AGILE: GIScience Series, 7, 18, 2026. https://doi.org/10.5194/agile-giss-7-18-2026
Proceedings of the 29th AGILE Conference on Geographic Information Science, 16–19 June 2026.
Eds.: Evelyn Uuemaa, Alexander Kmoch, Holger Virro, and Simon Scheider
This contribution underwent peer review based on a full paper submission.

Towards National-Scale Ecological Applications: Harmonised Framework for LiDAR Point Cloud Processing for Vegetation Metrics Calculation

Eveli Sisas [1], Holger Virro [1], Wai Tik Chan [1], Alexander Kmoch [1], Aveliina Helm [2], and Evelyn Uuemaa [1]

[1]Landscape Geoinformatics Lab, Institute of Ecology and Earth Sciences, University of Tartu, Estonia
[2]Landscape Biodiversity Group, Institute of Ecology and Earth Sciences, University of Tartu, Estonia

Correspondence: Eveli Sisas (eveli.sisas@ut.ee)

Abstract.

National airborne laser scanning (ALS) datasets provide essential three-dimensional information for large-scale vegetation metric calculation, but their ecological use is constrained by multi-year acquisition cycles, seasonal variation, heterogeneous point densities, and inconsistent point classification. These factors reduce the reliability and comparability of vegetation metrics derived directly from national ALS archives.

We developed a harmonised, reproducible point-cloud preprocessing framework for national-scale ecological applications of ALS data. The framework corrects metadata and file-format inconsistencies, removes overlap-related sampling artefacts, applies digital terrain model (DTM)-based height normalisation, and performs deterministic rule-based reclassification using height information, seasonal normalised difference vegetation index, and ancillary vector data. The workflow was applied to selected 2019–2024 acquisition campaigns of the Estonian national ALS archive, comprising approximately 132 000 LAZ files from leaf-off and leaf-on surveys.

Harmonised preprocessing increased thematic consistency of vegetation, ground, building, and water classes across acquisition campaigns by correcting systematic classification and sampling artefacts. Removal of overlap-flagged points reduced local point-density inflation by approximately 25 %, thereby improving comparability of later grid-based vegetation metrics. Rule-based reclassification reassigned a median of 32 % of previously unclassified returns to vegetation, increasing completeness and seasonal consistency of vegetation representation. DTM-based height normalisation produced stable and interpretable height distributions across campaigns while retaining expected seasonal structural differences.

By providing a more consistent point-level representation of vegetation structure, the framework offers a robust basis for vegetation metric calculation and national-scale ecological analysis using heterogeneous ALS archives.

Submission Type. software; case study

BoK Concepts. [GD2-2] Remote sensing; [GD4-1] Data quality

Keywords. airborne laser scanning, LiDAR harmonisation, vegetation structure, national-scale analysis, reproducible workflows

1 Introduction

Light detection and ranging (LiDAR) is widely used for characterising three-dimensional vegetation structure across ecosystems. By capturing the vertical distribution of canopy and sub-canopy elements, airborne laser scanning (ALS) enables the quantification of vegetation height, cover, foliage stratification, and structural complexity that are not directly measurable from two-dimensional spectral imagery alone (Lefsky et al., 2002). These structural descriptors are closely linked to habitat heterogeneity and have repeatedly been shown to improve the characterisation and modelling of biodiversity patterns across taxa and habitat types (Moeslund et al., 2019; Acebes et al., 2021). At landscape and habitat-assessment scales, LiDAR-derived measures of canopy height and structural heterogeneity complement field-based habitat mapping and support the discrimination of forest habitat types in conservation contexts (Zellweger et al., 2014; Bässler et al., 2011). Beyond broad cross-taxon evidence, LiDAR-derived structural variables have been shown to predict vascular plant species richness and,

in some contexts, to act as proxies for fungal richness and composition in temperate forests, reflecting the sensitivity of biodiversity patterns to vertical vegetation structure (Lopatin et al., 2016; Thers et al., 2017). Even ALS datasets with moderate pulse densities can yield ecologically meaningful structure descriptors; however, many height- and density-based metrics remain sensitive to both between- and within-survey variation in pulse-density, including local increases caused by overlapping flight strips (Campbell et al., 2018; Toivonen et al., 2023; Roussel et al., 2017).

The increasing availability of national-scale ALS data has enabled ecological analyses at broad spatial extents, but their uptake has long been limited by substantial computational demands and the need for specialist processing expertise (Assmann et al., 2022; Kissling et al., 2022). In practice, biodiversity and forest-structure studies increasingly rely on governmental ALS datasets originally acquired for mapping purposes rather than on dedicated ecological flight campaigns (Bald et al., 2024). However, most national ALS programmes are designed primarily for topographic rather than ecological applications, and the resulting archives are often heterogeneous with respect to acquisition year, season, flight altitude, scan angle, sensor configuration, and point density, an issue explicitly noted in recent ecological modelling studies (Bald et al., 2024). This heterogeneity affects canopy penetration, understorey detectability, and ground classification reliability, with direct consequences for vegetation metrics that assume consistent sampling and height-reference conditions (Toivonen et al., 2023). Although repeated national ALS surveys are becoming more common, analysis-ready ecosystem-structure products remain comparatively scarce because raw point-cloud archives still require extensive preprocessing before integration into ecological workflows (Assmann et al., 2022). This gap highlights the need for scalable, reproducible processing frameworks that can transform multi-terabyte ALS holdings into transparent, reusable structure descriptors (Kissling et al., 2022; Meijer et al., 2020).

In addition to acquisition heterogeneity, point classification practices introduce further complexity. Although the American Society for Photogrammetry and Remote Sensing (ASPRS) LAS 1.4 standard defines a common set of point classes (American Society for Photogrammetry and Remote Sensing, 2019), its implementation varies across data providers and acquisition cycles. In many national ALS datasets, classification efforts prioritise terrain and built structures, while vegetation classes are applied less consistently. As a result, substantial proportions of vegetation returns may remain assigned to the *Unclassified* class, propagating uncertainty into downstream vegetation metrics and limiting their ecological interpretability (Valbuena et al., 2020; Kissling et al., 2022). This limitation is also reflected in national ecosystem-structure products that rely on provider classifications for class-based layers, such as point counts by class or canopy openness, underscoring the importance of explicitly documenting classification assumptions (Assmann et al., 2022). Improving vegetation class consistency may therefore require combining height-based rules with seasonal spectral indices and ancillary spatial data.

Height normalisation represents a further recurring source of uncertainty in multi-temporal ALS datasets derived from repeated national surveys (Tompalski et al., 2021; Riofrío et al., 2022; Shi et al., 2025). In regions where acquisitions span both leaf-off and leaf-on conditions, reliable ground reference estimation is particularly challenging because dense canopy cover can limit terrain visibility (Kobler et al., 2007; Roussel et al., 2017). Differences in acquisition season, pulse density, and vegetation structure can therefore introduce systematic inconsistencies between campaigns (Roussel et al., 2017; Marinelli et al., 2018). These artefacts propagate into canopy height products and LiDAR-derived vegetation metrics and may obscure or mimic true structural change rather than reflect ecological dynamics. Evidence from multi-temporal LiDAR change-detection studies further shows that variation in sampling density, point distribution, and acquisition configuration can complicate the interpretation of structural change, emphasising the need to distinguish ecological signals from acquisition- and processing-related artefacts in repeated surveys (Marinelli et al., 2018; Shi et al., 2025).

In response to these challenges, we present a harmonised and reproducible preprocessing framework for national-scale ecological applications of ALS data. National ALS archives commonly vary in acquisition strategy and classification practice across campaigns, while overlapping flight strips can introduce substantial local point-density artefacts that complicate the direct ecological use of raw point clouds (Kissling et al., 2022). We demonstrate the framework using Estonia, whose national ALS programme provides a representative example because it covers the entire country in a four-year cycle and combines leaf-off (spring) and leaf-on (summer) campaigns acquired at different altitudes, resulting in variable point densities (Estonian Land and Spatial Development Board, 2025d). These characteristics exemplify the broader challenges of reusing national ALS data for ecological applications and underscore the need for harmonised preprocessing.

This study investigates how a heterogeneous national airborne laser-scanning archive can be transformed into a more consistent and ecologically usable dataset. Using the Estonian national ALS archive, we develop a reproducible workflow that combines metadata harmonisation, overlap handling, digital terrain model (DTM)-based height normalisation, and deterministic rule-based reclassification. We evaluate its effects using internal diagnostic measures that quantify changes in point-density consistency, class transitions, and height above ground (HAG) behaviour. The study therefore

addresses the following research question: *To what extent does a harmonised preprocessing workflow improve the internal consistency, completeness, and interpretability of a heterogeneous national ALS archive, relative to the raw source data, for vegetation-metric calculation?* In doing so, it positions point-cloud harmonisation as a necessary intermediate step between mapping-oriented ALS acquisition and downstream vegetation metric calculation for ecological analysis.

