

APPLICATION PAPER  

Leveraging heterogeneous LiDAR data to model successional stages at tree species level in temperate forests

Lisa Bald¹ , Alice Ziegler¹, Jannis Gottwald⁶, Tiziana L. Koch^{2,3}, Marvin Ludwig⁴, Hanna Meyer⁴, Stephan Wöllauer^{1,5}, Dirk Zeuss¹ and Nicolas Frieß¹

¹Environmental Informatics, Faculty of Geography, Philipps-University Marburg, Marburg, Germany

²Remote Sensing Laboratories, Department of Geography, University of Zürich, Zürich, Switzerland

³Land Change Science, Swiss Federal Institute for Forest Snow and Landscape Research WSL, Birmensdorf, Switzerland

⁴Institute Landscape Ecology, University of Münster, Münster, Germany

⁵Faculty of Resource Management, HAWK University of Applied Sciences and Arts Goettingen, Goettingen, Germany

⁶tRackIT Systems GmbH, Cölbe, Germany

Corresponding author: Lisa Bald; Email: lisa.bald@uni-marburg.de

Received: 22 November 2023; **Revised:** 12 July 2024; **Accepted:** 09 August 2024

Keywords: heterogeneous data; LiDAR; readily available data; successional stages; tree species

Abstract

In the context of the ongoing biodiversity crisis, understanding forest ecosystems, their tree species composition, and especially the successional stages of their development is crucial. They collectively shape the biodiversity within forests and thereby influence the ecosystem services that forests provide, yet this information is not readily available on a large scale. Remote sensing techniques offer promising solutions for obtaining area-wide information on tree species composition and their successional stages. While optical data are often freely available in appropriate quality over large scales, obtaining light detection and ranging (LiDAR) data, which provide valuable information about forest structure, is more challenging. LiDAR data are mostly acquired by public authorities across several years and therefore heterogeneous in quality. This study aims to assess if heterogeneous LiDAR data can support area-wide modeling of forest successional stages at the tree species group level. Different combinations of spectral satellite data (Sentinel-2) and heterogeneous airborne LiDAR data, collected by the federal government of Rhineland-Palatinate, Germany, were utilized to model up to three different successional stages of seven tree species groups. When incorporating heterogeneous LiDAR data into random forest models with spatial variable selection and spatial cross-validation, significant accuracy improvements of up to 0.23 were observed. This study shows the potential of not dismissing initially seemingly unusable heterogeneous LiDAR data for ecological studies. We advocate for a thorough examination to determine its usefulness for model enhancement. A practical application of this approach is demonstrated, in the context of mapping successional stages of tree species groups at a regional level.

Impact Statement

Data sets obtained on a large regional or even national scale, not recorded for a specific study, often present a significant heterogeneity, requiring extensive preprocessing efforts. Despite these challenges, these data sets can reveal valuable ecological information and can be used as readily available data sets. This study shows the advantages of using heterogeneous light detection and ranging (LiDAR) data for ecological modeling and



This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

mapping. The study emphasizes the benefits of exploiting so-called “data-at-hand”, rather than dismissing those in anticipation of more refined data sources.

1. Introduction

Forests provide a variety of indispensable ecosystem services, such as water storage and purification, regulation of air quality, and climate regulation by functioning as a sink and source for greenhouse gases, as well as recreation and provision of raw materials and food (TEEB, 2010). Thus, forests and their biodiversity are indispensable for mitigating the effects of climate change e.g., as carbon sinks (Hisano et al., 2018). In addition, forests with rich vegetation diversity and structural complexity offer various positive effects on biodiversity, including the promotion of animal species richness (Felix et al., 2004; Heidrich et al., 2023; Macarthur and Macarthur, 1961; Stein et al., 2014; Zellweger et al., 2013). To preserve the ecosystem services and functions that forests provide, and to secure their climate mitigation potential, comprehensive information on the state and diversity of their ecosystems is needed to inform decision-making. An important component in this context is the accurate assessment of tree species composition (Berg, 1997; Cavard et al., 2011; Felton et al., 2020; Gamfeldt et al., 2013; Seidl et al., 2011) and additionally, the classification of forest successional stages at the tree species level. Forest successional stages typically describe the development of the forest ecosystem after a disturbance in several phases, which are different in forest structure and can thus serve as indicators for forest biodiversity (Wilson and Peter, 1988). For example, early successional forest ecosystems can provide complex structures of herbs and shrubs, that support high species diversity, and provide valuable habitat for many arthropods as well as numerous rare species (Swanson et al., 2011). Hilmers et al. (2018) found that the early and late successional stages support high biodiversity in temperate forests. Such a high level of biodiversity also enhances forest resilience to climate change, as it is linked to the functioning of the ecosystem (Hisano et al., 2018). However, the effects of climate change, such as fires, storms, and the introduction of new species, can also alter processes of forest succession (Dale et al., 2001). Therefore, knowledge of forest successional stages and their associated ecological processes is crucial for understanding and mitigating for example climate change or anthropogenic disturbances (Corona et al., 2011; Poorter et al., 2023). Such an understanding can furthermore improve monitoring and is fundamental for the development of adequate conservation strategies (Hilmers et al., 2018; Tew et al., 2022).

The monitoring of forests and their successional stages is one of the main goals of extensive manual forest inventories, which usually only provide point-based information (Vidal et al., 2016). Comprehensive area-wide information on their spatial distribution and proportions can be of help for near-natural forest management (Hilmers et al., 2018). Remote sensing can also contribute to enhancing traditional forest inventories (White et al., 2016). Multispectral remote sensing has been found to be a feasible approach to classify tree species in numerous studies (Grabska et al., 2019; Hemmerling et al., 2021; Hościło and Lewandowska, 2019; Immitzer et al., 2016; Welle et al., 2022; Wessel et al., 2018; Xi et al., 2021). Several studies have utilized remote sensing, particularly light detection and ranging (LiDAR) data, for area-wide classification of successional stages and the age of forest stands (Berveglieri et al., 2018; Cao et al., 2015; Duan et al., 2023; Falkowski et al., 2009; Fujiki et al., 2016; Maltman et al., 2023; Zhao et al., 2021). However, those studies analyze the successional stages across large areas without differentiating between tree species. Up-to-date studies classifying tree species-specific successional stages are still rare (Stoffels et al., 2015), but would contribute to recognizing the distinct differences associated with each tree species. Additionally, these studies were performed in rather small areas with temporally aligned LiDAR data, typically collected through dedicated flight campaigns. Unfortunately, LiDAR surveys are still very cost- and labor-intensive and therefore often not directly commissioned by ecological monitoring programs.

Even though costs for LiDAR flight campaigns are high, Germany and large parts of Europe benefit from abundant LiDAR data collected through statewide governmental campaigns. The complete coverage of a federal state in Germany through multiple flights typically spans several years. For instance, regions such as Hesse (HVBG, 2023) and Saxony (GeoSN, 2023) have intervals of 6 years, North Rhine-

Westphalia (Geobasis NRW, 2023) is covered every 5 years, and Rhineland-Palatinate (LVermGeo, 2023) every 4 years. Similar circumstances are found in other European countries, for example in Finland (NLS, 2023) or Spain (MITMA, 2023) where governmental LiDAR data are collected at intervals of approximately 6 years, or in Estonia with updates every 4 years (Maa-amet, 2023). Moreover, data availability is unsystematically documented with no common standard or database. As the data are collected over multiple flights, there are e.g., inconsistencies in flight dates and technical scanning properties. Furthermore, also the already rather low point resolution of LiDAR data can vary as there are ongoing developments in sensor technologies (see Figure 4). As a consequence, these data sets are often viewed as not reliable enough for ecological research purposes at larger scales and in some cases, it is even documented that studies refrained from using LiDAR data sets for modeling due to their presumed poor quality (Stoffels et al., 2015). However, obtaining an exact overview of when the data were not used for this reason is difficult, as the majority of studies do not report instances of unused data.

This study evaluates the potential of typically available heterogeneous LiDAR data in Germany and many parts of Europe for mapping temperate forest successional stages at the tree species level. Instead of only mapping e.g. tree species or age distribution of a forest, this present study explicitly focuses on classifying and mapping forest successional stages for individual tree species. A comparative analysis of models is conducted, employing different combinations of variables, which were derived from optical satellite data (Sentinel-2) and heterogeneous LiDAR data. Random forest models were used with a modeling approach that takes spatial auto-correlation into account by using spatial variable selection and spatial cross-validation techniques (Meyer et al., 2019; Ploton et al., 2020). In a hierarchical modeling approach, first a large-scale map for the seven most common tree species groups (Douglas fir, larch, pine, spruce, beech, oak, and other deciduous trees) was generated for the entire federal state of Rhineland-Palatinate, Germany. Subsequently, for each mapped tree species group up to three successional stages (qualification, dimensioning, and maturing) were modeled in three modeling approaches utilizing different variable sets. In doing so, the aim of this study is to determine whether the utilization of heterogeneous LiDAR data can positively influence model outcomes for forest successional stages at the tree species level.

2. Materials and methods

In the following sections, the modeling of tree species group-specific forest successional stages is presented in detail (see Sections 2.1–2.3.4 and Figure 1). The methodology involves training different models with varying combinations of Sentinel-2 and/or LiDAR data to predict forest successional stages utilizing the forest inventory of Rhineland-Palatinate as reference data. Through the application of spatial variable selection and spatial validation techniques, the potential of the heterogeneous LiDAR data was evaluated. The successional stages of the different tree species groups were mapped with a hierarchical approach. A tree species group model of Rhineland-Palatinate formed the basis for the successional stages models (see Section 2.3.4). All data processing and modeling was done in R version 4.2.3 (R Core Team, 2023).

2.1. Study area

The federal state of Rhineland-Palatinate with an area of 19,858 km² (see Figure 2) is one of the especially forest-rich regions in Germany, with 42% of its area covered by temperate forest (BMEL, 2018). Only 25.6% are state-owned and surveyed by regular forest inventory campaigns. Most of the forests in Rhineland-Palatinate (46.1%) are owned by public corporations (e.g., local administration) or privately owned (26.7%) and therefore, no centralized information on the state of all forests is available (Thünen-Institut, 2012b). The majority of the forests are mixed forests and only 17.7% of the forests are pure stands (Thünen-Institut, 2012a). Overall, deciduous forests predominate, with shares ranging from 54.8% to 64.6% of the forest, depending on the ownership structure (Thünen-Institut, 2012c).

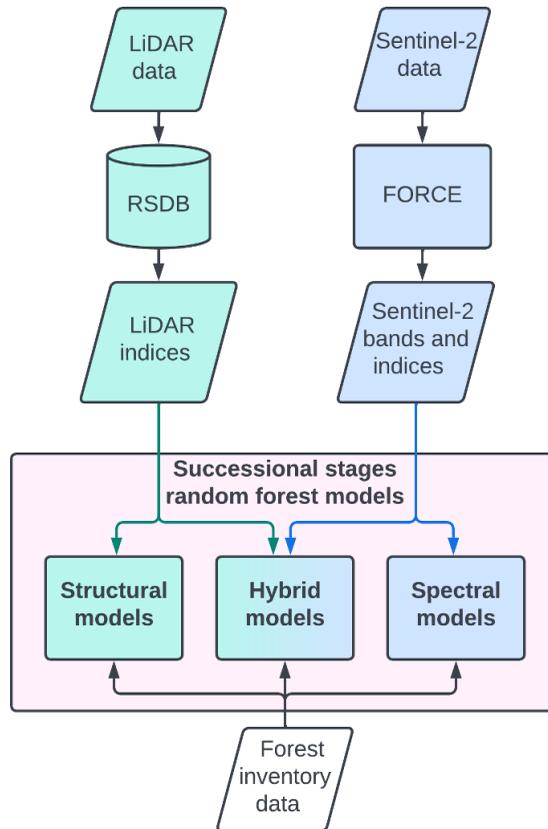


Figure 1. General modeling workflow that is applied to every tree species group to model the specific successional stages. The modeling makes use of two different data sets for prediction. On the one hand LiDAR data, on the other hand, multispectral Sentinel-2 data. From the LiDAR data various indices are derived in the Remote Sensing Database (RSDB; Wöllauer et al., 2020). The Sentinel-2 data were processed within the “Software Framework for Operational Radiometric Correction for Environmental Monitoring” (FORCE; version 3.7.10; Frantz, 2019) and additionally spectral indices were calculated. As reference data and therefore response variable of the models the data from the forest inventory were used. We conducted three different model types, namely structural, hybrid, and spectral models using different combinations of the predictive data sets.

2.2. Databases

2.2.1. Forest inventory data

In this study, the official forest inventory of RhinelandPalatinate was used as reference data, encompassing stand information from state-owned forests. Each forest stand in varying size and shape is recorded in polygons, which led to approximately 170,000 polygons (see Figure 3). From these forest inventory polygons (Landesforsten Rheinland-Pfalz, 2014) information about the forest successional stage, the most common tree species group and the species purity were utilized. The polygons were filtered to a purity of the most common tree species group of at least 80%. Following this filtering process, seven tree species groups with at least 50 polygons for two to three successional stages (see Table A5 in the appendix) remained for model training: Douglas fir, larch, pine, spruce, beech, oak and other deciduous trees. Three successional stages were considered for all tree species groups except for larch and pine: the qualification stage (I), represents the early growth phase, which begins as the young trees outgrow competition vegetation (Landesforsten Rheinland-Pfalz, 2023). Following this, the dimensioning stage (II) develops, characterized by a notable decline in the height and lateral growth of the tree crown. The

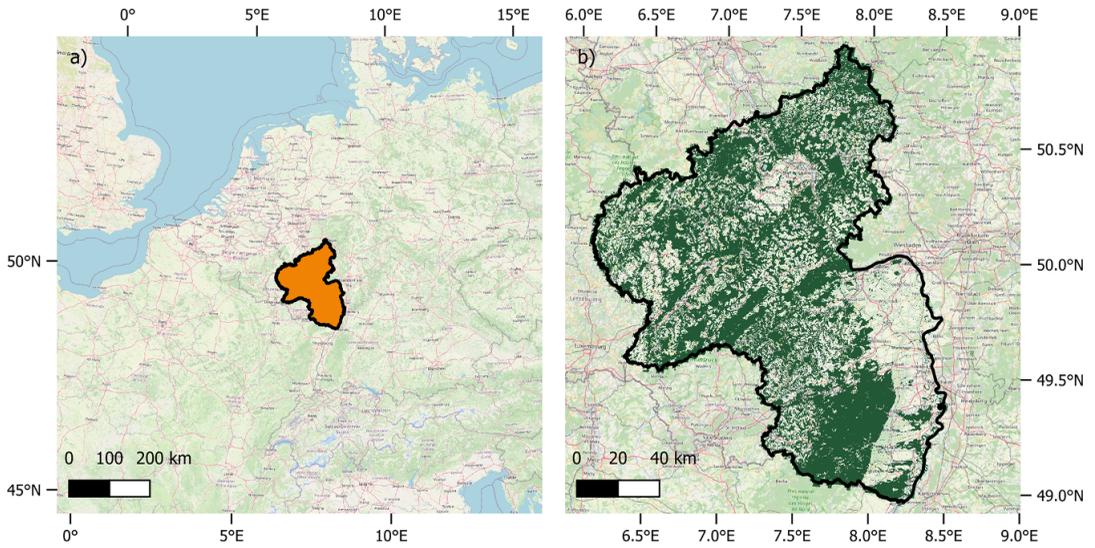


Figure 2. (a) Location of the study area (orange) in Europe. (b) The study area is confined to the forest mask (dark green) derived from the Copernicus high-resolution layer. Data: EEA, 2022; GeoBasis-DE/LVermGeoRP, 2022; OpenStreetMap, 2023.

