Philipps Universität Marburg

LiDAR Insights into Biodiversity Patterns

Evaluating Vegetation Structure's Role in Ecosystem Assessments

kumulative Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

im Fachbereich Geographie
der Philipps-Universität Marburg
vorgelegt von

M.Sc. Alice Ziegler
aus Aschaffenburg

Gutachter:
Prof. Dr. Thomas Nauss
Prof. Dr. Jörg Bendix
Einreichungstermin:
22.11.2023
Prüfungstermin:
26.2.2024
Erscheinungsort:
Marburg
Erscheinungsjahr:
2024
Hochschulkennziffer:
1180

Contents

Chapter 1

Introduction

1.1 Motivation

1.1.1 Relevance of biodiversity and its monitoring

The inventory, monitoring and strategic observation of biodiversity and its loss is of major importance for nature conservation. As documented in the United Nations Convention on Biological Diversity in 1992, this includes genetic, species and ecosystem diversity (United Nations Environmental Programme, 1992). One of the Sustainable Development Goals of the United Nations in the context of the protection and restoration of terrestrial ecosystems is the halt of the biodiversity loss, with a special focus on forest and mountain ecosystems (United Nations, 2015). The environment is exposed to the multi-faceted threats of climate change, pollution and deforestation, causing an escalation in a loss of species (United Nations Department of Economic and Social Affairs, 2023). The future development depends on how ecosystems in all their complexity can respond to these stressors (Trumbore et al., 2015). As a baseline to comprehend biodiversity loss and to delineate conservation strategies, it is crucial to develop and install comprehensive monitoring approaches to measure or model biodiversity (Gao et al., 2014; Myers et al., 2000).

For the high-quality assessment of biodiversity, extensive in-situ studies are necessary. Besides project-based field studies for specific research questions, regular inventories (e.g. governmental forest inventories) provide continued statistically-designed procedures to monitor several biodiversity indicators like plant species or structural features in given areas (Corona et al., 2011). In contrast, field-based studies normally only contain information for a limited spatial coverage. To draw conclusions on larger scales, remote sensing techniques are valuable tools, which have led to a rapid evolution in the field of remote sensing for ecological purposes in the last decades (Cavender-Bares et al., 2022; Reddy et al., 2021). To enhance comparability between initiatives, there are efforts to establish essential biodiversity variables that can be generated through remote sensing (Pettorelli et al., 2016).
With the freely available optical data from ESAs (European Space Agency) Copernicus Sentinel-2 mission in 2014, spatial resolution reached a quality (10 m) that is sufficient for many regional ecological questions including tree species composition, grassland plant communities or invasive species mapping (Ahmed et al., 2021; Grabska et al., 2019; Hemmerling et al., 2021; Hościło & Lewandowska, 2019; Immitzer et al., 2016; Lange et al., 2022; Masemola et al., 2020; Rapinel et al., 2019; Welle et al., 2022; Wessel et al., 2018; Xi et al., 2021). Even prior to Sentinel, Landsat imagery with its 30 m spatial resolution opened up possibilities for different ecological questions, including biodiversity studies and now even offer the great possibility for time series reaching back to the 1980s (Berveglieri et al., 2021; Graf et al., 2019; Margono et al., 2012; Parisi et al., 2022). This offers valuable information for documenting changes in the environment.
Apart from capturing vegetation and surface properties directly from the images, the integration of area-wide remotely-sensed data with field observations (e.g. species composition) offer the possibility to model the occurrence of small vegetation species or even animal species by assuming that the satellite imagery provides a comprehensive view on the ecosystem and hence on potential habitats. Point-based observations can be used as training data for models and in combination with remote sensing data it is possible to apply the models to unsampled areas which allows mapping the target variable.

Remote sensing is a well established but still evolving method for measuring and modeling biodiversity (Pettorelli et al., 2016; Reddy et al., 2021). However, the use of optical sensors alone lack the possibility to look deeper than the uppermost vegetation surface. One important component of environmental heterogeneity is the structural heterogeneity of the vegetation consisting of horizontal heterogeneity, vertical stratification and a diverse species composition (Gao et al., 2014). Especially in forests the structure below the canopy surface is an indicator for species richness as more species-rich forests typically feature a higher structural diversity. At the same time, heterogeneous forest structures provide a higher number of niches and as a consequence, plant and animal species richness in general tends to be higher in complex than in simpler forest systems (Brokaw & Lent, 1999; Heidrich et al., 2023; Macarthur & Macarthur, 1961; Polechová & Storch, 2019; Stein et al., 2014; Wilson & Peter, 1988). For example, Ozanne et al. (2003) have shown that canopies with their structural complexity hold 40 % of all global species. Even though some studies show no or even negative correlations between measures of vegetation structure and the species richness of specific arthropod groups (Humphrey et al., 1999; Lassau & Hochuli, 2004; Müller et al., 2018), the positive relationship between the complexity of vegetation structure and biodiversity is generally accepted (Stein et al., 2014; Tews et al., 2004) leading to a need to quantify not only the horizontal, but also the vertical structure. Horizontal complexity of the vegetation surface can be measured by widely available spectral satellite remote sensing as outlined above. The vertical structure, however, requires different methods. LiDAR (Light Detection And Ranging) with its three-dimensional point clouds is able to penetrate different vegetation layers and therefore facilitates remotely-sensed insights into this important part of vegetation structure and its heterogeneity.

1.1.2 LiDAR data: remote sensing based indicators for structure

Structural data derived from LiDAR sensors have become more and more prominent during the last decades (Bergen et al., 2009; Davies & Asner, 2014; Simonson et al., 2014; Toivonen et al., 2023). LiDAR data, with the possibility to directly derive structure in both the horizontal and vertical direction, offers great capabilities to assess biodiversity (Toivonen et al., 2023; Zellweger et al., 2013). A lot of originally field-based measures for habitat heterogeneity that are considered to be relevant for species diversity can be measured and quantified by LiDAR data (Simonson et al., 2014). LiDAR is an active technology that sends out laser beams from the infrared spectrum (Lohani & Ghosh, 2017). For remote sensing purposes it is commonly mounted either airborne on airplanes, helicopters as well as UAVs (unmanned aerial vehicles) or spaceborne on satellites or space stations. The sensors detect the time between emitting and receiving a beam after the reflection at the surface at a specific location. As the signal is only partly penetrating vegetation layers, several reflected returns can be detected (see figure 1.1), the first one representing the uppermost surface and the last one ideally the ground (Senf, 2022). Returns in between can be traced back to reflections at different obstacles, such as vegetation layers. This results in a point cloud showing the three-dimensional distribution of different surfaces (see figure 1.1). This point cloud needs to be processed to derive further information. Different indices can be calculated and aggregated to raster layers of different pixel sizes to make use of the point cloud and retrieve proxies such as vegetation structure in forests. Metrics are used to characterize either the total vegetation (e.g. leaf area index, aboveground biomass), single trees (e.g. crown diameter), single layers (e.g. canopy height, mean understorey height) or multiple layers (e.g. penetration rates, return densities, height percentiles) (Bakx et al., 2019; Bergen et al., 2009). One of the simpler calculated metrics would be using all last returns reflected at the ground to create a digital elevation model or, respectively, all first returns to derive a canopy height model. Also aggregated metrics like the mean height of returns per pixel are possible (see figure 1.1). Some LiDAR indices that are computed specifically for distinct vegetation strata or for a designated height above the ground would use, for example, a ratio of the number of returns in several layers (Wöllauer et al., 2020).
Currently mainly airborne LiDAR data are used for ecological research and offer spatial resolutions of typically from around four pulses per square meter for governmental flight campaigns across a entire federal state (e.g. Landesamt für Vermessungs- und Katasterverwaltung Rheinland-Pflaz, 2021) to over 40 pulses per square meter for project-specific high-resolution campaigns (e.g. Mandl et al., 2023). Since 2019 the new GEDI (Global Ecosystem Dynamics Investigation) mission offers new possibilities of spaceborne LiDAR data. Compared to former spaceborne LiDAR missions that were primarily used for climate modeling

(Fouladinejad et al., 2019), GEDIs main purpose is to measure vertical vegetation structure (Dubayah et al., 2020). Compared to airborne LiDAR missions that typically cover areas wall-to-wall, GEDI offers scattered footprints with a fixed spacing along its track. Areas are covered, depending on their longitude, every couple of days to weeks and the exact coordinates of the footprints might shift with each overflight.

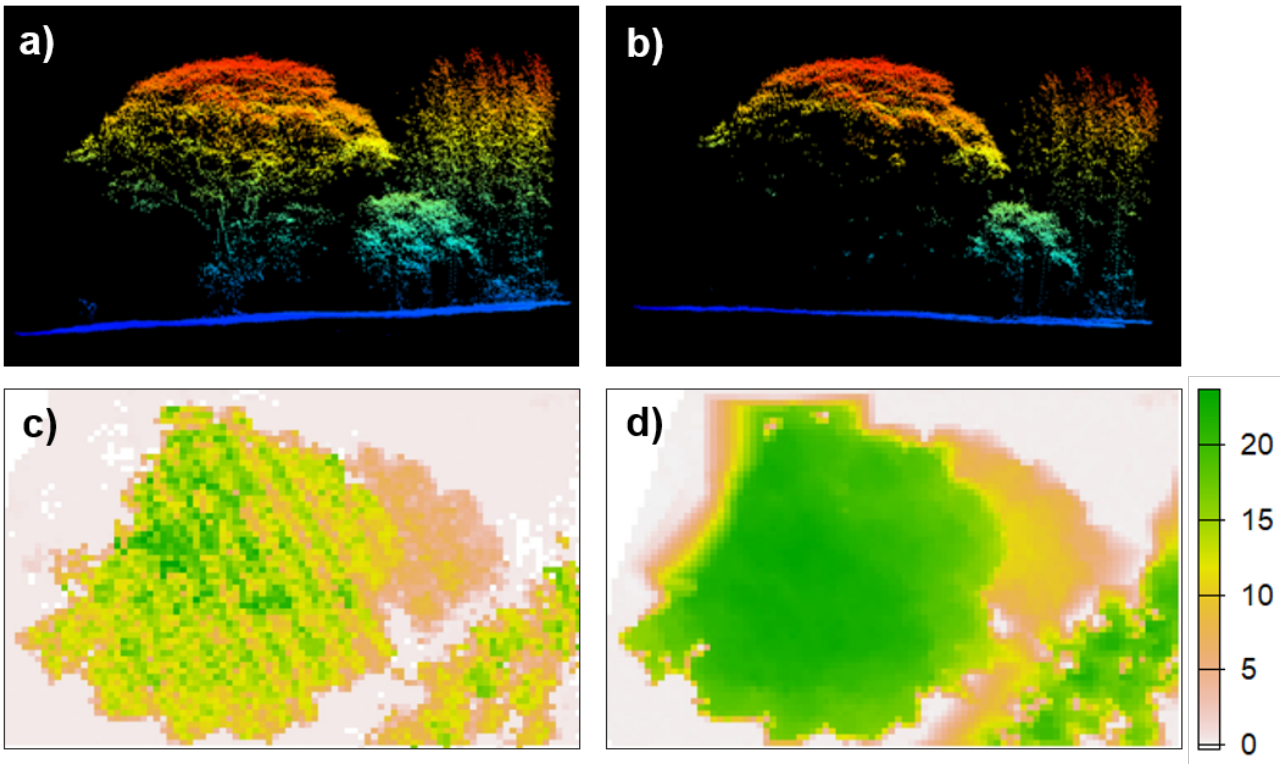


Figure 1.1: Example of a LiDAR data set at Mount Kilimanjaro showing a three dimensional point cloud and different levels of retrieved information. Plot a) shows a point cloud from side view including all returns, while b) shows all „first returns". The plots at the bottom show rasterized information derived from the point cloud from top view. Plot c) visualizes the mean vegetation height per pixel in meter, while plot d) shows the canopy height in meter. Data were collected during an airborne flight campaign at Mount Kilimanjaro in 2015 (Ziegler et al., 2022), processed using the remote sensing data base (RSDB) (Wöllauer et al., 2020) and visualized using the lidR package (Roussel et al., 2020).