2 Data and Study Area

2.1 Study Area

Estonia covers approximately 45 000 km^2 and is located in northeastern Europe. The country is characterised by low topographic relief, with elevations ranging from sea level to 318 m. Forests dominate the landscape, interspersed with wetlands, peatlands, agricultural areas, and semi-natural habitats, including alvars, coastal and wooded meadows, floodplains, mires, dunes, and a dense network of lakes and rivers. The combination of glacial landforms, heterogeneous soils, and an indented coastline results in high structural and habitat diversity, posing challenges for consistent ALS-based vegetation analysis across land-cover types and acquisition seasons.

The climate is humid continental (Köppen Dfb), with cold winters and moderately warm summers (Beck et al., 2018). Seasonal contrasts influence vegetation phenology, canopy closure, and ground visibility, which in turn affect ALS signal penetration and ground-reference estimation. Forests cover approximately 50 % of the national territory. Together with open and semi-natural ecosystems, this provides a suitable setting for evaluating LiDAR-derived vegetation metrics under contrasting seasonal conditions.

2.2 ALS Data

National airborne laser scanning in Estonia is conducted on a continuous four-year acquisition cycle, resulting in complete territorial coverage during each period. Each year, approximately one quarter of the country is surveyed through two campaigns: a spring campaign under leaf-off conditions and a summer campaign under leaf-on conditions (Estonian Land and Spatial Development Board, 2025d). The data are publicly accessible via the Estonian Land and Spatial Development Board (ELSDB) Geoportal (Estonian Land and Spatial Development Board, 2025c).

This study uses data from the two most recent acquisition cycles: 2017–2020 (ALS III) and 2021–2024 (ALS IV). From ALS III, only the 2019 and 2020 campaigns (two spring and two summer) were included, reflecting initial availability requirements for vegetation metrics. All campaigns from ALS IV were used, comprising four spring and four summer acquisitions. Hereafter, references to ALS III and ALS IV refer only to the acquisition years included in this study, unless stated otherwise.

In this paper, we use the term *national ALS archive* to refer to the multi-year collection of ALS point-cloud files distributed by ELSDB via the Geoportal, and *dataset used in this study* to refer to the subset of that archive processed and analysed here (2019–2024).

All data were collected by ELSDB using a Riegl VQ-1560i airborne laser scanner, ensuring consistent sensor type and acquisition technology (Estonian Land and Spatial Development Board, 2025d). Each seasonal campaign consists of approximately 11 000 spatial tiles, resulting in approximately 132 000 LAZ files processed in this study.

Each LAZ file corresponds to a 1×1 km tile aligned with the national 1:2 000 topographic map sheet grid. Vertical coordinates are referenced to the Estonian Height System 2000 (EH2000; based on the European Vertical Reference System *EPSG:5621*), and horizontal coordinates follow the L-EST97 reference system (*EPSG:3301*). The point clouds conform to the ASPRS LAS 1.4 specification (American Society for Photogrammetry and Remote Sensing, 2019) and include standard attributes such as three-dimensional coordinates, return information, intensity, scan angle rank, GPS time, and classification codes.

Points located within flight-strip overlap zones are flagged using the LAS 1.4 `overlap` flag. Such overlaps are a well-known source of fine-scale pulse-density variation that can produce striping patterns and locally inflated sampling intensity unless treated explicitly during preprocessing or downstream processing (Roussel et al., 2017).

The two seasonal campaigns differ substantially in flight parameters and point density. The spring campaign is optimised for terrain modelling and is flown at an average altitude of approximately 2600 m, yielding a mean point density of about 2.1 points m^{-2}. The summer campaign prioritises aerial imagery for forestry applications and is flown at a higher altitude of approximately 3100 m, resulting in a lower mean point density of about 0.8 points m^{-2} (Estonian Land and Spatial Development Board, 2025a). These differences affect canopy penetration, ground-return availability, and the stability of height- and density-based vegetation metrics across campaigns.

The raw point clouds include preclassified classes for *Unclassified* (Class 1), *Ground* (Class 2), *High vegetation* (Class 5), *Buildings* (Class 6), *Water* (Class 9), and noise (Classes 7 and 18). ELSDB classification practices prioritise accurate representation of ground and built structures. Most high and medium vegetation (first and intermediate returns) is assigned to the *High vegetation* class, while low vegetation is frequently assigned to *Unclassified* or, in some cases, to *Ground*, particularly under dense leaf-on canopy conditions (Estonian Land and

Spatial Development Board, 2025b). In the reclassified data, Class 5 is used as a unified vegetation category (Section 3.5).

LiDAR filenames encode the spatial tile identifier, acquisition year, and campaign type, where *tava* denotes spring (leaf-off) and *mets* denotes summer (leaf-on) acquisitions (e.g. `377650_2023_tava.laz`).

2.3 Ancillary Data

To support height normalisation and reclassification, two LiDAR-derived DTMs with 1 m resolution were used as ancillary data (Estonian Land and Spatial Development Board, 2025c). The DTMs are mosaics compiled by ELSDB from multiple ALS campaigns, predominantly acquired under leaf-off conditions, with additional high-accuracy low-altitude data in urban areas (Estonian Land and Spatial Development Board, 2025d). For consistency with the temporal structure of the LiDAR data used in this study, the DTM corresponding to the 2017–2020 acquisition cycle was applied to the 2019–2020 campaigns, and the DTM corresponding to the 2021–2024 cycle was applied to the later campaigns.

Topographic vector data were obtained from the Estonian Topographic Database (ETAK; Eesti topograafia andmekogu), including layers describing buildings and other constructions, inland and coastal waterbodies, and overhead electricity transmission lines (Estonian Land and Spatial Development Board, 2025e). For each LiDAR acquisition year, the temporally corresponding ETAK release was used to minimise discrepancies between vector and point-cloud data (Estonian Land and Spatial Development Board, 2025f).

Seasonal normalised difference vegetation index (NDVI) composites derived from ESA Copernicus Sentinel-2 Level-2A imagery were used to support vegetation-related reclassification. Spring composites represent April-May conditions and summer composites represent June-August conditions for each year from 2019 to 2024. These NDVI datasets were produced prior to the present study by the Landscape Geoinformatics Lab using Google Earth Engine and Python and were used here as auxiliary input data. All composites are aligned to the Estonian national grid and cover the entire terrestrial area of Estonia, including a 500 m coastal buffer. NDVI composites were provided by the Landscape Geoinformatics Lab; information on their availability is provided in Section 3.7.

3 Methods

This section describes a reproducible preprocessing framework for harmonising national-scale ALS point clouds for ecological analysis using vegetation metrics. It addresses metadata and file-format inconsistencies, overlap-related sampling artefacts, height-reference instability, and recurring classification errors through five stages: data download, quality control, data harmonisation, height normalisation, and reclassification.

Fig. 1 summarises these stages, which are described in the following subsections.

The workflow was implemented in Python using Laspy and PDAL and executed on the University of Tartu High Performance Computing (HPC) platform, with processing parallelised across individual tile-based LAZ files. Information on software availability, computational infrastructure, and reproducibility scope is provided in Section 3.7. Full implementation details are documented in the archived code repository.

3.1 Data Download

We developed an automated script to batch-download ALS LAZ files and the corresponding DTMs from the ELSDB Geoportal (Estonian Land and Spatial Development Board, 2025c). The script programmatically interacted with the portal's HTML search interface and queried available tiles based on the national map sheet grid. For each map-sheet grid cell covering land areas, it retrieved the available spring (leaf-off) and summer (leaf-on) point-cloud files (Section 2.2), stored them in object storage, and indexed them in a processing database. File integrity was verified using size and checksum validation.

During HPC processing, file metadata and processing status were tracked in a PostgreSQL database to support parallel execution, recovery, and quality control.

This automated approach reduced manual handling errors and ensured consistent file naming, systematic data organisation, and reproducible large-scale processing.

3.2 Quality Control

We performed quality control before large-scale processing to identify classification, spatial-alignment, and metadata issues. Because the full range of problems was not known in advance, we began with an exploratory visual inspection. Based on ELSDB documentation, expected artefacts included locally increased point densities in flight-strip overlap zones and vegetation returns assigned to *Unclassified* or misclassified as *Ground* under dense leaf-on vegetation conditions (Section 2.2).

We visually inspected 120 tiles using QGIS 3.40, examining 10 tiles per acquisition campaign (12 campaigns in total). Tiles were selected to represent major land-cover types in Estonia, including forests, semi-natural grasslands, agricultural areas, and wetlands. This same 120-tile subset was subsequently used for workflow evaluation (Section 3.6).