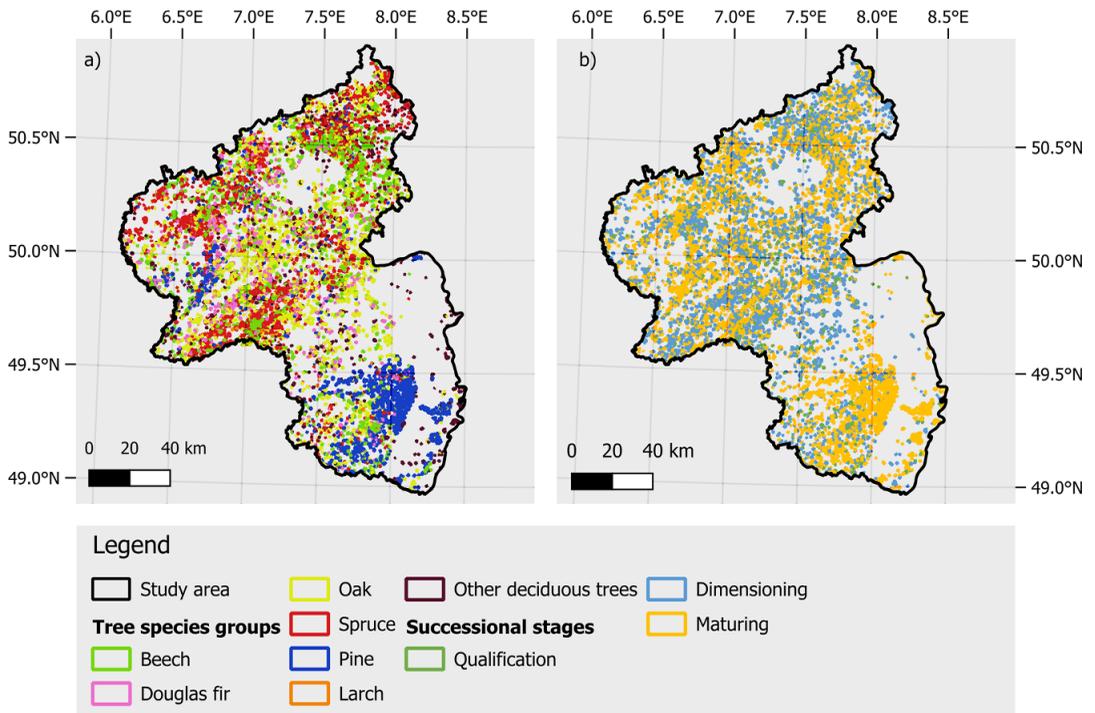


Figure 3. Spatial distribution of the tree species groups (a) and successional stages (b) from the forest inventory. As the forest inventory only surveys state-owned forests the depicted polygons represent only a subset within the whole study area. Data: Landesforsten Rheinland-Pfalz (2014).

oldest successional stage considered in this study is the maturing stage (III), where the tree surpasses 75–80% of its final height, resulting in a deceleration of height growth. Since the number of polygons available for the qualification stage of both pine (16 polygons) and larch (1 polygon) was insufficient to provide representative information, the focus for these two tree species groups was directed solely on the dimensioning and maturing stages.

2.2.2. Sentinel-2 data

Multispectral optical data are proven to be adequate for tree species classifications (Grabska et al., 2019; Hemmerling et al., 2021; Hościło and Lewandowska, 2019; Immitzer et al., 2016; Wessel et al., 2018; Xi et al., 2021), making them an essential component in this study as well. ESA's Sentinel-2 data provided the spectral predictors for the models and were processed using the "Software Framework for Operational Radiometric Correction for Environmental Monitoring" (FORCE; version 3.7.10; Frantz, 2019). With FORCE, the Sentinel-2 data from 2019 to 2021 were downloaded at level 1C and further atmospherically as well as topographically corrected. Within FORCE near-infrared Landsat data were used to correct the spatial position of the Sentinel-2 images and, thus, decreasing the spatial error across satellite images (Rufin et al., 2021). The Sentinel-2 images of the 3 years were used to create high-quality gap-free monthly mean composites for the entire state of Rhineland-Palatinate at a resampled spatial resolution of 10 m. To cover the whole phenological development, one image for winter (January), four covering the fast-changing period from deciduous leaf-unfolding to establishing the canopy (March, April, May, and June) and two images for leaf senescence (September and October) were created. In addition to the original bands, multiple spectral indices reflecting vegetation properties were calculated. Table 1 shows the Sentinel-2 spectral bands and indices that were used in this study. Refer to Bhandari et al. (2024) for a more detailed and comprehensive description of the workflow for processing the Sentinel-2 data.

2.2.3. LiDAR data

The LiDAR data are the key elements of this study, as its aim is to identify the potential of these heterogeneous data in contributing to the classification of successional stages for individual tree species groups. The LiDAR data utilized in this study were collected by the department for Surveying and Geographic Information of Rhineland-Palatinate (GeoBasis-DE/LVermGeoRP, 2022). Recently, the acquisition interval for LiDAR data in Rhineland-Palatinate was increased from a collection over 9 years to only 4 years. As the transition is still ongoing, in this study data from a 7-year interval from 2014 to 2021 covering the whole state were used (LVermGeo, 2023). Since the data result from many different flights, there are variations in data point density across the acquisition dates (see Figure 4). Study areas consisting of different flight campaigns tend to also vary notably in technical properties such as scan angle and flight altitude (Næsset, 2009; Ørka et al., 2010; Solberg et al., 2009). However, often detailed metadata on these parameters are not available for freely available data. For the LiDAR data used in this study information was provided only on the sensors used each year and beam divergence, which can be found in Table A3 in the appendix. In total, 29 indices were computed from the LiDAR data using the Remote Sensing Database (RSDB) of Wöllauer et al. (2020; see Table 2). The indices were computed for all areas identified as deciduous or coniferous forests according to the Copernicus high-resolution layer for forest types of 2018 with a spatial resolution of 10 m (EEA, 2022). The calculated indices represent different categories of forest structure including canopy characteristics (e.g., canopy height), vegetation structure (e.g., the penetration rate of different vegetation layers), overall vegetation properties (e.g., aboveground biomass, vegetation coverage, and leaf area index), and terrain features (e.g., elevation; see Table 2).

2.3. Methods

2.3.1. Matching data

For each tree species group, the successional stages from the forest inventory data were used as response variables, while either Sentinel-2, LiDAR, or Sentinel-2 and LiDAR variables were used as predictors in

Table 1. Sentinel-2 bands and indices used in this study. Images for these bands and indices were calculated from monthly composites for 2019 to 2021 for January, March, April, May, June, September, and October. See Table A2 in the appendix for the complete formulas to calculate the spectral indices

Name	Description	Reference/central wavelength
Visible bands		
B2	Blue	490 nm
B3	Green	560 nm
B4	Red	665 nm
Near infrared bands		
B5	Red edge 1	705 nm
B6	Red edge 2	740 nm
B7	Red edge 3	783 nm
B8	Near infrared	842 nm
B8a	Broad near infrared	865 nm
Shortwave bands		
B11	Shortwaved infrared 1	1610 nm
B12	Shortwaved infrared 2	2190 nm
Vegetation indices		
VI1	Chlorophyll index red edge	Gitelson et al. (2003)
VI2	Enhanced vegetation index	Huete et al. (2002)
VI3	Kernel NDVI	Camps-Valls et al. (2021)
VI4	Modified normalized difference water index	Xu (2006)
VI5	Modified simple ratio red edge	Chen (1996)
VI6	Modified simple ratio red edge narrow	Fernández-Manso et al. (2016)
VI7	Normalized difference moisture index	Gao (1996)
VI8	Normalized difference red edge index 1	A. Gitelson and Merzlyak (1994)
VI9	Normalized difference red edge index 2	Barnes et al. (2000)
VI10	Normalized difference vegetation index	Tucker (1979)
VI11	Normalized difference vegetation index red edge 1	Gitelson and Merzlyak (1994)
VI12	Normalized difference vegetation index red edge 2	Fernández-Manso et al. (2016)
VI13	Normalized difference vegetation index red edge 3	Fernández-Manso et al. (2016)
VI14	Normalized difference vegetation index red-edge 1 narrow	Fernández-Manso et al. (2016)
VI15	Normalized difference vegetation index red-edge 2 narrow	Fernández-Manso et al. (2016)
VI16	Normalized difference vegetation index red-edge 3 narrow	Fernández-Manso et al. (2016)
VI17	Normalized difference water index	McFeeters (1996)
VI18	Soil adjusted and atm. resistant vegetation index	Kaufman and Tanre (1992)
VI19	Soil adjusted vegetation index	Huete (1988)

different models (see Figure 1). To process the polygons from the forest inventory data, all intersecting pixels from the Sentinel-2 and LiDAR variables were extracted. To prevent confusion with adjacent areas, a 10 m negative buffer was applied at the edges of the polygons to exclude the border areas of the polygons.

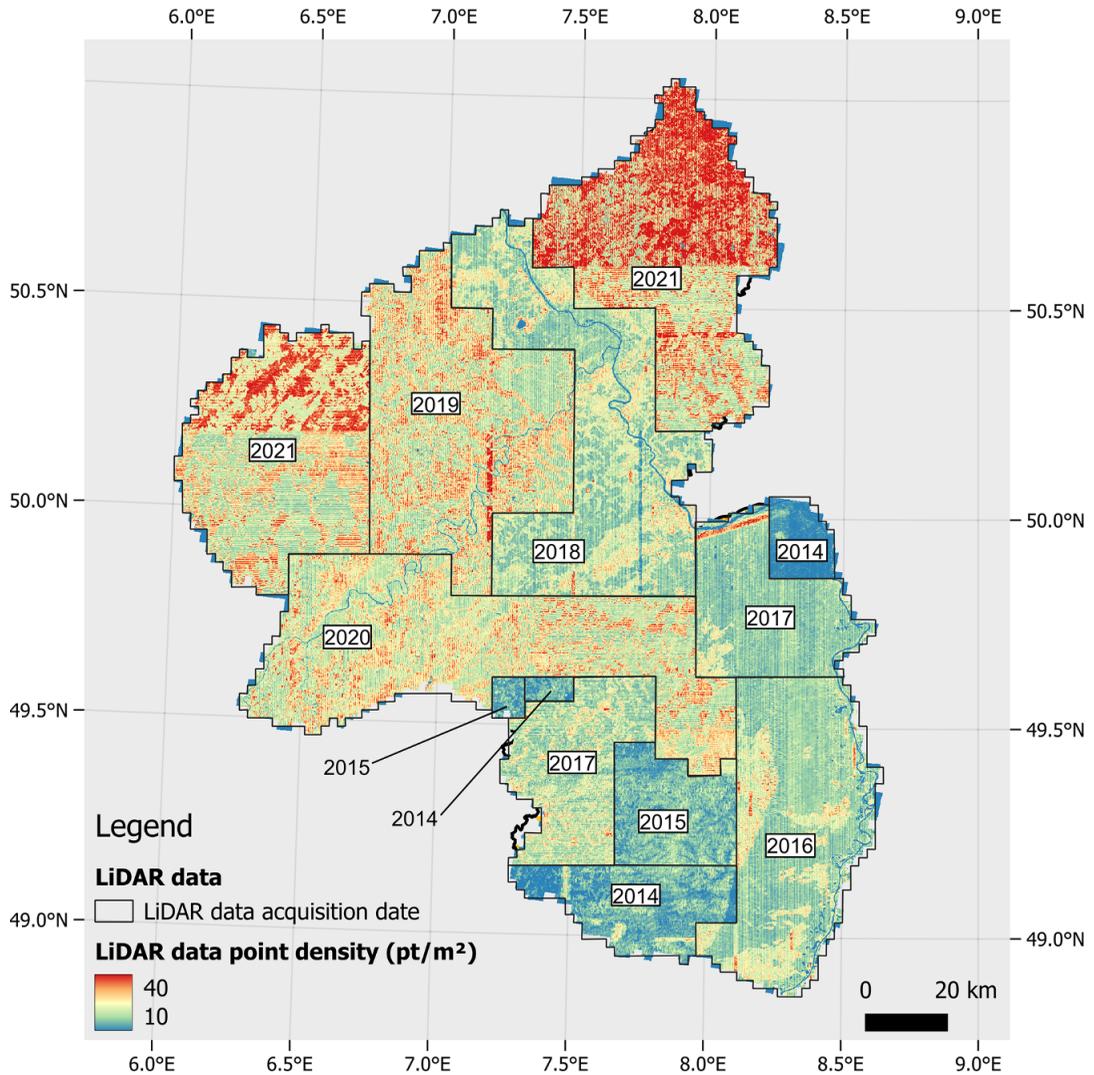


Figure 4. Properties of heterogeneous LiDAR data of Rhineland-Palatinate. The year in which the data were recorded as well as the calculated point density derived directly from the LiDAR data set, based on 100 m pixel are depicted. It is visible that there is a transition from lower point densities in earlier years to higher point densities in later years. Data: GeoBasis-DE/LVermGeoRP (2022).

2.3.2. Balancing data and splitting into testing and training data

Reference data were balanced to ensure that all classes of successional stages of tree species groups were treated equally in the modeling process, finding a trade-off between as much training data as possible and equal distributions across classes. The data were balanced to ensure that for each tree species group (I) the same number of polygons from each successional stage was used and at the same time (II) the same number of pixels from each polygon was randomly sampled. If a small number of pixels from within each polygon is chosen, many of the available polygons can be included (also very small polygons). However, only very few pixels are used from each polygon, even from those that are very large. On the other hand, sampling a large number of pixels from each polygon results in more polygons being excluded from consideration, but it allows for more pixels to be sampled from the remaining polygons. This sampling of the data set was therefore optimized individually for each tree species group producing a balanced data set as large as

Table 2. Overview of LiDAR indices characterizing the vegetation calculated with the Remote Sensing Database (RSDB; Wöllauer et al., 2020; see Appendix A4 for RSDB labels)

Name	Description
Canopy	
CH (canopy height) max	Maximum canopy height
CHM (canopy height model) max	Highest surface above ground - canopy height model (CHM) raster based
CH mean	Mean top-of-canopy height
CHM mean	Mean of the surface above ground - CHM raster based
CH SD	Standard deviation of canopy height
CHM SD	Standard deviation of the surface above ground - CHM raster based
CH median	Median canopy height
CH skew	Skewness of the canopy height distribution
CH kurtosis	Excess kurtosis of the canopy height distribution
CH perc 30	30% percentile of canopy heights
CH perc 70	70% percentile of canopy heights
Vegetation structure	
PR (penetration rate) canopy	Penetration rate of canopy vegetation layer
PR regeneration	Penetration rate of the regeneration vegetation layer
PR understory	Penetration rate of the understory vegetation layer
RD (return density) canopy	Return density of canopy vegetation layer
RD regeneration	Return density of regeneration vegetation layer
RD understory	Return density of understory vegetation layer
VDR	Vertical distribution ratio
Vegetation	
AGB	Aboveground biomass
LAI	Leaf area index
FHD	Foliage height diversity
VC (vegetation coverage) 1 m	Vegetation coverage in 1 m height
VC 2 m	Vegetation coverage in 2 m height
VC 5 m	Vegetation coverage in 5 m height
VC 10 m	Vegetation coverage in 10 m height
Terrain	
Elev (elevation) max	Highest ground a.s.l.
Elev mean	Mean elevation
Elev SD	Standard deviation of ground a.s.l.
Elev slope	Mean slope

possible (for more details see Appendix [Figure A1](#) or the R code in the data availability statement). From this data set, 20% of the polygons from each class were retained for external testing. The remaining data were used for model training and validation (see [Tables A1](#) and [A5](#) in the appendix). From the training data sets for the successional stages, a data set for the tree species group model was created. This data set was later used as a base for the hierarchical mapping of the successional stages. The same balancing process as for each of the successional stages models was done for the tree species group model on this data set.

