ARPHA Preprints

PENSOFT

Project Report

Author-formatted document posted on 13/02/2026

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e188540

D6.2 Geospatial data products of habitat metrics derived from LiDAR point clouds at national scales for multiple countries and time periods, findable and accessible through data portals, relevant catalogues or digital repositories

W. Daniel Kissling, Jinhu Wang

ARPHA Preprints Author-formatted document posted on 13/02/2026. DOI: https://doi.org/10.3897/arphapreprints.e188540

MAMBO
MODERN APPROACHES TO THE MONITORING OF BIODIVERSITY

D6.2 Geospatial data products of habitat metrics derived from LiDAR point clouds at national scales for multiple countries and time periods, findable and accessible through data portals, relevant catalogues or digital repositories

12/02/2026

Lead beneficiary: University of Amsterdam (UvA)

Authors: W. Daniel Kissling, Jinhu Wang

Reviewer/s: Yifang Shi, France Gerard

Funded by the European Union

Prepared under contract from the European Commission
Grant agreement No. 101060639, EU Horizon Europe Research and Innovation Action

Project acronym:	**MAMBO**
Project full title:	**Modern Approaches to the Monitoring of Biodiversity**
Project duration:	01.09.2022 – 31.08.2026 (48 months)
Project coordinator:	Dr. Toke Thomas Høye, Aarhus University (AU)
Call:	HORIZON-CL6-2021-BIODIV-01
Deliverable title:	D6.2 Geospatial data products of habitat metrics derived from LiDAR point clouds at national scales for multiple countries and time periods, findable and accessible through data portals, relevant catalogues or digital repositories
Deliverable n°:	D6.2
WP responsible:	WP6
Nature of the deliverable:	Data
Dissemination level:	Public
Lead beneficiary:	UvA
Due date of deliverable:	M42
Actual submission date:	M42

Deliverable status:

Version	Status	Date	Author/reviewer
0.1	First outline	27/01/2026	W. Daniel Kissling (UvA)
0.2	Preface, summary, appendix	28/01/2026	W. Daniel Kissling (UvA)
0.3	Complete draft	05/02/2026	W. Daniel Kissling (UvA)
0.4	Input section 8.3	07/02/2026	Jinhu Wang (UvA)
0.5	Review 1	11/02/2026	Yifang Shi (UvA)
0.6	Review 2	11/02/2026	France Gerard (UKCEH)
0.7	Incorporating feedback	11/02/2026	W. Daniel Kissling (UvA)
0.8	Final version	12/02/2026	W. Daniel Kissling (UvA)

Table of Contents

1 Preface

This document is a data deliverable (D6.2) prepared for the *Modern Approaches to the Monitoring of BiOdiversity* (MAMBO) project, funded by the European Commission through an EU Horizon Europe Research and Innovation Action (Grant Agreement No. 101060639). The MAMBO project aims to support EU biodiversity policy and address key knowledge gaps by developing and applying innovative tools, workflows, and data products for biodiversity monitoring at multiple spatial and temporal scales.

Deliverable D6.2 documents the geospatial research data products of vegetation structure metrics derived from airborne Light Detection and Ranging (LiDAR) point clouds that have been generated within the MAMBO project. The deliverable is of type 'DATA' and focuses on the actual datasets produced or collected during the project, rather than on the analysis or interpretation of scientific results. Its primary purpose is to support the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR) by providing a structured overview of the datasets, their scope, and the platforms through which they are made publicly available.

The datasets described in this deliverable comprise national-scale, multi-temporal vegetation structure metrics derived from country-wide airborne laser scanning surveys, LiDAR-based vegetation structure metric datasets for multiple MAMBO demonstration sites across Europe, and a set of thematic and methodological data products supporting habitat condition assessment. These include datasets on ungulate trail extraction, robustness testing of LiDAR-derived vegetation metrics, and the extraction of trees and shrubs in wetland environments. All datasets are published in open-access repositories, assigned persistent identifiers, and accompanied by openly available processing workflows and metadata records to support transparent reuse.

By consolidating information on these LiDAR-derived geospatial data products and their access points, this deliverable contributes to the long-term availability and reuse of MAMBO research data. It supports transparent, reproducible, and scalable habitat condition assessments and provides a data legacy that can be reused by researchers, practitioners, and policy-relevant initiatives beyond the lifetime of the MAMBO project.

2 Executive summary

This deliverable (D6.2) documents the geospatial data products of vegetation structure derived from airborne Light Detection and Ranging (LiDAR) point clouds that have been generated and published in the context of the *Modern Approaches to the Monitoring of BiOdiversity* (MAMBO) project. The deliverable is of type 'DATA' and focuses on making research data Findable, Accessible, Interoperable, and Reusable (FAIR) by providing an overview of the datasets, their scope, and their availability through established data portals, catalogues, and digital repositories.

Within MAMBO, airborne laser scanning (ALS) data have been processed into harmonised, high-resolution geospatial metrics describing vegetation structure and related habitat characteristics. These metrics are relevant for assessing habitat condition across a wide range of ecosystems, including forests, woodlands, grasslands, wetlands, marshes, reedbeds, scrublands, agricultural landscapes, and Mediterranean habitats. The datasets cover multiple spatial scales, ranging from full national coverage to site-based demonstration areas, and include both single-date and multi-temporal data products derived from repeated ALS surveys.

At the national scale, MAMBO has produced multi-temporal, country-wide data products of vegetation structure for the Netherlands, based on successive national airborne laser scanning campaigns. These datasets provide consistent, spatially explicit metrics at a 10 m resolution across the entire country and enable large-scale and long-term analyses of vegetation structure and change. The data are accompanied by peer-reviewed data publications, persistent identifiers (DOIs), and openly available processing code to ensure transparency and reproducibility.

In addition, LiDAR-derived vegetation structure metric datasets have been produced for seven MAMBO demonstration sites across five European countries. These datasets were generated using harmonised workflows applied to country-wide ALS data and encompass a broad diversity of habitat types. The processed data products are publicly available through open repositories, while the underlying raw LiDAR point clouds and the full processing workflows, implemented as Jupyter Notebooks, have been made accessible through dedicated research workflow platforms.

Beyond these core datasets, MAMBO has generated a set of specialised LiDAR-derived data products that support specific aspects of habitat condition assessment. These include datasets for extracting ungulate trail networks from ALS point clouds as spatial indicators of grazing and trampling pressure, datasets for evaluating the robustness of LiDAR-derived vegetation metrics across different point densities and habitat types, and datasets focusing on the extraction of trees and shrubs in wetland environments. All associated data products and processing workflows are published openly and linked to peer-reviewed scientific publications, supporting transparency, reproducibility, and reuse.

To enhance discoverability and long-term reuse, all datasets documented in this deliverable have been deposited in open research repositories, primarily Zenodo, and grouped within the MAMBO community, which is part of the EU Open Research Repository. Furthermore, all datasets have been registered in the LifeWatch Metadata Catalogue, providing rich, standardised metadata to support data discovery, interoperability, and reuse by both humans and machines. The datasets and related publications are also discoverable through the MAMBO RIO topical collection.

This deliverable provides a consolidated overview of the airborne LiDAR-derived geospatial data products generated by MAMBO to date, including their geographic and temporal coverage, access points, and metadata registration. By documenting these datasets in a structured and transparent manner, D6.2 contributes to the long-term availability and reuse of MAMBO data products and supports the uptake of LiDAR-based vegetation structure metrics in European biodiversity monitoring, assessment, and policy-relevant applications.

3 List of abbreviations

ALS	Airborne Laser Scanning
EC	European Commission
EPSG	European Petroleum Survey Group
EU	European Union
FAIR	Findability, Accessibility, Interoperability, and Reusability
LAS	Standard file format for the interchange and archiving of LiDAR point cloud data
LAZ	Compressed version of the LAS file format
LiDAR	Light Detection and Ranging
MAMBO	Modern Approaches to the Monitoring of BiOdiversity
DOI	Digital Object Identifier

4 Introduction

4.1 Role of LiDAR-derived vegetation structure metrics in MAMBO

Assessing and monitoring **habitat condition** is a key requirement for understanding biodiversity patterns, detecting environmental change, and supporting evidence-based conservation and policy decisions in the European Union (European Environment Agency, 2020; Kissling et al., 2024). Habitat condition reflects a combination of biotic, abiotic, and structural characteristics of ecosystems, including vegetation height, vertical complexity, openness, and spatial heterogeneity, which are closely linked to species occurrence, ecosystem functioning, and ecological processes (Maes et al., 2023). Quantitative, spatially explicit metrics describing these characteristics are therefore essential for consistent habitat condition assessments across sites, regions, and time periods.

Within the Modern Approaches to the Monitoring of BiOdiversity (MAMBO) project, **airborne Light Detection and Ranging (LiDAR)** has been applied as a core remote-sensing technology for deriving habitat condition metrics (Høye et al., 2023). Airborne laser scanning (ALS) provides three-dimensional point clouds with high spatial resolution and wall-to-wall coverage, enabling the systematic quantification of vegetation structure and terrain properties over large areas (Assmann et al., 2022; Kissling et al., 2023). Compared to many satellite-based products, LiDAR-derived metrics capture fine-scale structural attributes that are directly relevant for habitat condition assessments, while still being suitable for upscaling to national and cross-national levels (Moudrý et al., 2026).