The inspection confirmed the expected classification-related artefacts and identified two recurring issue categories relevant for preprocessing. First, metadata

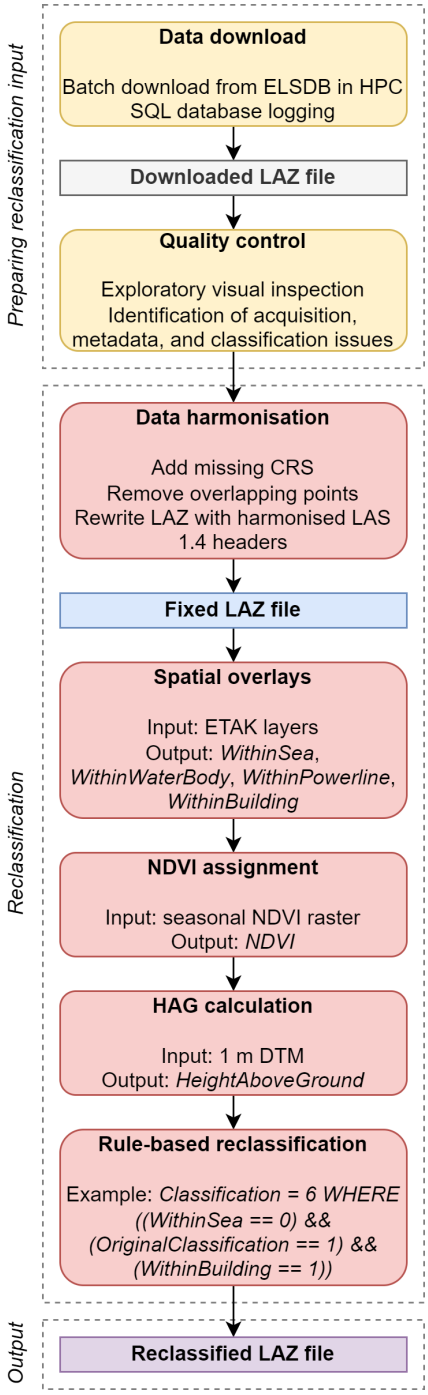


Figure 1. Overview of the preprocessing and reclassification workflow applied to national-scale ALS point clouds. Preparatory stages include automated data download, exploratory quality control, and file harmonisation. Reclassification integrates spatial overlays, seasonal NDVI assignment, and DTM-based height normalisation to derive height above ground (HAG) values, supporting deterministic rule-based class updates and improved cross-campaign consistency.

and format inconsistencies were observed, including missing coordinate reference system (CRS) information and mismatches between declared point data formats and extra-byte specifications. Second, classification inconsistencies were detected, including vegetation assigned to the *Unclassified* class, vegetation misclassified as *Ground* under dense leaf-on vegetation, and occasional misclassification of buildings or linear infrastructure. These findings directly informed the harmonisation and rule-based reclassification procedures described below.

Quality control continued during workflow development through inspection of intermediate outputs, and thresholds were refined iteratively to improve cross-campaign consistency.

3.3 Data Harmonisation

We harmonised tile-based ALS LAZ files before HAG computation and reclassification to standardise spatial referencing, point sampling, and file structure.

The harmonisation workflow comprised three steps. First, we assigned a CRS to files lacking spatial reference information by explicitly encoding the national reference system L-EST97. Second, we removed points carrying the LAS 1.4 `overlap` flag. These points represent repeated sampling in flight-strip overlap zones caused by acquisition geometry rather than additional ecological structure. Because the subsequent products are vegetation metrics calculated in regular 10 m horizontal grid cells, retaining overlap-flagged points would oversample some cells for technical rather than ecological reasons. We therefore removed points with `overlap`=1 using the provider's explicit point-level overlap annotation while leaving non-overlap observations unchanged. Third, we rewrote the LAZ files using Laspy to regenerate consistent headers and point-dimension registration, including correction of inconsistencies between declared point-data formats and registered extra-byte specifications, thereby enabling reliable downstream reading of all attributes in PDAL.

Together, these steps standardised metadata, point structure, and sampling characteristics across campaigns while preserving point coordinates and the geometric accuracy of the source data.

3.4 Height Normalisation

As shown in Fig. 1, height normalisation was implemented by computing HAG as the difference between each point's raw elevation (z) and the elevation of the corresponding 1 m DTM pixel. This provided a stable ground reference and supported consistent application of height-based thresholds across acquisition campaigns.

During preliminary testing on representative tiles, we evaluated two alternative height normalisation strategies: (i) Laserchicken's lowest-point subcell approach (Kissling et al., 2022) and (ii) PDAL's `hag_delaunay` interpolation. Both produced unstable height estimates for leaf-on datasets because dense summer vegetation often prevented ground returns from

being identified as local minima, inflating local HAG values and reducing spatial consistency across areas with dense vegetation cover. Such sensitivity of height metrics to sampling configuration, including pulse density, is well documented and can confound multi-temporal interpretation if acquisition-driven differences are not controlled (Roussel et al., 2017; Marinelli et al., 2018).

We therefore selected DTM-based normalisation for national-scale processing. By using an external terrain reference for both leaf-off and leaf-on campaigns, this approach reduced sensitivity to seasonal variation in ground-return availability and supported consistent application of height thresholds across years.

3.5 Reclassification

The final stage was a deterministic, rule-based reclassification of ALS point classes. Based on the issues identified during quality control, we designed the procedure to improve thematic consistency across acquisition campaigns, with particular emphasis on vegetation and ground classes, while also correcting recurring errors involving buildings, waterbodies, and transmission-line corridors.

A simplified summary of the reclassification logic is shown in Fig. 2. The procedure combined spatial overlays, seasonal NDVI information, DTM-based HAG values, preservation of the original class labels in the `OriginalClassification` attribute, and deterministic class reassignment rules. Points within sea polygons, as well as points outside the DTM extent, were assigned HAG = 0 to avoid undefined values during rule application.

The main target errors involved vegetation assigned to *Unclassified* (Class 1), especially low and intermediate returns, and vegetation misclassified as *Ground* (Class 2) under dense leaf-on vegetation. Additional recurring errors included vegetation incorrectly classified as *Building* (Class 6), mainly in earlier campaigns, transmission-line structures influencing vegetation classification, non-vegetation features remaining in *Unclassified*, and, in some cases, ground misclassified as *Water* (Class 9).

These issues were corrected using HAG thresholds, seasonal NDVI thresholds, and spatial overlay operations based on ancillary vector data, including building footprints, waterbodies, and buffers around high-voltage transmission lines (Table 1). A buffer width of 13 m was applied to transmission line geometries to identify areas influenced by transmission line structures and associated vegetation clearance. Thresholds were determined through iterative visual validation and empirical testing. A height tolerance of 0.20 m was adopted to account for vertical discrepancies across DTM mosaics and seasonal acquisitions. NDVI thresholds of 0.33 and 0.40 were selected on the basis of NDVI distribution inspection to balance vegetation detection against misclassification risk.

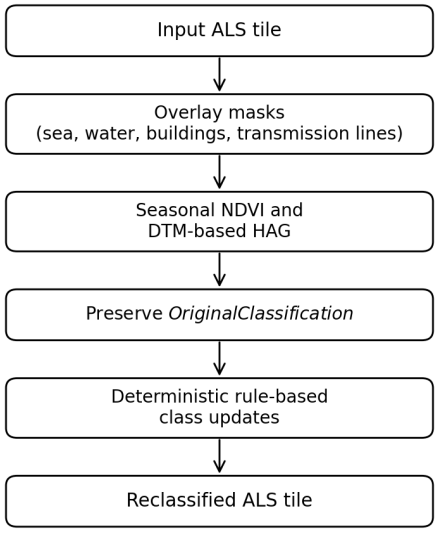


Figure 2. Simplified summary of the deterministic rule-based reclassification workflow. Spatial overlays, seasonal NDVI, and DTM-based height above ground (HAG) values are combined to update point classes while preserving the `OriginalClassification` attribute. The detailed class-transition rules are given in Table 1.

In the reclassified point clouds, the *High vegetation* class (Class 5) was used as a unified vegetation category encompassing all vegetation returns above the ground surface. This reflects the structure of the source data, in which Class 5 is the only vegetation-specific ASPRS class consistently used by the data provider (Section 2.2). Vegetation identified during reclassification, including low and intermediate returns previously assigned to *Unclassified* or *Ground*, was therefore aggregated into Class 5 regardless of vegetation height class, while the original class labels were preserved separately.

Ancillary layers representing the most temporally appropriate land-cover state were used throughout. For each LiDAR acquisition year, the vector dataset released at the beginning of the following year was applied. Spring NDVI composites were used for spring acquisitions and summer composites for summer acquisitions. Exact ancillary layer identifiers and their roles in the workflow are documented in the archived repository.