1.1.3 State of the art: Modeling biodiversity with LiDAR data

As elaborated earlier, there are plentiful applications where LiDAR data can be beneficial for biodiversity studies. With the evolving opportunities of especially LiDAR remote sensing and a growing number of different sensors and data sets, the meaningful combination of data sets is becoming increasingly important. Airborne LiDAR missions improved technically during the last years and are now able to provide reliable detailed area-wide measures on regional scales (Guo et al., 2021; Luo et al., 2018; Sumnall et al., 2016; Wang & Fang, 2020). Additionally, especially the new spaceborne GEDI data opened up great possibilities for repeated large-scale or even global retrieval of vegetation structure like canopy height or aboveground biomass (Duncanson et al., 2022; Potapov et al., 2021). Those structural data sets, along with the continuously expanding spectral remote sensing database, provide a multitude of combinatorial possibilities for ecosystem monitoring (Francini et al., 2022; Labenski et al., 2023; Li et al., 2020; Potapov et al., 2021). The advanced integration of readily available large scale data sets based on the knowledge and perspectives that can be compiled from the individual aspects from former studies can yield practical and applicable insights in a level of detail that was previously inaccessible for ecosystem conservation (Cavender-Bares et al., 2022).
Generally, there are three different common approaches to derive area-wide information from structural airborne LiDAR data that will be explained in the following paragraphs. Insights can be used to assess

either (i) vegetation measures, (ii) directly derived vegetation related properties or (iii) indirectly linked properties.
Metrics that can be directly measured from the LiDAR point cloud (i) include canopy height and vegetation layer distribution (Kalinicheva et al., 2022; Khosravipour et al., 2014; Liu et al., 2021; Luo et al., 2018; Whitehurst et al., 2013). Airborne LiDAR offers the area-wide retrieval of those measures at smaller scales and since the launch of the GEDI mission in 2019 there is even the possibility for almost global standardized retrieval of those metrics. Apart from metrics that are inherently included in the point cloud (e.g. canopy height), directly and indirectly derived metrics require ground truth data and modeling approaches to link the LiDAR-derived information to the target variable. A subsequent application of models to entire LiDAR data sets allows for an area-wide mapping of the trained biodiversity measures (e.g. maps of tree species groups, above ground biomass, species distribution) (Burns et al., 2020; Saarela et al., 2020; Welle et al., 2022).
Metrics directly linked to the structure detected by LiDAR technology (ii) are vegetation properties like tree species, leaf area index (LAI), aboveground biomass or forest successional stages (Duan et al., 2023; Duncanson et al., 2022; Falkowski et al., 2009; Li et al., 2013; Michałowska & Rapiński, 2021; Richardson et al., 2009; Torre-Tojal et al., 2022; Wang & Fang, 2020; Yan et al., 2019; Zhu et al., 2020). Those properties are derived by using the three-dimensional point cloud as a proxy.
In addition to the direct modeling of vegetation, which is visible in LiDAR data, there is also the potential for an indirect utilization (iii) through the association of structural heterogeneity with biodiversity measures like animal species richness. As animal species richness is generally linked to heterogeneity in vegetation structure, the vegetation indices derived from LiDAR data can be used as predictors for modeling animal species richness (Cavender-Bares et al., 2022; Simonson et al., 2014). Especially for the prediction of avian species like birds and bats (Bakx et al., 2019; Clawges et al., 2008; Flashpohler et al., 2010; Froidevaux et al., 2016; Goetz et al., 2007; Lesak et al., 2011; Melin et al., 2018; Müller & Vierling, 2014; Müller et al., 2009, 2010; Rechsteiner et al., 2017; Tew et al., 2022; Vogeler et al., 2014; Zellweger et al., 2013) but also other vertebrates and invertebrates (Müller & Brandl, 2009; Müller et al., 2018; Simonson et al., 2014), structural metrics derived from LiDAR data have proven to be reliable predictors.

1.1.4 Current challenges

The state of the art for remote sensing based biodiversity models, as outlined above, shows that LiDAR data have already proven to be a relevant and applicable source of information for delineation and modeling of biodiversity. There are, however, still considerable challenges that require further research in order to fully leverage the potential of LiDAR data. These challenges include, but are not limited to:

Disentangling of structural LiDAR information from co-factors

To draw conclusions on the relevance of LiDAR data for modeling, for example, animal species diversity, the effect needs to be assessed in comparison to other sources. This is a complex task, as vegetation structure might be correlated to other environmental influences such as elevation gradients. As areas of high biodiversity tend to be situated in mountainous areas (Myers et al., 2000), they are especially important for natural conservation strategies. The elevation itself however also is a prominent driver for other abiotic and biotic factors, including climate or land cover, which in return influences the LiDAR signal (Sandel & Smith, 2009; Stein et al., 2014). If this chain of effects is not considered, it might lead to an overestimation of the importance of the actual vegetation structure and thereby may lead to false conclusions on the relevance of LiDAR data. Therefore, it is important to disentangle the predictive power of LiDAR-derived structural variables to other factors (Müller & Brandl, 2009; Vierling et al., 2011).

Lack of systematic tests across taxa

With a growing expertise in the merged fields of ecology and remote sensing, an increasing number of essential biodiversity variables can be reliably predicted through remote sensing (Pettorelli et al., 2016). Nevertheless, progress still needs to be made and especially the comparability between methods as well as results of studies needs improvement. For flying organisms it has been shown that information derived from LiDAR data delivers suitable indicators for singular assemblages (Clawges et al., 2008; Flashpohler et al., 2010; Goetz et al., 2007; Jung et al., 2012; Lesak et al., 2011; Müller et al., 2009, 2010; Rechsteiner et al., 2017; Vogeler et al., 2014; Zellweger et al., 2013). But as environmental heterogeneity and therefore also vegetation structure lead to niches and higher expected species diversity in general (Heidrich et al., 2023; Macarthur & Macarthur, 1961; Stein et al., 2014; Tews et al., 2004), multi-taxa approaches are needed. Especially when using structural LiDAR remote sensing for the indirect prediction of animal species richness, work needs to be put in the comparability between different taxa to be able to eventually bridge the gap between local and global knowledge (Senf, 2022). This is not limited to the remote sensing part of the monitoring, but at least equally important in field campaigns, as collecting ground truth will always be necessary and provides the basis for the training and testing of models especially on larger scales. The availability of those approaches would facilitate drawing more general conclusions that are comparable across taxa and finding mechanisms that are not taxon-specific but applicable on a higher, more general level.

Uncertainty about the effective utilization of irregular LiDAR data sources

The continuous monitoring in both space and time is of great importance to track changes in biodiversity. Yet, most airborne LiDAR missions are singular campaigns or, in the case of governmental campaigns, are typically repeated only after several years. Different data acquisitions tend to vary in technical details like flight altitude, scan angle or phenological time of the year of acquisition, which hinders the comparability between campaigns (Næsset, 2009; Ørka et al., 2010; Solberg et al., 2009). Therefore, there is a need for approaches to make use of heterogeneous data sets. In contrast to airborne LiDAR missions, the spaceborne GEDI mission covers regions multiple times a month, which holds advantages, but the scattered footprints do not necessarily spatially overlap. Even though the irregular repetition offers new possibilities, an actual observation of changes for specific areas is not feasible. Furthermore, GEDI was planned to operate for only two years and even though extensions of operational time were granted, there will be at least a several-month gap inbetween. The discontinuity of LiDAR data poses a major challenge for ongoing biodiversity monitoring, highlighting the need for modeling strategies that are adapted to data sets with a temporal mismatch or bias.

Linking ecological process understanding to appropriate remote sensing data sets

The now available and growing variety of options and qualities of different missions and systems for LiDAR data allow for the connection of different existing components, variables and indicators of biodiversity (Reddy, 2021). Reddy et al. (2021) state that several biophysical properties, including floristic composition and different vegetation strata, can be quantified especially well by the integration of spectral and LiDAR data. But even with technological improvements, there are challenges that demand thorough investigations and further developments. As biodiversity has to be monitored specifically in time and space depending on the underlying process, a deep understanding of processes and their representation in adequate data sets is the base for effective and meaningful monitoring strategies. Cavender-Bares et al. (2022) even argue that further developments are not as dependent on advancements in technology but on the collaborative work in transferring knowledge and understanding biodiversity mechanisms. There is a need to overlay aspects of biodiversity and remote sensing, especially LiDAR, and examine the applications on different scales to identify where the process understanding has to be matched with the utilization of data. Therefore, different scales relevant for remote sensing based modeling approaches have to be considered and matched to

yield optimal results for the corresponding application (Wu & Li, 2009). Natural patterns evolve from different underlying natural processes that can even be working on different scales than the patterns emerging (Estes et al., 2018; Reddy et al., 2021; Sandel & Smith, 2009). For a more comprehensive understanding of processes at the interface of species diversity and ecosystem structure, further interdisciplinary research is essential, emphasizing the integration of multi-platform and multi-scale data sets (Acebes et al., 2021).

1.2 Aim of this thesis

In this thesis, I aim at making meaningful contributions towards addressing the challenges listed above. With LiDAR remote sensing, it is possible to capture the structural components of vegetation which play a vital role for biodiversity. In a mountainous study area, I want to disentangle the effects of vegetation structure derived by LiDAR data and the impact of elevation on animal species richness. In working towards systematic comparability across taxa, the goal is to offer results that can be transferred across various taxa. From a data-driven perspective, quality and resolution of data sets are constantly improving and new possibilities like global temporally and spatially scattered spaceborne LiDAR data are becoming operational. As an objective of this thesis I want to investigate whether such supposedly superior data consistently improve the potential for the mapping of biodiversity variables. Specifically, I want to provide a systematic comparison of the ability of LiDAR data for predicting animal species richness across taxa, mapping tree species specific successional stages and modeling seasonal leaf area index. With three different case studies varying by scale, data and ecological process, I want to deepen the understanding of the utility of LiDAR data at various scales for the mapping of biodiversity in the context of different data sources. The three case studies are intended to illuminate the following aspects:

1. Understanding the relation of vegetation structure in mountainous areas using LiDAR-derived vegetation structure as a proxy for animal species richness across multiple taxa,
2. Leveraging heterogeneous LiDAR data for monitoring of biodiversity processes in forest ecosystems,
3. Understanding the potential of scattered spaceborne LiDAR footprints as provided by the GEDI mission, for a comprehensive and seasonally differentiated assessment of leaf area index changes and evaluating their potential for time periods outside the acquisition window.

1.3 Case studies

To approach these aims, I present three biodiversity studies in this thesis, focusing on different ecological questions. Airborne as well as spaceborne LiDAR data sets with different spatial and temporal resolutions were used to emphasize different aspects of the utility of LiDAR data in biodiversity research.

1.3.1 Case study 1: Comparison of multi-taxa animal species richness at Mount Kilimanjaro predicted from airborne LiDAR data and elevational information

As the monitoring of animal species and functional diversity is of increasing relevance for the development of strategies for the conservation and management of biodiversity, the opportunities of LiDAR data in the observation for modeling occurrences of animal species were tested in this case study. Reliable estimates of the performance of monitoring techniques across taxa are important to provide a base for systematic comparison. Using a unique data set, this study investigated the potential of airborne LiDAR-derived variables characterizing vegetation structure as a predictor for animal species richness in a mountainous study area. To disentangle the structural LiDAR information from co-factors related to elevational vegetation zones, LiDAR-based models were compared to the predictive power of elevation models. 17 taxa and four feeding guilds were modeled and the standardized study design allowed for a systematic comparison across the assemblages.

1.3.2 Case study 2: Using heterogeneous LiDAR data for the modeling of tree species group specific successional stages

Monitoring the forest is one key element for actions against biodiversity loss. The component of species composition is satisfactorily distinguishable using spectral remote sensing. The important element of classifying forest succession for those tree species, however, is still a challenge but quite relevant for their habitat quality. Both factors collectively shape the biodiversity within forests and thereby influence the ecosystem services that forests provide. Remote sensing techniques offer promising solutions for obtaining area-wide information on tree species composition and their successional stages. As successional stages are closely linked to vegetation structure, LiDAR data are especially prone to providing additional information for this part of biodiversity research. While optical data are freely available in appropriate quality, acquiring LiDAR data, which provide valuable information about forest structure, present challenges due to its high financial and labor cost. Nonetheless, LiDAR data are often collected by public authorities but are typically recorded across several years and therefore are heterogeneous in quality. This poses challenges when attempting to model tree species and their successional stages. In this case study, it was tested if and how a heterogeneous data set from a governmental flight campaign obtained over several years can improve modeling tree species group specific successional stages of forest vegetation. Different combinations of multi-spectral optical satellite data (Sentinel-2) and heterogeneous airborne LiDAR data, collected by a federal government, were utilized to model successional stages of seven tree species groups (douglas fir, larch, pine, spruce, beech, oak, other deciduous trees) in Rhineland-Palatinate, Germany. Ground truth information was derived from forest inventory plots. Random forest models with spatial variable selection and spatial cross-validation were used to model and map successional stages of tree species groups at regional level.