MAMBO builds on the increasing availability of **open, country-wide ALS datasets** across Europe to generate harmonised geospatial data products of vegetation structure metrics. These products include metrics describing vegetation height distribution, canopy cover, and structural complexity, which can be consistently derived across a wide range of habitat types, including forests, woodlands, grasslands, wetlands, scrublands, dunes, agricultural landscapes, and Mediterranean ecosystems (Kissling et al., 2025; Shi et al., 2025). By applying reproducible, scalable workflows to national and sub-national LiDAR point clouds (Kissling et al., 2022), MAMBO demonstrates how structurally meaningful habitat indicators can be generated at spatial resolutions suitable for both site-based management (e.g. protected areas) and large-scale biodiversity monitoring.

A central objective of MAMBO is the production and publication of reusable **geospatial data products** that support long-term monitoring and upscaling. LiDAR-derived vegetation structure metrics generated within the project are therefore treated as research data assets. They are designed to enable spatial comparisons among regions, temporal analyses across repeated ALS surveys, and integration with complementary biodiversity data sources, including field observations and ecological monitoring schemes (de Vries et al., 2021; Koma et al., 2021; Kissling et al., 2025; Shi et al., 2025). This data-centric approach facilitates reuse beyond individual case studies and contributes to the development of consistent, large-scale biodiversity monitoring frameworks in Europe.

In this context, Deliverable D6.2 documents the LiDAR-derived geospatial data products produced within MAMBO and their role in supporting quantitative, scalable, and FAIR habitat condition assessments. By making these datasets findable, accessible, interoperable, and

ARPHA Preprints Author-formatted document posted on 13/02/2026. DOI: https://doi.org/10.3897/arphapreprints.e188540

reusable, MAMBO contributes to establishing a robust and transparent data foundation for LiDAR-based habitat monitoring across Europe.

4.2 Purpose and scope of Deliverable D6.2

The purpose of Deliverable D6.2 is to document the geospatial research data products of vegetation structure metrics derived from airborne LiDAR point clouds associated FAIR-enabling research objects that have been generated and published within the MAMBO project (see overview in Annex Table A1). The deliverable is of type '**DATA**' and focuses on the actual datasets produced or collected during the project, rather than on methodological development, data analysis, or interpretation of scientific results.

The **scope of D6.2** includes LiDAR-derived geospatial datasets describing vegetation structure and related habitat characteristics at multiple spatial and temporal scales. These comprise national-scale, multi-temporal data products derived from country-wide airborne laser scanning surveys; site-based vegetation structure metric datasets produced for MAMBO demonstration sites across several European countries; and thematic datasets supporting specific aspects of habitat condition assessment, such as indicators of grazing and trampling pressure or the robustness of LiDAR-derived vegetation metrics to varying data characteristics. The deliverable covers both processed data products (e.g. raster and vector layers of vegetation structure metrics) and, where relevant, the associated raw data sources that underpin their generation.

A key objective of this deliverable is to support the implementation of the **FAIR data principles** (Wilkinson et al., 2016) by providing a structured overview of each dataset's content, spatial and temporal coverage, access conditions, and persistent identifiers. Accordingly, D6.2 places emphasis on documenting where and how datasets can be discovered and accessed, including their publication in open research repositories, registration in metadata catalogues, and linkage to openly available processing workflows and documentation, insofar as these are essential for data reuse, reproducibility, and interoperability.

Deliverable D6.2 does not introduce other data products beyond those generated within MAMBO, nor does it evaluate the scientific performance or ecological interpretation of the datasets. Instead, it provides a transparent inventory and summary of the LiDAR-derived geospatial data assets produced by the project, forming part of the MAMBO data legacy and facilitating their reuse by researchers, practitioners, and policy-relevant initiatives beyond the lifetime of the project.

5 Overview of data products

5.1 General characteristics

The data products documented in this deliverable consist of **geospatial datasets of vegetation structure** derived from airborne LiDAR point clouds generated within the MAMBO project. They are primarily processed research data products derived from country-wide airborne laser scanning surveys and are designed to support quantitative assessments of habitat condition across a wide range of ecosystems.

The core datasets comprise spatially explicit metrics describing **vegetation structure** and related habitat characteristics. Derived from three-dimensional LiDAR point clouds, these metrics capture key structural properties of ecosystems, including vegetation height distribution, canopy cover and openness, and measures of vertical structural complexity (Davies and Asner, 2014; Bakx et al., 2019; Kissling and Shi, 2023). Such metrics quantify biotic and structural attributes that are closely linked to species composition, habitat suitability, and ecosystem functioning.

Most processed data products are provided as **raster datasets** at a consistent spatial resolution, typically 10 m for national-scale products, enabling wall-to-wall coverage and spatial comparability across large areas. Some datasets are available at finer spatial resolutions, such as raster trail networks representing animal trails (approximately 10 cm resolution). **Additional data products** include labelled point clouds (e.g. ground truth plots) and tabular data associated with methodological evaluations, such as robustness testing of LiDAR-derived vegetation metrics. Where relevant, links are provided to the underlying raw LiDAR point cloud datasets, which are typically stored in standard open formats (LAS or LAZ) and hosted in public repositories.

The datasets cover a broad **diversity of habitat types**, including forests and woodlands, dry and wet grasslands, marshes and reedbeds, scrublands, dunes, agricultural landscapes, and Mediterranean habitats. They span multiple spatial scales, from full national coverage to specific demonstration sites, and include both single-date and multi-temporal data products derived from repeated ALS surveys. Together, these characteristics form a coherent set of LiDAR-derived geospatial data assets suitable for spatial comparisons among regions and temporal analyses of vegetation structure and change.

All data products documented in this deliverable have been published through **open-access repositories** and assigned persistent identifiers to ensure long-term availability. They are accompanied by descriptive metadata specifying content, spatial and temporal extent, formats, and licensing conditions. This reflects a data production approach that prioritises consistency, transparency, and reusability, enabling the use of MAMBO data products beyond the duration of the project.

5.2 Spatial and temporal coverage

The LiDAR-derived geospatial data products documented in this deliverable cover a wide range of spatial and temporal scales, reflecting the multi-level approach of the MAMBO project to habitat condition assessment and biodiversity monitoring.

At the largest **spatial scale**, the datasets include country-wide coverage derived from national airborne laser scanning surveys. In particular, multi-temporal vegetation structure data products have been generated for the Netherlands based on successive national ALS campaigns (Shi et al., 2025). These datasets provide wall-to-wall coverage at a consistent spatial resolution and support spatially continuous analyses of vegetation structure across the entire country. The availability of repeated ALS surveys enables temporal comparisons and the assessment of structural changes in habitats over time.

In addition to national-scale products, LiDAR-derived vegetation structure metric datasets have been produced for seven MAMBO demonstration sites distributed across five European countries (Kissling et al., 2025). These sites represent diverse biogeographical regions and land-use contexts and illustrate the generation of harmonised LiDAR-based vegetation structure metrics across contrasting ecosystems. Spatial extents range from individual protected areas to larger landscape units, depending on the coverage of the underlying ALS data. Although smaller in extent than the national products, these datasets provide detailed, site-specific examples relevant for local and regional habitat condition assessments.

Temporal coverage varies according to the availability of ALS data. Some datasets represent single-date snapshots derived from individual surveys (e.g. Kissling et al., 2025; Wang et al., 2025), while others are explicitly multi-temporal and based on repeated surveys spanning multiple years or decades (Shi et al., 2025). Some datasets also include auxiliary layers, such as digital terrain models (DTMs), digital surface models (DSMs), and point or pulse density rasters, to support temporal consistency and minimise biases in comparative analyses.

Across all datasets, spatial and temporal coverage as well as characteristics of raw ALS point clouds (e.g. point density, flight dates, classification) are explicitly documented through metadata records, including information on geographic extent, coordinate reference systems, survey dates, and temporal resolution. This enables users to assess the suitability of each dataset for applications ranging from site-level habitat monitoring to national-scale assessments and comparative analyses across regions and time periods.

6 National-scale multi-temporal LiDAR metrics (Netherlands)

6.1 Dataset description

The national-scale dataset of **multi-temporal LiDAR-derived vegetation structure metrics** for the Netherlands was generated from four successive country-wide ALS surveys (AHN1–AHN4) and provides spatially explicit information on vegetation structure across the entire national territory (Figure 1). The dataset represents a harmonised set of geospatial data products derived from multiple ALS campaigns conducted over several decades (1996–2022), enabling consistent large-scale and long-term assessments of vegetation structure (Shi et al., 2025).

For each ALS survey, the data products consist of 25 raster layers capturing **key metrics of vegetation structure**, including vegetation height distribution, canopy cover and openness, and measures of vertical structural complexity (Figure 1). These metrics were derived from three-dimensional LiDAR point clouds and calculated at a spatial resolution of 10 m, providing wall-to-wall coverage across the Netherlands. This resolution represents a balance between preserving ecologically meaningful structural detail and ensuring robustness and comparability across different ALS surveys with varying point densities and acquisition characteristics.