2.3.3. Model specifications

Random forest models (Breiman, 2001) were trained with a forward feature selection (FFS) from the R package CAST (version 0.7.1; Meyer et al., 2023). The FFS trains the models with each possible two-variable combination, keeps the best performing one and adds more predictor variables until none

decreases the error of the current best model. This allowed the recognition and removal of variables that lead to overfitting (Meyer et al., 2018). As a result, only a small proportion of the variables prepared in this study, specifically those that are relevant to the models, are actually used.

Spatial cross-validation was used during variable selection and model tuning to evaluate which variables and hyperparameters lead to the highest ability to make predictions for new spatial locations within the study area. The polygons were used as spatial units and were randomly split into ten different folds for spatial cross-validation (Meyer et al., 2018; Ploton et al., 2020). The final models were then tested on 20% of the polygons that were held out for spatially independent testing (see section 2.3.2) to evaluate the potential of LiDAR data for classifying successional stages.

2.3.4. Modeling approach

To analyze the utility of LiDAR variables to classify and map forest successional stages, models were trained on different combinations of Sentinel-2 and LiDAR variables using a variable selection algorithm. Three models were trained for each of the tree species groups Douglas fir, larch, pine, spruce, beech, oak, and other deciduous trees, to predict the successional stages. The models solely using Sentinel-2 variables are hereafter referred to as the “spectral models”, the models incorporating Sentinel-2 and LiDAR variables will be denoted as “hybrid models” in the following and the models exclusively trained on LiDAR variables as “structural models”. The comparison focused solely on models for successional stages assuming the tree species group as known.

To show the applicability of the successional stage models, an area-wide map with a resolution of 10 m for all forested areas of Rhineland-Palatinate was generated. To achieve this, the Copernicus high-resolution layer forest-type data from 2018 were used as a forest mask (EEA, 2022). To map the successional stages a tree species groups model was used as a baseline in a hierarchical modeling approach. This entailed a two-step process: first, modeling all tree species groups across the entire area as a baseline, and second, modeling successional stages based on the predicted tree species groups. The modeling approach (either spectral, structural, or hybrid) that performed best across all tree species groups on the test data, was used for mapping. The tree species groups model was based on 360 training polygons from the forest inventory data, each with 180 pixels, and tested on 96 polygons. The tree species groups were modeled using the same modeling approach and the performance was tested using the same testing data set as for all successional stages models. These data sets were never considered during model training, neither for the tree species model nor for the successional stages models and were spatially independent from the training data.

3. Results

This section presents the study’s findings on the performance of three different modeling approaches for modeling the successional stages of tree species groups. The potential of using heterogeneous LiDAR data was assessed by comparing the results of spectral, structural, and hybrid models. Additionally, the variable selection of the different models and the area-wide prediction of tree species groups specific successional stages throughout Rhineland-Palatinate were analyzed.

3.1. Model performance

The structural models (accuracy from 0.4 to 0.68) and hybrid models (accuracy from 0.43 to 0.78) performed notably better than the spectral models (accuracy from 0.33 to 0.63) for all tree species groups. The hybrid models and structural models were quite similar in performance with the hybrid models performance always being slightly superior except for pine, where model performances were the same (see Figure A2 in the appendix). Therefore, and for more clear comparability, the following analyses were limited to the comparison of the spectral models to the hybrid models. The results of the structural model can be found in Figure A2 in the appendix. The two left columns of Figure 5 show the test results of each successional stage and model. The right column shows the difference in the proportion of correctly classified pixels between models. For the spectral models, an overall accuracy between 0.33 for the group

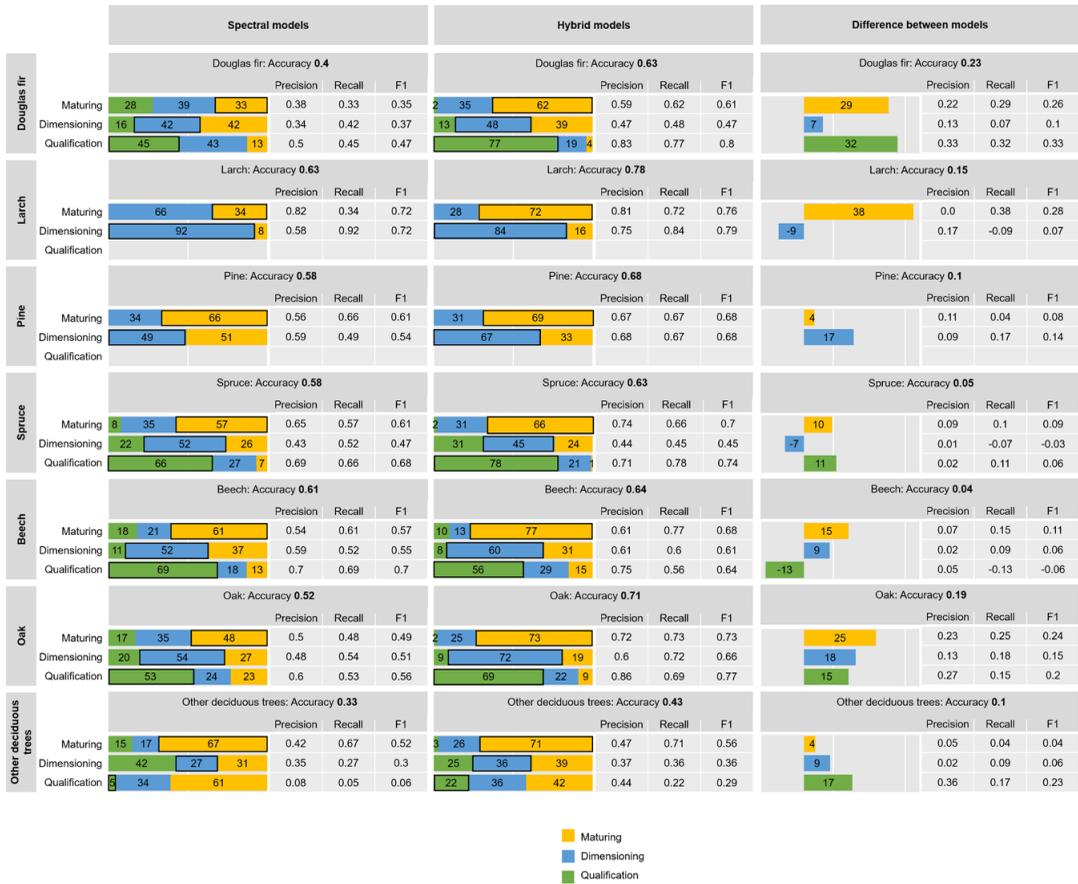


Figure 5. Model performances. The left column of the plots shows the results of models applied to the test data sets only using Sentinel-2 variables (spectral models), and the middle column shows the results using Sentinel-2 and LiDAR variables (hybrid models). Each colored plot shows the confusion matrices of the testing for one tree species. Labels from the reference data are shown on the x-axis and the predicted values on the y-axis in percent. For example, the bar for the maturing phase (yellow), indicating the model classification, should be as large as possible in the first row (maturing) of each plot. All classifications in the same row, but in the other phases (blue and green) are misclassified. The right column shows the differences in accuracy for each class between the spectral models and the hybrid models. All values are rounded to two decimals.

other deciduous trees and 0.63 for larch could be achieved. With the additional LiDAR variables in the hybrid models, the overall accuracies could be increased to between 0.43 for other deciduous trees to 0.78 for larch. However, the models for spruce and beech only gained very little improvements in overall accuracy (0.05 and 0.04) compared to the spectral models and therefore, the additional use of LiDAR variables (hybrid model) could not notably improve those performances.

The largest increase per tree species group in overall accuracy by adding LiDAR variables occurred for Douglas fir with an increase of 0.23, followed by oak and larch with an increase of 0.19 and 0.15, respectively. Overall, for individual successional stages, only the performances of the maturing stage of larch and beech decreased, all other stages benefited from the additional LiDAR variables. To investigate whether one successional stage profited more from the availability of LiDAR variables than another, an analysis of variance was performed. The differences in gain of accuracy between the modeled stages across all tree species groups showed no significant trend (p -value 0.29). A t -test was conducted to determine if the increase in accuracy differs between deciduous and coniferous forests, however, no

significant difference existed (p -value 0.72). Generally, confusion matrices indicated that confusion predominantly occurred among adjacent successional stages. The only exception is the qualification stage of other deciduous trees, where only 5% of pixels were classified correctly. Here, the most misclassifications were not in the adjacent stage (dimensioning stage) but in the maturing stage, which led to low evaluation scores (precision, recall, and $f1$). In the hybrid model, the scores of the most diverse tree species group of other deciduous trees clearly improved but still had the weakest performance with an accuracy of 0.43 and poor recall (qualification: 0.22 and dimensioning: 0.36) and precision values below 0.4 (dimensioning: 0.37). Therefore, it appeared inappropriate to use this group and its classified stages for mapping and consequently, it was excluded from area-wide mapping (see Section 4.3). For all other hybrid models, overall accuracies were at least above 0.6 and the gain in accuracy through the usage of LiDAR variables was 0.13 on average (see Figure 5).

3.2. Contribution of predictor variables

During model training, the feature selection process optimized the selection of variables to create the optimal model. As described in more detail in Section 2.3.4, the variable selection of the FFS starts with a combination of two variables and adds the variable to the model that improves the current model the most until no further improvement occurs (Meyer et al., 2018). As a result, from a multitude of variables, only those deemed important for the models were selected and used in the models, as a side effect of this process correlated variables are only considered once. The assessment of the variable importance of each variable and each model is provided in Figure A4 in the appendix. As the hybrid model additionally used LiDAR-derived variables it was expected that the composition of variables changed for each hybrid model compared to the spectral model. As there might be similarities between variables (especially between the vegetation indices), all variables were categorized into groups by their information content to enable a comparison (see Tables 1 and 2). Figure 6 displays boxplots for both the spectral and the hybrid models containing the ranks of the variables per group as determined by the FFS. A smaller rank indicates an earlier selection and therefore, a stronger improvement and contribution to the model. The boxplot of the spectral model shows the lowest ranks for the variables of the shortwave infrared group (median rank 2), followed by the visible (median rank 4). The near-infrared and the group of the vegetation indices both were selected on average at rank 6. For the spectral models, variables from all groups but the visible bands were selected for the first variable combination (Note: As the FFS chooses a combination of two variables to start the feature selection with, rank 1 exists twice for each model). In every model at least one variable from the group of vegetation indices and for the hybrid models additionally, one variable from the group canopy was selected. For all other variable groups, at least two models did not select any variables from the respective variable group (see Table 3).

For the hybrid models, there was a clear shift in variable selection. During variable selection, the group of canopy variables, containing different properties of the canopy, was selected the earliest. This is represented by the median on the first rank, which differed significantly from the other variable groups (see appendix Table A6) with median ranks ranging from four to six. The vegetation indices and the group of near-infrared bands had on average a lower rank in the hybrid model than in the spectral model. In all spectral models except for Douglas fir, at least one variable from the group of vegetation indices was used as the initial variable combination. Only variables of canopy and vegetation indices were used in the initial variable combination (rank 1) in the hybrid model. Each model used one of those two groups for the initial combination, except for beech, where even two variables from the canopy properties group were used. The number of selected variables did not significantly differ between the spectral and the hybrid models (t -test p -value = 0.48).

3.3. Area-wide mapping

To assess the applicability of the models, successional stages for all forested areas in Rhineland-Palatinate were mapped, allowing for the approximation of a comprehensive spatial cross-validation error and the visual testing of the plausibility of spatial patterns. The tree species groups model reached an accuracy of

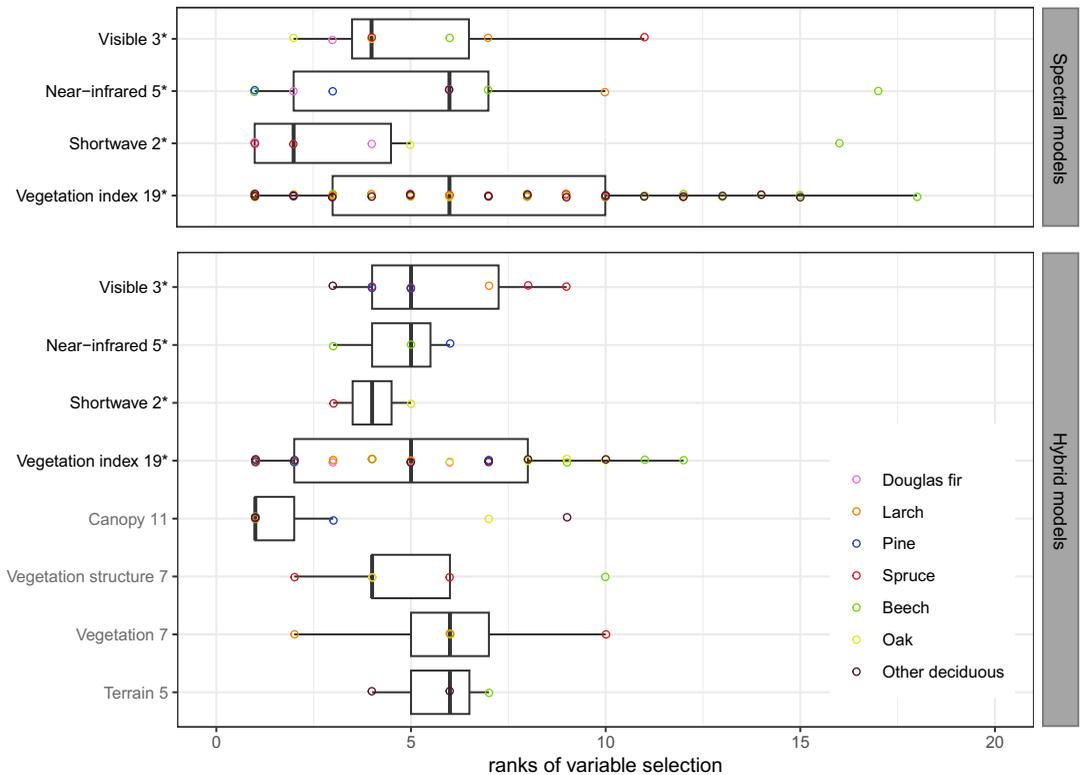


Figure 6. The ranks of variable groups from variable selection. Boxplots display the ranks of variables selected during feature selection for different variable groups. Colored dots show the rank separated by tree species group. The boxes in the plot show the interquartile range of the ranks, with the median marked by a vertical line within each box. The whiskers extend to the minimum and maximum ranks without outliers. The data within the boxes indicate the average rank at which variables from each variable group were selected during the feature selection process in model tuning. As each variable group consists of several variables (see Tables 1 and 2), each model might be represented in each group multiple times. Numbers in y-axis labels indicate how many variables belong to each specific group. Black y-axis labels indicate that variables are Sentinel-2 variables, while gray labels are LiDAR-derived variables. Variable importance for each variable individually for each model is provided in Figure A4 in the appendix. *Each Sentinel-2 variable is available for 7 months.