1.3.3 Case study 3: Integration of Sentinel and GEDI data for seasonal explicit modeling of the leaf area index

The leaf area index is defined as the ratio of leaf area to ground area and it is solely calculated from structural features. It is an important metric for several fields in ecology. As field measurements are time-consuming and do not scale to landscapes, it depends on model-based air- or spaceborne remote sensing methods to document larger areas. As flight campaigns are expensive though, repetition rates are low and only situational snapshots are available for larger areas. As leaf area index massively changes across seasons, it is necessary to increase repetitions. Satellite-based remote sensing seems like a possible solution and with the new spaceborne GEDI LiDAR sensor new opportunities arise, but in the technical trade-off for temporal resolution, spatial resolution must be lowered immensely. As of 2019, NASA's GEDI mission delivered a point-based leaf area index product with 25 m footprints and periodical repetition. This opens up new possibilities in integrating GEDI as frequently generated training samples with high resolution (spectral) sensors. However, as the GEDI mission only has a limited lifespan, it was tested how temporally limited data sets can be used for ongoing monitoring approaches. The landscape of Hesse, Germany, with its pronounced seasonal changes was investigated. Assuming a relationship between GEDI's plant area index and Sentinel-1 and -2 data, a random forest approach together with spatial variable selection was used to make predictions for new Sentinel scenes. The model was trained with two years of GEDI's plant area index data and validated against a third year to provide a robust and temporally independent model validation. This ensured the applicability of the validation for years outside the training period.

Chapter 2

Potential of Airborne LiDAR Derived Vegetation Structure for the Prediction of Animal Species Richness at Mount Kilimanjaro

This chapter has been published as:

Ziegler, A., Meyer, H., Otte, I., Peters, M. K., Appelhans, T., Behler, C., Böhning-Gaese, K., Classen, A., Detsch, F., Deckert, J., Eardley, C. D., Ferger, S. W., Fischer, M., Gebert, F., Haas, M., Helbig-Bonitz, M., Hemp, A., Hemp, C., Kakengi, V., Mayr, A. V., Ngereza, C., Reudenbach, C., Röder, J., Rutten, G., Schellenberger Costa, D., Schleuning, M., Ssymank, A., Steffan-Dewenter, I., Tardanico, J., Tschapka, M., Vollstädt, M. G. R., Wöllauer, S., Zhang, J., Brandl, R. and Nauss, T. **(2022)**, 'Potential of Airborne LiDAR Derived Vegetation Structure for the Prediction of Animal Species Richness at Mount Kilimanjaro', *Remote Sensing* 14(3), 786. Remote Sens. 2022, 14, 786. https://doi.org/10.3390/rs14030786.

Abstract
The monitoring of species and functional diversity is of increasing relevance for the development of strategies for the conservation and management of biodiversity. Therefore, reliable estimates of the performance of monitoring techniques across taxa become important. Using a unique dataset, this study investigates the potential of airborne LiDAR-derived variables characterizing vegetation structure as predictors for animal species richness at the southern slopes of Mount Kilimanjaro. To disentangle the structural LiDAR information from co-factors related to elevational vegetation zones, LiDAR-based models were compared to the predictive power of elevation models. 17 taxa and 4 feeding guilds were modeled and the standardized study design allowed for a comparison across the assemblages. Results show that most taxa (14) and feeding guilds (3) can be predicted best by elevation with normalized RMSE values but only for three of those taxa and two of those feeding guilds the difference to other models is significant. Generally, modeling performances between different models vary only slightly for each assemblage. For the remaining, structural information at most showed little additional contribution to the performance. In summary, LiDAR observations can be used for animal species prediction. However, the effort and cost of aerial surveys are not always in proportion with the prediction quality, especially when the species distribution follows zonal patterns, elevation information yields similar results.

2.1 Introduction

We are facing a decrease of global biodiversity (Ceballos et al., 2017; Pereira et al., 2012) and the rate of this loss is accelerating with ongoing climate change (Warren et al., 2018) as well as the rapid transformation of natural habitats by human landuse (Newbold et al., 2015). To mitigate the effects of this biodiversity loss on the functionality of ecosystems (Hooper et al., 2005; Loreau et al., 2001; Soliveres et al., 2016), monitoring of species and functional diversity is an important prerequisite for focused management strategies (Nagendra et al., 2013; Noss, 1990; Wiens, 2009). To facilitate a unified monitoring system, a set of essential biodiversity variables (EBVs) were developed during the last years (e.g. Haase et al., 2018; Jetz et al., 2019; Pereira et al., 2013). However, gathering those variables during field campaigns is only possible in a number of limited situations, as area-wide coverage is unfeasible due to high costs as well as the lack of experts. This is particularly true for surveys across large areas with a steep elevation gradient, as complex terrain hinders accessibility (Müller & Brandl, 2009).

Facing this challenge of comprehensive mappings of EBVs (Kissling et al., 2018; Proença et al., 2017), Pettorelli et al. (2016) proposed a subset of the EBVs that could potentially be surveyed using satellite remote sensing. Some of these remote sensing based EBVs already meet reasonable quality requirements (e.g. land cover, leaf area index or phenology), others like species occurrences or taxonomic diversity require further research. The retrieval of the latter is often based on correlations between land-cover properties and information on taxonomic richness from field studies that is used to build remote sensing models (Froidevaux et al., 2016; Macarthur & Macarthur, 1961; Martins et al., 2017; Melin et al., 2018; Rocchini et al., 2010). Multispectral or hyperspectral retrievals primarily rely on compositional information (e.g. Asner and Martin, 2016; Mairota et al., 2015; Nagendra and Rocchini, 2008; Wallis et al., 2017, reviewed in Pettorelli et al., 2014). LiDAR retrievals utilize information on the vertical structure of the ecosystem (e.g. Lefsky et al., 2002; Müller & Vierling, 2014) and seem to be particularly successful for flying organisms (Clawges et al., 2008; Flashpohler et al., 2010; Goetz et al., 2007; Jung et al., 2012; Lesak et al., 2011; Müller et al., 2009, 2010; Rechsteiner et al., 2017; Vogeler et al., 2014; Zellweger et al., 2013), although the biological background of this observation is poorly understood.

To advance the application of species richness related EBVs from remote sensing, performance must be compared across many taxa. While meta-analyses across different case studies allow some conclusions, the individual study design, different computations of error estimates and the uniqueness of the study regions make it difficult to actually compare the results regarding the model performance for different taxa. This becomes even more challenging if the study has been conducted in mountainous terrain, which is common for ecological space-for-time approaches as global biodiversity hotspots often tend to be in mountainous areas (Myers et al., 2000). Here, elevational change of the land cover in combination with a fixed amount of work force generally limits the number of ground samples that can be collected within one land cover zone. These individual limitations in training and testing datasets lead to a variety of testing approaches with varying degrees of reliability and comparability of error estimates.

This study analyzes the predictive performance of airborne LiDAR-derived variables for mapping the species richness of 17 taxonomic groups from four feeding guilds in a comprehensive manner. The study area is located at the southern slopes of Kilimanjaro (figure 2.1) and field observations stretch from an elevation of 800 to 4400 metres. Since the taxonomic assemblages cannot be directly observed using LiDAR, vegetation structure is used as a surrogate for species richness. It is assumed that the taxa or the aggregated feeding guilds can be predicted differently well by LiDAR data. For example, the "plant diversity hypothesis", links consumer richness (especially herbivores) to plant diversity (Basset et al., 2012; Novotny et al., 2006; Peters et al., 2016). Therefore, it is expected, that the performances of structural models decline from herbivores to predators (Müller & Brandl, 2009; Vehmas et al., 2009), as the distance of their position in the food chain to plants increases. Furthermore, as structural properties, at scales accessible for remote sensing, tend to be more relevant for animals with larger body-sizes (Borthagaray et al., 2012; Morse et al., 1985; Siemann et al., 1999; Stanska & Stanski, 2017), it is expected that the performance of species richness models incresase with increasing body size. Similar consideration of correlation between structure and species traits apply for flying taxa that perceive the landscape with a grain not as detailed as walking taxa (environmental grain hypothesis, eg. Kaspari and Weiser (1999) and Sarty et al. (2006)). The unique dataset also allows to critically evaluate whether LiDAR-derived information brings any gain

at all compared to models that rely solely on the known decrease of species richness with elevation or elevation correlated environmental properties (Peters et al., 2016). This study investigates the potential of LiDAR-derived variables characterizing vegetation structure as predictors for animal species richness at the southern slopes of Mount Kilimanjaro. To disentangle the structural LiDAR information from co-factors related to elevational vegetation zones, LiDAR-based models were compared to the predictive power of elevation models.

2.2 Materials and Methods

2.2.1 Study Area and Sampling Design

The study area are the southern slopes of Mount Kilimanjaro. In the framework of a comprehensive research unit, study plots of 50 m x 50 m were established across 12 land cover zones with five replica plots per zone. The 60 plots describe two ecological gradients: an elevational gradient from 800 m a.s.l. to 4400 m a.s.l. and a disturbance gradient from (near) natural to anthropogenic land-cover types within elevation zones (figure 2.1 and table 2.A1).

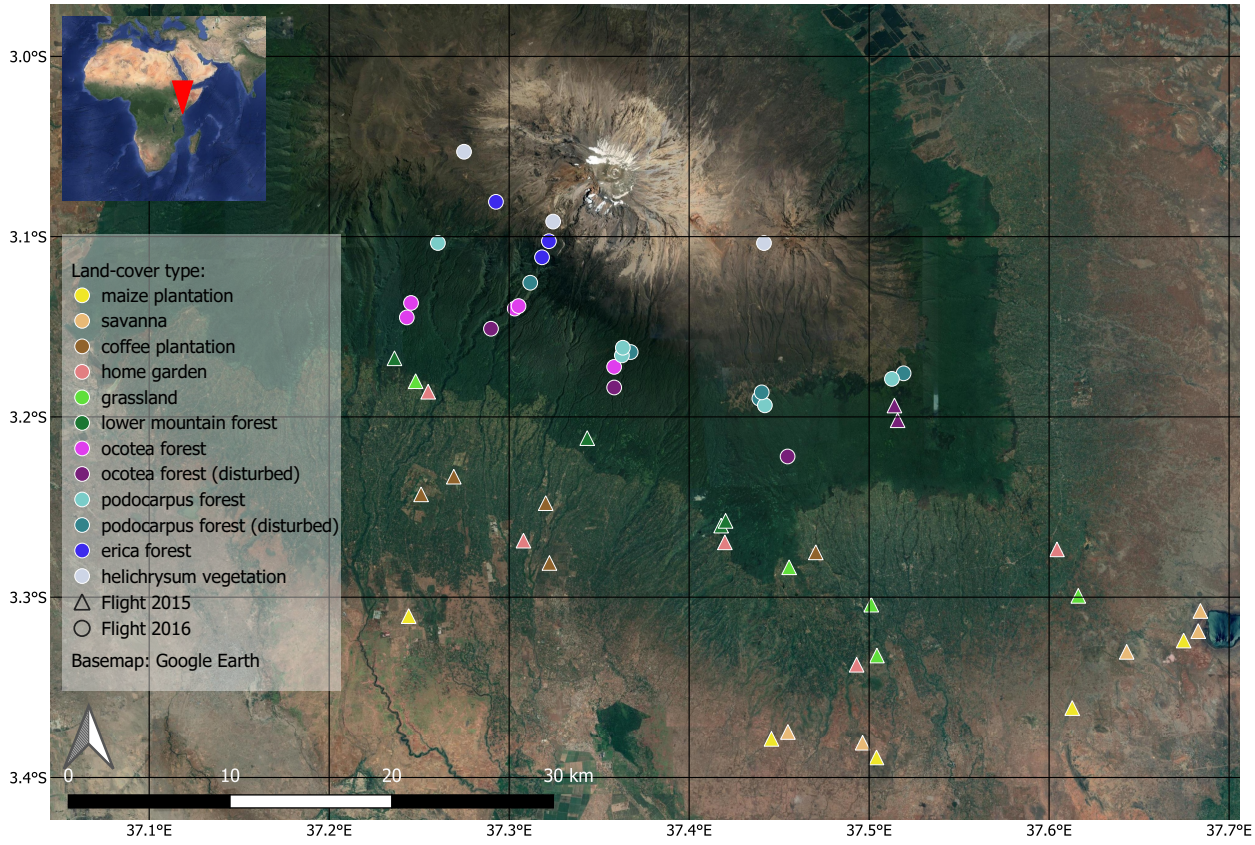


Figure 2.1: Study area with sampling plots. Colors of symbols show different land covers, shapes show the different flight missions from 2015 and 2016. The background image indicates the large-scale vegetation zones along the elevational gradient (background: Google Maps (2020)).