The dataset is explicitly multi-temporal, incorporating information from **successive national ALS campaigns** acquired between the late 1990s and the early 2020s. By applying consistent processing and metric definitions across survey periods, these data products enable temporal comparisons of vegetation structure and support analyses of structural change in habitats over time. This includes applications such as long-term monitoring of woodland development, changes in habitat openness, and structural dynamics in managed and semi-natural landscapes.

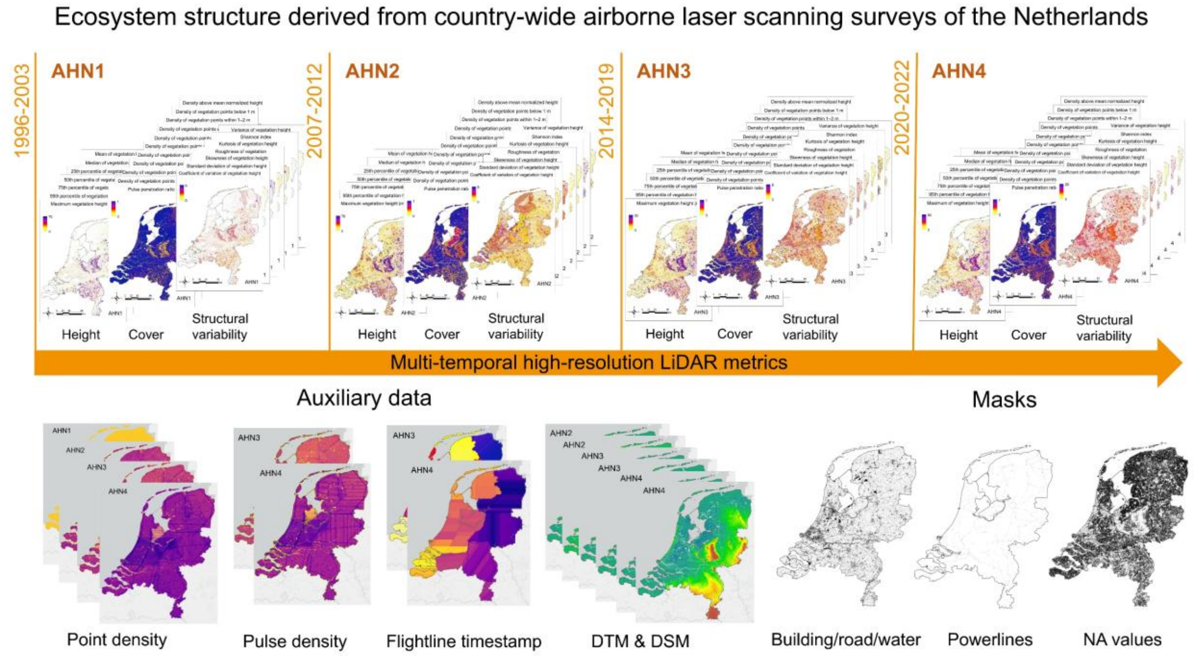

Figure 1: Geospatial data products of national-scale, multi-temporal LiDAR-derived vegetation structure metrics for the Netherlands. High-resolution (10 m) raster layers of ecosystem structure were generated from four successive country-wide airborne laser scanning surveys (AHN1–AHN4), together with auxiliary layers and mask products supporting interpretation and temporal comparability (from Shi et al., 2025).

All **raster layers** are provided in standard geospatial formats and use a consistent coordinate reference system, thereby facilitating integration with other spatial datasets and environmental information systems and providing ready-to-use vegetation structural information for ecological and ecosystem research. In addition to the core vegetation structure metrics, a set of **auxiliary layers** is provided where relevant (Figure 1), including point density, pulse density, and flightline timestamp rasters, DTMs, and DSMs, which support interpretation and ensure temporal consistency across successive ALS surveys. Mask layers are further included to minimise the influence of buildings, roads, water surfaces, powerlines, and no-data values on the derived vegetation structure metrics (Figure 1).

To support the reliability and reuse of the dataset, additional data products document **quality assurance and robustness assessments**. These include datasets evaluating the sensitivity of LiDAR-derived vegetation metrics to variations in pulse density across different habitat types, as well as datasets assessing the effects of terrain filtering and pre-classification accuracy in the AHN surveys on vegetation change detection. All such supporting data products are made openly available alongside the core metrics. Comprehensive metadata accompany the dataset, documenting spatial and temporal coverage, data provenance, processing assumptions, and licensing conditions, thereby supporting transparency, reproducibility, and long-term reuse.

The national-scale multi-temporal LiDAR-derived vegetation structure metrics dataset for the Netherlands constitutes a foundational data asset for habitat condition assessment within MAMBO. By providing harmonised, reproducible, and openly accessible data products at a national extent, the dataset supports spatially continuous analyses, cross-regional comparisons, and long-term monitoring of vegetation structure. Moreover, it can be readily reused within broader biodiversity monitoring and assessment frameworks beyond the lifetime of the project and serves as a **blueprint for upscaling vegetation structure metrics** derived from ALS data across Europe (Moudrý et al., 2026).

6.2 Data access and persistent identifiers

The national-scale multi-temporal LiDAR-derived vegetation structure metrics for the Netherlands are publicly available through the open-access research data repository **Zenodo**, ensuring long-term preservation, discoverability, and reuse. The processed geospatial data products are grouped as a coherent dataset and accompanied by descriptive documentation detailing their content, spatial and temporal coverage, and licensing conditions.

The dataset is assigned persistent **Digital Object Identifiers (DOIs)**, providing stable and citable references for the data products and enabling reliable linking between data products, related publications, and associated research objects. Zenodo supports explicit versioning, allowing users to identify, access, and cite specific dataset releases and ensuring transparency in the case of updates or revisions. Currently, a concept DOI representing all versions of the dataset is available (https://doi.org/10.5281/zenodo.13940846), together with DOIs for individual released versions (version v1: https://doi.org/10.5281/zenodo.13940847; version v2: https://doi.org/10.5281/zenodo.15261042; version v3: https://doi.org/10.5281/zenodo.18499735).

The Zenodo repository record provides the authoritative access point for the processed vegetation structure metrics and serves as the primary reference for citation and reuse of the dataset in subsequent studies, applications, and policy-relevant assessments.

6.3 Code and reproducibility

To support transparency, reproducibility, and reuse, all **code** used to generate the national-scale multi-temporal LiDAR-derived vegetation structure metrics for the Netherlands is made openly available. The processing workflows are implemented as Jupyter Notebooks and are published in a public GitHub repository (https://github.com/ShiYifang/AHN), which provides the complete source code required to reproduce the data products directly from the underlying raw ALS point clouds.

The **GitHub repository** documents the full processing chain, including scripts for data preparation, normalization, metric calculation, and rasterization steps. It further includes scripts and links for accessing the underlying raw ALS point cloud data from the Dutch national airborne laser scanning programme (Actueel Hoogtebestand Nederland, AHN), which are distributed via public data portals. Specifically, the raw LiDAR point clouds are provided through the AHN viewer (https://www.ahn.nl/ahn-viewer) or AHN data room (https://www.ahn.nl/dataroom), and hosted in the Dutch National Georegister (https://www.nationaalgeoregister.nl). Although the raw point cloud data are stored and managed separately due to their size and structure, their access points and provenance are clearly documented in the code repository and associated dataset metadata, enabling users to trace the complete data lineage from raw observations to derived vegetation structure metrics.

Reproducibility is further supported through explicit **references** in the Zenodo dataset record to the processing workflows and software dependencies used. These include links to the high-throughput Laserfarm workflow for large-scale LiDAR data processing (Kissling et al., 2022) and to the underlying open-source software components used for metric calculation (Meijer et al., 2020). In addition, the dataset record links to peer-reviewed data publications that describe the data products and processing assumptions in detail (Shi et al., 2025), as well as to related country-wide LiDAR-derived data products generated using comparable methods (Kissling et al., 2023).

Together, the openly available code, documented workflows, and explicit links between raw data, processing steps, and derived products ensure that the national-scale LiDAR-derived vegetation structure metrics for the Netherlands are fully reproducible. This approach enables users to replicate the data generation process, adapt the workflows to other regions or time periods, and build upon the MAMBO data products in future habitat condition assessments and biodiversity monitoring applications beyond the lifetime of the project.

7 LiDAR vegetation structure metrics for MAMBO demonstration sites

7.1 Sites and habitat types

LiDAR-derived vegetation structure metric datasets have been generated for seven **MAMBO demonstration sites** distributed across five European countries (Figure 2). The sites were selected to represent a broad range of biogeographical regions, land-use contexts, and habitat types, and to demonstrate the applicability of harmonised LiDAR-based vegetation structure metrics across contrasting European ecosystems (Kissling et al., 2025).