3.6 Evaluation Framework

The preprocessing workflow was evaluated using internal diagnostic measures of sampling consistency, thematic completeness, and height interpretability. These correspond to the three main issues addressed by the workflow: overlap-related density artefacts, instability in height reference, and recurring thematic misclassification.

For evaluation, we used a stratified subset of 120 tiles, comprising equal numbers of spring and summer acquisitions drawn from ALS III (2019–2020 campaigns used here) and ALS IV (2021–2024). Tiles were selected

Table 1. Rule-based reclassification logic applied to ALS point classes. Transitions indicate original and updated ASPRS classification codes.

Class transition	Rule description	Applied to
$1 \rightarrow 6$	Point located within building footprints	All files
$1 \rightarrow 5$	HAG $>$ 0.20 m and NDVI $>$ 0.33 outside 13 m transmission line buffers; or within 13 m buffers with HAG $>$ 0.20 m, HAG $<$ 6 m and NDVI $>$ 0.33	All files
$1 \rightarrow 2$	HAG $<$ 0.20 m	All files
$1 \rightarrow 9$	Point within waterbody polygons and HAG $<$ 0.20 m	All files
$6 \rightarrow 5$	Outside building footprints and NDVI $>$ 0.40	2019 campaigns and spring 2020 (applied conditionally by campaign)
$2 \rightarrow 5$	HAG $>$ 0.20 m and NDVI $>$ 0.33	Files from all summer campaigns
$5 \rightarrow 1$	Within 13 m transmission line buffers and HAG $>$ 6 m	All files
$9 \rightarrow 2$	Outside waterbody polygons	All files

to represent the main land-cover contexts relevant to the workflow, including forests, open and semi-natural habitats, arable land, wetlands, and built environments with linear infrastructure such as transmission lines. Each tile was processed using the full harmonisation workflow.

Sampling consistency was assessed through the effects of overlap-flagged point removal. Because flight-strip overlaps introduce local inflation of sampling intensity that can bias downstream density-sensitive vegetation metrics, we quantified how overlap removal changed the spatial distribution of point density in 10 m grid cells. Evaluation included comparisons of pre- and post-processing density patterns, density histograms, and the coefficient of variation (CV) of point counts per grid cell for a representative tile with clear flight-strip overlap. Density was expressed in points m^{-2} because the artefact appears as horizontal overrepresentation within the raster cells used in subsequent ecological analysis. The key question was therefore not whether all land-cover types should exhibit equal point densities, but whether cells with similar vegetation structure received comparable sampling support irrespective of flight-strip geometry. Voxel-based thinning was not adopted because it addresses point-cloud homogenisation in 3D space rather than the specific horizontal oversampling artefact introduced by overlapping flight strips. For relatively sparse national ALS data, especially under leaf-on conditions, voxelisation would also require voxel-size choices that may alter vertical sampling structure beyond overlap zones.

Thematic completeness was assessed through class-transition statistics derived from the preserved `OriginalClassification` attribute and the final harmonised `Classification` field. Particular attention was given to vegetation-related transitions, especially reassignment from *Unclassified* (Class 1) and *Ground* (Class 2) to vegetation (Class 5), because these changes are most relevant for subsequent vegetation metric calculation. Supporting transitions among *Ground*, *Water*, *Unclassified*, and other non-vegetation classes were also summarised to document broader effects of the deterministic rule set. Medians and interquartile ranges across the evaluation tiles were used to characterise the frequency and variability of these corrections.

Height interpretability was assessed through the behaviour of DTM-based HAG values after height normalisation. A representative seasonal tile pair was used for visual assessment because pooling all tiles would mix heterogeneous land-cover contexts and obscure interpretable patterns.

These diagnostics do not constitute a formal external accuracy assessment based on independent reference labels. Rather, they provide a transparent and reproducible basis for assessing whether the preprocessing workflow improved the internal coherence and ecological usability of the ALS archive relative to the raw source data.

3.7 Data and Software Availability

3.7.1 Input Data

Research data supporting this publication include both publicly accessible third-party datasets and auxiliary NDVI composites produced by the authors' research team. ALS point clouds and the LiDAR-derived 1 m DTMs used for height normalisation were downloaded from the ELSDB Geoportal (Estonian Land and Spatial Development Board, 2025c); the subset analysed in this study is described in Section 2.2. Topographic vector layers used for reclassification (building footprints and other constructions, inland and coastal waterbodies, and high-voltage transmission lines) were obtained from ETAK via the ELSDB Geoportal (Estonian Land and Spatial Development Board, 2025f), using the ETAK release corresponding to each acquisition year. Seasonal NDVI composites (spring: April–May; summer: June–August) were produced by the Landscape Geoinformatics Lab prior to this study and were used as auxiliary inputs for rule-based reclassification (Section 2.3). Because the ALS, DTM, and ETAK datasets are maintained by external providers, they are cited rather than re-hosted here; exact layer identifiers, temporal matching rules, and access routes for the NDVI composites are documented

in the archived code repository and in the exemplary reproducible workflow deposit described below.

3.7.2 Derived Data Products

The harmonised and reclassified LAZ files generated for this study are not currently redistributed by the authors because sustainable public hosting for the full set of derived outputs is not yet available. To support conditional reproducibility, the archived repository provides the complete reclassification rule set (Table 1), the ancillary-layer identifiers used in processing, and the scripts and configuration templates required to reproduce the workflow from the same ELSDB/ETAK and NDVI inputs. An exemplary reproducible workflow, including a small input dataset and the expected final LAZ output, is archived on Zenodo (DOI: 10.5281/zenodo.18172978; Sisas et al., 2026b).

3.7.3 Software and Code

The computational workflow supporting this publication is implemented as open-source Python code and released under the MIT License. The source repository is available on GitHub and the archived release used for this paper is available on Zenodo (DOI: 10.5281/zenodo.17867739; Sisas et al., 2026a). The archive contains the executable workflow, environment specification, configuration templates, example commands, and supplementary technical documentation required to reproduce the preprocessing pipeline. Core point-cloud processing relies on *Laspy* (v2.5.4; Laspy Contributors, 2025) and *PDAL* (v3.4.5; PDAL Contributors, 2025).

3.7.4 Computational Infrastructure and Reproducibility Scope

Large-scale processing was carried out on the University of Tartu HPC platform (University of Tartu, 2018) using embarrassingly parallel processing of individual tile-based LAZ files. The workflow is deterministic given identical input point clouds, ancillary layers, and configuration settings. Full national-scale reproduction therefore requires access to the same ELSDB/ETAK source data and NDVI inputs, together with the archived code and configuration files. The Zenodo example deposit provides a smaller demonstration dataset for practical reproducibility testing.

Table 2. Summary statistics of 10 m grid cell point densities before and after removal of overlap-flagged points for tile 585701_2022_tava, including the coefficient of variation (CV).

Metric	Before	After	Δ	Δ (%)
Min density (points m^{-2})	1.0	1.0	0.0	0.0
Max density (points m^{-2})	11.5	7.8	-3.7	-32.0
Mean density (points m^{-2})	4.3	3.2	-1.1	-26.3
CV (%)	38.4	35.0	-3.4	-8.9

4 Results

4.1 Data Harmonisation: Metadata, Format, and Overlap Correction

The harmonisation workflow corrected missing CRS metadata, removed overlap-flagged points, and standardised file structure across the 2019–2024 subset.

We restored missing CRS metadata in approximately 51 000 LAZ files from the 2022–2024 campaigns by encoding L-EST97 and metric units in the file headers. This update affected header metadata only and did not alter point coordinates or spatial accuracy. We removed points flagged with the LAS 1.4 `overlap` flag (quantified in Section 4.2) to suppress density inflation in flight-strip overlap zones. Finally, we rewrote harmonised LAZ files to a consistent LAS 1.4 structure using `Laspy`. Approximately 23 000 files from the 2019 spring and summer campaigns required correction because declared point-data formats did not match the registered extra-byte specifications; rewriting resolved these inconsistencies and ensured correct registration of all point attributes.

4.2 Effects of Overlap-Flagged Point Removal

Overlap-flagged points were concentrated in flight-strip overlap zones and produced linear bands of elevated point density in the raw tiles (Section 3.3).

Removing these points reduced density peaks and produced a more homogeneous spatial distribution (Fig. 3a,b). For tile 585701_2022_tava, the maximum 10 m-cell point density decreased from 11.5 to 7.8 points m^{-2} (1150 to 780 points per cell), and the CV decreased from 38.4 % to 35.0 % (Table 2). Point removal was confined to the overlap zone (Fig. 3c).

Density histograms (Fig. 4) show that extremely high-density cells were largely eliminated and that the overall distribution narrowed after harmonisation.