0.81. Details of the model and its variable importance are provided in the appendix (see Figure A3 and Table A7). The area-wide map of tree species groups specific successional stages for the entire state of Rhineland-Palatinate using the tree species groups model as well as the hybrid models for the successional stages achieved an overall accuracy of 0.6 on the test data sets. For detailed confusion matrices see Tables A7 and A8 in the appendix.

Figure 7 shows the map of tree species groups specific successional stages for the entire federal state of Rhineland-Palatinate. On this map, general spatial patterns of distributions are visible. In the Southeast (Palatinate forest), pine trees dominate, while in the North (Westerwald) and West (Eifel and Hunsrück), spruce trees are predominantly present. In the areas around the rivers (e.g., Moselle and Rhine), mainly various classes of deciduous trees are found. For a more detailed view visit the digital map at <https://envima.github.io/LidarForestModeling/>. Artifacts caused by the heterogeneous LiDAR data were not detected.

Figure 8 shows two areas of the area-wide map in more detail, which are located directly at the survey borders of different LiDAR scenes (see Figure 8b) with up to 6 years of temporal difference. Figure 8a

Table 3. Count of chosen variables during feature selection. Numbers in row names indicate how many variables belong to the specific variable group. Black variables are Sentinel-2 variables while gray variables are LiDAR-derived variables. (a) Spectral models. (b) Hybrid models. *Each Sentinel-2 variable is available for 7 month

(a)								
	Douglas fir	Larch	Pine	Spruce	Beech	Oak	Other deciduous trees	Total times chosen
Visible bands (3*)	1	1	0	2	1	2	0	7
Near infrared bands (5*)	2	1	2	0	3	0	1	9
Shortwave bands (2*)	3	0	0	2	1	1	0	7
Vegetation indices (19*)	1	9	2	9	14	7	15	57
Total number of vars per model	7	11	4	13	19	10	16	
(b)								
	Douglas fir	Larch	Pine	Spruce	Beech	Oak	Other deciduous trees	Total times chosen
Visible bands (3*)	0	1	2	4	0	0	1	8
Near infrared bands (5*)	0	0	1	0	2	0	0	3
Shortwave bands (2*)	0	0	0	1	0	1	0	2
Vegetation indices (19*)	5	4	3	2	6	7	6	33
Canopy (11)	1	1	2	1	2	2	2	11
Vegetation structure (7)	1	0	0	2	1	1	0	5
Vegetation (7)	0	2	0	1	1	0	0	4
Terrain (5)	0	0	0	0	1	0	2	3
Total number of vars per model	7	8	8	11	13	11	11	

shows an area dominated by deciduous species, while [Figure 8c](#) illustrates an area where predominantly coniferous forests are located. In the detailed maps of these border areas, no patterns are identifiable that can be attributed to artifacts of the LiDAR data.

4. Discussion

Although there have been attempts to identify either only tree species (Breidenbach et al., 2021; Hemmerling et al., 2021) or tree species in combination with successional stages (Stoffels et al., 2015) on large-scale recently, the identification of the successional stages of tree species remains a major challenge (Fassnacht et al., 2016). In this context, the successional stages of seven distinct tree species groups were modeled using different combinations of input variables and a variable selection approach. The best results were obtained through the combined use of Sentinel-2 and LiDAR data, even though the LiDAR data were of heterogeneous quality. This approach illustrated the potential of incorporating heterogeneous LiDAR data sources in varying quality as typically available from governmental sources for ecological mapping and monitoring.

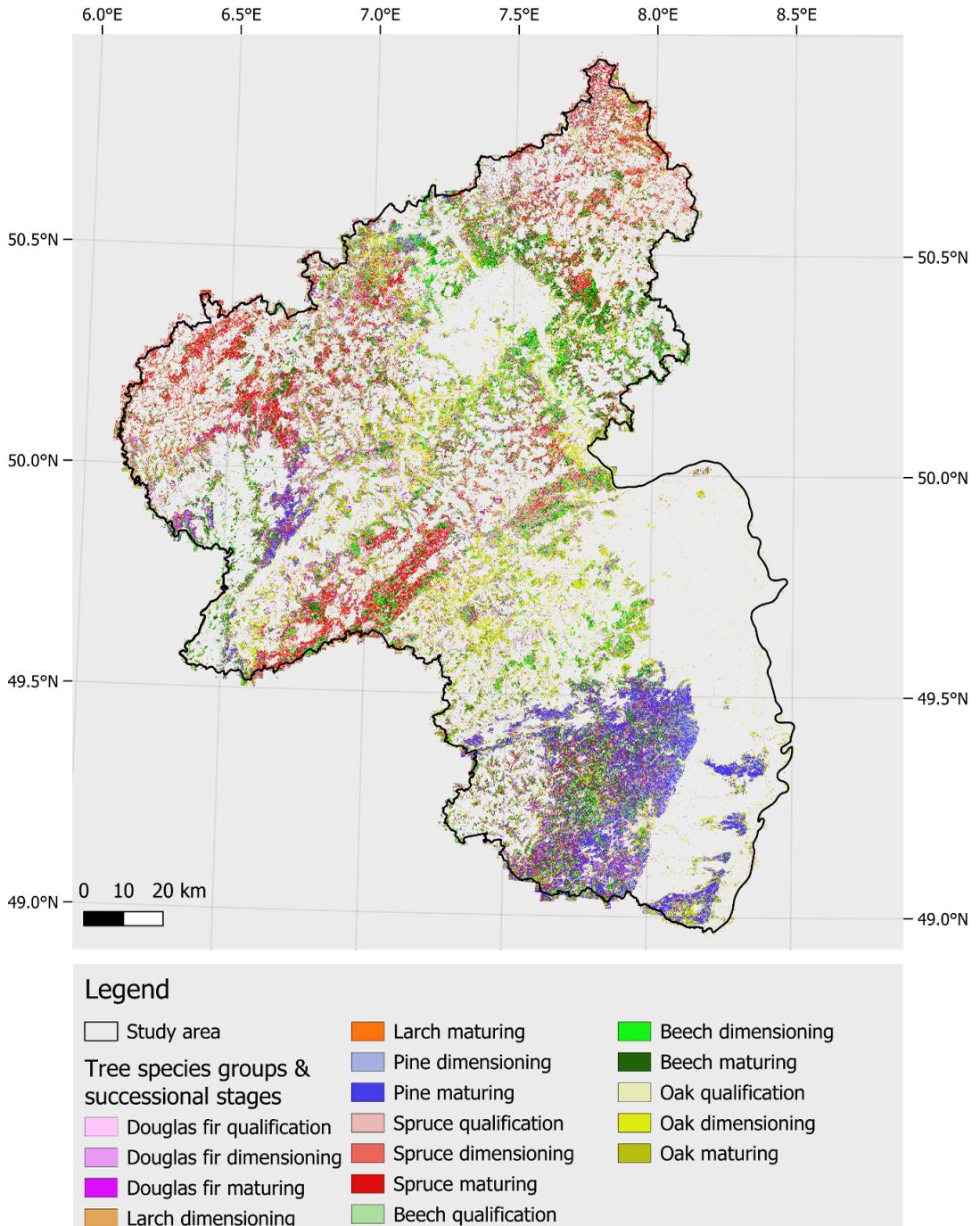


Figure 7. Area-wide map of tree species groups and its successional stages for Rhineland-Palatinate. Leaflet available at: <https://envima.github.io/LidarForestModeling/>.

4.1. Modeling of tree species groups specific successional stages

The results of the study highlight that models of tree species groups specific successional stages benefit from additional structural LiDAR variables regardless of the tree species group. Only in three models, the recall of singular successional stages decreased with the additional use of LiDAR

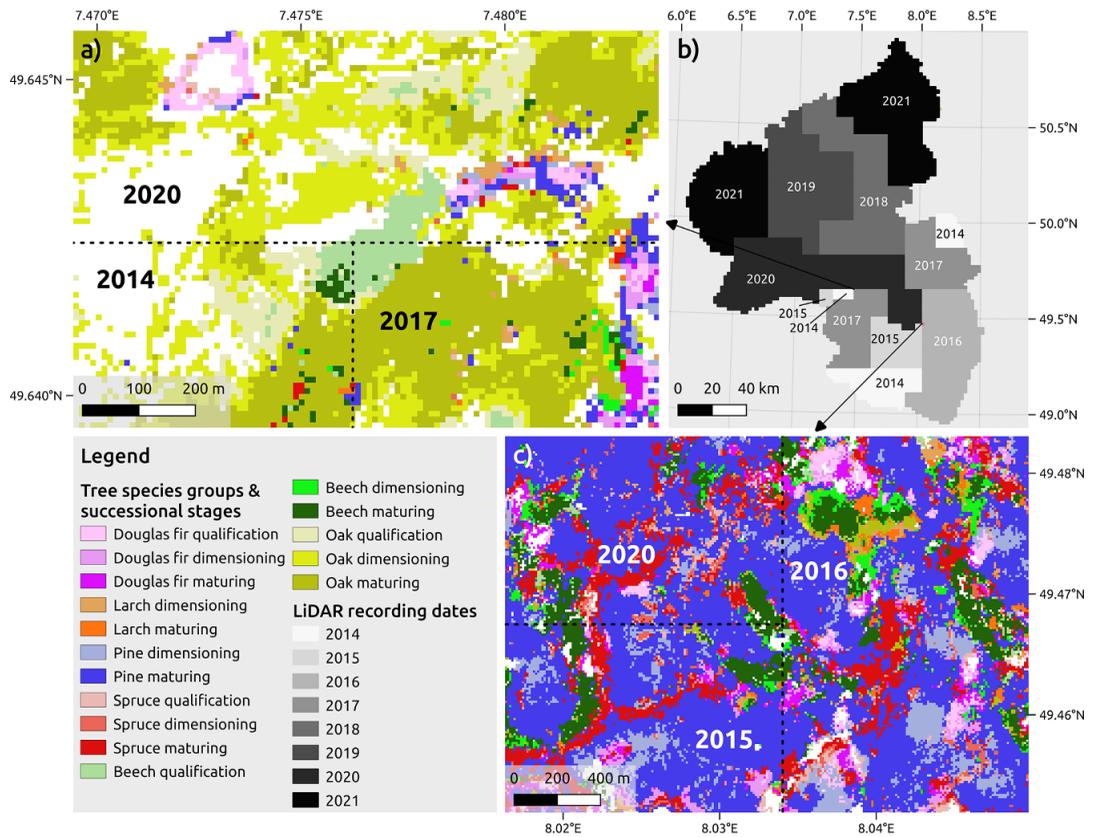


Figure 8. Detailed map of LiDAR survey borders. Variations of the tree species group specific successional stages are indicated with colors in two map sections in plot (a) and (c). These exemplary areas were chosen at boundaries of LiDAR scenes with large temporal differences. On plot (b) the years of acquisition for the LiDAR data set can be identified with lighter colors for older data and darker colors for more recent data.

variables, while at the same time the overall model performance of the particular tree species groups specific successional stages model was increased. This confirms that heterogeneous LiDAR data can supplement models based on multispectral satellite data for modeling tree species groups specific successional stages.

Several of the hybrid models predicted the tree species groups specific successional stages with high precision, but there were still limitations. Especially the successional stages model for the tree species group of other deciduous trees showed a rather poor performance. Even though its accuracy increased by 0.1, from 0.33 to 0.43, with the additional LiDAR variables, the performance still seemed not sufficient to use this model for accurate area-wide mapping. Therefore the group of other deciduous tree species was excluded from mapping. One potential factor causing the poor performance was likely to be the highly heterogeneous composition of this class. While polygons were filtered to have at least 80% purity, a large number of different tree species were grouped together in this class (see appendix Table A9). Individual tree species as cherry, birch, or willow were available in the data set with extremely limited amounts of polygons preventing a meaningful modeling of these groups independently. As these species are occurring less frequently in the study area, only enhanced field surveys targeting these species could enable effective modeling of these species. All other classes yielded overall accuracies above 0.6 for the hybrid models and were therefore appropriate for the use of mapping (see the use-case in section 3.3). Larch achieved the best performance with an overall accuracy of 0.78. However, due to limited data

availability, larch was one of the two tree species groups where only two successional stages were modeled, which increased the possibility of random correct classification.

An area-wide map of tree species specific successional stages can be used for the identification of the habitat suitability for a certain species (e.g. Felix et al., 2004). However, the reliability of such maps is determined by the quality of the remote sensing and especially the forest inventory data as well as by the modeling approach. Forest inventory data of higher spatial resolution, potentially collected at individual tree level, could improve spatial mapping and could overcome limitations of data availability and quality for accurate and transferable models (Yates et al., 2018). Despite these challenges, mapping successional stages has great potential. In managed forests, the successional stages with the highest biodiversity (early and late successional stages) are often the least represented (Hilmers et al., 2018). Large-scale regional mapping can provide planners with a comprehensive overview of the current state of the forest, that extends beyond the information from the forest inventory data. Such maps of successional stages can be utilized to identify particularly valuable habitats for conservation efforts (Hilmers et al., 2018; Reif et al., 2013) and guide forest management on which areas have high potential for future biodiversity enhancement and restoration initiatives.

The mapping of tree species groups specific successional stages in this study not only served as an end in itself but will form a baseline for more indirect biodiversity mapping. Specifically, this study is a component of a broader project that incorporated this information for modeling the habitat of endangered forest-dwelling bat species (Bald et al., 2024). This direct application underscores how such readily available but heterogeneous LiDAR data can contribute to nature conservation efforts. The availability of governmental LiDAR data is high in Europe, and the successful utilization of LiDAR data in biodiversity research was proven (Reddy et al., 2021; Toivonen et al., 2023) for various ecological domains. However, the unsystematic accessibility and inherent heterogeneity in acquisition years and pulse densities of large-scale governmental data sets for entire federal states remain challenging and time-consuming. Greater focus must be placed on the preprocessing of data and the adjustment of modeling techniques, which increases workload substantially. Nonetheless, in this study the rather slow development of successional stages in forest ecosystems was investigated and the additional value of heterogeneous LiDAR data was shown. The data set used in this study, illustrates a typical temporal and spatial imbalance of data often faced. Advocating for thoroughly analyzing and, if applicable, using such heterogeneous and “old” data with appropriate training and validation procedures rather than dismissing it prematurely. In order to meet the growing requirements on conservation monitoring, those readily available but highly heterogeneous data should not be neglected (Vanden Borre et al., 2011). These data offer valuable insights into the three-dimensional forest structure, which passive sensors cannot fully substitute, which makes them especially valuable for forest ecosystem monitoring. Although high-resolution LiDAR data that are acquired close to the time of the conducted study, as used by Falkowski et al. (2009), should be preferable and likely yield better results, such options are too cost-intensive for most practical ecosystem monitoring applications. Governmental LiDAR data are more and more freely accessible (at least for scientific projects), and when combined with publicly available Sentinel-2 data, they can provide a valuable and cost-effective data set. Therefore, researchers and practitioners are encouraged to utilize the available heterogeneous data to advance the understanding of ecosystems.