The demonstration sites encompass a wide **diversity of natural and semi-natural habitats**, including forests and woodlands (broadleaf and coniferous), small forest plantations, dry and wet grasslands, marshes and reedbeds, scrublands, arable fields and farmland, as well as Mediterranean habitats such as garigue (Kissling et al., 2025). These habitats are representative of ecosystem types that are central to European biodiversity monitoring and are sensitive to pressures such as land-use change, management practices, hydrological alterations, and grazing.

Site extents range from approximately 0.08 km^2 to 54 km^2 and include individual protected areas and nature reserves as well as larger landscape units. All sites are covered by country-wide or regional ALS surveys, enabling the derivation of comparable vegetation structure metrics through harmonised processing workflows. Together, the demonstration sites complement the national-scale datasets by providing ecologically diverse, site-level examples that support site management, regional assessments, and cross-site comparisons within and beyond the MAMBO project.

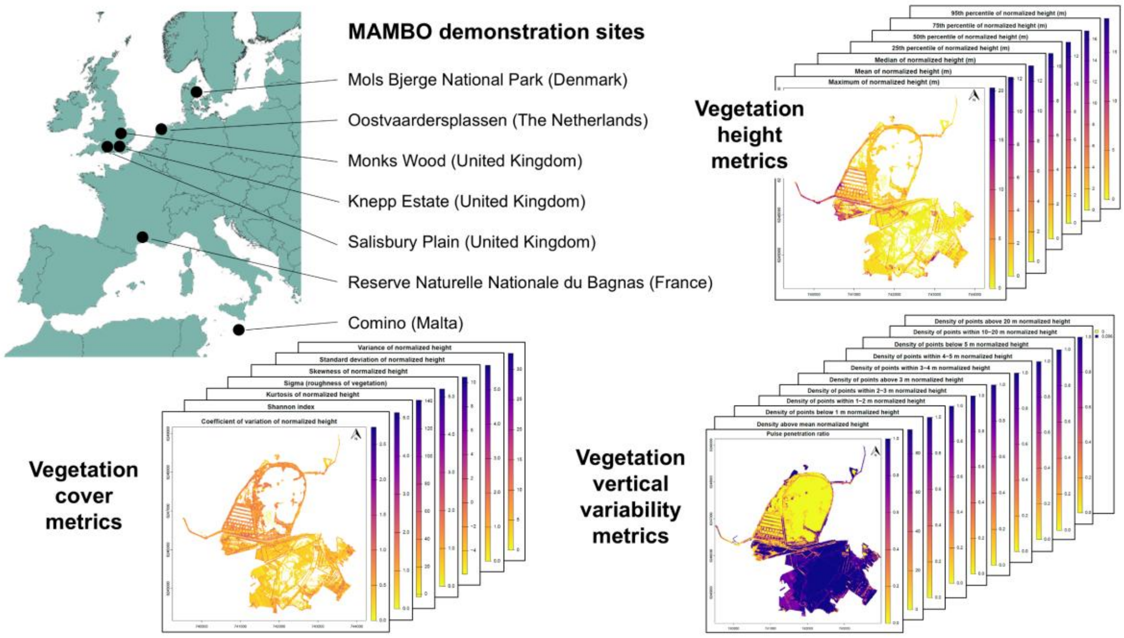

Figure 2: Geospatial data products of site-specific LiDAR-derived vegetation structure metrics for MAMBO demonstration sites. Harmonised high-resolution (10 m) raster layers of habitat condition metrics were generated for seven demonstration sites distributed across five European countries. The examples shown illustrate vegetation structure raster layers produced for the Réserve Naturelle Nationale du Bagnas in France (see Kissling et al., 2025).

7.2 Dataset description

The LiDAR-derived vegetation structure metrics for the MAMBO demonstration sites comprise a harmonised set of **site-specific geospatial data products** derived from national or regional ALS surveys (Kissling et al., 2025). The datasets demonstrate how comparable vegetation structure metrics can be generated from heterogeneous national LiDAR point cloud datasets using reproducible and scalable workflows.

For each site, the dataset includes processed **raster layers** describing key aspects of vegetation structure and habitat condition, such as vegetation height distribution, canopy cover and openness, and measures of vertical structural complexity. Metrics were derived from three-dimensional ALS point clouds and calculated at a spatial resolution of 10 m, providing a balance between ecological relevance and robustness to variations in point density and acquisition characteristics among national ALS surveys. The resulting raster layers provide spatially continuous coverage within each site boundary and are directly suitable for site-level habitat condition assessments.

Vegetation structure metrics were generated by clipping national or regional **ALS point clouds** to site-specific boundary polygons provided as geospatial vector data, ensuring alignment with the management or study units of each demonstration site. The underlying ALS datasets originate from different national repositories and vary in point density, pre-classification schemes, and acquisition parameters, reflecting the diversity of ALS data characteristics across Europe (Kissling et al., 2025).

All datasets are provided in **standard geospatial formats** and are accompanied by detailed metadata describing spatial extent, coordinate reference systems, acquisition periods, processing assumptions, and licensing conditions. Together, the demonstration site datasets provide methodologically consistent and ecologically diverse examples of LiDAR-derived habitat condition metrics that complement the national-scale data products documented in this deliverable.

7.3 Data, workflows and raw inputs

The LiDAR-derived vegetation structure metrics for the MAMBO demonstration sites are made publicly available as a comprehensive and well-documented data package comprising processed datasets, processing workflows, and the corresponding raw input data. This integrated approach supports transparency, reproducibility, and reuse.

Processed datasets

All processed data products are published in the open-access research data repository **Zenodo** under a single concept DOI representing all dataset versions (https://doi.org/10.5281/zenodo.14745309). Version-specific DOIs are also provided (v1: https://zenodo.org/records/14745310; v2: https://doi.org/10.5281/zenodo.17406711), ensuring transparent versioning and long-term traceability.

For each demonstration site, the Zenodo repository contains a dedicated **ZIP archive** including:

- Six Jupyter Notebooks implementing the Laserfarm workflow, together with text files specifying the coordinate reference system (EPSG code) and re-tiling grid configuration;
- GeoTIFF raster layers containing 35 LiDAR-derived vegetation structure metrics at 10 m resolution;
- PDF map visualisations illustrating the spatial distribution of all derived metrics.

Additional files include a ZIP archive with GIS shapefiles defining site boundaries, a machine-readable JSON metadata file, a README describing dataset structure and content, and a detailed methodological PDF. The methodological documentation provides background on the Laserfarm workflow, LiDAR pre-processing steps for the demonstration sites, and implementation details, including the high-performance computing environment and the four main workflow modules.

Workflows

All Jupyter Notebooks used to generate the demonstration site datasets are also published independently on **WorkflowHub** (https://workflowhub.eu/projects/302#workflows). WorkflowHub provides persistent identification of the workflows and facilitates discovery, inspection, adaptation, and reuse in other computational environments.

Raw LiDAR inputs

The **clipped ALS point clouds** used as input for each demonstration site are available from a publicly accessible repository (https://doi.org/10.48546/workflowhub.datafile.5.1). Providing access to these site-specific raw inputs enables users to reproduce the full processing chain from raw point clouds to derived vegetation structure metrics or to apply alternative processing approaches.

8 Thematic and methodological LiDAR datasets

8.1 Ungulate trail extraction dataset

In addition to vegetation structure metrics, MAMBO has generated a thematic LiDAR-derived dataset focusing on the extraction of **ungulate trails as spatial indicators** of animal space use and trampling pressure (Figure 3). The dataset demonstrates how high-resolution three-dimensional ALS point clouds can be used to detect fine-scale linear features associated with large herbivore activity and to support habitat condition assessment in grazed wetland ecosystems.

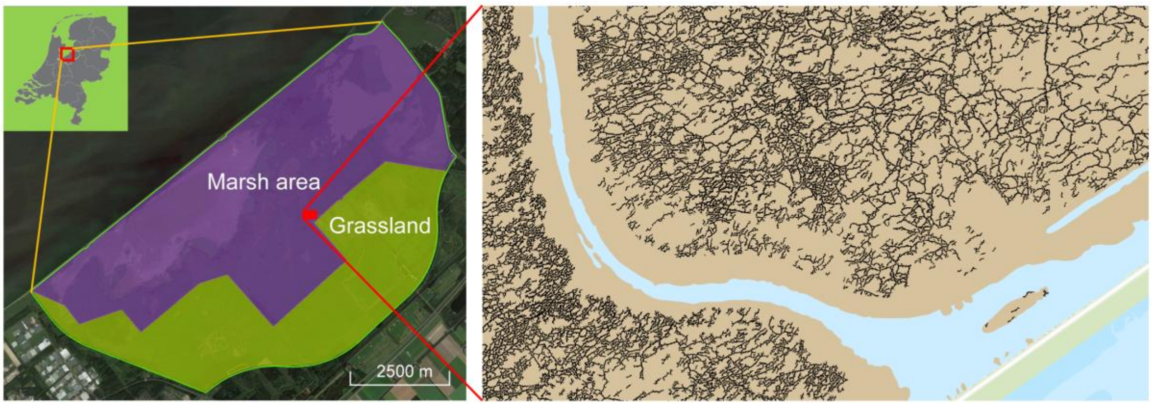

Figure 3: Ungulate trail extraction from airborne laser scanning in a Dutch nature reserve. Left: Location of the Oostvaardersplassen nature reserve in the Netherlands, showing marsh and grassland habitats. Right: Example of LiDAR-derived trail extraction results illustrating red deer (*Cervus elaphus*) trails in reedbeds within the extensive marsh area. For methodological details, see Wang et al. (2025).