2.2.2 Data and preprocessing

Diversity Data

On 59 of the 60 study plots, data for estimating the diversity of 17 taxa was available (see table 2.3). Sampling followed standardized approaches as described in detail in Peters et al. (2016) and Peters et al.

(2019). For an overview of the taxa and sampling methods see table 2.A2. The species richness data was aggregated (sum of species) to four feeding guilds. See table 2.A3 and 2.A4 for the proportional allocation of species per taxon to feeding guild. As true bugs, spiders and springtails were not identified to species level, the entire group of spiders was assumed to be predators, springtails count as decomposers and true bugs were ignored when estimating species richness of feeding guilds.

LiDAR Data

The LiDAR data set was acquired during two missions with a Riegl LMS-Q780 sensor carried by an Airbus Helicopter at an altitude between 850 m and 1750 m above ground level. The northern (high land) plots were sampled in March 2015 and the southern (low land) plots in November 2016 (figure 2.1). The temporal distance was assumed to be negligible since both acquisition dates fall into the early rainy season and plots of the same land-cover type are covered within the same flight campaign except for disturbed ocotea forests. The LiDAR pulses contain between one and seven returns with a vertical accuracy of 0.15 m and a horizontal accuracy of 0.20 m (95 % confidence interval). The mean point density is 34 points per square meter but varies due to terrain and flight conditions. Outliers were removed and points were classified into ground and non-ground following the standard procedure using the LAStools preprocessing software (Isenburg, n.d.). Rasterized LiDAR layers (e.g. digital terrain model (DTM), digital surface model (DSM), canopy height model (CHM)) were generated by the open source remote sensing data base (RSDB) at a resolution of 1 m (Wöllauer et al., 2020). The resulting DTM preserves fine details in regions with high ground point densities and plausible elevation estimates in regions with low ground point densities (e.g. dense forest).
To derive a set of potential predictor variables from the LiDAR observations, several indices, which characterize structural properties of the 50 m x 50 m areas, were computed for each of the plots using the RSDB (see table 2.1) (Wöllauer et al., 2020). The compiled 97 LiDAR metrics used included e.g. canopy height metrics (maximum, standard deviation, median, quartiles, etc.), return number metrics (maximum, standard deviation of different layers, etc.) and ecological estimates (leaf area index, above-ground biomass and gap fraction, etc., for the complete list of variables see table 2.1).
The land covers in the study area can be grouped into forest and non-forest (see figure 2.1 and table 2.A1 for details). Due to their complex multi-layered structure, forested plots appear considerably different to non-forested plots hence the sets of LiDAR variables for these two types differ slightly. For the current study this means that on non-forested plots, variables describing vegetation layers reached a maximum of 8 m height, and on forest plots, vegetation height reached a maximum of 29 m (indicated in table 2.1). The two thresholds correspond to variables where at least 50% of the plots had vegetation in this height. In the following, all modeling approaches were always carried out for forested and non-forested plots separately, to account for these fundamental differences.

2.2.3 Predictive modeling of diversity

The computations and analyses in this study were performed using the R environment 3.5 in conjunction with the caret package (Kuhn, 2018; R Core Team, 2018). Partial least squares regression (PLSR) is useful for models in data settings with a smaller number of observations relative to the number of predictor variables. It can also handle multicolinearity, a situation unavoidable, when using LiDAR-derived variables (Carrascal et al., 2009; Laurin et al., 2014; Luo et al., 2017; Næsset et al., 2005). To reduce the impact of overfitting caused by correlated variables, a forward feature selection (FFS) implemented in the CAST package (Meyer, 2018) was used, which ensures a more stable variable selection than recursive feature elimination approaches (Meyer et al., 2018).
To distinguish between effects on species richness predictability based on a pure elevation gradient versus habitat structure, three different model groups (elevation and its square only, structural metrics only and structural metrics to predict residuals of elevation based model) were established. Then, within each of these three groups of models, an individual prediction model was separately built for each taxon and feeding guild for forested and non-forested areas. The same combinations of plots for training and testing were used across all models (table 2.2 for overview).

Table 2.1: Overview of structural variables characterizing the vegetation and a description on their calculation. Most Indices were calculated on the LiDAR point cloud of each plot (50 m x 50 m), only a few were calculated on 1 m x 1 m cells of the canopy height model (marked: based on CHM).

Name	**Explanation**
canopy	
maximum canopy height (CH)	Maximum canopy height
mean CH	Mean canopy height
median CH	Median canopy height
percentile of CH	10% Percentile of canopy heights
... (in 10% steps)	x% Percentile of canopy heights
standard deviation CH	Standard Deviation of canopy height
skewness CH	Skewness of canopy height Distribution
variance CH	Variance of canopy height
curtosis CH	Excess Kurtosis of canopy height Distribution
coefficient of variation CH	Coefficient of Variation of canopy height
area ratio	area ratio of raster pixels (based on CHM) (Jenness, 2004)
vegetation structure	
Return Density (RD) of different layers	Return density of 1 meter layer
... (1 meter steps up to 8 m/29 m)	Return density of x meter layer
RD canopy ($>$ 5 m)	Return density of canopy vegetation layer
RD regeneration (2 - 5 m)	Return density of regeneration vegetation layer
RD understory ($<$ 2 m)	Return density of understory vegetation layer
RD ground	Return density of ground layer
penetration rate (PR) of layers	Penetration rate of 1 meter layer
... (1 meter steps up to 8 m/29 m)	Penetration rate of x meter layer
PR canopy ($>$ 5 m)	Penetration rate of canopy vegetation layer
PR regeneration (2 - 5 m)	Penetration rate of regeneration vegetation layer
PR understory ($<$ 2 m)	Penetration rate of understory vegetation layer
maximum returns	highest number of return points per LiDAR laser pulse
mean returns	mean of return points per LiDAR laser pulse
standard deviation returns	standard deviation of return points per LiDAR laser pulse
median returns	median of return points per LiDAR laser pulse
standard deviation first return	standard deviation of first return points per plot
vegetation	
above ground biomass	Aboveground Biomass (6.85 * $TCH^{0.952}$) (top-of-canopy height (TCH) based on CHM) (Getzin et al., 2017)
foliage height diversity	Foliage height diversity (Shannon Index grouped by layers of understory, regeneration and canopy) (Macarthur & Macarthur, 1961)
leaf area index	Leaf-area index, with k=0.3, h.bin=1, GR.threshold=5 (based on CHM) (Getzin et al., 2017)
vegetation coverage (VC) of different layers	vegetation coverage in 1 meter height (based on CHM)
... (for 2,5, and 10 meter)	vegetation coverage in x meter height (based on CHM)
gap fraction	fraction of clear area above 10 m compared to whole area (to be detected as gap: minimum of 9 cells, based on CHM)

Table 2.2: Overview of the different models calculated in this study.

model name	independent variables (predictors)	target variables (response)
elevation	elevation, elevation^2	species richness
structure	structural LiDAR variables	species richness
residuals	structural LiDAR variables	residuals from elevation model
combination	sum of elevation model and structural model	species richness

In the first group of models, only the elevation and its square were used to predict species richness ("elevation model"). In the second group of models, only the structural metrics derived from LiDAR (no elevation) were considered ("structure model"). In the third group of models, the same structural metrics were used to predict the residuals of the elevation model ("residual model"). Hence the predictions of the residual model do not represent the complete species richness, but only that part which cannot be explained by the elevation model. Therefore, to be able to compare the results of this model with the elevation and structure model, the results of the residual model were added to the elevation model ("combination model"). Even though it is not a separate model in a strict sense, but only the sum of the elevation model and the residual model, this mixed approach will be called "combination model" in the following. The pure residual model, on the other hand, can be used to compare the plain structure dependence of taxa and feeding guilds without effects of elevation. Prediction results from forested and non-forested areas were assembled to one error estimation per response variable, to compare the general model performance of taxa and feeding guilds for the whole study area.
To test if the model performance depends on species traits, the correlation of model performance of each taxa and feeding guild to the respective body size and the mode of movement were tested. For body size the Spearman rank correlation coefficient was calculated. Groups were sorted by body size from large (large mammals with up to 1.7 m length) to small groups (parasitoid wasps with only a few millimeters). For the test between model performance and flying/non-flying groups of organisms the Mann and Whitney U-Test was performed.

Validation strategy and model tuning

Due to the limited number of observation samples per taxonomic group, choosing an appropriate tuning and testing strategy of the various models was of major importance. As illustrated in figure 2.2, model training and testing consisted of two separate cross validation cycles. The outer 20-fold-cross validation withholds one random plot of each land-cover type in every resample. Those samples were held back from model training to qualify them for estimating the model performance for new locations in the study region. This repeated approach allows for more stable validation results given the limited number of plots. The inner cross validation was embedded within the PLSR machine learning approach. It uses the same method of leaving one plot per land-cover type out in each resample. The inner cross validation was used for model tuning and variable selection only. Tuning affected the number of principal components used in the PLSR and varied between one and two. Feature selection was implemented according to Meyer et al. (2018).
For quantifying the predictive performance, the root mean squared error (RMSE) was computed for each fold of the outer cross validation. Previously, the results from forested and non-forested areas were combined, to be able to make a general statement per taxon and feeding guild. Since species richness varied considerably across the taxonomic groups, the RMSE of each group was normalized with the standard deviation of the species richness per group of the plots used in each model.

2.3 Results

The elevation model performs best for 14 out of 17 taxa (figure 2.3a and table 2.3). On average, the RMSE/sd values for these 14 taxa are 0.21 lower than in the structural and 0.23 lower than in the combined model. The structural model performs better only for parasitoid wasps and the combined model is the best for aculeate wasps and insectivorous bats.

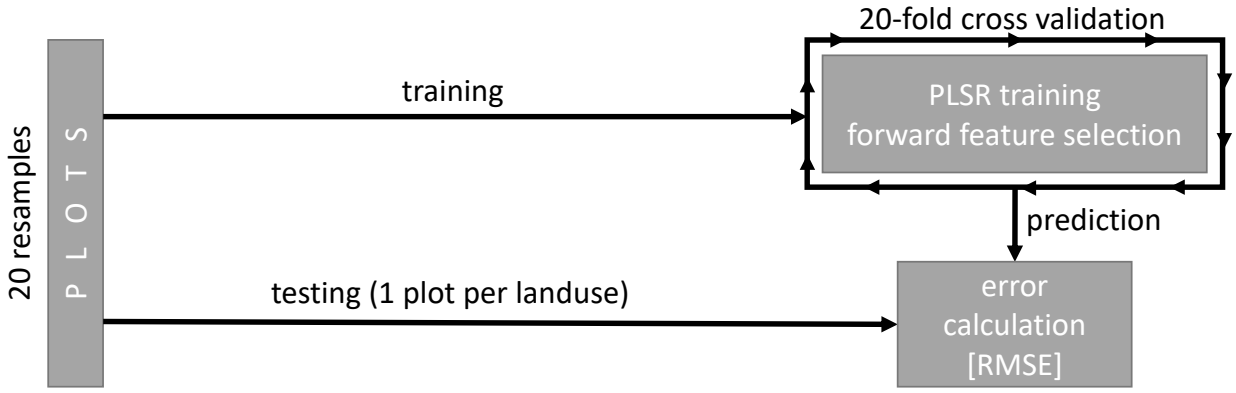
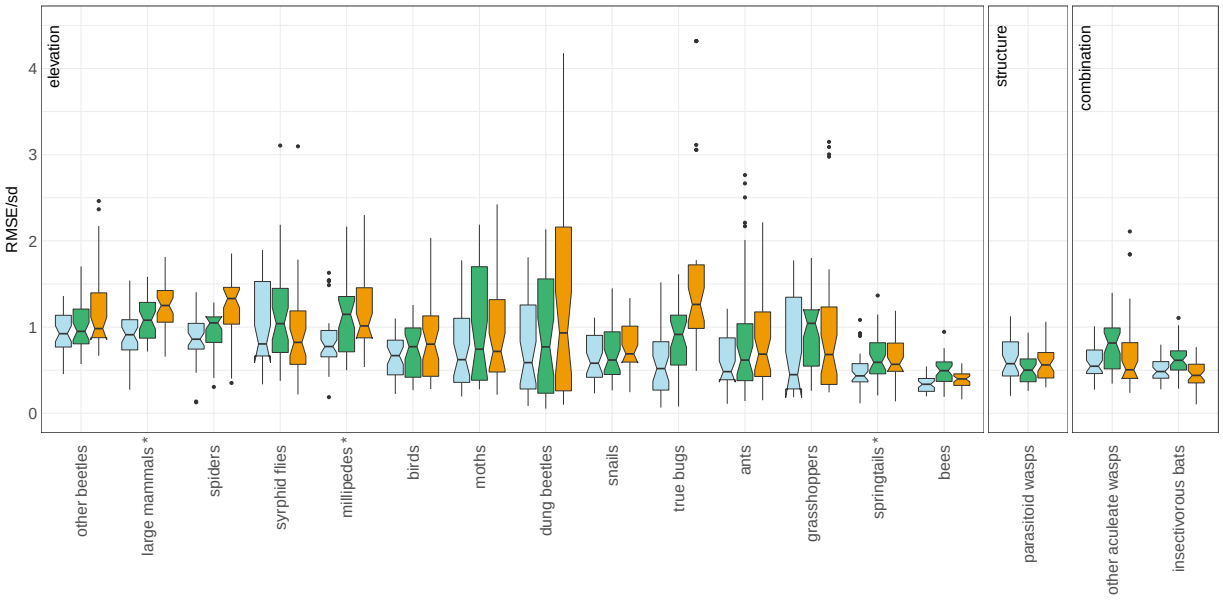


Figure 2.2: The model training (upper right loop) uses a partial least squares regression (PLSR) and a forward feature selection with a 20-fold cross validation. Validation is carried out by predicting the values of the testing plots. The division of testing and training plots (outer loop) follows a repeated stratified sampling approach, with randomly chosen resamples of one plot per land cover for the testing, leaving the rest of the plots for training. Validation is based on the median root mean square error (RMSE) of the individual resamples, normalized by the standard deviation of these RMSE values.