4.2. Change in variable selection

To assess the potential of heterogeneous LiDAR data for modeling of tree species groups specific successional stages, it is of interest how the variable composition changes, when LiDAR variables are available for selection. Figure 6 clearly shows that with the availability of LiDAR variables, the canopy properties became important predictors. For the group of vegetation indices, interpretation is twofold. Except for the spectral model for douglas fir and the hybrid model for beech, at least one variable from the group of vegetation indices was used for every initial variable combination, indicating vegetation indices can facilitate the prediction of successional stages (see Table 3). However, the rather low average rank shows that vegetation indices were also often selected rather late in the FFS, indicating only little

improvement of modeling performance. Vegetation indices form a strong modeling base, with different other variables in the spectral model depending on the tree species group. In the hybrid model these fluctuations were uniformly replaced by canopy properties adding to a strong combination of canopy properties and vegetation indices for the prediction of successional stages in all tree species groups. Apart from the initial variable combination, the median ranks of all variables but the canopy properties (rank 1) ranged between 4 and 6 in the hybrid models. In the hybrid model, the variables of the canopy properties also seem to replace the early usage of single Sentinel-2 bands in the spectral model. This means that the combination of structural and optical features form a great baseline for the differentiation between successional stages. The importance of structural information is reasonable as during succession the growth of vegetation is a key component. Therefore, it seems plausible that canopy variables are more crucial. In particular, canopy height (see [Figure A4](#) in the appendix) was often selected as one of the first and, therefore, most important variables.

4.3. Area-wide mapping

According to Holzwarth et al. (2020) and to our state of knowledge, large-scale mapping of tree species groups specific successional stages has so far only been carried out once in Germany by Stoffels et al. (2015). Here, the tree species groups specific successional stages for the entirety of Rhineland-Palatinate were mapped. The hybrid models, which demonstrated superior performance in modeling successional stages compared to the other models (see [Section 3](#)) supported by a preceding tree species groups model were utilized for mapping. The accuracy as derived from the test data was 0.6 for all 16 classes (this excludes the successional stages classes of other deciduous trees) and is therefore comparable with the results of Stoffels et al. (2015; Accuracy 0.55).

While the direct comparability of results of other studies is limited due to slightly different classes and the spatially independent validation and testing approach, a rough comparison of the magnitudes of their performances is permissible given the similarity. Both models demonstrate similar qualities, with the extended scope in the approach used here of including the larch tree species group into modeling. Notably, both approaches also show confusion mainly among adjacent successional stages. One advantage of the approach of this study is the utilization of free Sentinel-2 data in combination with readily available LiDAR data. However, even though typically LiDAR data are being collected and available almost everywhere across Europe, its documentation and accessibility vary significantly, often requiring case-based inquiries with the relevant authorities to obtain access. Nevertheless, depending on local regulations, the LiDAR data are often freely available for research and monitoring purposes, enabling monitoring regardless of financial capabilities. However, not only the availability and accessibility of LiDAR data are often unsatisfactorily documented, the metadata are also often lacking and incomplete. The LiDAR data used in this study were provided with very little detailed metadata (see [Table A3](#) in the appendix) and information on pulse density, wavelength or footprint were lacking. Despite our efforts to obtain more details by contacting the data providers and the federal state office, metadata were not available to, or known by the authorities. Therefore, we advocate for more standardized documentation practices and improved metadata transparency when LiDAR data are collected and distributed. This would further enhance their broader usability and facilitate greater comparability across studies. The hierarchical modeling approach additionally features flexibility to improve the quality depending on the research question. Especially when interested in specific tree species or tree species groups this approach delivers the possibility to develop or use specialized tree species groups models and add successional stages models. This study also demonstrates that model accuracy of certain tree species groups (beech and spruce) do not significantly benefit from the additional LiDAR data. In cases where studies specifically focus on one of these tree species groups for the modeling of successional stages, potentially, no further benefit can be derived by adding LiDAR data for these species.

Not only the careful quantitative testing but also the visual analysis of the map yielded convincing results. [Figure 7](#) shows general spatial patterns of tree species that align with the actual forest patterns in Rhineland-Palatinate (see [Figure 3](#); PEFC—Arbeitsgruppe Rheinland-Pfalz, 2010). If the heterogeneity

of the LiDAR data had posed a problem for the models, this would be expected to be revealed by the observation of rectangular areas mirroring the LiDAR flight scenes across the map. No artifacts are visible on the map, and even at boundaries between LiDAR aerial surveys that are furthest apart in time as shown in Figure 8 do not exhibit any distorting patterns. At the intersection of the 2014, 2017 and 2020 flight campaign boundaries in Figure 8, there is an area classified as beech in the qualification stage spreading across the borders of all three LiDAR scenes, without showing any artificial linear structures that could originate from these abrupt transitions. Nevertheless, we acknowledge that differences in acquisition dates have an impact on the data, due to changes in forest structure over time, such as the transition of trees between successional stages. However, the absence of visible artifacts and distorting patterns at these boundaries suggests that any temporal changes did not greatly affect the model outputs in our study.

This study focused on forests, where changes occur rather slowly compared to other vegetation types. It was demonstrated that even heterogeneous LiDAR data can be valuable for mapping tree species groups specific forest successional stages. However, there might be limits with faster growing vegetation that should be explored further.

5. Conclusion

In the ongoing biodiversity crisis, the monitoring of forests is of high importance. Traditional field-based inventories are not able to provide comprehensive, area-wide coverage of information over large areas due to their cost and labor-intensive nature. Remote sensing is a promising solution to develop efficient area-wide monitoring strategies. However, capturing structural data, such as LiDAR information for modeling of successional stages, necessitates expensive flight campaigns to acquire current high-resolution data. Such resources are often unavailable for regional monitoring purposes or minor nature conservation projects. In Germany, federal states commonly conduct smaller LiDAR flight campaigns, covering the same region approximately every 5 to 10 years. Consequently, this results in heterogeneous data sets that are often viewed with skepticism regarding their utility, leading to their exclusion from modeling efforts. The present study reveals that these highly heterogeneous LiDAR data improved the modeling of tree species groups specific successional stages considerably. Therefore, it can be concluded that the potential of LiDAR data should not be underestimated and at least a thorough analysis of their potential benefit for ecological studies should be conducted. The effort of adapting preprocessing and modeling can lead to improved results that can be valuable to nature conservation approaches. It is expected that more recent and higher-quality LiDAR data would improve model results further, however, such instances are rarely encountered in reality for large-scale studies. Improving the available LiDAR data was not within the scope of this study but it might be possible that utilizing current data sources like GEDI (Global Ecosystem Dynamics Investigation), could potentially optimize the use of heterogeneous LiDAR data in the future. During this study, it was found that even heterogeneous LiDAR data were evidently helpful for modeling tree species groups specific successional stages and should not be neglected. While public authorities collect LiDAR data almost everywhere in Germany and also other European countries, the direct availability and documentation are highly heterogeneous, incomplete, and disorganized. Therefore, we advocate for relevant authorities to make the data more accessible or at least visible in a structured manner and provide comprehensive metadata whenever data are made publicly accessible. As this study showed, this could provide a valuable contribution to ecosystem research and, subsequently, to the preservation of forest ecosystem services.

Open peer review. To view the open peer review materials for this article, please visit <http://doi.org/10.1017/eds.2024.31>.

Acknowledgments. We would like to acknowledge the assistance of the AI language model ChatGPT, based on GPT-3.5, developed by OpenAI, which assisted in enhancing English writing and grammar checking. ChatGPT can be assessed via <https://chat.openai.com/>.

Author contribution. Conceptualization: L.B.(Equal); A.Z.(Equal); J.G.(Equal). Supervision: N.F.(Lead). Funding acquisition: J.G.(Lead). Project administration: J.G.(Lead). Resources: J.G.(Lead). Data curation: L.B.(Lead). Formal analysis: L.B.(Equal); A.Z.(Equal). Investigation: L.B. (Equal); A.Z.(Equal). Methodology: L.B. (Equal); A.Z.(Equal); J.G.(Equal); M.L.(Equal); H.M.

(Equal). Software: L.B.(Lead); M.L.(Supporting); H.M.(Supporting); S.W.(Supporting). Validation: L.B.(Lead); A.Z.(Supporting). Visualization: L.B.(Lead); A.Z.(Supporting). Writing original draft: L.B.(Equal); A.Z.(Equal). Writing—review & editing: L.B.(Equal); A.Z.(Equal); J.G.(Equal); T.K.(Equal); M.L.(Equal); H.M.(Equal); D.Z.(Equal); N.F.(Equal). All authors approved the final submitted draft.

Competing interest. The authors have no conflict of interest to declare.

Data availability statement. R scripts used for this study are available under a GPL 3.0 license as Git repository at github.com. A release of the Git repository to reproduce the results of the study is available at <https://github.com/envima/LidarForestModeling> accessed on June 25, 2024. To reproduce the study, information on the data and sample datasets is available at the Open Science Framework (OSF): <https://doi.org/10.17605/OSF.IO/CEK5J>. The Sentinel-2 and LiDAR datasets used in this study are freely available; however, due to their large size (totaling over 2 TB), only the workflow to reproduce these datasets and small sample datasets are provided on OSF. The forest inventory data were made available to us by “Landesforsten Rheinland-Pfalz” specifically for this study and are not publicly available. Consequently, a dummy dataset, structured identically to the original dataset, is provided in the repository. Additionally, this study utilized the border of Rhineland-Palatinate to crop the data to the study area and a forest mask processed from Copernicus high-resolution forest type data. Both datasets are freely available for download and are provided on OSF.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Funding statement. These tree species groups and successional stages models are part of the project “Development of forest structure-based habitat models for forest bats” funded by the Rhineland-Palatinate State Office for the Environment (Landesamt für Umwelt Rheinland-Pfalz). The LiDAR data and forestry data were also provided as part of this project. The development of improved methods for remote sensing based classification of forests was conducted within the Natur 4.0 project funded by the Hessian state offensive for the development of scientific-economic excellence (LOEWE).

References

- Bald L, Gottwald J, Hillen J, Adorf F and Zeuss D** (2024) The devil is in the detail: environmental variables frequently used for habitat suitability modeling lack information for forest-dwelling bats in Germany. *Ecology and Evolution*, *14*, e11571. <https://doi.org/10.1002/ece3.11571>
- Barnes E, Clarke T, Richards S, Colaizzi P, Haberland J, Kostrzewski M, Waller P, Choi C, Riley E and Thompson T** (2000) Coincident detection of crop water stress, nitrogen status, and canopy density using ground based multispectral data. *Proceedings of the fifth international conference on precision Agriculture*.
- Berg Å** (1997) Diversity and abundance of birds in relation to forest fragmentation, habitat quality and heterogeneity. *Bird Study*, *44*, 355–366. <https://doi.org/10.1080/00063659709461071>
- Berveglieri A, Imai NN, Tommaselli AM, Casagrande B and Honkavaara E** (2018). Successional stages and their evolution in tropical forests using multi-temporal photogrammetric surface models and superpixels. *ISPRS Journal of Photogrammetry and Remote Sensing*, *146*, 548–558. <https://doi.org/10.1016/j.isprsjprs.2018.11.002>
- Bhandari N, Bald L, Wraase L and Zeuss D** (2024) Multispectral analysis-ready satellite data for three east African mountain ecosystems. *Scientific Data*, *11*, 473. <https://doi.org/10.1038/s41597-024-03283-3>
- BMEL—Bundesministerium für Ernährung und Landwirtschaft** (2018) Der Wald in Deutschland. ausgewählte Ergebnisse der dritten Bundeswaldinventur. Retrieved January 26, 2021, from https://www.bmel.de/SharedDocs/Downloads/DE/Broschueren/bundeswaldinventur3.pdf?sessionid=1A403B5E29DCE9F51A42A4DC2D4366AD.intranet922?__blob=publicationFile&v=3
- Breidenbach J, Waser LT, Debella-Gilo M, Schumacher J, Rahlf J, Hauglin M, Puliti S and Astrup R** (2021) National mapping and estimation of forest area by dominant tree species using Sentinel-2 data. *Canadian Journal of Forest Research*, *51*, 365–379. <https://doi.org/10.1139/cjfr-2020-0170>
- Breiman L** (2001) Random forests. *Machine Learning*, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Camps-Valls G, Campos-Taberner M, Moreno-Martínez Á, Walther S, Duveiller G, Cescatti A, Mahecha MD, Muñoz-Mari J, García-Haro FJ, Guanter L, Jung M, Gamon JA, Reichstein M and Running SW** (2021) A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, *7*, eabc7447. <https://doi.org/10.1126/sciadv.abc7447>
- Cao S, Yu Q, Sanchez-Azofeifa A, Feng J, Rivard B and Gu Z** (2015) Mapping tropical dry forest succession using multiple criteria spectral mixture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *109*, 17–29. <https://doi.org/10.1016/j.isprsjprs.2015.08.009>
- Cavard X, Macdonald SE, Bergeron Y and Chen HY** (2011) Importance of mixedwoods for biodiversity conservation: evidence for understory plants, songbirds, soil fauna, and ectomycorrhizae in northern forests. *Environmental Reviews*, *19*, 142–161. <https://doi.org/10.1139/a11-004>
- Chen JM** (1996) Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, *22*, 229–242. <https://doi.org/10.1080/07038992.1996.10855178>