The dataset is illustrated using a 36 km² wetland and marsh area within the **Oostvaardersplassen nature reserve** in the Netherlands (Figure 3) and is based on country-wide ALS data (AHN4). A dedicated workflow was developed to extract ungulate trails formed by repeated trampling (Wang et al., 2025), particularly in reedbed habitats, resulting in spatially explicit representations of trail networks (Figure 3) that are difficult to capture through conventional field-based monitoring at such a large spatial extent.

All data products are made publicly available through the open-access research data repository **Zenodo**. The dataset is assigned a concept DOI representing all versions (https://doi.org/10.5281/zenodo.13889768), as well as version-specific DOIs (v1: https://doi.org/10.5281/zenodo.13889769; v2: https://doi.org/10.5281/zenodo.14986928; v3: https://doi.org/10.5281/zenodo.14987391; v4: https://doi.org/10.5281/zenodo.14987500), ensuring persistent access and transparent versioning.

The Zenodo repository contains a **structured data package** comprising (i) geospatial shapefiles defining the study area and grazing regimes, (ii) trail extraction results provided in both LiDAR (LAZ) and raster (GeoTIFF) formats at 10 cm spatial resolution, and (iii) manually created validation data, including ground-truth plots and point-level annotations distinguishing trail and non-trail features. In addition, plot-wise accuracy files are provided to support quantitative evaluation of the extraction results. A README file documents the dataset structure and content in detail.

The processing workflow used to generate the dataset is made openly available through a public **GitHub repository** (https://github.com/Jinhu-Wang/Extracting-ungulate-trails-in-wetlands-using-3D-point-clouds-obtained-from-airborne-laser-scanning) under an Apache 2.0 licence. The workflow implements four main processing stages—pre-processing, near-terrain filtering, digital terrain model generation, and trail extraction—and includes all required source code and dependencies to reproduce the results from the raw ALS point clouds.

The methodological details of the workflow, including algorithm design, validation strategy, and performance assessment, are described in a **peer-reviewed scientific publication** (Wang et al., 2025). Together, the openly available data, validation material, processing code, and scientific documentation provide a complete and reproducible thematic dataset that complements the core LiDAR-derived vegetation structure metrics in MAMBO. The ungulate trail extraction dataset extends the scope of LiDAR-based habitat condition assessment by providing spatially explicit indicators of grazing-related disturbance and animal space use patterns that can be reused and further developed beyond the lifetime of the project.

8.2 Robustness of LiDAR metrics datasets

To support the upscaling and cross-country application of LiDAR-derived vegetation structure metrics, MAMBO has generated two methodological datasets designed to assess the **robustness of vegetation structure metrics** to variations in ALS point density and spatial resolution. Together, these datasets address a key source of uncertainty in large-scale habitat condition assessments: the comparability of LiDAR-derived metrics across national ALS surveys with differing acquisition characteristics.

Both datasets are based on the Dutch national airborne laser scanning dataset AHN4 (2020–2022), which has an original point density of approximately 20–30 points/m^2. Using this dataset, a consistent set of **25 LiDAR vegetation metrics** describing vegetation height, vegetation cover, and vertical structural complexity was calculated within Natura 2000 sites in the Netherlands.

The first dataset focuses on woodland habitats and evaluates metric **robustness across multiple spatial resolutions and point densities** (Supplementary Data 2 in Kissling et al., 2024). Metrics were calculated for plots of 1 × 1 m, 2 × 2 m, 5 × 5 m, and 10 × 10 m, using the original AHN4 point density and six systematically down-sampled point clouds retaining 5%, 10%, 20%, 40%, 60%, and 80% of the original points (see example in Figure 4). This design enables a quantitative assessment of metric sensitivity to both spatial resolution and point density.

The second dataset extends the robustness analysis across **multiple habitat types**, including dunes, marshes, shrublands, grasslands, and woodlands (Appendix A and B in Shi et al., 2025). In this dataset, the spatial resolution was fixed at 10 × 10 m, while pulse densities (instead of point densities, as suggested by a reviewer) were systematically reduced from the original AHN4 density to levels comparable to earlier Dutch ALS surveys (AHN1–AHN3). This enables a direct comparison of metric behaviour across different habitats and ALS generations, ensuring temporal consistency in long-term monitoring.

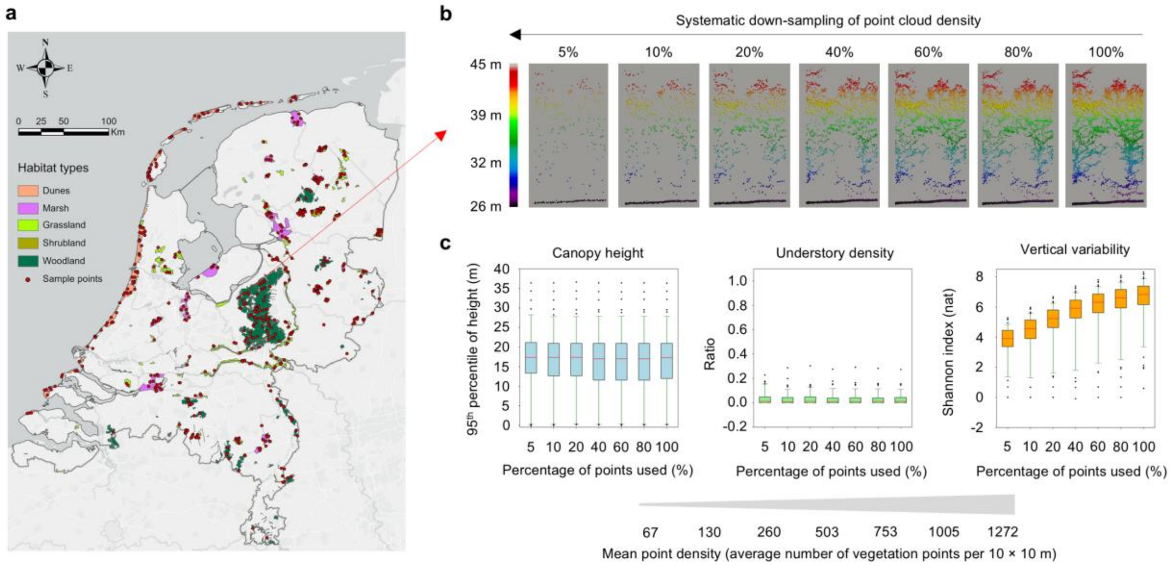

Figure 4: Assessing the robustness of LiDAR-derived vegetation structure metrics to variations in airborne laser scanning point density. (**a**) Distribution of sample plots across Dutch Natura 2000 sites representing dunes, marshes, grasslands, shrublands, and woodlands (adapted from Shi et al., 2025). (**b**) Illustration of systematic down-sampling of a LiDAR point cloud for a woodland plot, from the original point density of the Dutch national airborne laser scanning dataset (AHN4; 100%) to reduced point densities retaining 5%, 10%, 20%, 40%, 60%, and 80% of the points. (**c**) Effect of point cloud down-sampling on the robustness of three vegetation structure metrics representing canopy height, understory density, and vertical variability across randomly located 10 × 10 m woodland plots in the Netherlands (n = 94), with mean point density across plots indicated (adapted from Kissling et al., 2024).

The first robustness dataset is publicly available through an open-access repository on **Zenodo**, with a concept DOI representing all versions (https://doi.org/10.5281/zenodo.13619387) and multiple version-specific DOIs (v1.0: https://zenodo.org/records/13619388; v1.1: https://doi.org/10.5281/zenodo.13628041; v3: https://doi.org/10.5281/zenodo.18495708; v3.1: https://doi.org/10.5281/zenodo.18506099). The repository includes geolocation data and plot definitions, clipped and down-sampled LiDAR point clouds, computed vegetation metrics, scripts for metric computation and visualisation, and graphical summaries of the results. In addition, the full processing and analysis workflow is made openly available via a public GitHub repository (https://github.com/Jinhu-Wang/Testing-the-robustness-of-LiDAR-vegetation-metrics-to-varying-point-densities), supporting transparency, reproducibility, and reuse.

The second robustness dataset is included in the national-scale multi-temporal LiDAR data repository for the Netherlands on **Zenodo** (concept DOI: https://doi.org/10.5281/zenodo.13940846). Within this repository, the folder '8_Sensitivity_analysis.zip' contains the centroids, polygons, and clipped point clouds of the sampling plots, as well as the vegetation metrics generated from point clouds with original and reduced pulse densities for each habitat type. The folder further includes the scripts and figures used in the analysis. Version-specific releases of this repository are available (v1: https://doi.org/10.5281/zenodo.13940847; v2: https://doi.org/10.5281/zenodo.15261042;

v3: https://doi.org/10.5281/zenodo.18499735), ensuring persistent access and transparent versioning.

