

Metadata form of Silva Fennica

This form is designed for writing the elements of metadata, which are used in the description of research materials such as data and codes. The form is based on the work done in the Work Group “Description of research materials” under the Finnish Open Science Coordination.

Item	Description	Responsible
<i>Name of the data / code</i>	Research data: Segments_and_calculated_features.csv Computing code (4 files): 01_feature_calculation.R, 02_smote_optimisation.R, 03_tuning_and_classification.R, and 04_reanalysis.R README_file.txt	Author
<i>Author & ORCID</i>	Hardenbol Alwin A. (https://orcid.org/0000-0002-0615-505X) Ørka Hans Ole (https://orcid.org/0000-0002-7492-8608) Gobakken Terje (https://orcid.org/0000-0001-5534-049X) Kostensalo Joel (https://orcid.org/0000-0001-9883-1256)	Author
<i>Authors' affiliation(s)</i>	Natural Resources Institute Finland (Luke) (https://ror.org/02hb7bm88) Norwegian University of Life Sciences (NMBU) (https://ror.org/04a1mvv97)	Author
<i>Owner of the material</i>	Ørka Hans Ole (https://orcid.org/0000-0002-7492-8608) Gobakken Terje (https://orcid.org/0000-0001-5534-049X) Norwegian University of Life Sciences (NMBU) (https://ror.org/04a1mvv97)	Author
<i>Publisher</i>	Natural Resources Institute Finland (Luke) (https://ror.org/02hb7bm88) Zenodo (https://ror.org/01ggx4157)	Author
<i>Funder</i>	Ørka Hans Ole (https://orcid.org/0000-0002-7492-8608) Gobakken Terje (https://orcid.org/0000-0001-5534-049X) Kostensalo Joel (https://orcid.org/0000-0001-9883-1256) Natural Resources Institute Finland (Luke) (https://ror.org/02hb7bm88) Norwegian University of Life Sciences (NMBU) (https://ror.org/04a1mvv97)	Author
<i>Description</i>	The material was collected to support the development, evaluation, and reuse of methods for detecting and classifying European aspen (<i>Populus tremula</i> L.) in boreonemoral forests using high-density airborne laser scanning (ALS) and high-resolution multispectral aerial imagery. The primary aim was to assess how well rare tree species can be identified at the individual tree level and to evaluate classifier performance under realistic base rate conditions. The material is suitable for research on individual tree detection, tree species classification, feature engineering from ALS and multispectral data, and evaluation of classification uncertainty for rare species. The material consists of (i) field inventory data from 253 circular plots (400 m² each) including spatially explicit measurements of all living trees with DBH ≥ 4 cm, (ii) wall-to-wall ALS point clouds with very high point density and associated spectral information derived from multispectral imagery, (iii) individual tree segments derived from ALS data, (iv)	Author

	engineered tree- and plot-level features, and (v) classification outputs and performance estimates. The research object is standing living trees in boreonemoral forests, and the primary unit of observation is the individual tree, with additional aggregation to plot level for cross-validation and performance assessment. Each tree is represented by field-measured attributes, remotely sensed features, and (where applicable) a matched ALS-derived tree segment, enabling reuse of the material for method development, benchmarking, and comparative studies. Please note, however, that the raw point clouds and coordinates of the segments have not been shared. As such, the script is functional starting from 02_smote_optimisation.R.	
<i>Methods</i>	Link to article: https://doi.org/10.14214/sf.26012 We combined detailed field inventories with high-density airborne laser scanning (ALS) and high-resolution multispectral aerial imagery to detect and classify European aspen (<i>Populus tremula</i>) at the individual-tree level in boreonemoral forests in southeastern Norway. Field data were collected from 253 fixed circular plots (400 m²), recording tree species, diameter at breast height, spatial position, and tree height, with heights measured using a Hagl�f Vertex hypsometer and modelled where necessary. Tree volumes were calculated using species-specific taper models implemented in the R package <i>TaperNOR</i>. ALS data (~250 points m⁻²) were acquired using a RIEGL LMS-VQ-1560II-S laser scanner, and multispectral imagery (RGB + NIR, 16.5 cm ground sampling distance) was collected with an UltraCam Osprey 4.1 camera. Spectral information was fused with ALS point clouds using photogrammetric collinearity equations. Ground classification and height normalisation were performed using TerraScan and the R package <i>lidR</i>. Individual trees were segmented using the deep-learning framework ForAINet, and ALS-derived trees were algorithmically matched to field-measured trees. A total of 98 ALS- and spectral-derived features were engineered at tree and plot levels. Species classification (aspen vs. non-aspen) was conducted using random forest models implemented in the R package <i>ranger</i>, with class imbalance addressed using the synthetic minority oversampling technique (SMOTE; <i>performanceEstimation</i> package). Model performance was evaluated using tree- and plot-level cross-validation. Bayesian estimation with Beta priors was used to quantify uncertainty in key performance metrics, particularly false discovery rate and false negative rate, supplemented by non-parametric bootstrap methods for volume-based estimates. In addition, we conducted a systematic literature review of tree species classification studies including aspen in Fennoscandia, followed by a re-analysis of published confusion matrices. These were evaluated under a standardised assumption of low aspen prevalence using Bayesian estimators, enabling direct comparison of classifier behaviour and breaking points across studies.	Author
<i>Variables</i>	We provided a full list in Supplementary file S3	Author
<i>Author keywords</i>	Airborne laser scanning; Base rate problem; Bayesian statistics; Broadleaves; European aspen classification; Individual tree detection; Lidar; Model transferability; <i>Populus tremula</i>	Author
<i>Vocabulary keywords (community standard)</i>	-	Author

<i>Discipline</i>	Remote sensing; Ecology, Forestry, Conservation biology	Archive/Repository/Publisher
<i>Type of material</i>	Research data and computational code	Author
<i>Language</i>	Eng	Author
<i>Time range covered</i>	2022-06-16 – 2023-08-12	Author
<i>Geographic region</i>	Våler municipality in Østfold county, Norway, about 50 km southeast of Oslo (approximate centre coordinates of plots: 59°31'N, 10°54'E [EPSG: 4326 – WGS 84];	Author
<i>Version</i>	-	Author
<i>File format(s)</i>	.txt, .csv, and .R	Author
<i>Availability of the materials (open, embargo, registration, limited, registration required)</i>	Open	Author
<i>Justification for access restrictions</i>	-	Author
<i>Licence</i>	CC BY 4.0	Author
<i>Connections with other research materials</i>	The research data is based on non-open raw data	Author
<i>Access to the connected research materials</i>	Available at Zenodo (https://ror.org/01ggx4157)	Author
<i>Codes only: hardware/software requirements for running the code</i>	Windows 10/11, R version 4.5.2 (https://cran.r-project.org/bin/windows/base/) with packages shown in the computational code	Author
<i>Connections to other products of research</i>	-	Author
<i>Personal data</i>	-	Author
<i>Confidential or secret data</i>	-	Author
<i>Publication date</i>	2026-09-30	Archive/Repository/Publisher
<i>Preservation policy</i>	Permanent	Author
<i>Permanent identifier (PID)</i>	DOI: https://doi.org/10.5281/zenodo.22736953	Archive/Repository/Publisher