- Corona P, Chirici G, McRoberts RE, Winter S and Barbati A** (2011). Contribution of large-scale forest inventories to biodiversity assessment and monitoring. *Forest Ecology and Management*, 262, 2061–2069. <https://doi.org/10.1016/j.foreco.2011.08.044>
- Dale VH, Joyce LA, McNulty S, Neilson RP, Ayres MP, Flannigan MD, Hanson PJ, Irland LC, Lugo AE, Peterson CJ, Simberloff D, Swanson FJ, Stocks BJ and Wotton BM** (2001) Climate change and forest disturbances: climate change can affect forests by altering the frequency, intensity, duration, and timing of fire, drought, introduced species, insect and pathogen outbreaks, hurricanes, windstorms, ice storms, or landslides. *BioScience*, 51, 723–734. [https://doi.org/10.1641/0006-3568\(2001\)051\[0723:CCAFD\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0723:CCAFD]2.0.CO;2)
- Duan M, Bax C, Laakso K, Mashhadi N, Mattie N and Sanchez-Azofeifa A** (2023) Characterizing transitions between successional stages in a tropical dry forest using LiDAR techniques. *Remote Sensing*, 15, 479. <https://doi.org/10.3390/rs15020479>
- EEA—European Union, Copernicus Land Monitoring Service, European Environment Agency** (2022) *Copernicus land monitoring service. high resolution land cover characteristics. forest type 2018*. <https://land.copernicus.eu/pan-european/high-resolution-layers/forests/forest-type-1/status-maps/forest-type-2018>
- Falkowski MJ, Evans JS, Martinuzzi S, Gessler PE and Hudak AT** (2009) Characterizing forest succession with lidar data: an evaluation for the inland northwest, USA. *Remote Sensing of Environment*, 113, 946–956. <https://doi.org/10.1016/j.rse.2009.01.003>
- Fassnacht FE, Latifi H, Stereńczak K, Modzelewska A, Lefsky M, Waser LT, Straub C and Ghosh A** (2016) Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>
- Felix AB, Campa H, Millenbah KF, Winterstein SR and Moritz WE** (2004) Development of landscape-scale habitat potential models for forest wildlife planning and management. *Wildlife Society Bulletin*, 32, 795–806. [https://doi.org/10.2193/0091-7648\(2004\)032\[0795:DOLHMF\]2.0.CO;2](https://doi.org/10.2193/0091-7648(2004)032[0795:DOLHMF]2.0.CO;2)
- Felton A, Petersson L, Nilsson O, Witzell J, Cleary M, Felton AM, Björkman C, Sang ÅO, Jonsell M, Holmström E, Nilsson U, Rönnberg J, Kalén C and Lindbladh M** (2020) The tree species matters: biodiversity and ecosystem service implications of replacing scots pine production stands with norway spruce. *Ambio*, 49, 1035–1049. <https://doi.org/10.1007/s13280-019-01259-x>
- Fernández-Manso A, Fernández-Manso O and Quintano C** (2016) Sentinel-2a red-edge spectral indices suitability for discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation*, 50, 170–175. <https://doi.org/10.1016/j.jag.2016.03.005>
- Frantz D** (2019) FORCE—landsat + Sentinel-2 analysis ready data and beyond. *Remote Sensing*, 11, 1124. <https://doi.org/10.3390/rs11091124>
- Fujiki S, Okada K-i, Nishio S and Kitayama K** (2016). Estimation of the stand ages of tropical secondary forests after shifting cultivation based on the combination of WorldView-2 and time-series landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119, 280–293. <https://doi.org/10.1016/j.isprsjprs.2016.06.008>
- Gamfeldt L, Snäll T, Bagchi R, Jonsson M, Gustafsson L, Kjellander P, Ruiz-Jaen MC, Fröberg M, Stendahl J, Philipson CD, Mikusiński G, Andersson E, Westerlund B, Andrén H, Moberg F, Moen J and Bengtsson J** (2013) Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications*, 4, 1340. <https://doi.org/10.1038/ncomms2328>
- Gao B** (1996) NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Geobasis NRW—Bezirksregierung Köln** (2023) *3D-messdaten. die bezirksregierung köln, geobasis NRW, erfasst flächendeckend 3d-messdaten des geländes und der oberfläche aus flugzeuggestütztem laserscanning*. Retrieved July 31, 2023, from <https://www.bezreg-koeln.nrw.de/geobasis-nrw/produkte-und-dienste/hoehenmodelle/3d-messdaten>
- GeoBasis-DE/LVermGeoRP** (2022) Landesamt für vermessung und geobasisinformation rheinland-pfalz. koblenz, Germany.
- GeoSN—Landesamt für Geobasisinformation Sachsen** (2023) *Primärdaten/laserscandaten der laserscanner-messaufnahme einschließlich klassifizierung—freistaat sachsen*. Retrieved July 31, 2023, from <https://geomis.sachsen.de/geomisclient/?lang=de#/datasets/iso/99815cb4-f6de-4bef-ac7f-e8e005e3fd1d>
- Gitelson A and Merzlyak MN** (1994) Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, 143, 286–292. [https://doi.org/10.1016/S0176-1617\(11\)81633-0](https://doi.org/10.1016/S0176-1617(11)81633-0)
- Gitelson AA, Gritz Y and Merzlyak MN** (2003) Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160, 271–282. <https://doi.org/10.1078/0176-1617-00887>
- Grabska E, Hostert P, Pflugmacher D and Ostapowicz K** (2019) Forest stand species mapping using the Sentinel-2 time series. *Remote Sensing*, 11, 1197. <https://doi.org/10.3390/rs11101197>
- Heidrich L, Brandt R, Ammer C, Bae S, Bäessler C, Doerfler I, Fischer M, Gossner MM, Heurich M, Heibl C, Jung K, Krzystek P, Levick S, Magdon P, Schall P, Schulze E-D, Seibold S, Simons NK, Thorn S ... Müller J** (2023) Effects of heterogeneity on the ecological diversity and redundancy of forest fauna. *Basic and Applied Ecology*, 73, 72–79. <https://doi.org/10.1016/j.baae.2023.10.005>
- Hemmerling J, Pflugmacher D and Hostert P** (2021) Mapping temperate forest tree species using dense Sentinel-2 time series. *Remote Sensing of Environment*, 267, 112743. <https://doi.org/10.1016/j.rse.2021.112743>

- Hilmers T, Friess N, Bässler C, Heurich M, Brandl R, Pretzsch H, Seidl R and Müller J (2018) Biodiversity along temperate forest succession. *Journal of Applied Ecology*, 55, 2756–2766. <https://doi.org/10.1111/1365-2664.13238>
- Hisano M, Searle EB and Chen HYH (2018) Biodiversity as a solution to mitigate climate change impacts on the functioning of forest ecosystems. *Biological Reviews of the Cambridge Philosophical Society*, 93, 439–456. <https://doi.org/10.1111/brv.12351>
- Holzwarth S, Thonfeld F, Abdullahi S, Asam S, Da Ponte Canova E, Gessner U, Huth J, Kraus T, Leutner B and Kuenzer C (2020) Earth observation based monitoring of forests in Germany: a review. *Remote Sensing*, 12, 3570. <https://doi.org/10.3390/rs12213570>
- Hoščilo A and Lewandowska A (2019) Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data. *Remote Sensing*, 11, 929. <https://doi.org/10.3390/rs11080929>
- Huete A, Didan K, Miura T, Rodriguez E, Gao X and Ferreira L (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Huete A (1988) A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- HVBG—Hessische Verwaltung für Bodenmanagement und Geoinformation (2023) *Airborne laserscanning. luftgestützte messverfahren*. Retrieved July 31, 2023, from <https://hvb.g.hessen.de/landesvermessung/geotopographie/3d-daten/airbornelaser-scanning>
- Immitzer M, Vuolo F and Atzberger C (2016) First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, 8, 166. <https://doi.org/10.3390/rs8030166>
- Kaufman Y and Tanre D (1992) Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 261–270. <https://doi.org/10.1109/36.134076>
- Landesforsten Rheinland-Pfalz (2014) *Forsteinrichtung: Geo—und sachdaten*.
- Landesforsten Rheinland-Pfalz (2023) *Qualifizieren—dimensionieren: Eine naturnahe wertholzerzeugung*. Retrieved July 28, 2023, from <https://www.wald-rlp.de/de/nutzen/naturnahe-waldbewirtschaftung/qualifizieren-dimensionieren/>
- LVerGeo—Landesamt für Vermessung und Geobasisinformation Rheinland-Pfalz (2023) *Laserpunkte—produktbeschreibung*. Retrieved August 1, 2023, from https://lvermgeo.rlp.de/fileadmin/lvermgeo/pdf/produktblaetter/ProduktbeschreibungRP_Laserpunkte.pdf
- Maa-amet—Republic of Estonia Land Board Geoportaal (2023) *Elevation data*. Retrieved July 31, 2023, from <https://geoportaal.maaamet.ee/eng/Spatial-Data/Elevation-data-p308.html>
- Macarthur R and Macarthur JW (1961) On bird species diversity. *Ecology*, 42, 594–598. <https://doi.org/10.2307/1932254>
- Maltman JC, Hermosilla T, Wulder MA, Coops NC and White JC (2023) Estimating and mapping forest age across Canada's forested ecosystems. *Remote Sensing of Environment*, 290, 113529. <https://doi.org/10.1016/j.rse.2023.113529>
- McFeeters SK (1996) The use of the normalized difference water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17, 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Meyer H, Milà C, Ludwig M and Linnenbrink J (2023). *CAST: 'caret' applications for spatial-temporal models. r package version 0.8.1*. <https://CRAN.R-project.org/package=CAST>
- Meyer H, Reudenbach C, Hengl T, Katurji M and Nauss T (2018) Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, 101, 1–9. <https://doi.org/10.1016/j.envsoft.2017.12.001>
- Meyer H, Reudenbach C, Wöllauer S and Nauss T (2019) Importance of spatial predictor variable selection in machine learning applications—moving from data reproduction to spatial prediction. *Ecological Modelling*, 411, 108815. <https://doi.org/10.1016/j.ecolmodel.2019.108815>
- MITMA—Ministry of Transport, Mobility and Urban Agenda National Geographic Institute (2023) *National plan for aerial orthophotography. national plan for territory observation*. Retrieved July 31, 2023, from <https://pnoa.ign.es/web/portal/pnoa-lidar/estado-del-proyecto>
- Næsset E (2009) Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, 113, 148–159. <https://doi.org/10.1016/j.rse.2008.09.001>
- NLS—National Land Survey of Finland (2023) *Laser scanning and aerial photography*. Retrieved July 31, 2023, from <https://www.maanmittauslaitos.fi/laserkeilaus-ja-ilmakuvaus>
- OpenStreetMap (2023) Base map from OpenStreetMap. <https://www.openstreetmap.org/copyright>
- Ørka HO, Næsset E and Bollandås OM (2010) Effects of different sensors and leaf-on and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. *Remote Sensing of Environment*, 114, 1445–1461. <https://doi.org/10.1016/j.rse.2010.01.024>
- PEFC—Arbeitsgruppe Rheinland-Pfalz (2010) 3. Regionaler Waldbericht Rheinland-Pfalz, 49.
- Ploton P, Mortier F, Réjou-Méchain M, Barbier N, Picard N, Rossi V, Dormann C, Cornu G, Viennois G, Bayol N, Lyapustin A, Gourlet-Fleury S and Péliissier R (2020) Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, 11, 4540. <https://doi.org/10.1038/s41467-020-18321-y>
- Poorter L, Amissah L, Bongers F, Hordijk I, Kok J, Laurance SGW, Lohbeck M, Martínez-Ramos M, Matsuo T, Meave JA, Muñoz R, Peña-Claros M and van der Sande MT (2023) Successional theories. *Biological Reviews*, n/a. <https://doi.org/10.1111/brv.12995>

- R Core Team** (2023) *R: A language and environment for statistical computing. R foundation for statistical computing*, Vienna, Austria. <https://www.R-project.org/>
- Reddy CS, Kurian A, Srivastava G, Singhal J, Varghese AO, Padalia H, Ayyappan N, Rajashekar G, Jha CS and Rao PVN** (2021). Remote sensing enabled essential biodiversity variables for biodiversity assessment and monitoring: Technological advancement and potentials. *Biodiversity and Conservation*, 30, 1–14. <https://doi.org/10.1007/s10531-020-02073-8>
- Reif J, Marhouf P and Koptik J** (2013) Bird communities in habitats along a successional gradient: Divergent patterns of species richness, specialization and threat. *Basic and Applied Ecology*, 14, 423–431. <https://doi.org/10.1016/j.baec.2013.05.007>
- Rufin P, Frantz D, Yan L and Hostert P** (2021) Operational coregistration of the Sentinel-2a/b image archive using multitemporal Landsat spectral averages. *IEEE Geoscience and Remote Sensing Letters*, 18, 712–716. <https://doi.org/10.1109/LGRS.2020.2982245>
- Seidl R, Schelhaas M-J and Lexer MJ** (2011) Unraveling the drivers of intensifying forest disturbance regimes in Europe: drivers of forest disturbance intensification. *Global Change Biology*, 17, 2842–2852. <https://doi.org/10.1111/j.1365-2486.2011.02452.x>
- Solberg S, Brunner A, Hanssen KH, Lange H, Næsset E, Rautiainen M and Stenberg P** (2009) Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sensing of Environment*, 113, 2317–2327. <https://doi.org/10.1016/j.rse.2009.06.010>
- Stein A, Gerstner K and Kreft H** (2014) Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters*, 17, 866–880. <https://doi.org/10.1111/ele.12277>
- Stoffels J, Hill J, Sachtleber T, Mader S, Buddenbaum H, Stern O, Langshausen J, Dietz J and Ontrup G** (2015). Satellite-based derivation of high-resolution forest information layers for operational forest management. *Forests*, 6, 1982–2013. <https://doi.org/10.3390/f6061982>
- Swanson ME, Franklin JF, Beschta RL, Crisafulli CM, DellaSala DA, Hutto RL, Lindenmayer DB and Swanson FJ** (2011) The forgotten stage of forest succession: early successional ecosystems on forest sites. *Frontiers in Ecology and the Environment*, 9, 117–125. <https://doi.org/10.1890/090157>
- TEEB—The Economics of Ecosystems and Biodiversity** (2010) *Ecological and economic foundations*. In Kumar P (ed.), Earthscan, London and Washington.
- Tew ER, Conway GJ, Hendersen IG, Milodowski DT, Swinfield T and Sutherland WJ** (2022) Recommendations to enhance breeding bird diversity in managed plantation forests determined using LiDAR. *Ecological Applications*, 32, e2678. <https://doi.org/10.1002/eap.2678>
- Thünen-Institut** (2012a) Order code: 69z1ji_1321of_2012_1322, archiving date: 2014-8-28 13:44:57.920, title: Waldfläche [ha] nach bestockungstyp und beimischung, filter: Land=rheinland-pfalz; jahr=2012. <https://bwi.info>, request on: 03.12.2021. dritte bundeswaldinventur—ergeb. [2012d].
- Thünen-Institut** (2012b) Order code: 69z1ji_1321of_2012_1322, archiving date: 2014-8-28 13:44:57.920, title: Waldfläche [ha] nach bestockungstyp und beimischung, filter: Land=rheinland-pfalz;jahr=2012. <https://bwi.info>, request on: 03.12.2021. *Dritte Bundeswaldinventur—Ergebnisdatenbank*. Retrieved December 3, 2021, from <https://bwi.info/Tabellenauswahl.aspx>
- Thünen-Institut** (2012c) Order code: 69z1ji_1321of_2012_1322, archiving date: 2014-8-28 13:44:57.920, title: Waldfläche [ha] nach eigentumsart und beimischung, filter: Land=rheinland-pfalz; jahr=2012. <https://bwi.info>, Request on: December 03, 2021. dritte bundeswaldinventur—ergeb.
- Toivonen J, Kangas A, Maltamo M, Kukkonen M and Packalen P** (2023) Assessing biodiversity using forest structure indicators based on airborne laser scanning data. *Forest Ecology and Management*, 546, 121376. <https://doi.org/10.1016/j.foreco.2023.121376>
- Tucker CJ** (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8, 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Vanden Borre J, Paelinckx D, Múcher CA, Kooistra L, Haest B, De Blust G and Schmidt AM** (2011) Integrating remote sensing in natura 2000 habitat monitoring: Prospects on the way forward. *Journal for Nature Conservation*, 19, 116–125. <https://doi.org/10.1016/j.jnc.2010.07.003>
- Vidal C, Alberdi I, Redmond J, Vestman M, Lanz A and Schadauer K** (2016) The role of European national forest inventories for international forestry reporting. *Annals of Forest Science*, 73, 793–806. <https://doi.org/10.1007/s13595-016-0545-6>
- Welle T, Aschenbrenner L, Kuonath K, Kirmaier S and Franke J** (2022) Mapping dominant tree species of German forests. *Remote Sensing*, 14, 3330. <https://doi.org/10.3390/rs14143330>
- Wessel M, Brandmeier M and Tiede D** (2018) Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data. *Remote Sensing*, 10, 1419. <https://doi.org/10.3390/rs10091419>
- White JC, Coops NC, Wulder MA, Vastaranta M, Hilker T and Tompalski P** (2016) Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, 42, 619–641. <https://doi.org/10.1080/07038992.2016.1207484>
- Wilson EO and Peter FM** (1988) Structural and functional diversity in temperate forests. In *Biodiversity*. USA: National Academies Press. Retrieved October 26, 2023, from <https://www.ncbi.nlm.nih.gov/books/NBK219319/>
- Wöllauer S, Zeuss D, Magdon P and Nauss T** (2020) RSDB: an easy to deploy open-source web platform for remote sensing raster and point cloud data management, exploration and processing. *Ecography*, 44, 414–426. <https://doi.org/10.1111/ecog.05266>
- Xi Y, Ren C, Tian Q, Ren Y, Dong X and Zhang Z** (2021) Exploitation of time series Sentinel-2 data and different machine learning algorithms for detailed tree species classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 7589–7603. <https://doi.org/10.1109/JSTARS.2021.3098817>