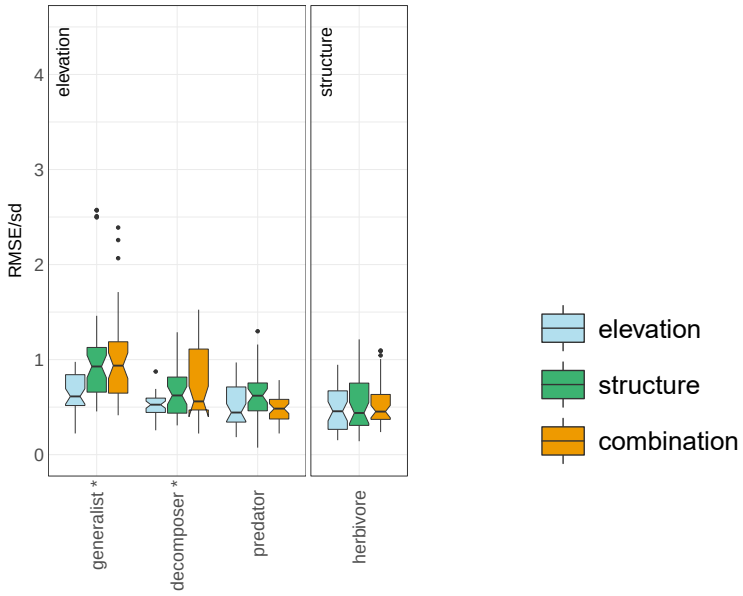
For all three model types, the interquartile range (IQR) of large mammals, springtails, bees, parasitoid wasps and insectivorous bats is rather small, while syrphid flies, moths, dung beetles and grasshoppers show large variations of RMSE/sd values. Only large mammals, millipedes and springtails show a siginificant superior model performance for the best model (here elevation) compared to both other models (Tukey test). For an individual taxon, a median performance of the RMSE/sd of half the standard deviation or better is only reached for ants, grasshoppers, springtails, bees, parasitoid wasps, other aculeate wasps and insectivorous bats. Bees reach the best model results across all taxa and model types with an RMSE/sd of 0.34 (elevation model).
Regarding feeding guilds, the elevation model performs best for generalists, decomposers and predators with an RMSE/sd value that is 0.20 and 0.13 lower on average than the structural and the combined model respectively (figure 2.3b and table 2.3). Only for generalists and decomposers the best model (elevation) was performing significantly better than the other two. The structural model performs best for herbivores but only with slight differences in the RMSE/sd to the combined (0.01) and the elevation model (0.02).
To explore the potential of modeling species richness outside the gradient of Mount Kilimanjaro, figure 2.4 shows a comparison of the plain residual models (RMSE/sd). These results are independent of the elevation and do not model species richness, but the residuals of the elevation model. Therefore it is possible to compare the ranking of species performance as it would be suspected if it was only dependent on structure without an superimposing elevational gradient. Taxa and feeding guilds are sorted by their median error estimates which range between 1.1 and 2.5 (table 2.3). Smaller values here mean a closer relationship to structural metrics. Value ranges of model performances within each group lie within the same magnitude, except for dung beetles which show a high variation. The RMSE/sd of the residual model shows a ranking of the feeding guilds from predators (1.2), over herbivores (1.3), to decomposers (1.8) and generalists (2.2). The analysis of the best subsets of prediction variables, could not identify regular patterns (figures 2.A1 to 2.A4).
There is no statistical relationship between model performance and body size of the assemblages (table 2.4). However, there is a difference for the combined and residual model with a better performance for the flying than for the non-flying taxonomic groups (see table 2.4).

2.4 Discussion

The study evaluated the potential of LiDAR data to predict species richness at Mount Kilimanjaro. The influence of the respective effects of elevation and vegetation structure on species richness were investigated



(a)



(b)

Figure 2.3: Modeling performances for each taxon (a) and feeding guild (b) in terms of the root mean square error normalized by standard deviation (RMSE/sd). Smaller values show a better model performance. Colors represent the different model types. Taxa are grouped into "elevation", "structure" and "combination" depending on which of the three models shows the best median RMSE/sd. Stars indicate if the best model is significantly better than both of the other models. Within the groups, taxa and feeding guilds are sorted by descending RMSE/sd. The boxes include the median and the inter quartile range (IQR) with notches indicating roughly the 95% confidence interval. Whiskers are extending to ± 1.5 times the IQR and points indicate single error values outside of this range.

Table 2.3: Mean ± standard deviation across all plots of observed species richness per plot and median of the root mean square error normalized by standard deviation of field samples per taxon and feeding guild ± the standard deviation of model runs. Lower RMSE/sd values mean a better model performance.

taxon/ feeding guild	species richness per plot	elevation model	structure model	combined model	residual model
	mean ± standard deviation	median RMSE/sd ± standard deviation			
ants	2.7 ± 3.4	0.48 ±0.29	0.62±0.72	0.69 ±0.57	1.5 ±1.3
bees	6.7 ± 5.9	0.34 ±0.1	0.49±0.17	0.4 ±0.11	1.5 ±0.4
birds	16 ± 8.6	0.67 ±0.25	0.77±0.31	0.8 ±0.45	1.9 ±1
dung beetles	5.2 ± 7.7	0.59 ±0.5	0.77±0.73	0.93 ±1.3	1.6 ±2.2
grasshoppers (locusts, crickets)	10 ± 14	0.45 ±0.54	1 ±0.41	0.68 ±0.82	1.1 ±1.3
insectivorous bats	5.5 ± 3.3	0.48 ±0.13	0.62±0.18	0.44 ±0.14	1.4 ±0.46
large mammals	2.3 ± 1.8	0.91 ±0.28	1.1 ±0.24	1.3 ±0.26	2.2 ±0.45
millipedes	1.2 ± 1.7	0.77 ±0.34	1.1 ±0.48	1 ±0.43	1.5 ±0.65
moths	9.6 ± 11	0.62 ±0.48	0.74±0.69	0.72 ±0.61	1.2 ±1
other aculeate wasps	3.2 ± 4	0.55 ±0.21	0.82±0.26	0.5 ±0.45	1.2 ±1.1
other beetles	8.6 ± 4.9	0.92 ±0.24	0.95±0.27	0.98 ±0.46	1.7 ±0.81
parasitoid wasps	16 ± 14	0.58 ±0.26	0.5 ±0.18	0.56 ±0.19	1.1 ±0.38
snails (slugs)	6.8 ± 5.6	0.58 ±0.26	0.62±0.31	0.69 ±0.28	1.5 ±0.61
spiders	5 ± 2.2	0.86 ±0.26	1 ±0.23	1.3 ±0.39	2.5 ±0.72
springtails	4.6 ± 2.4	0.43 ±0.2	0.59±0.27	0.57 ±0.25	1.4 ±0.63
syrphid flies	3.5 ± 3.3	0.8 ±0.47	1 ±0.55	0.82 ±0.5	1.2 ±0.74
true bugs	2.1 ± 2.3	0.57 ±0.39	0.91±0.37	1.3 ±1.1	2.4 ±2.2
decomposer	12 ± 7.8	0.53 ±0.13	0.62±0.25	0.56 ±0.38	1.8 ±1.2
generalist	6.8 ± 3.8	0.61 ±0.21	0.93±0.57	0.94 ±0.48	2.2 ±1.2
herbivore	40 ± 30	0.46 ±0.23	0.44±0.3	0.45 ±0.26	1.3 ±0.77
predator	41 ± 22	0.44 ±0.22	0.62±0.23	0.48 ±0.15	1.2 ±0.37

Table 2.4: Results of rank tests comparing the performances of models measured by RMSE/sd (as shown in figure 2.3 and figure 2.4) with respect to selected traits (body size, mode of movement) of the taxa. For the tests between the performance of the models and body size the Spearman rank correlation coefficient (r) was used. Body size was sorted from large to small groups. For the test between flying and non-flying groups of organisms the Mann and Whitney U-Test was used. Significant results, in terms of the p-value are marked bold.

	Body size		mode of movement	
	r	p	p	
RMSE/sd elevation	-0.14	0.58	0.28	
RMSE/sd structure	-0.13	0.62	0.28	
RMSE/sd combination	-0.10	0.70	**0.011**	**(flying <non-flying)**
RMSE/sd residuals	-0.22	0.39	**0.006**	**(flying <non-flying)**

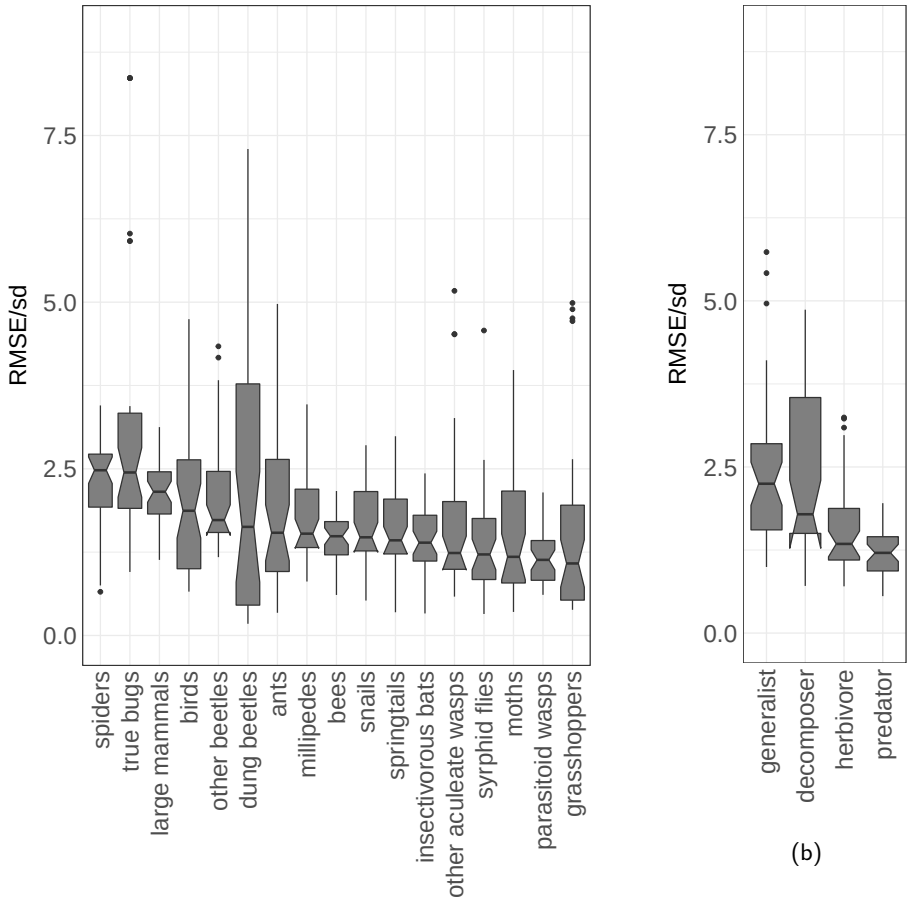


Figure 2.4: Modeling performance for the residuals of the elevation model for each taxon (a) or feeding guild (b) as root mean square error normalized by standard deviation [RMSE/sd]. Taxa (a) and feeding guilds (b) are sorted by increasing median modeling performance and therefore increasing influence of vegetation structure on the target variable (which means decreasing median RMSE/sd). For the description of plot elements see figure 2.3.