4.3 Height Normalisation

Fig. 5 illustrates seasonal separation in HAG distributions using all points from one representative spring–summer tile pair. We present a paired example for visual illustration, because pooling tiles would mix

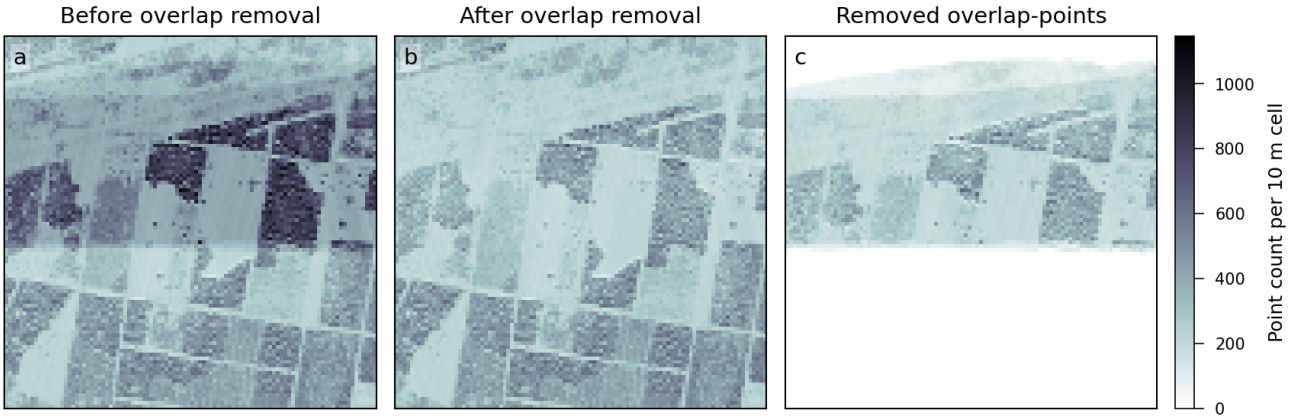


Figure 3. Point count per 10 m grid cell for the example tile 585701_2022_tava, shown before (a) and after (b) removal of overlap-flagged points; panel (c) shows the number of points removed per cell (i.e. (a–b)). All panels share the same colour scale to enable direct comparison. In panel (c), cells with zero removed points are shown in white; the extensive white area corresponds to regions outside the flight-strip overlap zones, where no point removal occurred. High-density artefacts associated with flight-strip overlap zones are reduced in the harmonised tile.

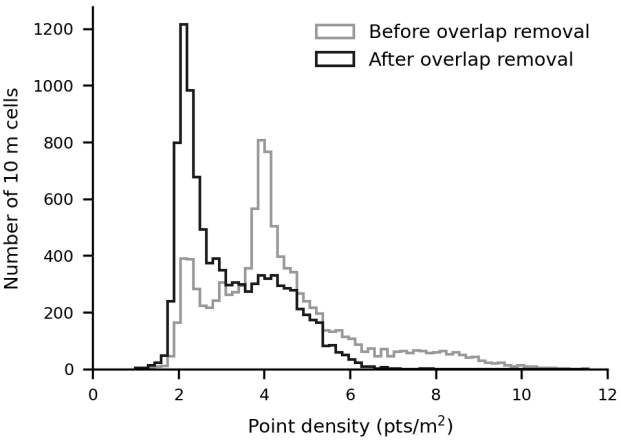


Figure 4. Distribution of 10 m grid cell point densities before and after removal of overlap-flagged points for tile 585701_2022_tava. Overlaid step histograms show that extreme high-density cells associated with flight-strip overlaps are substantially reduced after overlap correction.

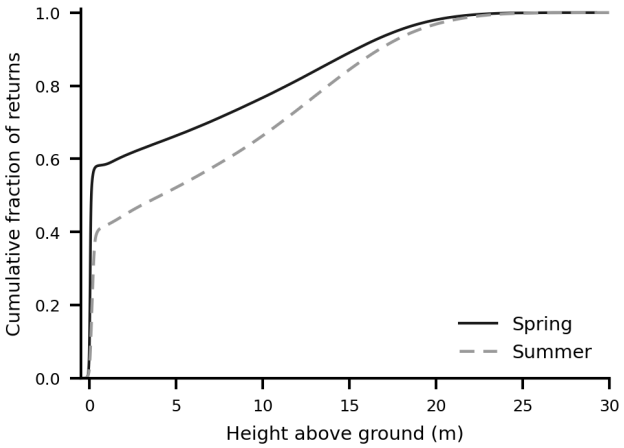


Figure 5. Empirical cumulative distribution functions (ECDFs) of height above ground (HAG) values for a representative spring–summer tile pair following DTM-based height normalisation. The distributions show similar overall ranges with systematic seasonal separation.

heterogeneous land-cover contexts. The seasonal distributions showed similar overall ranges but differed systematically in the lower and mid-height ranges, with the summer curve shifted towards higher HAG values. This seasonal separation is consistent with the expected effects of reduced ground-return availability and denser foliage under leaf-on conditions.

4.4 Reclassification Outcomes

Rule-based reclassification shifted point-class distributions across years and seasons, with the ecologically most relevant changes involving reassignment to vegetation (Class 5). Table 3 summarises all observed transitions, and Fig. 6 highlights their relative importance using median per-tile percentages. These percentages do not represent externally validated classification error rates before and after correction; rather, they quantify the proportion of source-data class assignments that were changed by the deterministic harmonisation workflow.

Across the 120 evaluation tiles, the most consequential transition for ecological applications was from *Unclassified* (Class 1) to vegetation: a median of 32.47 % (IQR 34.44) of previously unclassified points were reassigned to vegetation, reflecting low and intermediate vegetation returns not explicitly labelled in the source data. In addition, a median of 4.00 % (IQR 7.61) of points originally classified as *Ground* (Class 2) were reassigned to vegetation, primarily in summer acquisitions where dense understory reduced ground visibility and led to systematic misclassification.

Table 3. Reclassification statistics across 120 evaluation tiles. Reported values include the number of tiles in which each transition occurred, the median and interquartile range (IQR) of per-tile percentages, and total affected points.

Reclassification	Tiles	Median (%)	IQR	Total points
$1 \rightarrow 5$	120	32.47	34.44	14 615 610
$2 \rightarrow 5$	60	4.00	7.61	1 886 189
$6 \rightarrow 5$	26	17.96	23.22	45 426
$5 \rightarrow 1$	16	0.04	0.10	23 848
$1 \rightarrow 2$	120	65.42	32.73	33 014 212
$9 \rightarrow 2$	89	0.26	0.53	41 480
$1 \rightarrow 6$	86	0.26	0.74	249 669
$1 \rightarrow 9$	98	0.06	0.20	121 100

Reclassification of points originally labelled as *Building* (Class 6) to vegetation occurred in 26 tiles (median 17.96 %, IQR 23.22), primarily in earlier acquisition campaigns. Together, these transitions increased the proportion of points assigned to vegetation across seasons and acquisition campaigns.

A comparatively small proportion of points originally classified as vegetation were reassigned to *Unclassified* (Class 1) within buffers around high-voltage transmission lines (Table 3). Supporting corrections primarily affected non-vegetation classes. A large proportion of *Unclassified* points were reassigned to *Ground* (median 65.42 %, IQR 32.73), while water-related corrections affected comparatively small proportions of points and mainly ensured consistency between *Unclassified*, *Water*, and *Ground* classes in areas with open water surfaces.

Example tiles illustrating classification changes are shown in Fig. 7 and Fig. 8.

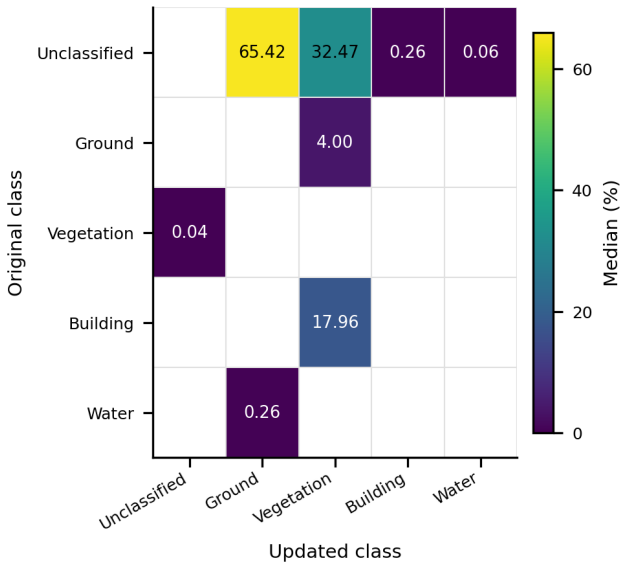


Figure 6. Heatmap of median per-tile class-transition percentages across the 120 evaluation tiles. Rows indicate original classes and columns updated classes. The dominant transitions were from *Unclassified* to *Ground* and from *Unclassified* to *Vegetation*, indicating substantial correction of thematic incompleteness in the source classifications.

