperspectral data, International Journal of Applied Earth Observation and Geoinformation, 73, 207-219, https://doi.org/10.1016/j.jag.2018.06.018, 2018b.

Shoot, C., Andersen, H.-E., Moskal, L. M., Babcock, C., Cook, B. D., and Morton, D. C.: Classifying Forest Type in the National Forest Inventory Context with Airborne Hyperspectral and Lidar Data, Remote Sensing, 13, 1863, https://doi.org/10.3390/rs13101863, 2021.

Shugart, H. H., Saatchi, S., and Hall, F. G.: Importance of structure and its measurement in quantifying function of forest ecosystems, Journal of Geophysical Research: Biogeosciences, 115, https://doi.org/10.1029/2009JG000993, 2010.

Swart, L.: How the Up-to-date Height Model of the Netherlands (AHN) became a massive point data cloud, NCG KNAW, 17, 17-32, 2010.

Toivonen, J., Kangas, A., Maltamo, M., Kukkonen, M., and Packalen, P.: Assessing biodiversity using forest structure indicators based on airborne laser scanning data, Forest Ecology and Management, 546, 121376, https://doi.org/10.1016/j.foreco.2023.121376, 2023.

Tompalski, P., Coops, N. C., White, J. C., Goodbody, T. R. H., Hennigar, C. R., Wulder, M. A., Socha, J., and Woods, M. E.: Estimating Changes in Forest Attributes and Enhancing Growth Projections: a Review of Existing Approaches and Future Directions Using Airborne 3D Point Cloud Data, Current Forestry Reports, 7, 1-24, https://doi.org/10.1007/s40725-021-00135-w, 2021.

Trotto, T., Coops, N. C., Achim, A., Gergel, S. E., and Roeser, D.: Characterizing forest structural changes in response to non-stand replacing disturbances using bitemporal airborne laser scanning data, Science of Remote Sensing, 10, 100160, https://doi.org/10.1016/j.srs.2024.100160, 2024.

Valbuena, R., O'Connor, B., Zellweger, F., Simonson, W., Vihervaara, P., Maltamo, M., Silva, C. A., Almeida, D. R. A., Danks, F., Morsdorf, F., Chirici, G., Lucas, R., Coomes, D. A., and Coops, N. C.: Standardizing ecosystem morphological traits from 3D information sources, Trends in Ecology & Evolution, 35, 656-667, https://doi.org/10.1016/j.tree.2020.03.006, 2020.

Vastaranta, M., Wulder, M. A., White, J. C., Pekkarinen, A., Tuominen, S., Ginzler, C., Kankare, V., Holopainen, M., Hyyppä, J., and Hyyppä, H.: Airborne laser scanning and digital stereo imagery measures of forest structure: comparative results and implications to forest mapping and inventory update, Canadian Journal of Remote Sensing, 39, 382-395, https://doi.org/10.5589/m13-046, 2013.

Vepakomma, U., St-Onge, B., and Kneeshaw, D.: Response of a boreal forest to canopy opening: assessing vertical and lateral tree growth with multi-temporal lidar data, Ecological Applications, 21, 99-121, https://doi.org/10.1890/09-0896.1, 2011.

Vierling, K. T., Vierling, L. A., Gould, W. A., Martinuzzi, S., and Clawges, R. M.: Lidar: shedding new light on habitat characterization and modeling, Frontiers in Ecology and the Environment, 6, 90-98, https://doi.org/10.1890/070001, 2008.

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., and Tompalski, P.: Remote Sensing Technologies for Enhancing Forest Inventories: A Review, Canadian Journal of Remote Sensing, 42, 619-641, https://doi.org/10.1080/07038992.2016.1207484, 2016.

https://doi.org/10.5194/essd-2024-488
Preprint. Discussion started: 25 November 2024

Wilcoxon, F., Katti, S., and Wilcox, R. A.: Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test, Selected tables in mathematical statistics, 1, 171-259, 1970.

Wolter, P. T., Townsend, P. A., and Sturtevant, B. R.: Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data, Remote Sensing of Environment, 113, 2019-2036, https://doi.org/10.1016/j.rse.2009.05.009, 2009.

Xu, H., Cheng, L., Li, M., Chen, Y., and Zhong, L.: Using Octrees to Detect Changes to Buildings and Trees in the Urban Environment from Airborne LiDAR Data, Remote Sensing, 7, 9682-9704, https://doi.org/10.3390/rs70809682, 2015.

Zellweger, F., Braunisch, V., Baltensweiler, A., and Bollmann, K.: Remotely sensed forest structural complexity predicts multi species occurrence at the landscape scale, Forest Ecology and Management, 307, 303-312, https://doi.org/10.1016/j.foreco.2013.07.023, 2013.

Zellweger, F., Roth, T., Bugmann, H., and Bollmann, K.: Beta diversity of plants, birds and butterflies is closely associated with climate and habitat structure, Global Ecology and Biogeography, 26, 898-906, https://doi.org/10.1111/geb.12598, 2017.

Zenner, E. K. and Hibbs, D. E.: A new method for modeling the heterogeneity of forest structure, Forest Ecology and Management, 129, 75-87, https://doi.org/10.1016/S0378-1127(99)00140-1, 2000.

Zhao, K., Suarez, J. C., Garcia, M., Hu, T., Wang, C., and Londo, A.: Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux, Remote Sensing of Environment, 204, 883-897, https://doi.org/10.1016/j.rse.2017.09.007, 2018.

Zheng, Z., Zeng, Y., Schneider, F. D., Zhao, Y., Zhao, D., Schmid, B., Schaepman, M. E., and Morsdorf, F.: Mapping functional diversity using individual tree-based morphological and physiological traits in a subtropical forest, Remote Sensing of Environment, 252, 112170, https://doi.org/10.1016/j.rse.2020.112170, 2021.