by comparing the model performances of models that used elevation as the only predictor, models that used LiDAR variables only and models that used LiDAR variables to predict the residuals of the elevation models.
Generally, performances of the different models varied only slightly within each taxon, with no significant difference of the best performing model to both other models, except for three taxa (large mammals, millipedes, springtails). However, there is a trend, indicating that the elevation model performs best for 14 out of 17 taxa. All taxa which do perform significantly better with one specific model, belong to that group. In the cases where the structure or combined model performs best, the performances differ only marginally and differences are not significant.
As expected, results of the model performances of feeding guilds indicate, that herbivores are influenced more by structure than generalists and decomposers (figure 2.4b). However, considering the feeding guilds, generalists and decomposers are the only groups for which the best of the three models (elevation model) is significantly better than the other two. It is ecologically reasonable, that generalists, which obviously use a wide variety of food, are, at least for feeding reasons, not specifically connected to the vegetation structure. Opposed to that, for herbivores structure was suspected to be the most relevant predictor as they rely solely on vegetation and therefore structure should influence feeding patterns and the occurrence and diversity of species. Even though performances improve slightly with the structure model, differences were not significant. Decomposers rely on the existence of organic material. Still, as long as the supply of organic material is given, it seems reasonable that other environmental factors, which are linked to elevation, would have a greater impact. In conclusion, prediction results for the feeding guilds show a tendency of the hypothesized correlation between a lower feeding guild and a higher dependency on the structure. However, these differences are very small and therefore not convincing. The model performances of the feeding guilds are generally comparable with the ones of the individual taxa. Nevertheless, as more field samples are included in the feeding guilds, sampling uncertainties are partly leveled out by the higher number of sampled individuals.
This study further aimed at comparing the general potential of LiDAR-derived variables for the prediction of the structurally dependent proportion of species richness for different taxa and feeding guilds. For this comparison the elevation corrected residual model provides the relevant information (figure 2.4). In line with the discussion about the best model type, generalists and decomposers seem to be the group not tightly connected to habitat structure, whereas herbivores seem to depend more on vegetation structure 2.3b. Along with the other models, there is no notable difference in the overall performances between taxa and feeding guilds for the residual models.
A comparison between the model types allows for drawing conclusions about the influence of elevation and structure as relevant predictors for biodiversity. In their study at Mount Kilimanjaro about diversity gradients at different levels of aggregation, Peters et al. (2016) already showed that mean annual temperature is the most important variable to predict animal species richness in the region. In the present study, some taxa are significantly more influenced by elevation than by properties of the structure itself, but generally, median performances between models differ only slightly. The residual model attempts to illuminate patterns within the remaining structural properties that are not attributable to the strong gradient in the study area. However, only samples from four to five replica plots per land-cover type have been available for model training which limits the performance. This might promote over-fitted models for structural properties leading to larger prediction errors when applied to unknown locations, whereas the elevation model is able to find a general pattern within all plots as they are well distributed along the elevation gradient (figure 2.1). Still, even a slightly worse structural or combined model compared to the elevation model validates the general usability of LiDAR data for predicting species richness, even though the effort of LiDAR missions then seems questionable. In the variable selection of the LiDAR metrics, no patterns emerge (figures 2.A1 to 2.A4). Neither individual variables nor variable groups (table 2.1) appear in clear patterns across models. The LiDAR variables were calculated for individual plots (50m x 50 m). For some taxa it might be beneficial to account for the structure of a larger spatial environment. Therefore, in future studies, it could be tested whether variable cell sizes of the LiDAR metrics can improve prediction models.
The hypothesized positive correlation between body size and the modeling performance is not supported by the data. However, the mode of movement significantly correlates with the prediction performance in

the combined and the residual models. Especially in the residual model, flying taxa outperform the others. The 6 taxa with the smallest median error are species with the ability to fly. Only the flying taxa bees and birds (Rank 9 and 14 out of 17) lie within the worse performing half of taxa, showing the generally poorer performance of non-flying taxa. The comparably poor performance of predicting birds with structural metrics alone is rather surprising, as birds are the most studied taxonomic group in species-habitat structure relationships (Acebes et al., 2021) and there are many studies that demonstrate promising correlations of bird diversity and different structural features (e.g. Müller et al. (2009), Smart et al. (2012), and Vogeler et al. (2013) or see the detailed review of Davies and Asner (2014)).

The results of this study are based on 59 plots. Even though the total number of study sites seem to be sufficient for modeling purposes, compared to similar studies, the different land-cover types that follow the elevation gradient, in addition to the necessary division into forest and non-forest areas for modeling, limit the number of repetitions. Hence, model building has to be carefully adjusted to the limited number of plots. The possibility of also using land-cover type as a categorical predictor variable was discarded due to the low number of replicates. However, land cover is indirectly included in the model by the natural orientation of land cover along the elevational gradient at Mt. Kilimanjaro.

In general, it is not easy to evaluate the results in the context of other studies, since a comparison of the results can only provide indications of the success of the modeling. This is because the studies were conducted in different landscapes, for different taxa, but most importantly, with different measures of biodiversity. Species richness, beta diversity, and other metrics are related but not identical. In previous studies, the role of elevation is handled in different ways. The studies of Müller and Brandl (2009) and Vierling et al. (2011) for example analyze the influence of LiDAR-derived variables compared to other abiotic and biotic variables for the prediction of spider species distribution and forest beetle assemblages. Results show a comparable or even much better performance of LiDAR variables to ground based measures (Vierling et al., 2011). As the variable elevation is a by-product of LiDAR point clouds, these studies included elevation in the group of LiDAR-derived variables, with elevation being a rather important variable. However, elevation changes within the study area are limited to about 800 meters with a rather homogeneous forest cover.

The studies of Zellweger et al. (2013) and Rechsteiner et al. (2017) are situated in a more mountainous terrain (>1200 m elevation difference), however, the LiDAR derived variables are limited to structural ones and elevation is not used as a predictor, although elevation is intrinsically included in structural variables at least along elevational vegetation zones. With similar complex terrain (around 4000 m elevation difference) Zellweger et al. (2017) used structural as well as topographic and climate variables. Even though elevation was not used directly, climatic variables (including temperature) showed the highest importance for modeling beta-diversity of birds and butterflies, even exceeding results when vegetation structure was included in the models. This seems consistent with the results of the present study, given that elevation is a main proxy for temperature (McCain & Grytnes, 2010). Overall, all these studies show clearly that the elevation gradient might be able to explain a major part of structural variables. An observation in a similar sense is made by Acebes et al. (2021) in their review of 173 papers. They find, especially for forested areas, that while canopy height is most commonly used as a LiDAR metric to model species-habitat structure, canopy cover and terrain topography performed better overall when they where used.

The study of Müller et al. (2014) covers an elevation gradient of around 800 m and does only take vertical profile metrics derived from LiDAR data into account. They could show that canopy arthropod diversity is driven by different structural features in the vegetation. Using a similar number of study plots as in our study, but exclusively in spruce forest, the overall vegetation structure is much more homogeneous than at Mt.Kilimanjaro. Thus, finer differences are likely to be masked by the large variability in vegetation structure in the models. The same is true for the study of Schooler and Zald (2019), who analyzed the predictability of small mammals diversity in temperate mixed forest and found that it could be predicted by LiDAR derived structural metrics.

Therefore, when using LiDAR data, non-structural properties (e.g. elevation, temperature, or other abiotic variables - depending on the study area) should be investigated separately to avoid false conclusions concerning the effect of LiDAR-derived vegetation structure. Those abiotic conditions are relevant for modeling and therefore models from one study area are not necessarily representative in other areas (Vogeler et al., 2014). Using a separate residual model shows great potential to avoid spurious correlation that leads to

erroneous predictions when the model is applied to new locations.
At Mount Kilimanjaro, with its substantial elevation gradient, the utilization of LiDAR data does not significantly improve modeling results. A larger sampling size per land cover is required to further improve the robustness of conclusions drawn for the selection of models. To approach the long-term goal of comprehensive mapping of EBVs like species occurrence or taxonomic diversity with the use of remotely sensed data, areas with a less complex land-cover gradient in homogenous landscapes need to be adressed in future studies to understand the influence of structure better.
To provide comparable results, further studies need to be conducted on multi-taxon approaches with field surveys and data sets of similar granularity. Study areas with different terrain complexities should be considered. In doing so, a solid base for valuable model-building strategies can be generated and can assist the research community in quantifying EBVs in the future.

Contributions

Conceptualization, A.Z., H.M., C.R., R.B. and T.N.; Data curation, I.O., M.P., T.A., C.B., A.C., F.D., J.D., C.E., S.F., F.G., M.H., M.H.-B., C.H., V.K., A.M., C.N., J.R., G.R., D.S.C., A.S., J.T., M.T., M.V. and S.W.; Formal analysis, A.Z., H.M., R.B. and T.N.; Funding acquisition, K.B.-G., M.F., A.H., M.S., I.S.-D., R.B. and T.N.; Methodology, A.Z., H.M., R.B. and T.N.; Project administration, R.B. and T.N.; Software, A.Z., S.W. and J.Z.; Supervision, H.M., C.R., R.B. and T.N.; Validation, A.Z., H.M. and T.N.; Visualization, A.Z., T.A. and F.D.; Writing – original draft, A.Z.; Writing – review & editing, A.Z., H.M., I.O., M.P., A.H., C.H., C.R., J.R., G.R., M.S., A.S., S.W., R.B. and T.N. All authors have read and agreed to the published version of the manuscript.

Funding

This work was carried out in the frame work of the DFG Research Unit 1246 KiLi—"Kilimanjaro ecosystems under global change: Linking biodiversity, biotic interactions and biogeochemical ecosystem processes" funded by the German Research Foundation (DFG, funding ID Ap 243/1-2 and Na 783/5-2).

Institutional review

The study was conducted with the least possible disturbance for nature and the environment. As we aimed at collecting data on the real ecosystems, we did not experimentally modify any vegetation. Vertebrates were sampled by acoustic and visual detection, which did not cause any disturbance. To assess biodiversity of arthropods we had to collect, kill and preserve them. However, we are confident that this did not have a major impact on the species populations. During the study we recorded 53 mammals, 202 birds and 1909 arthropods. Mammals and birds were identified by visual and acoustic detection; they were not captured nor killed. Arthropods were collected using ethanol or ethylenglycol for killing and preservation of specimens. Arthropods were mounted and labelled to be stored in public museum collections in Germany and Tanzania.

Data availability statement

The data that support the findings of this study are documented and archived in the central project database of the DFG-Research Unit FOR1246 (https://www.kilimanjaro.biozentrum.uni-wuerzburg.de), and are available from data owners upon reasonable request. Most data are already published or will be published shortly via GFBio (https://www.gfbio.org/), following the Rules of Procedure of the German Research Foundation (DFG) and the DFG-Research Unit FOR1246. The R scripts used within the study are available under a GPL 3.0 license as Git repository at https://github.com. A release of the Git repository to reproduce the results of the study is also available at https://github.com/envima/Kili_src.

Acknowledgements

We would like to thank our colleagues Andreas Ensslin, Sara B. Frederiksen, Elisabeth K.V. Kalko, William J. Kindeketa, Michael Kleyer, Ephraim Mwangomo, Ralph Peters, Marion Renner and Giulia Zancolli for their contribution to this study.