5 Discussion

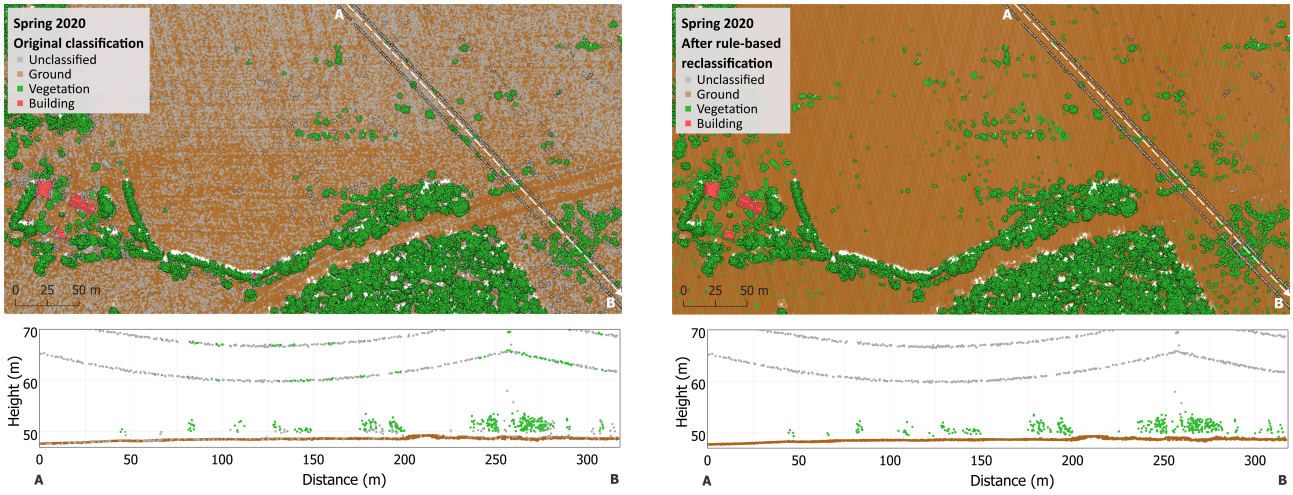
Our results show that inconsistencies in national ALS archives, including both campaign-level heterogeneity and recurring within-tile artefacts, can be reduced through point-cloud preprocessing in the 2019–2024 Estonian subset analysed here. This improves the stability and interpretability of downstream vegetation-structure inputs with respect to sampling density and height reference. The findings align with calls for explicit standardisation choices and transparent reporting of uncertainty in the ecological reuse of mapping-oriented 3D data (Valbuena et al., 2020), particularly when analyses require spatiotemporal comparison across acquisition campaigns (Tompalski et al., 2021).

The results should be interpreted in light of the study's scope as a preprocessing and harmonisation workflow rather than a full external accuracy assessment or downstream ecological model comparison. Our evaluation, therefore, focuses on whether the workflow improves the internal coherence and analytical readiness of the point-cloud archive for vegetation metric calculation, rather than on deriving absolute thematic accuracy estimates from independent reference labels.

National-scale descriptor products demonstrate demand for analysis-ready ecosystem-structure layers and show that barriers to their production are often computational and infrastructural rather than ecological (Assmann et al., 2022). EcoDes-DK15 illustrates how national ALS holdings can be condensed into compact raster descriptors usable in common ecological workflows, accompanied by auxiliary masks and acquisition-context layers that document exclusions (Assmann et al., 2022). Our contribution complements this approach by reducing common artefacts before metric calculation at the point-cloud level, thereby reducing the need to propagate such corrections through downstream analyses.

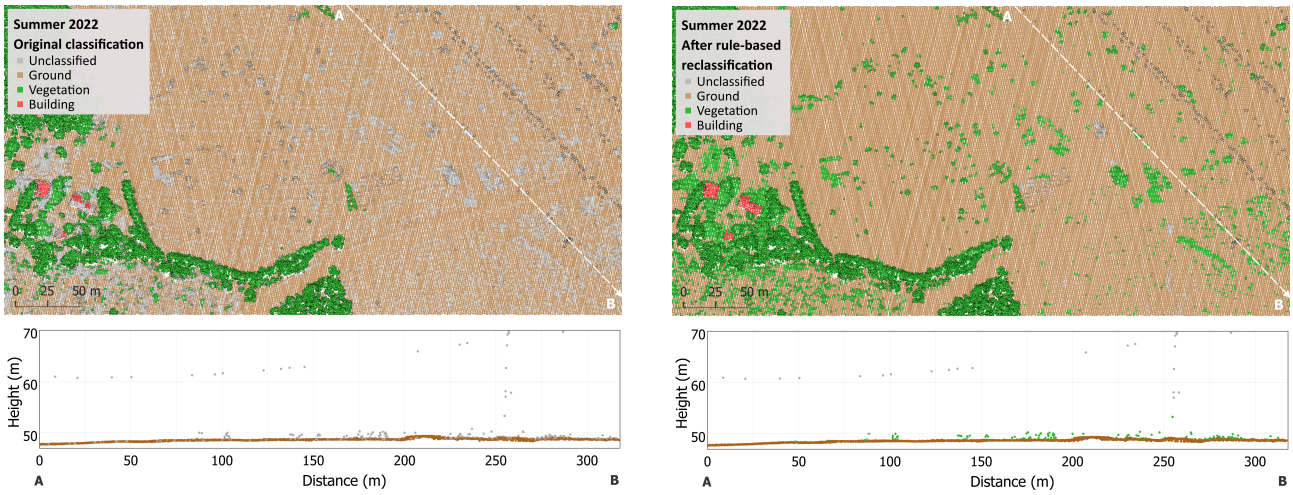
5.1 Acquisition Heterogeneity as a Constraint on Ecological Comparability

National ALS archives are shaped by multi-year acquisition logistics, evolving sensors, and changing flight configurations. Comparative studies across sensors, platforms, and acquisition settings show that differences in system design, point density, and recording mode can influence the representation of vegetation structure and the robustness of derived metrics, particularly for attributes related to foliage density and vertical variability (Kamoske et al., 2019; Koma et al., 2021; Lin et al., 2022). Reviews of multi-temporal ALS applications likewise identify acquisition heterogeneity as a key challenge (Tompalski et al., 2021). Differences in sampling density



(a) Original classification **(b) After the rule-based reclassification**

Figure 7. Representative examples of original (a) and harmonised (b) point classifications for the upper right corner of tile 568539, spring 2020.



(a) Original classification **(b) After the rule-based reclassification**

Figure 8. Representative examples of original (a) and harmonised (b) point classifications for the upper right corner of tile 568539, summer 2022.

and processing practice therefore make acquisition-aware preprocessing essential for accurately detecting and interpreting structural changes in ecological contexts.

In Estonia, this heterogeneity occurs both between and within campaigns. Between campaigns, data are acquired under leaf-off and leaf-on conditions at different flight altitudes and with different point densities. Within individual tiles, sampling intensity also varies because of flight-strip overlap zones, introducing localised deviations from nominal point density.

Flight-strip overlaps create linear bands of elevated sampling density and spatial striping that can propagate into density-sensitive vegetation metrics unless treated explicitly. Similar strip-pattern artefacts and locally elevated densities at flight-strip overlaps have been identified as a common source of within-dataset sampling variation that can affect height-derived metrics and produce spatial discontinuities (Roussel et al., 2017). In the representative tile analysed here, removal of overlap-flagged points reduced mean 10 m-cell point density by approximately 25 %, illustrating the magnitude of this within-tile artefact. Because the effect is expressed as horizontal sampling intensity within the raster cells from which vegetation metrics are commonly derived, we evaluate density in points m^{-2} rather than using volumetric density measures. Although voxel-based approaches may be useful for denser point clouds, the Estonian national ALS archive is relatively sparse, particularly in leaf-on campaigns, and voxel partitioning would produce many underpopulated or empty voxels unless large voxel sizes were used. Removing overlap-flagged points before height normalisation and reclassification therefore provides a transparent correction of a known acquisition artefact based on the provider's explicit point-level annotation and improves

the comparability of subsequent grid-based vegetation metrics (Tompalski et al., 2021). Although overlap-related density variation could in principle be handled only during downstream metric calculation, the aim of this study is to generate a harmonised point-cloud input suitable for a broad set of vegetation metrics rather than to optimise a single metric formulation. Correcting this artefact at the point-cloud stage therefore reduces the need to address overlap separately in each subsequent metric workflow.