Both datasets underpin **peer-reviewed scientific analyses** of LiDAR metric robustness (Kissling et al., 2024; Shi et al., 2025). Detailed methodological documentation, including experimental design, down-sampling strategies, and metric definitions, is provided in the supplementary materials of these publications. Together, the openly available data, code, and documentation provide a transparent and reusable methodological reference for selecting and applying LiDAR-derived vegetation metrics in regions with heterogeneous ALS data characteristics.

By explicitly quantifying metric sensitivity to point/pulse density and spatial resolution across habitats, these datasets complement the core vegetation structure metric products generated within MAMBO. They support informed metric selection, increase confidence in cross-site and cross-country comparisons, and contribute to the development of consistent, scalable, and FAIR LiDAR-based habitat condition assessments beyond the lifetime of the project.

8.3 Tree and shrub extraction dataset

In addition to vegetation structure metrics and ungulate trail extraction, MAMBO has generated a thematic LiDAR-derived dataset focusing on the extraction and mapping of **woody vegetation**, specifically trees and shrubs, in wetland environments (Figure 5). The dataset demonstrates how high-resolution three-dimensional ALS point clouds can be used to identify and delineate woody vegetation structures in extensive and largely inaccessible marsh and reedbed habitats, where field-based mapping is logistically challenging.

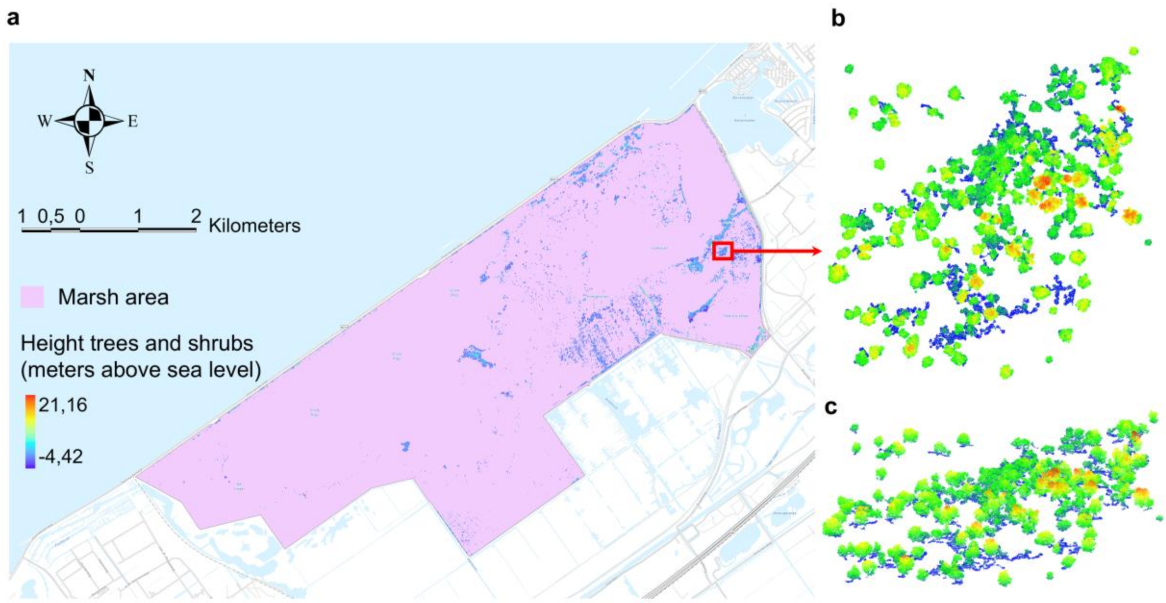

Figure 5: Extraction of trees and shrubs from airborne laser scanning (AHN4) point clouds in the Oostvaardersplassen nature reserve, the Netherlands. (**a**) Wall-to-wall extraction results showing the spatial distribution of trees and shrubs across the study area. (**b**–**c**) Top and side views of the extracted woody vegetation within the highlighted area, illustrating the three-dimensional structure of the identified trees and shrubs.

The dataset is illustrated using the Oostvaardersplassen nature reserve in the Netherlands and is derived from country-wide ALS data (AHN4). Within this wetland system, woody vegetation such as willow (*Salix* spp.) and elder (*Sambucus nigra*) forms an important structural component of the reedbed ecosystem and plays a key role in ecological succession and grazing dynamics. A dedicated workflow was developed to classify and extract woody vegetation points from ALS point clouds in marsh areas, where trees and shrubs co-occur with dense reed vegetation. The resulting data products provide explicit spatial representations of trees and shrubs that are relevant for assessing vegetation dynamics, grazing impacts, and habitat development.

All data products are made publicly available through the open-access research data repository **Zenodo**. The dataset is assigned a persistent concept DOI representing all versions (https://doi.org/10.5281/zenodo.18511137), as well as a version-specific DOI for the initial release (v1.0: https://doi.org/10.5281/zenodo.18511138; v2: https://doi.org/10.5281/zenodo.18594751), ensuring long-term accessibility and transparent versioning.

The Zenodo repository contains a **structured data package** comprising: (i) vegetation-classified ALS point clouds from AHN4 for the Oostvaardersplassen study area, provided in standard LAS format; (ii) extracted tree and shrub point clouds, also provided in LAS format; (iii) Python scripts for converting the extracted woody vegetation points to raster GeoTIFF products and for deriving spatial boundaries from these rasters; and (iv) a README file describing the structure, content, and intended use of the dataset.

The **processing workflow** used to generate the dataset is made openly available through a public GitHub repository (https://github.com/Jinhu-Wang/Tree_Classification_and_Individualization_In_Marsh_Area) under an Apache 2.0 licence. The workflow implements an iterative classification approach that progressively filters out reed vegetation points from the ALS point clouds while retaining woody vegetation points. The repository includes detailed documentation of software dependencies, usage instructions, example data, validation material, and sample code, enabling users to reproduce the results and adapt the workflow to other wetland areas or ALS datasets.

Together, the openly available data products, processing workflow, and documentation provide a reproducible thematic dataset that complements the ungulate trail extraction and robustness testing datasets within MAMBO. The tree and shrub extraction dataset extends the scope of LiDAR-based habitat condition assessment by enabling spatially explicit analysis of woody vegetation structure and dynamics in wetland ecosystems. In addition, it provides a methodological starting point for the further development and upscaling of tree classification and individualization workflows from site-based applications to larger spatial extents, including national and transnational airborne laser scanning datasets, supporting future large-scale habitat monitoring efforts beyond the lifetime of the project.

9 FAIR data implementation

9.1 Findability

The geospatial LiDAR-derived data products generated within the MAMBO project are made findable through the systematic use of persistent identifiers, rich metadata, and registration in established European and international data discovery services. Together, these measures ensure that datasets can be reliably discovered, cited, and linked to related research outputs by both humans and machines.

All datasets documented in this deliverable are published in open-access research data repositories and assigned persistent DOIs. The primary repository used is Zenodo, which forms part of the **EU Open Research Repository** infrastructure (https://zenodo.org/communities/eu/). For each dataset, a concept DOI represents all released versions, while version-specific DOIs enable precise citation of individual dataset releases, supporting long-term persistence, transparent versioning, and unambiguous dataset identification.

To further enhance discoverability, all datasets are included in the **MAMBO community on Zenodo** (https://zenodo.org/communities/mambo/records), providing a common project-level entry point. Zenodo records include descriptive titles, abstracts, keywords, licensing information, and links to related research objects such as publications, workflows, and source code repositories, supporting keyword-based search and machine indexing.

In addition, all datasets are registered in the **LifeWatch Metadata Catalogue** (https://metadatacatalogue.lifewatch.eu, search for 'MAMBO'), where they are described using structured metadata compliant with biodiversity and geospatial data standards. This registration increases visibility within the biodiversity research community and enables discovery through federated searches across European research infrastructures.

Processing workflows associated with the datasets, including Jupyter Notebooks implementing the Laserfarm workflow and other thematic pipelines, are published and indexed in **WorkflowHub** (https://workflowhub.eu/projects/302#workflows) or **GitHub**, and linked to the corresponding datasets via persistent identifiers. Finally, datasets and their associated data publications are aggregated in the **MAMBO RIO topical collection** (https://riojournal.com/topical_collection/223/), providing a curated, project-specific discovery context. Together, these measures ensure that MAMBO LiDAR-derived data products remain highly findable beyond the lifetime of the project.

9.2 Accessibility

The LiDAR-derived geospatial data products generated within MAMBO are made openly accessible through established, **long-term research data infrastructures** using standardised and well-documented access mechanisms.

All datasets are published in **open-access repositories**, primarily Zenodo, which provides free and unrestricted access to data files via persistent DOIs. Access does not require user registration, authentication, or proprietary software, and datasets can be downloaded directly using standard web protocols (HTTP/HTTPS), ensuring broad accessibility for researchers, practitioners, and other stakeholders.