- Xu H** (2006) Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27, 3025–3033. <https://doi.org/10.1080/01431160600589179>
- Yates KL, Bouchet PJ, Caley MJ, Mengersen K, Randin CF, Parnell S, Fielding AH, Bamford AJ, Ban S, Barbosa AM, Dormann CF, Elith J, Embling CB, Ervin GN, Fisher R, Gould S, Graf RF, Gregr EJ, Halpin PN ... Sequeira AM** (2018) Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33, 790–802. <https://doi.org/10.1016/j.tree.2018.08.001>
- Zellweger F, Braunisch V, Baltensweiler A and Bollmann K** (2013) Remotely sensed forest structural complexity predicts multi species occurrence at the landscape scale. *Forest Ecology and Management*, 307, 303–312. <https://doi.org/10.1016/j.foreco.2013.07.023>
- Zhao G, Sanchez-Azofeifa A, Laakso K, Sun C and Fei L** (2021) Hyperspectral and full-waveform LiDAR improve mapping of tropical dry forest's successional stages. *Remote Sensing*, 13, 3830. <https://doi.org/10.3390/rs13193830>

A. Appendix

Table A1. Number of available pixels and polygons per tree species group after balancing

Tree species group	Training pixels	Training polygons	Pixel per polygon	Test pixels	Test polygons
Douglas fir	10,380	60	151	3114	18
Larch	7596	36	211	2110	10
Pine	12,844	26	494	3952	8
Spruce	30,810	237	130	7410	57
Beech	61,152	168	364	14,560	40
Oak	18,573	123	151	4983	33
Other deciduous trees	12,714	78	163	3423	21

Table A2. Formulas for Sentinel-2 bands and indices used in this study. Images for these bands and indices were calculated from monthly composites for 2019–2021 for January, March, April, May, June, September, and October. For the full workflow to create the Sentinel-2 data refer to Bhandari et al. (2024)

Name	Description	Formula/central wavelength
Visible bands		
B2	Blue	490 nm
B3	Green	560 nm
B4	Red	665 nm
Near-infrared bands		
B5	Red edge 1	705 nm
B6	Red edge 2	740 nm
B7	Red edge 3	783 nm
B8	Near infrared	842 nm
B8a	Broad near infrared	865 nm
Shortwave bands		
B11	Shortwave infrared 1	1610 nm
B12	Shortwave infrared 2	2190 nm

Continued

Table A2. Continued

Name	Description	Formula/central wavelength
Vegetation indices		
VI1	Chlorophyll index red edge	$(B7/B5) - 1$
VI2	Enhanced vegetation index	$G \times ((B8a - B4)/(B8a + C1 \times B4 - C2 \times B2 + L))$ with $G = 2.5, L = 1, C1 = 6, C2 = 7.5$
VI3	Kernel NDVI	$(1 - k) / (1 + k)$ with $k = \exp(-(B8a - B4)^2 / (2 \times \sigma^2))$ with $\sigma = 0.5 \times (B8a + B4)$
VI4	Modified normalized difference water index	$(B3 - B11)/(B3 + B11)$
VI5	Modified simple ratio red edge	$((B8/B5) - 1)/\sqrt{((B8/B5) + 1)}$
VI6	Modified simple ratio red edge narrow	$((B8a/B5) - 1)/\sqrt{((B8a/B5) + 1)}$
VI7	Normalized difference moisture index	$(B8a - B11)/(B8a + B11)$
VI8	Normalized difference red edge index 1	$(B6 - B5)/(B6 + B5)$
VI9	Normalized difference red edge index 2	$(B7 - B5)/(B7 + B5)$
VI10	Normalized difference vegetation index	$(B8a - B4)/(B8a + B4)$
VI11	Normalized difference vegetation index red edge 1	$(B8 - B5)/(B8 + B5)$
VI12	Normalized difference vegetation index red edge 2	$(B8 - B6)/(B8 + B6)$
VI13	Normalized difference vegetation index red edge 3	$(B8 - B7)/(B8 + B7)$
VI14	Normalized difference vegetation index red-edge 1 narrow	$(B8a - B5)/(B8a + B5)$
VI15	Normalized difference vegetation index red-edge 2 narrow	$(B8a - B6)/(B8a + B6)$
VI16	Normalized difference vegetation index red-edge 3 narrow	$(B8a - B7)/(B8a + B7)$
VI17	Normalized difference water index	$(B3 - B8a)/(B3 + B8a)$
VI18	Soil adjusted and atm. Resistant vegetation index	$(B8a - RB)/(B8a + RB + L) \times (1 + L)$ with $RB = B4 - (B2 - B4)$ with $L = 0.5$
VI19	Soil adjusted vegetation index	$(B8a - B4) / (B8a + B4 + L) \times (1 + L)$ with $L = 0.5$

Table A3. Used sensors for the acquisition of LiDAR data for each year and beam divergence of the respective sensor. 0.18 mrad correspond to an increase of 18 cm of beam diameter per 1000 m distance

Time of data acquisition	Sensor	Beam divergence
2014 to 2016	Riegl LMS Q560 and Riegl LMS Q680i	≤ 0.5 mrad
2017 to 2019	Riegl LMS Q780	≤ 0.25 mrad
2020	Riegl LMS VQ780i	≤ 0.18 mrad
2021	Riegl LMS VQ780II	≤ 0.18 mrad

Table A4. Overview of structural LiDAR indices with the names of the Remote Sensing Database (RSDB; Wöllauer et al., 2020). The formulas of the calculation of each index can be found under their RSDB name at: <https://github.com/environmentalinformatics-marburg/rsdb/wiki/Point-cloud-indices>

Name	RSDB label
Canopy	
CH (canopy height) max	BE_H_MAX
CHM max	chm_height_max
CH mean	BE_H_MEAN
CHM mean	chm_height_mean
CH SD	BE_H_SD
CHM SD	chm_height_sd
CH median	BE_H_MEDIAN
CH skew	BE_H_SKEW
CH curtosis	BE_H_KURTOSIS
CH perc 30	BE_H_P30
CH perc 70	BE_H_P70
Vegetation structure	
PR (penetration rate) canopy	BE_PR_CAN
PR regeneration	BE_PR_REG
PR understory	BE_PR_UND
RD (return density) canopy	BE_RD_CAN
RD regeneration	BE_RD_REG
RD understory	BE_RD_UND
VDR	VDR
Vegetation	
AGB	AGB
LAI	LAI
FHD	BE_FHD
VC (vegetation coverage) 1 m	vegetation_coverage_01m_CHM
VC 2 m	vegetation_coverage_02m_CHM
VC 5 m	vegetation_coverage_05m_CHM
VC 10 m	vegetation_coverage_10m_CHM
Terrain	
Elev (elevation) max	dtm_elevation_max
Elev mean	BE_ELEV_MEAN
Elev sd	dtm_elevation_sd
Elev slope	BE_ELEV_SLOPE

function

```

density = count rows of data grouped by response
rename density columns to ("Var1", "Freq")
densityPlots = count rows of data grouped by polygonID
Rename desityPlots columns to ("Var1", "Freq")
balancingDF = create an empty data frame
for i in 1 to 500 (max number of pixel per polygon to consider)
    minID = get rows from densityPlots where Freq is greater than or equal to i
    locationID = unique values of minID Var1
    dataSubset = filter rows of data where polygonID is in locationID
    trainDensity = count rows of dataSubset grouped by response
    summarize trainDensity with numberDistinctLocations and store it in sampels and
    minSamples columns
    append trainDensity to balancingDF
end
balancer = get rows from balancingDF where sampels is the minimum value
balancerDF = get rows from balancingDF where response matches balancer
maxBalancer = get rows from balancerDF where sampels is the maximum value
if rows in maxBalancer are more than 1
    maxBalancer = get rows from maxBalancer where numberDistinctLocations is the
    maximum
end
balanceAll = get rows from balancingDF where minSamples matches maxBalancer
minID = get rows from densityPlots where Freq is greater than or equal to maxBalancer
minSamples
locationID = unique values of minID Var1
dataSubset = filter rows of data where polygonID is in locationID
IDs = empty list
for each class in unique response
    tmp = filter rows of dataSubset where response matches class
    ID = get unique values of tmp polygonID and sample a minimum of density
    numberDistinctLocations
    add ID to IDs
end
balancedData = filter Rows of dataSubset where polygonID is in IDs
return balancedData
end

```

Figure A1. Pseudocode balancing.

Table A5. Number of available forest inventory polygons for each successional stage and tree species group

Tree species group	Establishment	Qualification	Dimensioning	Maturing	Generational	
					change	Decay
Douglas fir	8	157	1104	98	0	0
Larch	0	1	78	88	0	0
Pine	0	16	101	1726	0	0
Spruce	6	255	1777	1320	0	2
Beech	10	283	695	1636	0	13
Oak	25	162	1135	1810	0	1
Other deciduous trees	13	157	600	462	0	2

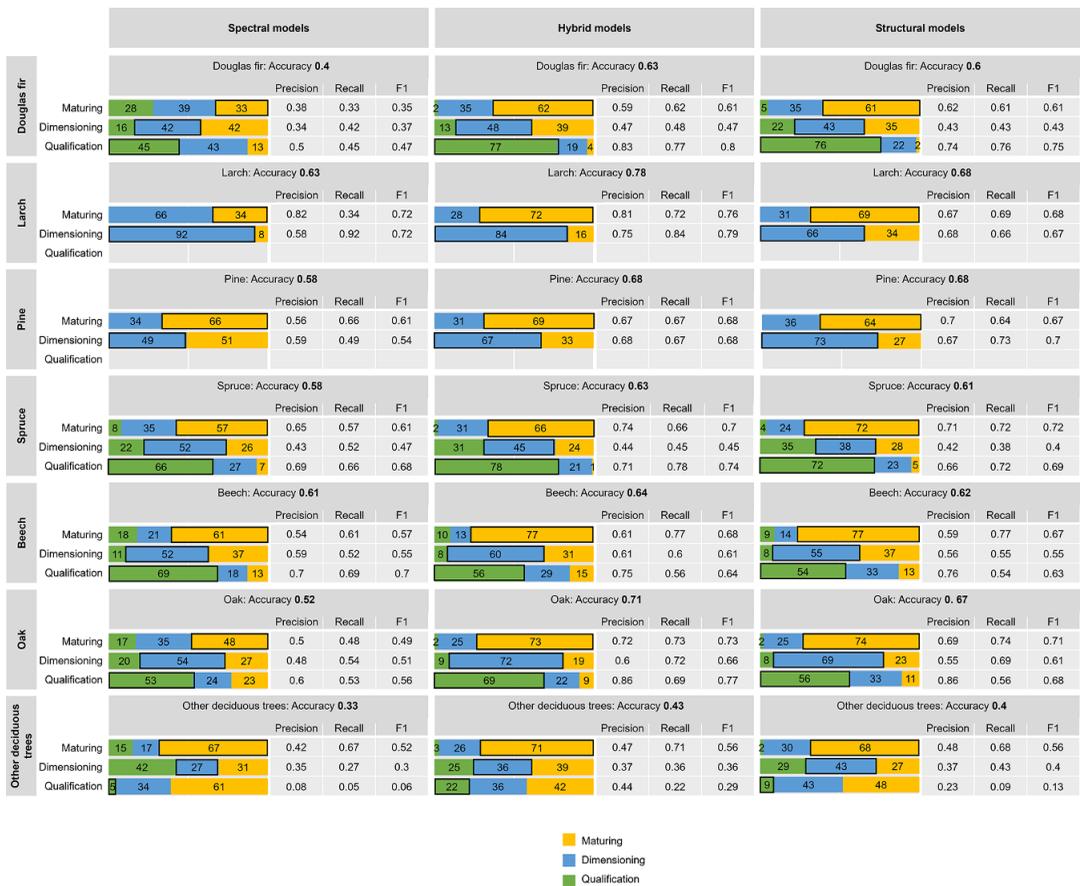


Figure A2. The left Column of the plots shows the test results of models only using Sentinel-2 variables (spectral models), middle column shows test results using Sentinel-2 and LiDAR variables (hybrid models). The right column shows the test results of the models using only LiDAR variables (structural model). Each plot shows the confusion matrices of the testing for one tree species group. Observed values are shown on the x-axis and the predicted values on the y-axis in percent. For example, the bar for the maturing phase (yellow) should be as large as possible in the first row (maturing) of each plot. All values that end up in the other phases (blue and green) are misclassified pixels. Values are rounded to two decimals.