In contrast to overlap-related sampling artefacts, leaf-off and leaf-on acquisitions introduce a between-campaign source of heterogeneity related to phenology and canopy closure rather than sampling geometry. These acquisition modes should therefore be treated as complementary observation configurations rather than interchangeable replicates. Retaining both seasons supports ecological applications across contrasting vegetation contexts, even though the interpretation of vegetation structure differs systematically between leaf-off and leaf-on conditions.

5.2 Metadata Integrity and Access as Foundations for Reproducibility

The prevalence of CRS-missing tiles and header/format inconsistencies identified during data harmonisation shows that ecological reuse can fail for infrastructural reasons even when the underlying algorithms are sound. Incomplete or ambiguously encoded metadata can lead to inconsistent processing outcomes, reinforcing calls to treat metadata integrity and uncertainty documentation as first-order requirements for standardised ecosystem-structure descriptors intended for temporal comparison and broad reuse (Valbuena et al., 2020).

A related issue is large-scale discoverability and access. In our case, bulk tile discovery relied on an interactive web interface, so we built an internal spatial footprint index (served as WMS/WFS) to support traceability, inspection, and reproducible subset definition. Improved machine-readable catalogues and support for programmatic and bulk access to national ALS holdings would reduce friction for reproducible processing and lower practical barriers to ecological reuse.

5.3 Height-Reference Stability in Multi-Season ALS Data

Our comparison of alternative normalisation approaches confirms a known challenge in dense vegetation. Reliable identification of the ground reference becomes increasingly difficult under closed canopies, particularly during leaf-on conditions (Kobler et al., 2007). In our tests, neighbourhood-minima-based normalisation frequently selected returns above the true ground surface where ground returns were occluded by dense vegetation, inflating HAG values and introducing spatially inconsistent height offsets across tiles. This issue is directly relevant for national multi-temporal ecosystem-structure products that rely on lowest-point normalisation, because such choices can introduce systematic biases (Shi et al., 2025). More broadly, multi-temporal analyses require detected change signals to reflect ecological dynamics rather than artefacts of acquisition timing or preprocessing decisions (Riofrío et al., 2022; Tompalski et al., 2021).

DTM-based normalisation provided a stable ground reference and supported consistent application of height thresholds across campaigns. However, DTMs are not error-free, and multi-temporal workflows can remain sensitive to small vertical inconsistencies between acquisitions and their terrain references (Riofrío et al., 2022). We therefore interpret residual spring–summer offsets in HAG distributions primarily as phenology and canopy-closure effects, while recognising that local DTM uncertainty, vertical misalignment, or temporal mismatch between terrain and point cloud acquisition may also contribute in specific locations.

5.4 Thematic Consistency: Point-Level Correction Versus Downstream Masking

Our results show that ecologically relevant returns, including vegetation and ground, can remain in *Unclassified* or be misclassified across campaigns, especially under dense leaf-on conditions. This indicates that provider classifications in mapping-oriented ALS programmes are not necessarily optimised for ecological applications and reinforces the need for explicit preprocessing and documentation when producing standardised structural descriptors (Kissling et al., 2022).

A key design choice in our workflow is to address recurring thematic inconsistencies at the point-cloud level through deterministic rule-based reclassification while preserving `OriginalClassification` for auditability. An alternative strategy, exemplified by national raster descriptor products, is to distribute vegetation metrics together with auxiliary layers such as water masks and acquisition-context information so that users can propagate exclusions and uncertainty through downstream analyses (Assmann et al., 2022). This approach is transparent and effective, but it can increase end-user complexity in multi-temporal studies, where meaningful comparison requires multiple masks and context layers to be carried through each analysis step (Shi et al., 2025).

We view these strategies as complementary rather than competing. Point-cloud correction reduces the need for repeated downstream masking of common systematic issues, whereas auxiliary layers remain valuable for uncertainty-aware analyses, specialised exclusions, and diagnosis of residual artefacts that no deterministic rule set can fully eliminate (Shi et al., 2025).

The observed reclassification changes are also likely to affect downstream class-based and height-related

vegetation metrics, because substantial proportions of previously *Unclassified* and *Ground* points were reassigned to vegetation. This is particularly important for metrics that depend on the completeness of vegetation-class representation or on the vertical distribution of returns above the ground surface. These downstream effects were not quantified directly, because the focus of the present study was preprocessing and harmonisation rather than formal metric-sensitivity analysis. However, the magnitude of the documented class transitions suggests that vegetation metrics derived from harmonised point clouds may differ meaningfully from those calculated from the raw archive, especially in summer acquisitions where vegetation misclassification was more pronounced. These differences are likely to affect class-based point-count metrics, vegetation-cover estimates, and height-distribution summaries derived from the point clouds, although formal metric-sensitivity analysis was beyond the scope of this study.

5.5 Limitations and Implications for Ecological Applications

Despite the improvements achieved, several limitations remain. First, acquisition-related differences, including campaign-to-campaign variation in point density and acquisition settings, may still influence the derivation of fine-scale vegetation metrics even after within-tile overlap artefacts are removed. The sensitivity of canopy height products and change analyses to point-density variation is well recognised in multi-temporal studies (Tompalski et al., 2021). We do not claim that removal of overlap-flagged points eliminates all density-related effects, nor that it is the only valid harmonisation strategy. Rather, it provides a transparent correction of one explicit and spatially localised acquisition artefact that would otherwise propagate into subsequent grid-based ecological metrics. These findings also have practical implications for national mapping agencies managing long-term ALS programmes: where feasible, machine-readable metadata on acquisition parameters, overlap information, processing history, point-classification logic, and terrain-model generation should be provided to support downstream harmonisation and help users distinguish acquisition-related artefacts from ecological signals.

Second, DTM-based normalisation reduces sensitivity to missing ground returns under leaf-on canopy, but residual uncertainty remains due to DTM error, vertical alignment, and temporal mismatch between terrain and point cloud acquisitions. These uncertainties can propagate into HAG values used for subsequent metric calculation (Riofrío et al., 2022). Seasonal acquisitions therefore remain complementary rather than directly interchangeable, even after technical harmonisation.

Third, our deterministic reclassification targets recurring error modes that can be addressed using the available ancillary data, but residual misclassifications are expected in complex environments. These may still require auxiliary masks or targeted quality assessment during later analyses, consistent with national products that distribute acquisition-context layers to support uncertainty-aware use (Shi et al., 2025). In addition, this work focuses on structural harmonisation of ALS point clouds; radiometric attributes (e.g. intensity) were not standardised and remain an important avenue for future research, consistent with broader calls for standardised, comparable ecosystem-structure descriptors derived from 3D data (Valbuena et al., 2020).

The present study also does not provide formal error rates for thematic classification changes or direct comparisons of downstream ecological analyses before and after harmonisation. Such assessments would require independent reference labels and systematic manual validation across the national archive, which were beyond the scope of this study. Instead, we report class-transition statistics and other internal diagnostic measures as indicators of how preprocessing changed thematic completeness, density consistency, and height-reference behaviour relative to the raw source data. The documented reduction in overlap-driven density artefacts, reassignment of substantial proportions of *Unclassified* and *Ground* points to vegetation, and stabilisation of HAG values nevertheless indicate that density-sensitive, class-based, and height-based vegetation metrics would change in non-trivial ways if calculated from harmonised rather than raw point clouds.

Future work should quantify how residual density variation, thematic uncertainty, and height-reference uncertainty propagate into vegetation metrics and downstream ecological analyses, ideally through before-and-after sensitivity analyses across representative habitat contexts and acquisition conditions. A useful next step for Estonian releases would be to package per-tile acquisition context and terrain-reference provenance alongside harmonised point clouds to support uncertainty-aware reuse. Extending the rule set with targeted validation in complex built and edge environments would further clarify how preprocessing choices affect downstream ecological modelling.

6 Conclusions

Our study presents a harmonised, reproducible preprocessing framework for national-scale ecological applications of airborne laser scanning data. By explicitly addressing metadata inconsistencies, overlap-related sampling artefacts, DTM-based height normalisation, and systematic point-classification errors, the workflow improves the internal coherence and ecological interpretability of LiDAR-derived inputs for vegetation metric calculation across heterogeneous acquisition campaigns.

Applied to the 2019–2024 subset of the Estonian national ALS archive, the framework shows that general-purpose ALS mapping data, despite multi-year and multi-season heterogeneity, can be systematically prepared for vegetation metric calculation and biodiversity-oriented ecological analysis. Harmonisation at the point-cloud stage provides a stable and transparent foundation for subsequent metric calculation and reduces the propagation of acquisition-related sampling and height-reference artefacts into ecological analyses. More broadly, the findings show that heterogeneous legacy national ALS datasets can be systematically harmonised and repurposed for ecological analysis rather than treated as unsuitable for biodiversity-oriented applications.