Access to **supporting research objects** is also provided to ensure transparency and reproducibility. Processing workflows are openly accessible through WorkflowHub, while source code repositories are hosted on public version-control platforms. These resources are directly linked from dataset records and can be accessed independently.

Where relevant, **access points to the underlying raw ALS point clouds** are documented. Although raw LiDAR data are often hosted separately due to their size and national data management policies, their locations in public data portals are clearly referenced in the metadata of the processed datasets, enabling users to retrieve original inputs subject to the conditions defined by the data providers.

All dataset records include clear **information on data availability, access conditions, and licensing**. By combining open repositories, standard access protocols, and transparent documentation, MAMBO ensures that its LiDAR-derived data products are accessible in a reliable and user-friendly manner throughout and beyond the project lifetime.

9.3 Interoperability

The LiDAR-derived geospatial data products generated within MAMBO are designed to be interoperable across software environments, research infrastructures, and application domains. **Interoperability** is achieved using open data formats, consistent spatial referencing, harmonised metric definitions, and structured metadata.

Processed data products are provided in widely used **open geospatial formats**, including GeoTIFF for raster data, ESRI Shapefile for vector data, and LAS or LAZ for LiDAR point clouds. These formats are supported by a broad range of GIS, remote sensing software, and programming libraries, facilitating integration with other spatial datasets.

Spatial interoperability is ensured through explicit documentation of **coordinate reference systems** using standard EPSG codes (https://epsg.io), enabling consistent spatial alignment and transformation across datasets from different countries and sources. Harmonised processing workflows are applied where datasets originate from different national ALS programmes, ensuring comparability of derived vegetation structure metrics despite differences in underlying data characteristics.

Semantic interoperability is supported through consistent and well-documented definitions of LiDAR-derived vegetation structure metrics. Metric definitions are aligned with those used in established open-source tools and workflows, including Laserfarm (Kissling et al., 2022), Laserchicken (Meijer et al., 2020), and lidR (Roussel et al., 2020), and are described in accompanying documentation and peer-reviewed data publications. This supports consistent interpretation and reuse across studies and regions.

Structured, machine-readable metadata accompany all datasets and are registered in domain-relevant catalogues such as the LifeWatch Metadata Catalogue, enabling harvesting and indexing by external discovery services. Together, these measures ensure that MAMBO's

LiDAR-derived data products can be readily integrated with other datasets and analytical pipelines for cross-site, cross-country, and cross-domain applications.

9.4 Reusability

The LiDAR-derived geospatial data products generated within MAMBO are designed to be reusable across a wide range of scientific, management, and policy-relevant applications. **Reusability** is supported through open licensing, comprehensive documentation, transparent data provenance, and the provision of reproducible workflows and validation materials.

All datasets are released through open-access repositories, primarily Zenodo, under licences that permit reuse, redistribution, and adaptation. **Licensing conditions** are clearly stated in dataset metadata, enabling users to understand the terms under which the data can be reused.

Each dataset is accompanied by **detailed documentation**, including README files, machine-readable metadata, and methodological descriptions provided through linked data publications. These resources document data provenance, processing assumptions, spatial and temporal coverage, and known limitations, facilitating correct interpretation and informed reuse.

Reusability is further enhanced through the availability of **openly accessible processing workflows, source code, tutorials and documented examples**. Jupyter Notebooks and scripts used to generate the LiDAR-derived vegetation structure metrics are published in WorkflowHub and public version-control repositories, enabling users to reproduce the data products, adapt workflows to new regions or time periods, and extend analyses using comparable methods.

Several datasets additionally include **validation data, quality assessment outputs, and robustness analyses** that quantify metric sensitivity to factors such as point density and spatial resolution. These materials provide essential context on data reliability and applicability. Together with links to peer-reviewed scientific publications describing the datasets and methodologies in detail, these measures ensure that MAMBO's LiDAR-derived data products are reusable beyond the lifetime of the project and can support future habitat condition assessments and biodiversity monitoring initiatives at national and European scales.

10 Conclusions

Deliverable D6.2 documents the geospatial airborne LiDAR-derived data products of vegetation structure generated within the MAMBO project and provides a consolidated overview of their content, scope, and availability. The deliverable focuses on the data assets themselves, including national-scale, site-based, thematic, and methodological datasets, and describes how these data products support quantitative, scalable assessments of habitat condition across multiple spatial and temporal scales.

The datasets presented in this deliverable demonstrate the applicability of airborne laser scanning for deriving harmonised vegetation structure metrics across heterogeneous national data sources and diverse European ecosystems. Together, the national-scale multi-temporal products, demonstration site datasets, and thematic and methodological datasets form a coherent data portfolio that supports spatially continuous monitoring, cross-site and cross-country comparisons, and the evaluation of structural habitat change over time. In addition, methodological datasets addressing robustness and sensitivity contribute to informed metric selection and strengthen confidence in the reuse of LiDAR-derived vegetation structure metrics across different airborne laser scanning datasets.

All datasets generated within MAMBO have been made openly available through established research data infrastructures and are accompanied by persistent identifiers, rich metadata, and openly accessible processing workflows. The implementation of the FAIR principles ensures that the data products are findable, accessible, interoperable, and reusable, facilitating their uptake by the scientific community, environmental agencies, and other stakeholders. By explicitly linking processed data products to raw inputs, workflows, validation materials, and peer-reviewed publications, the deliverable supports transparency, reproducibility, and long-term reuse.

By providing a structured and FAIR-compliant overview of LiDAR-derived vegetation structure metric datasets, Deliverable D6.2 contributes to the data legacy of the MAMBO project. The documented data products offer a robust foundation for future biodiversity monitoring, habitat condition assessment, and upscaling efforts at national and European levels, and support the continued use and further development of LiDAR-based approaches beyond the lifetime of the project.

11 Acknowledgements

We thank Yifang Shi and Wessel Mulder for generating some of the geospatial data products and associated code that are presented in this deliverable. Yifang Shi and France Gerard reviewed a draft version of this document. Lucia Vaira supported the upload of metadata to the LifeWatch Metadata Catalogue. We acknowledge the computing resources provided by SURF, the Dutch national facility for information and communication technology. The development of the data products was also supported by LifeWatch ERIC.

12 References

Assmann, J. J., Moeslund, J. E., Treier, U. A., & Normand, S. (2022). 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(2), 823-844. https://doi.org/10.5194/essd-14-823-2022.

Bakx, T. R. M., Koma, Z., Seijmonsbergen, A. C., & Kissling, W. D. (2019). Use and categorization of Light Detection and Ranging vegetation metrics in avian diversity and species distribution research. *Diversity and Distributions*, 25(7), 1045-1059. https://doi.org/10.1111/ddi.12915.

Davies, A. B., & Asner, G. P. (2014). Advances in animal ecology from 3D-LiDAR ecosystem mapping. *Trends in Ecology & Evolution*, 29(12), 681–691. https://doi.org/10.1016/j.tree.2014.10.005.

de Vries, J. P. R., Koma, Z., WallisDeVries, M. F., & Kissling, W. D. (2021). Identifying fine-scale habitat preferences of threatened butterflies using airborne laser scanning. *Diversity and Distributions*, 27(7), 1251-1264. https://doi.org/10.1111/ddi.13272.

European Environment Agency (2020) 'State of nature in the EU: Results from reporting under the nature directives 2013-2018'. Luxembourg: Publications Office of the European Union.

Høye, T. T., August, T., Balzan, M. V., Biesmeijer, K., Bonnet, P., Breeze, T. D., et al. (2023). Modern Approaches to the Monitoring of Biodiversity (MAMBO). *Research Ideas and Outcomes*, 9, e116951. https://doi.org/10.3897/rio.9.e116951.

Kissling, W. D., Mulder, W., Wang, J., & Shi, Y. (2025). Data of vegetation structure metrics retrieved from airborne laser scanning surveys for European demonstration sites. *Data in Brief*, 60, 111548. https://doi.org/10.1016/j.dib.2025.111548.

Kissling, W. D., & Shi, Y. (2023). Which metrics derived from airborne laser scanning are essential to measure the vertical profile of ecosystems? *Diversity and Distributions*, 29(10), 1315-1320. https://doi.org/10.1111/ddi.13760.

Kissling, W. D., Shi, Y., Koma, Z., Meijer, C., Ku, O., Nattino, F., et al. (2022). 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.

Kissling, W. D., Shi, Y., Koma, Z., Meijer, C., Ku, O., Nattino, F., et al. (2023). 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.

Kissling, W. D., Shi, Y., Wang, J., Walicka, A., George, C., Moeslund, J. E., et al. (2024). Towards consistently measuring and monitoring habitat condition with airborne laser scanning and unmanned aerial vehicles. *Ecological Indicators*, 169, 112970. https://doi.org/10.1016/j.ecolind.2024.112970.

Koma, Z., Grootes, M. W., Meijer, C. W., Nattino, F., Seijmonsbergen, A. C., Sierdsema, H., et al. (2021). Niche separation of wetland birds revealed from airborne laser scanning. *Ecography*, 44(6), 907-918. https://doi.org/10.1111/ecog.05371.