Appendix

Table 2.A1: Overview of land-cover types with elevation range of their occurrence, the number of plots in each land cover and the basic ecosystem category (forest or non-forest).

land cover	elevation[m a.s.l]	number of plots	ecosystem
maize plantation	866-1009	5	non-forest
savanna	871-1153	5	non-forest
coffe plantation	1124-1648	5	non-forest
homegarden	1169-1788	5	non-forest
grassland	1303-1748	5	non-forest
lower mountain forest	1560-2040	5	forest
ocotea forest	2120-2750	5	forest
ocotea forest (disturbed)	2220-2560	5	forest
podocarpus forest	2720-2970	5	forest
podocarpus forest (disturbed)	2770-3060	5	forest
erica forest	3500-3880	5	forest
helichrysum vegetation	3880-4390	4	non-forest

Table 2.A2: General sampling methods as well as some details and the calculation of the species richness of the biodiversity data. Further details on the sampling approaches can be found in Peters et al. (2016) and Peters et al. (2019). If not indicated otherwise, the species richness is calculated as the total (cumulative) number of species per study site (equal sampling effort for all sites)

Taxon	general sampling method	sampling specifics	species richness calculation
ants	plastic tubes with diverse set of resource baits on the ground	2 hours at times of peak ant activity	
bees	pan traps in different vegetation heights	48 hours with 3 sampling rounds	
birds	audiovisual point counts	15 min before sunrise, completed before 9 am	
dung beetles	baited pitfall trap	72 hours sampling time	
grass hoppers	sightings and two rounds of sweep net sampling	non-forested ares: repeatedly 1.5 hours walking on parallel tracks, forested areas: 1.5 hours shaking understory vegetation	sampling effort per site was adjusted to measure asymptotic species richness
insectivorous bats	acoustic monitoring (point stop method)	every corner of plot visited for 5 minutes	averaging species richness values of single surveys
large mammals	camera trap, analysis of dung remains	5 camera trapsper site, 70 trap-days per site	number of all non-domestic mammal species recorded on study site
millipedes	pitfall traps and sightings	sightings: 2 hours	
moths	light traps	20 minutes with 4 sampling rounds on obstacle-free branch in 1.5 - 2 metres height	averaging species richness values of single surveys
other aculeate wasps	pan traps in different vegetation heights	48 hours with 3 sampling rounds	
other beetles	pitfall traps	7 days sampling time with a total of 5 sampling rounds	
parasitoid wasps	pan traps in different vegetation heights	48 hours with 3 sampling rounds	
snails	sightings (large taxa) and collection of leaf litter (small taxa)	sightings: four rounds of fixed time surveys of 30 minutes, collection: 1 litre leaf litter	
spiders	pitfall traps	7 days sampling time	
springtails	pitfall traps	7 days sampling time	
syrphid flies	pan traps in different vegetation heights	48 hours with 3 sampling rounds	total (cumulative) number of species richness with varying number of samples among study sites
true bugs	sweep net sampling	100 sweeps along two 50 metre transects	

Table 2.A3: Fractional breakdown of feeding guilds. The numbers for each taxon represent the relative contribution to the species richness of a given feeding guild. Therefore, each column adds up to a total of 1. Taxa are listed in alphabetic order. True bugs were not identified to species level and are therefore not included.

	decomposer	generalist	herbivore	predator
ants	0	0.24	0.02	0.02
bees	0	0	0.14	0
birds	0	0.24	0.08	0.1
dung beetles	0.41	0	0	0
grasshoppers	0	0.03	0.17	0.01
insectivorous bats	0	0	0	0.02
large mammals	0	0.11	0.01	0
millipedes	0.16	0	0	0
moths	0	0	0.47	0
other aculeate wasps	0	0	0	0.12
other beetles	0.24	0.37	0.07	0.17
parasitoid wasps	0	0	0	0.51
snails	0.08	0.02	0.03	0.02
syrphid flies	0.11	0	0	0.02

Table 2.A4: Fractional breakdown of taxa. The numbers for each feeding guild represent the relative contribution of species richness to a given taxon. Therefore, each row adds up to a total of 1. Taxa are listed in alphabetic order. True bugs were not identified to species level and are therefore not included.

	decomposer	generalist	herbivore	predator
ants	0	0.47	0.24	0.29
bees	0	0	1	0
birds	0	0.18	0.31	0.51
dung beetles	1	0	0	0
grasshoppers	0	0.03	0.9	0.08
insectivorous bats	0	0	0	1
large mammals	0	0.52	0.33	0.15
millipedes	1	0	0	0
moths	0	0	1	0
other aculeate wasps	0	0	0	1
other beetles	0.13	0.17	0.17	0.53
parasitoid wasps	0	0	0	1
snails	0.24	0.05	0.41	0.31
syrphid flies	0.44	0	0.09	0.47

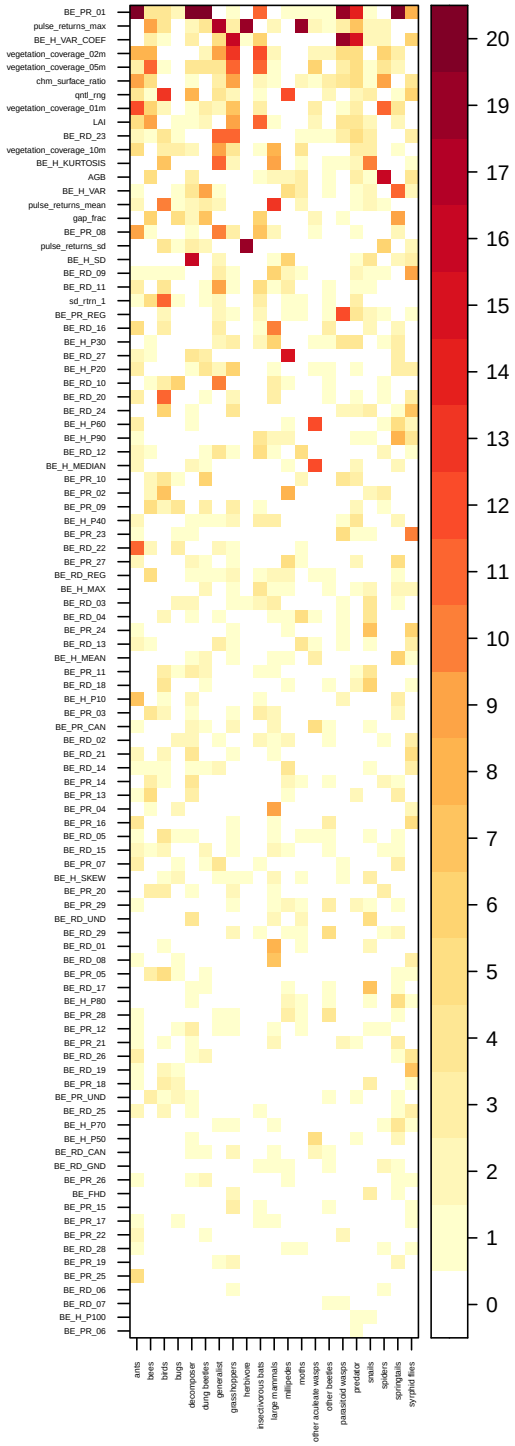


Figure 2.A1: Variable selection for the structure model of each taxon in forest. Colors show how often variables were included during 20-fold cross-validation. Structural variables are sorted by the total number of times they where selected. See Wöllauer et al. (2020) for variable details.

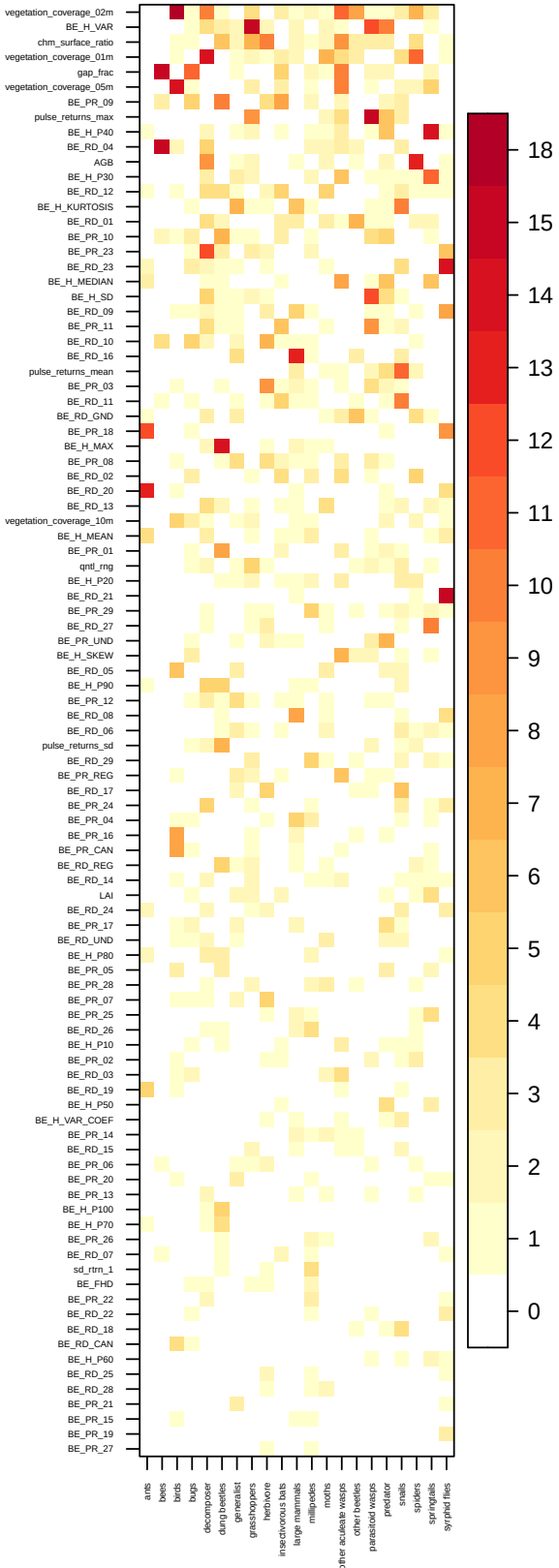


Figure 2.A2: Variable selection for structural variables for the residual model and combined model of each taxa in forest. Colors show how often variables were included during 20-fold cross-validation. Structural variables are sorted by the total number of times they where selected. See Wöllauer et al. (2020) for variable details.

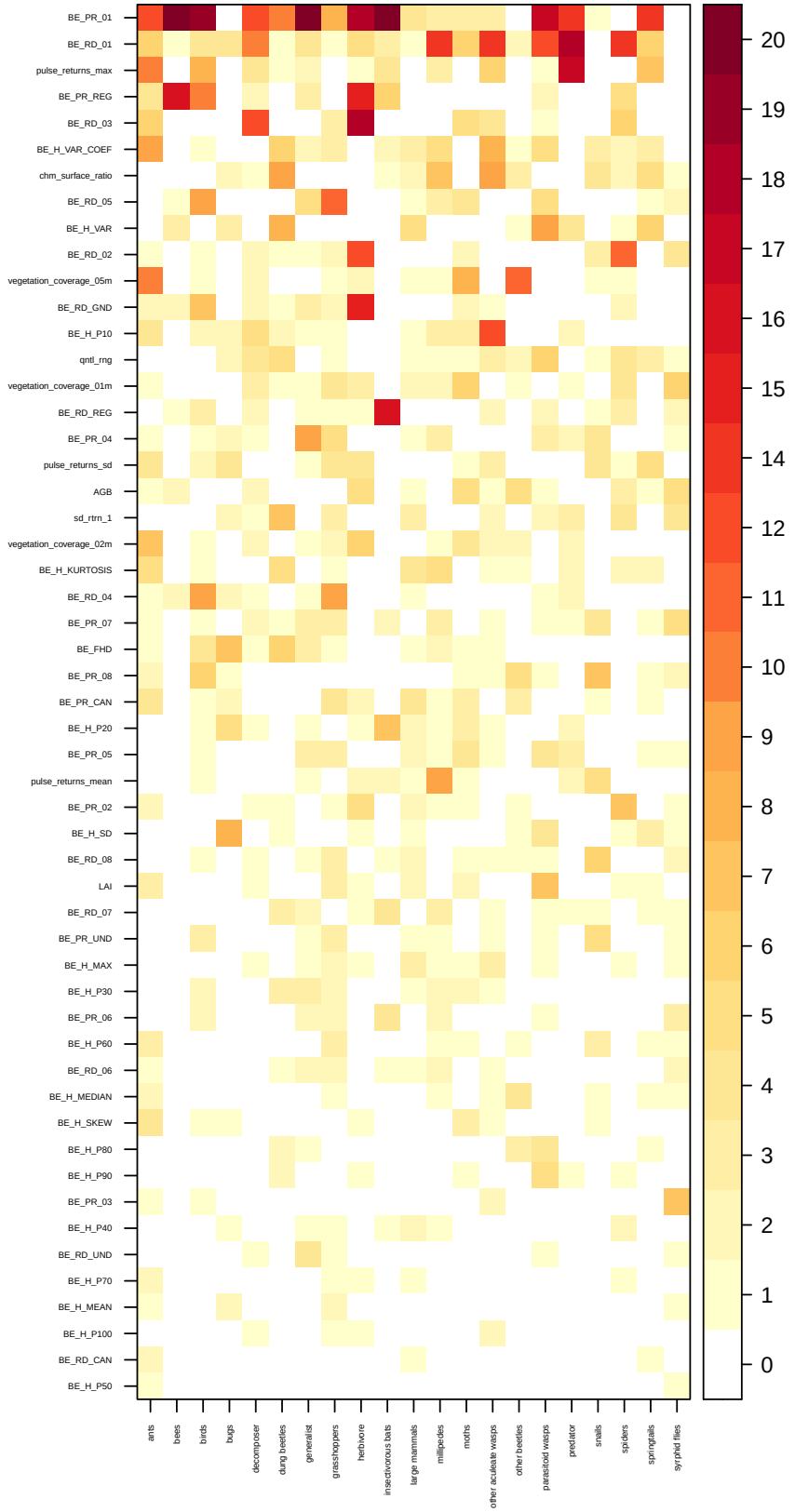


Figure 2.A3: Variable selection for the structure model of each taxon in non-forest. Colors show how often variables were included during 20-fold cross-validation. Structural variables are sorted by the total number of times they where selected. See Wöllauer et al. (2020) for variable details.