Our modular and open-source implementation supports transferability to other national LiDAR programmes facing similar challenges and enables reproducible large-scale processing. As national ALS archives continue to expand, such harmonised preprocessing workflows will be essential for unlocking their full potential for ecological modelling, monitoring, and long-term environmental assessment.

Declaration of Generative AI in Writing

Generative artificial intelligence (AI) tools were used during manuscript preparation to support language editing, improve clarity, and assist with consistency checks of the text. The AI tools were not used to generate scientific content, data, results, interpretations, figures, or citations. All methodological choices, analyses, and conclusions were developed by the authors, who take full responsibility for the content of the manuscript.

Acknowledgements

This work was funded by the Estonian Research Agency (grant number PRG1764, PSG841), Estonian Ministry of Education and Research (Centre of Excellence for Sustainable Land Use (TK232)), and by the European Union (ERC, WaterSmartLand, 101125476). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The authors also thank Adam Steer for preliminary harmonisation tests and technical troubleshooting that informed the development of the preprocessing workflow. The authors are grateful for technical support from the High Performance Computing Center of the University of Tartu (University of Tartu, 2018).

References

Acebes, P., Lillo, P., and Jaime-González, C.: Disentangling LiDAR Contribution in Modelling Species–Habitat Structure Relationships in Terrestrial Ecosystems Worldwide. A Systematic Review and Future Directions, Remote Sensing, 13, 3447, https://doi.org/10.3390/rs13173447, 2021.

American Society for Photogrammetry and Remote Sensing: LAS Specification Version 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 System Science Data, 14, 823–844, https://doi.org/10.5194/essd-14-823-2022, 2022.

Bald, L., Ziegler, A., Gottwald, J., Koch, T. L., Ludwig, M., Meyer, H., Wöllauer, S., Zeuss, D., and Frieß, N.: Leveraging heterogeneous LiDAR data to model successional stages at tree species level in temperate forests, Environmental Data Science, 3, e24, https://doi.org/10.1017/eds.2024.31, 2024.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Scientific Data, 5, 180 214, https://doi.org/10.1038/sdata.2018.214, 2018.

Bässler, C., Stadler, J., Müller, J., Förster, B., Göttlein, A., and Brandl, R.: LiDAR as a rapid tool to predict forest habitat types in Natura 2000 networks, Biodiversity and Conservation, 20, 465–481, https://doi.org/10.1007/s10531-010-9959-x, 2011.

Campbell, M. J., Dennison, P. E., Hudak, A. T., Parham, L. M., and Butler, B. W.: Quantifying understory vegetation density using small-footprint airborne lidar, Remote Sensing of Environment, 215, 330–342, https://doi.org/10.1016/j.rse.2018.06.023, 2018.

Estonian Land and Spatial Development Board: ALS III ring (2016) 2017–2020, available at: https://geoportaal.maaamet.ee/est/ruumiandmed/korgusandmed/aerolaserskaneerimise-korguspunktid/als-iii-ring-2016-20172020-p625.html, last access: 11 December 2025, 2025a.

Estonian Land and Spatial Development Board: Aerolaserskaneerimise kõrguspunktid, available at: https://geoportaal.maaamet.ee/est/ruumiandmed/korgusandmed/aerolaserskaneerimise-korguspunktid-p499.html, last access: 9 December 2025, 2025b.

Estonian Land and Spatial Development Board: Download Elevation Data, available at: https://geoportaal.maaamet.ee/eng/Maps-and-Data/Elevation-data/Download-Elevation-Data-p664.html, last access: 11 December 2025, 2025c.

Estonian Land and Spatial Development Board: Elevation Data, available at: https://geoportaal.maaamet.ee/eng/spatial-data/elevation-data-p308.html, last access: 10 December 2025, 2025d.

Estonian Land and Spatial Development Board: Estonian Topographic Database, available at: https://geoportaal.maaamet.ee/eng/spatial-data/

estonian-topographic-database-p305.html, last access: 12 December 2025, 2025e.

Estonian Land and Spatial Development Board: Eesti topograafia andmekogu vektorandmete ajalugu, available at: https://geoportaal.maaamet.ee/index.php?lang_id=1&page_id=940, last access: 12 December 2025, 2025f.

Kamoske, A. G., Dahlin, K. M., Stark, S. C., and Serbin, S. P.: Leaf area density from airborne LiDAR: Comparing sensors and resolutions in a temperate broadleaf forest ecosystem, Forest Ecology and Management, 433, 364–375, https://doi.org/10.1016/j.foreco.2018.11.017, 2019.

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, 101 836, https://doi.org/10.1016/j.ecoinf.2022.101836, 2022.

Kobler, A., Pfeifer, N., Ogrinc, P., Todorovski, L., Oštir, K., and Džeroski, S.: Repetitive interpolation: A robust algorithm for DTM generation from Aerial Laser Scanner Data in forested terrain, Remote Sensing of Environment, 108, 9–23, https://doi.org/10.1016/j.rse.2006.10.013, 2007.

Koma, Z., Zlinszky, A., Bekő, L., Burai, P., Seijmonsbergen, A. C., and Kissling, W. D.: Quantifying 3D vegetation structure in wetlands using differently measured airborne laser scanning data, Ecological Indicators, 127, 107 752, https://doi.org/10.1016/j.ecolind.2021.107752, 2021.

Laspy Contributors: Laspy, available at: https://github.com/laspy/laspy, last access: 10 December 2025, 2025.

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.

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.

Lopatin, J., Dolos, K., Hernández, H., Galleguillos, M., and Fassnacht, F.: Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile, Remote Sensing of Environment, 173, 200–210, https://doi.org/10.1016/j.rse.2015.11.029, 2016.

Marinelli, D., Paris, C., and Bruzzone, L.: A Novel Approach to 3-D Change Detection in Multitemporal LiDAR Data Acquired in Forest Areas, IEEE Transactions on Geoscience and Remote Sensing, 56, 3030–3046, https://doi.org/10.1109/TGRS.2018.2789660, 2018.

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, 100 626, https://doi.org/10.1016/j.softx.2020.100626, 2020.

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, e01 907, https://doi.org/10.1002/eap.1907, 2019.

PDAL Contributors: PDAL Point Data Abstraction Library, https://doi.org/10.5281/zenodo.10884408, 2025.

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., Caspersen, J., Béland, M., Thomas, S., and Achim, A.: Removing bias from LiDAR-based estimates of canopy height: Accounting for the effects of pulse density and footprint size, Remote Sensing of Environment, 198, 1–16, https://doi.org/10.1016/j.rse.2017.05.032, 2017.

Shi, Y., Wang, J., and Kissling, W. D.: Multi-temporal high-resolution data products of ecosystem structure derived from country-wide airborne laser scanning surveys of the Netherlands, Earth System Science Data, 17, 3641–3677, https://doi.org/10.5194/essd-17-3641-2025, 2025.

Sisas, E., Virro, H., Chan, W. T., Kmoch, A., and Uuemaa, E.: Code supplement: Towards National-Scale Ecological Applications: Harmonised framework for LiDAR point cloud processing for vegetation metrics calculation, https://doi.org/10.5281/zenodo.17867739, 2026a.

Sisas, E., Virro, H., Chan, W. T., Kmoch, A., and Uuemaa, E.: Data supplement: Towards National-Scale Ecological Applications: Harmonised framework for LiDAR point cloud processing for vegetation metrics calculation, https://doi.org/10.5281/zenodo.18172978, 2026b.

Thers, H., Brunbjerg, A. K., Læssøe, T., Ejrnæs, R., Bøcher, P. K., and Svenning, J.-C.: Lidar-derived variables as a proxy for fungal species richness and composition in temperate Northern Europe, Remote Sensing of Environment, 200, 102–113, https://doi.org/10.1016/j.rse.2017.08.011, 2017.

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, 121 376, https://doi.org/10.1016/j.foreco.2023.121376, 2023.

Tompalski, P., Coops, N. C., White, J. C., Goodbody, T. R., 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.

University of Tartu: UT Rocket, https://doi.org/10.23673/PH6N-0144, 2018.

Valbuena, R., O'Connor, B., Zellweger, F., Simonson, W., Vihervaara, P., Maltamo, M., Silva, C., Almeida, D., Danks, F., Morsdorf, F., Chirici, G., Lucas, R., Coomes, D., and Coops, N.: 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.

Zellweger, F., Morsdorf, F., Purves, R. S., Braunisch, V., and Bollmann, K.: Improved methods for measuring forest landscape structure: LiDAR complements field-based habitat assessment, Biodiversity and Conservation, 23, 289–307, https://doi.org/10.1007/s10531-013-0600-7, 2014.