Maes, J., Bruzón, A. G., Barredo, J. I., Vallecillo, S., Vogt, P., Rivero, I. M., et al. (2023). Accounting for forest condition in Europe based on an international statistical standard. *Nature Communications*, 14(1), 3723. https://doi.org/10.1038/s41467-023-39434-0.

Meijer, C., Grootes, M. W., Koma, Z., Dzigan, Y., Gonçalves, R., Andela, B., et al. (2020). 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.

Moudrý, V., Remelgado, R., Forkel, M., Torresani, M., Laurin, G. V., Šárovcová, E., et al. (2026). Spaceborne canopy height products should be complemented with airborne laser scanning data: Toward a European canopy height model. *Earth and Space Science*, 13(1), e2025EA004544. https://doi.org/10.1029/2025EA004544.

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

Shi, Y., Wang, J., & Kissling, W. D. (2025). 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(7), 3641-3677. https://doi.org/10.5194/essd-17-3641-2025.

Wang, J., Cornelissen, P., & Kissling, W. D. (2025). A workflow for extracting ungulate trails in wetlands using 3D point clouds obtained from airborne laser scanning (Methods). *Frontiers in Remote Sensing*, 6, 1599128. https://doi.org/10.3389/frsen.2025.1599128.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18.

MAMBO

MODERN APPROACHES TO THE MONITORING OF BIODIVERSITY

www.mambo-project.eu

Project partners

AARHUS UNIVERSITY

Naturalis Biodiversity Center

UFZ

University of Reading

Inria

PENSOFT

UK CEH

cirad

ECOSTACK INNOVATIONS LIMITED

13 Annex Table A1. Overview of LiDAR-derived geospatial data products and associated FAIR-enabling research objects

Table A1. Overview of LiDAR-derived geospatial datasets generated within the MAMBO project and their associated FAIR-enabling research objects, including processing workflows, metadata records, and discovery services. *

Dataset title	Geographic coverage	Temporal coverage	Spatial resolution	Habitat types	Data/object type	Repository/ access point	Persistent identifier (DOI/URL)	Related publication	Licence
LiDAR-derived geospatial datasets									
Multi-temporal high-resolution data products of ecosystem structure derived from country-wide airborne laser scanning surveys of the Netherlands	Netherlands (national coverage)	Multiple ALS surveys (1996–2022)	10 m	Forests, woodlands, grasslands, wetlands, agricultural land, scrub, urban etc.	Raster LiDAR vegetation structure metrics (GeoTIFF)	Zenodo	https://doi.org/10.5281/zenodo.13940846	Shi et al. (2025)	CC BY 4.0
LiDAR-derived vegetation structure metrics across seven MAMBO demonstration sites	Seven demonstration sites in five European countries	Site-specific ALS survey years	10 m	Forests, woodlands, grasslands (dry and wet), marshes, reedbeds, farmland, scrublands, Mediterranean garigue	Raster LiDAR vegetation structure metrics (GeoTIFF)	Zenodo	https://doi.org/10.5281/zenodo.14745309	Kissling et al. (2025)	CC BY 4.0
Raw airborne laser scanning point clouds for MAMBO demonstration sites	Seven demonstration sites in five European countries	Site-specific ALS survey years	Native ALS resolution (points)	Forests, woodlands, grasslands (dry and wet), marshes, reedbeds, farmland, scrublands, Mediterranean garigue	Raw LiDAR point clouds (LAZ)	Public LiDAR repository	https://public.spider.surfsara.nl/project/lidarac/MAMBO/	Kissling et al. (2025)	Open (specified per dataset)
Extracted ungulate trails from airborne laser scanning in wetlands	Oostvaardersplassen, Netherlands (~36 km²)	Single ALS survey	10 cm	Reedbeds, marshes	Raster trail networks + ground truth plots	Zenodo	https://doi.org/10.5281/zenodo.13889768	Wang et al. (2025)	CC BY 4.0
Datasets for testing the robustness of LiDAR vegetation metrics to	Netherlands (100 test plots)	Single ALS survey with simulated	1 × 1 m, 2 × 2 m, 5 × 5	Woodland habitats	Tabular LiDAR metrics, vector data (ESRI	Zenodo	https://doi.org/10.5281/zenodo.13619387	Kissling et al. (2024)	CC BY 4.0

varying point densities within woodlands		point densities	m, and 10 × 10 m		shapefiles), LiDAR point clouds (LAZ)				
Datasets for testing the robustness of LiDAR vegetation metrics to varying pulse densities across habitats	Netherlands (500 test plots, i.e. 100 plots per habitat type)	Single ALS survey with simulated pulse densities	10 × 10 m	Five habitat types (i.e. dunes, marshes, shrublands, grasslands, and woodlands)	Tabular LiDAR metrics, vector data (ESRI shapefiles), LiDAR point clouds (LAZ)	Zenodo	https://doi.org/10.5281/zenodo.13940846	Shi et al. (2025)	CC BY 4.0
Extracted trees and shrubs from airborne laser scanning in wetlands	Oostvaardersplassen, Netherlands	AHN4 ALS survey (2020–2022)	Native ALS resolution (points); derived raster products	Wetlands, marshes, reedbeds	LiDAR point clouds (LAS), derived raster products (GeoTIFF), vector data (ESRI shapefiles)	Zenodo	https://doi.org/10.5281/zenodo.18511137	Not applicable (dataset description and workflow documentation provided via GitHub)	CC BY 4.0
Associated FAIR-enabling research objects									
Laserfarm workflows for generating national-scale multi-temporal LiDAR vegetation metrics (Netherlands)	Netherlands (national coverage)	Multiple ALS surveys (1996–2022)	10 m (derived products)	Forests, woodlands, grasslands, wetlands, agricultural land, scrub, urban etc.	Jupyter Notebooks / processing workflow	GitHub	https://github.com/ShiYifang/AHN	Shi et al. (2025)	Apache-2.0
Laserfarm workflows for calculating LiDAR vegetation metrics (demonstration sites)	Seven demonstration sites in five European countries	Not applicable	Not applicable	Multiple habitat types	Jupyter Notebooks / workflows	WorkflowHub	https://workflowhub.eu/projects/302#workflows	Kissling et al. (2025)	Apache-2.0
Workflow for extracting ungulate trails from airborne laser scanning in wetlands	Oostvaardersplassen, Netherlands (~36 km²)	Single ALS survey (AHN4)	Native ALS resolution; derived raster outputs at 10 cm	Reedbeds, marshes	Processing workflow / source code	GitHub	https://github.com/Jinhu-Wang/Extracting-ungulate-trails-in-wetlands-using-3D-point-clouds-obtained-from-airborne-laser-scanning	Wang et al. (2025)	Apache-2.0
Processing and analysis workflow for testing	Netherlands (100 test plots)	Single ALS survey	1 × 1 m, 2 × 2 m, 5 × 5	Woodland habitats	Processing and analysis code	GitHub	https://github.com/Jinhu-	Kissling et al. (2024)	Apache-2.0

robustness of LiDAR vegetation metrics to varying point densities within woodlands		(AHN4, 2020–2022)	m, and 10 × 10 m		(R, C++, Python)		Wang/Testing-the-robustness-of-LiDAR-vegetation-metrics-to-varying-point-densities		
Processing and analysis scripts for testing robustness of LiDAR vegetation metrics to varying pulse densities across habitats	Netherlands (500 test plots; five habitat types)	Single ALS survey (AHN4) with simulated pulse densities	10 × 10 m	Dunes, marshes, shrublands, grasslands, woodlands	Processing and analysis scripts (R, C++, Python)	Zenodo	https://doi.org/10.5281/zenodo.13940846 (folder 8_Sensitivity_analysis.zip/8_4_scripts)	Shi et al. (2025)	CC BY 4.0
Workflow for tree and shrub extraction from airborne laser scanning in wetlands	Oostvaardersplassen, Netherlands	AHN4 ALS survey (2020–2022)	Native ALS resolution (points); derived raster outputs	Wetlands, marshes, reedbeds	Processing workflow / source code (C++)	GitHub	https://github.com/Jinhu-Wang/Tree_Classification_and_Individualization_In_Marsh_Area	Not applicable	Apache-2.0
Metadata records for LiDAR-derived datasets from MAMBO	Europe-wide	Not applicable	Not applicable	Multiple habitat types	Metadata records	LifeWatch Metadata Catalogue	https://metadatacatalogue.lifewatch.eu/ (search 'MAMBO')	Not applicable	Open metadata
Publication collection of MAMBO LiDAR datasets	Europe-wide	Not applicable	Not applicable	Multiple habitat types	Dataset and publication collection	RIO Journal	https://riojournal.com/topical_collection/223/	Not applicable	Open access

* The table includes both primary research datasets (raw and processed LiDAR-derived data products) and associated research objects that support their FAIR implementation, such as processing workflows, metadata catalogue records, and curated discovery collections. These accompanying resources are included because they are essential for the findability, reproducibility, and reuse of the datasets.