Table A6. Significance of t-test for each variable group

	Visible	Near-infrared	Shortwave	Vegetation index	Canopy	Structure	Vegetation	Terrain
Visible		0.6789	0.4249	0.4874	0.009882	0.7102	0.7979	1
Near-infrared	0.6789		0.7609	1	0.1009	1	0.5821	0.5066
Shortwave	0.4249	0.7609		0.8581	0.1761	0.8451	0.4811	0.4
Vegetation index	0.4874	1	0.8581		0.009464	0.8112	0.5548	0.666
Canopy	0.009882	0.1009	0.1761	0.009464		0.02926	0.03983	0.06984
Structure	0.7102	1	0.8451	0.8112	0.02926		0.705	0.6447
Vegetation	0.7979	0.5821	0.4811	0.5548	0.03983	0.705		1
Terrain	1	0.5066	0.4	0.666	0.06984	0.6447	1	

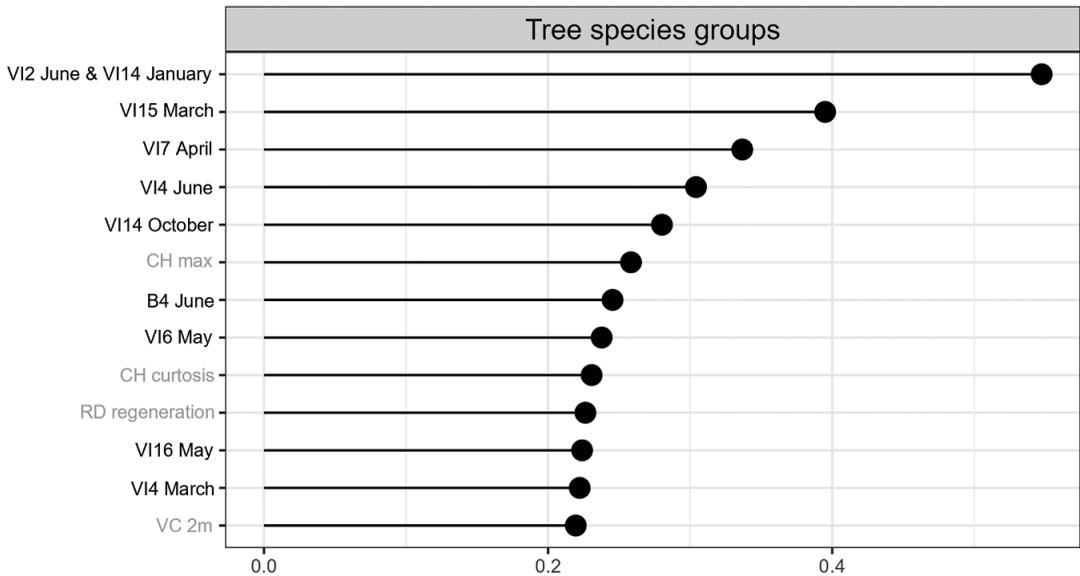


Figure A3. Variable importance of the tree species groups model. Sentinel-2 variables are labeled black and LiDAR variables are labeled gray at the y-axis.

Table A7. Confusion matrix for the tree species groups model with a total accuracy of 0.81 on test data. The values indicate the classified pixels as percentage

	Douglas fir	Larch	Pine	Spruce	Beech	Oak	Other deciduous trees
Douglas fir	85	0	4	2	0	1	2
Larch	1	86	2	2	1	2	4
Pine	5	3	90	2	5	1	2
Spruce	7	0	2	89	2	1	2
Beech	1	5	2	1	84	11	3
Oak	0	3	1	0	6	63	13
Other deciduous trees	1	2	0	3	1	23	74

Table A8. Confusion matrix for the hierarchical modeling of tree species groups and successional stages with letters indicating the successional stages (qualification: *q*, dimensioning: *d*, and maturing: *m*). The values indicate the classified pixels as a percentage

	Douglas fir			Larch		Pine		Spruce			Beech			Oak		
	<i>q</i>	<i>d</i>	<i>m</i>	<i>d</i>	<i>m</i>	<i>d</i>	<i>m</i>	<i>q</i>	<i>d</i>	<i>m</i>	<i>q</i>	<i>d</i>	<i>m</i>	<i>q</i>	<i>d</i>	<i>M</i>
Douglas fir <i>q</i>	65	12	2	0	0	0	0	2	0	0	0	0	0	1	0	0
Douglas fir <i>d</i>	14	41	31	0	0	0	1	0	0	1	0	0	0	1	0	0
Douglas fir <i>m</i>	3	36	56	0	0	2	3	0	0	2	0	0	0	0	0	0
Larch <i>d</i>	0	1	1	76	22	0	0	1	0	0	1	2	0	2	1	0
Larch <i>m</i>	0	0	1	16	63	3	0	0	0	0	1	1	0	0	0	3
Pine <i>d</i>	3	4	3	4	1	65	27	1	1	1	1	5	1	1	0	0
Pine <i>m</i>	0	1	2	0	1	26	61	0	1	1	1	0	1	2	0	1
Spruce <i>q</i>	12	1	0	0	0	0	0	72	26	2	3	1	1	2	0	0
Spruce <i>d</i>	1	2	1	0	0	0	0	21	47	29	0	1	0	0	0	0
Spruce <i>m</i>	0	2	3	0	0	0	4	1	22	62	1	1	1	0	1	0
Beech <i>q</i>	0	0	0	0	0	0	0	1	0	0	48	3	5	12	1	0
Beech <i>d</i>	0	1	0	3	2	1	1	0	0	0	25	53	12	1	4	0
Beech <i>m</i>	0	1	0	1	4	2	1	0	1	1	10	30	72	9	5	11
Oak <i>q</i>	0	0	0	0	0	0	0	1	0	0	5	1	4	54	9	0
Oak <i>d</i>	0	0	0	1	1	0	0	0	1	0	3	2	1	14	62	24
Oak <i>m</i>	0	0	0	0	4	1	0	0	0	0	3	1	3	1	18	61

Table A9. The most important species in the tree species groups in the forest inventory data

Douglas fir	
Douglas fir	<i>Pseudotsuga menziesii</i>
False cypress	<i>Chamaecyparis</i>
Juniper	<i>Juniperus</i>
Redwoods	<i>Sequoioideae</i>
Thuja	<i>Thuja</i>
Tsuga	<i>Tsuga</i>
Yew	<i>Taxus</i>
Larch	
European larch	<i>Larix decidua</i>
Japanese larch	<i>Larix kaempferi</i>
Pine	
Black pine	<i>Pinus nigra</i>
Pine	<i>Pinus nigra</i>
Ponderosa pine	<i>Pinus ponderosa</i>
Weymouth pine	<i>Pinus strobus</i>
Spruce	
Serbian spruce	<i>Picea omorika</i>
Sitka spruce	<i>Picea sitchensis</i>
Spruce	<i>Picea</i>

Continued

Table A9. Continued

Douglas fir	
<hr/>	
Beech	
Beech	Fagus
Oak	
Common oak	<i>Quercus robur</i>
Downy oak	<i>Quercus pubescens</i>
Sessile oak	<i>Quercus petraea</i>
Turkey oak	<i>Quercus cerris</i>
Other deciduous trees	
Short-lived deciduous trees	
Alder	Alnus
Aspen	<i>Populus tremula</i>
Balsam poplar	<i>Populus balsamifera</i>
Birch	Betula
Black poplar	<i>Populus nigra</i>
Cherry	Prunus
Downy birch	<i>Betula pubescens</i>
European crab apple	<i>Malus sylvestris</i>
European wild pear	<i>Pyrus pyraeaster</i>
Goat willow	<i>Salix caprea</i>
Poplar	Populus
Rowan	<i>Sorbus aucuparia</i>
Silver birch	<i>Betula pendula</i>
Long-lived deciduous trees	
Ash	<i>Fraxinus excelsior</i>
Black locust	<i>Robinia pseudoacacia</i>
Common hornbeam	<i>Carpinus betulus</i>
Common whitebeam	<i>Sorbus aria</i>
Eastern american black walnut	<i>Juglans nigra</i>
Elm	Ulmus
English walnut	<i>Juglans regia</i>
Field maple	<i>Acer campestre</i>
Large-leaved linden	<i>Tilia platyphyllos</i>
Linden	<i>T. platyphyllos</i>
Montpellier maple	<i>Acer monspessulanum</i>
Norway maple	<i>Acer platanoides</i>
Red oak	<i>Quercus rubra</i>
Service tree	<i>Sorbus domestica</i>
Shadbush	Amalanchier
Small-leaved linden	<i>Tilia cordata</i>
Sorbus species	Sorbus
Sweet chestnut	<i>Castanea sativa</i>
Sycamore maple	<i>Acer pseudoplatanus</i>
Wild service tree	<i>Sorbus torminalis</i>

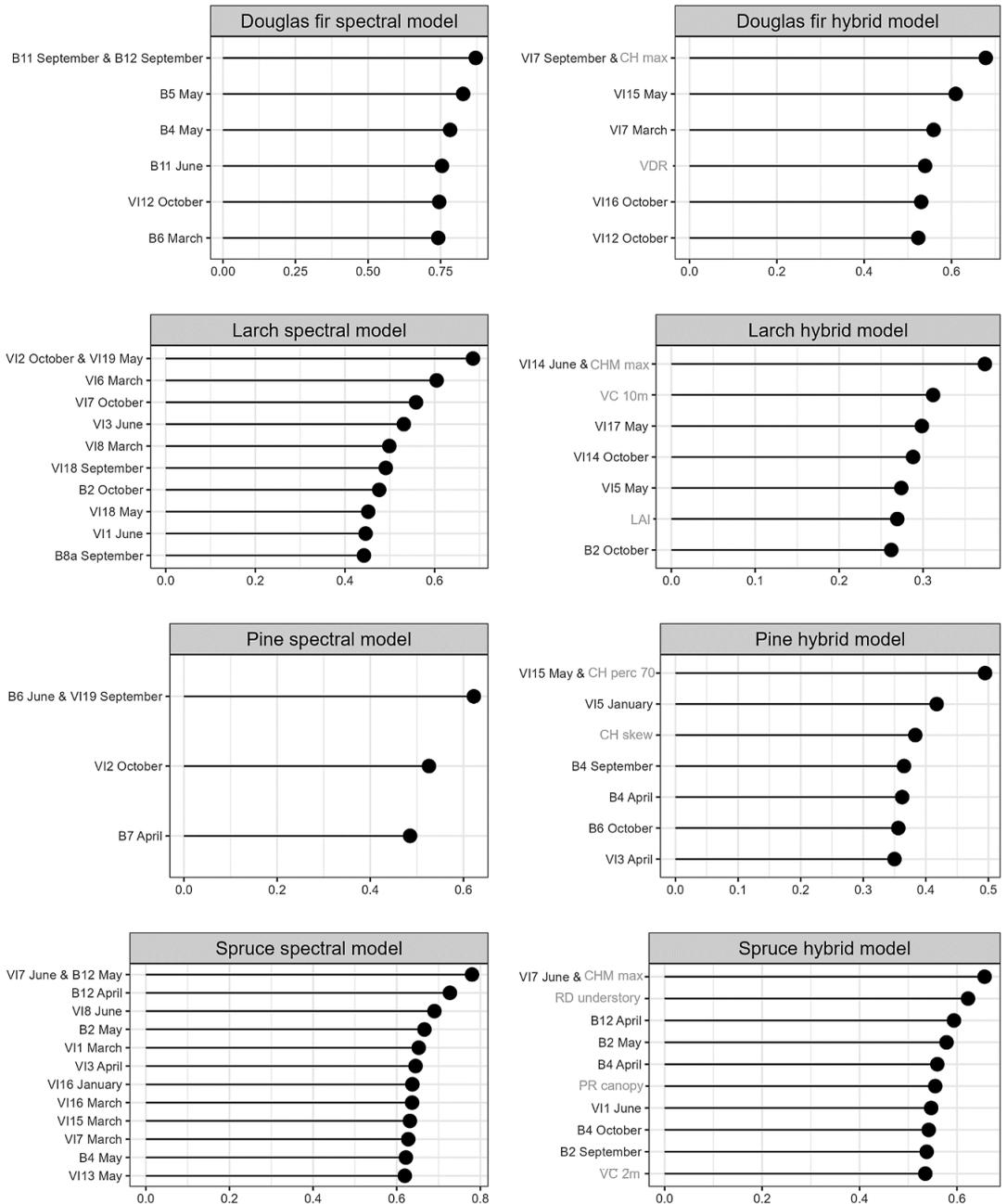


Figure A4. Variable importance of hybrid and spectral tree species groups specific successional stages models. For each tree species group one model only using Sentinel-2 data (spectral on the left) and one using Sentinel-2 and LiDAR (hybrid on the right) is depicted. Sentinel-2 variables are labeled black and LiDAR variables are labeled grey at the y-axis.

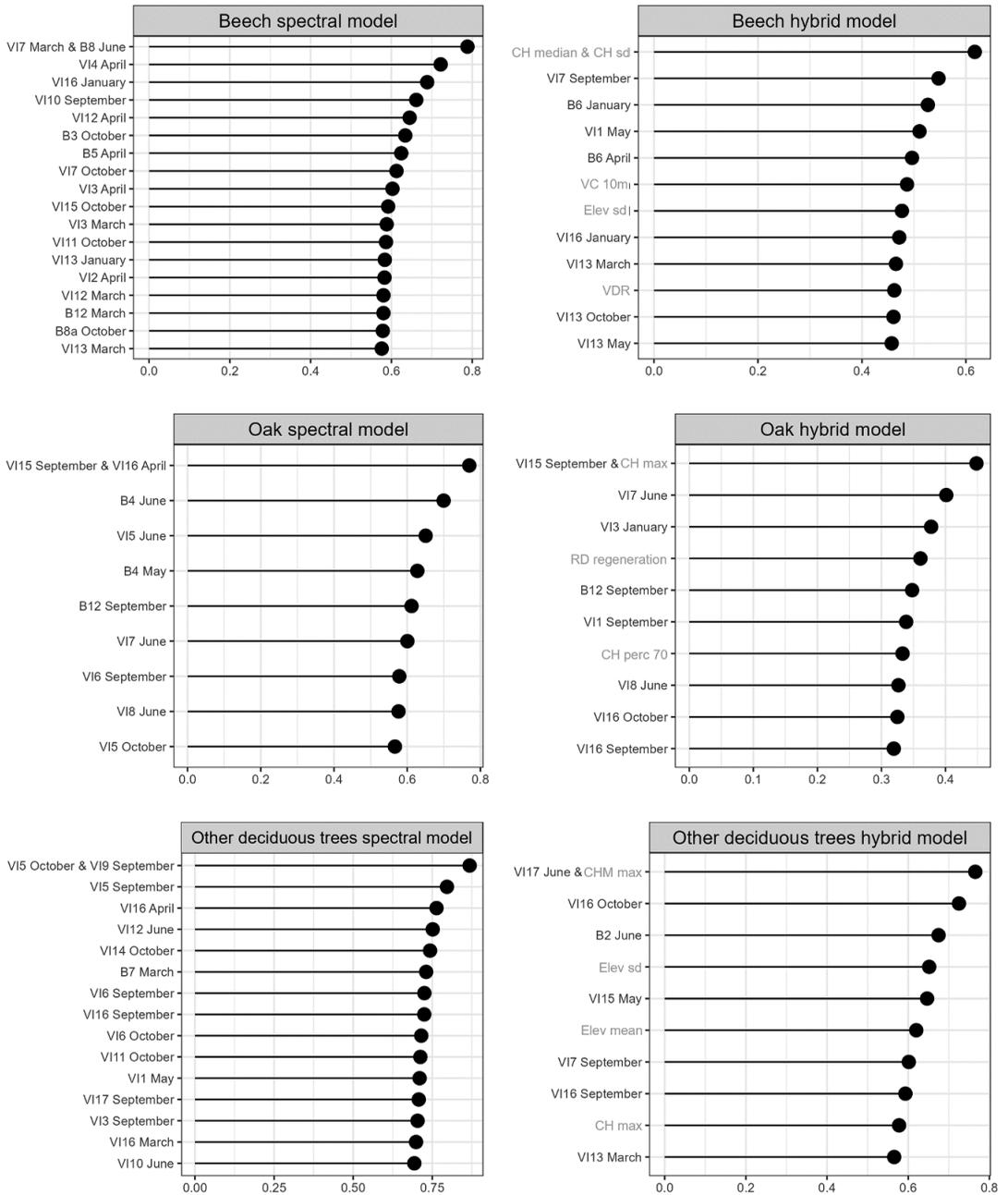


Figure A4. (continue)