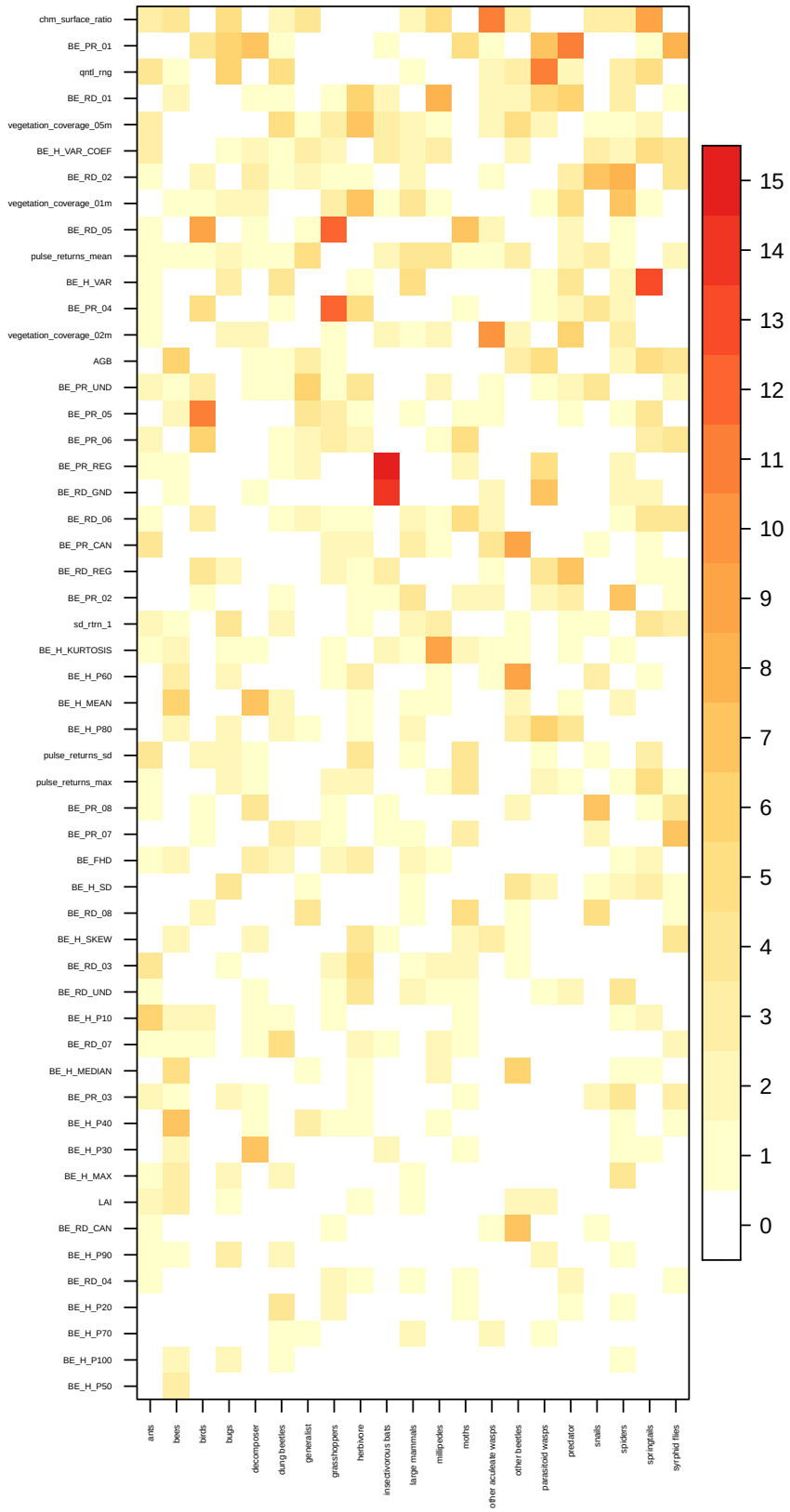


Figure 2.A4: Variable selection for structural variables for the residual model and combined model of each taxa in non-forest. Colors show how often variables were included during 20-fold cross-validation. Structural variables are sorted by the total number of times they where selected. See Wöllauer et al. (2020) for variable details.

Chapter 3

Leveraging readily available heterogeneous LiDAR data to enhance modeling of successional stages at tree species level in temperate forests

This chapter has been submitted as:

Bald, L., **Ziegler, A.**, Gottwald, J., Koch, T.L., Ludwig, M., Meyer, H., Zeuss, D., Frieß, N., 'Leveraging readily available heterogeneous LiDAR data to enhance modeling of successional stages at tree species level in temperate forests'
to
Environmental Data Science

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 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.

3.1 Introduction

Forests provide a variety of indispensable ecosystem services, such as water storage and purification, regulation of air quality, climate regulation by functioning as 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 & 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 tree species level. Forest successional stages describe the development of the forest ecosystem typically after a disturbance in several phases, which are different in forest structure and can thus serve as indicators for forest biodiversity (Wilson & Peter, 1988). The knowledge on 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) and remote sensing can contribute to enhance traditional forest inventories (White et al., 2016). Multi-spectral 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 & 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, however, those studies analyze the successional stages across large areas without differentiating between tree species (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). 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 with multiple flights typically spans several years, with for example an interval of six years for regions like the federal states of Hesse (HVBG, 2023) and Saxony (GeoSN, 2023), five years for North Rhine-Westphalia (Geobasis NRW, 2023) , and four years for Rhineland-Palatinate (LVermGeo, 2023). 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 six years, or in Estonia with updates every four years (Maa-amet, 2023). Moreover, data availability is unsystematically documented with no common standard or database. As the data is 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 3.4). 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).

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 with a modeling approach that takes spatial auto-correlation into account by using spatial variable selection and spatial cross-validation techniques were used (Meyer, Reudenbach, 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, 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, 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 influences model outcomes for forest successional stages at the tree species level.

3.2 Material and Methods

In the following sections, the modeling of tree species group specific forest successional stages is presented in detail (see sections 4.2.1 to 3.2.3 and figure 3.1). 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, using a tree species group model across the entire area of Rhineland-Palatinate as the basis for the successional stages models. All data processing and modeling was done in R version 4.2.3 (R Core Team, 2023).

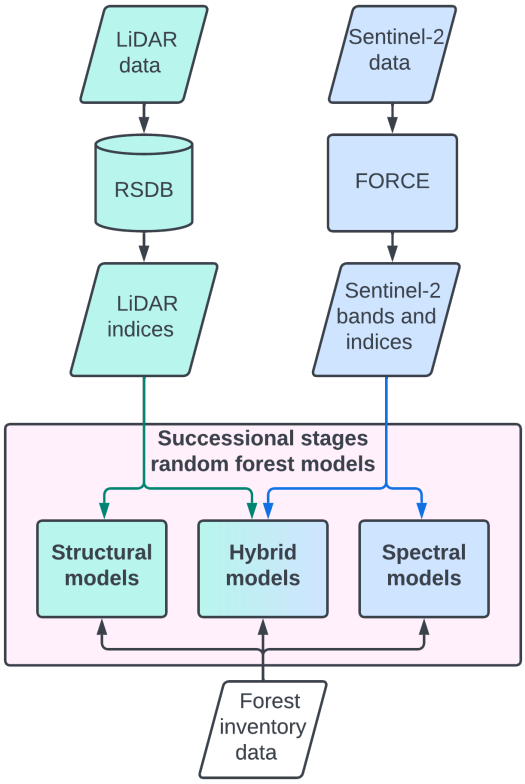


Figure 3.1: General modeling workflow that is applied to every tree species group to model the specific successional stages. The modeling makes use of two different data sets for prediction. On the one hand LiDAR data, on the other hand multi-spectral Sentinel-2 data. From the LiDAR data various indices are derived and from the Sentinel-2 data, which was processed within the software FORCE, the original bands and additionally spectral indices were calculated. As reference data and therefore response variable of the models the data from the forest inventory is used. We conducted three different model types, namely structural, hybrid and spectral models using different combinations of the predictive data sets.

3.2.1 Study area

The federal state of Rhineland-Palatinate with an area of 19 858 km^2 (see figure 3.2) 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, 2012a). The majority of the forests are mixed forests and only 17.7 % of the forests are pure stands (Thünen-Institut, 2012b). Overall, deciduous forests predominate, with shares ranging from 54.8 % to 64.6 % of the forest, depending on the ownership structure (Thünen-Institut, 2012c).

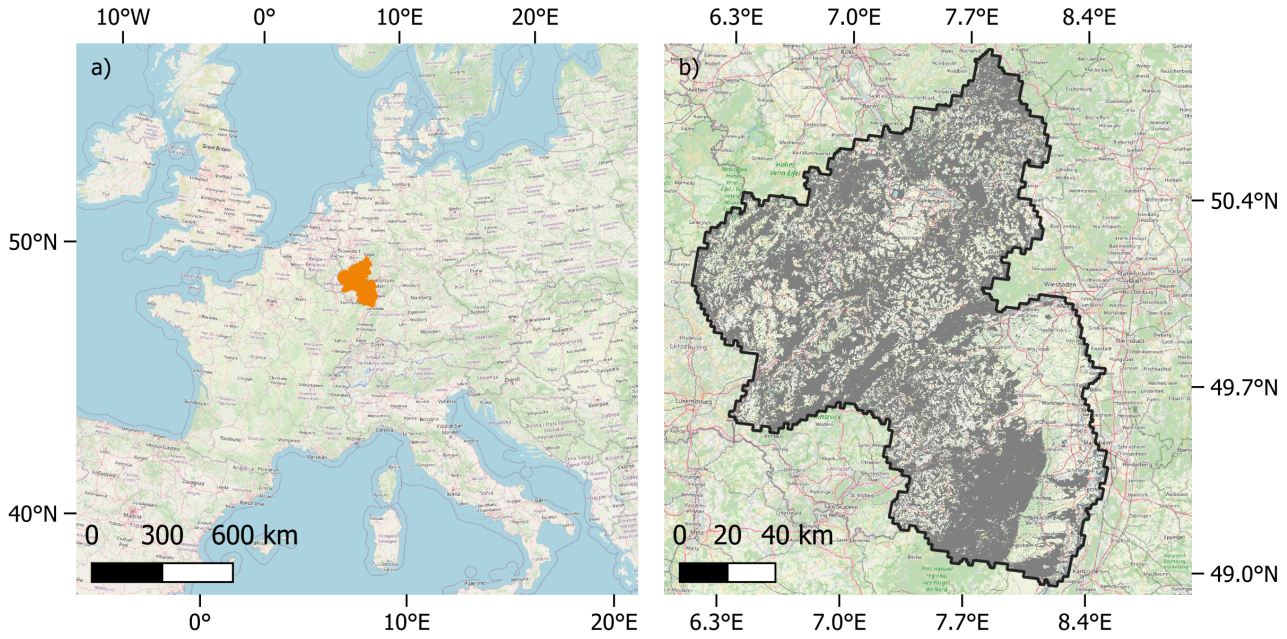


Figure 3.2: a) Location of the study area (orange) in Europe. b) Study area confined to the forest mask (grey) derived from Copernicus high resolution layer. Data: EEA, 2022; GeoBasis-DE / LVermGeoRP, 2022; OpenStreetMap, 2023.

3.2.2 Data bases

Forest inventory data

In this study, the official forest inventory of Rhineland Palatinate 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 3.3). 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 3.A3 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 oldest successional stage considered in this study is the maturing stage (III), where the tree surpasses 75 to 80 percent 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.

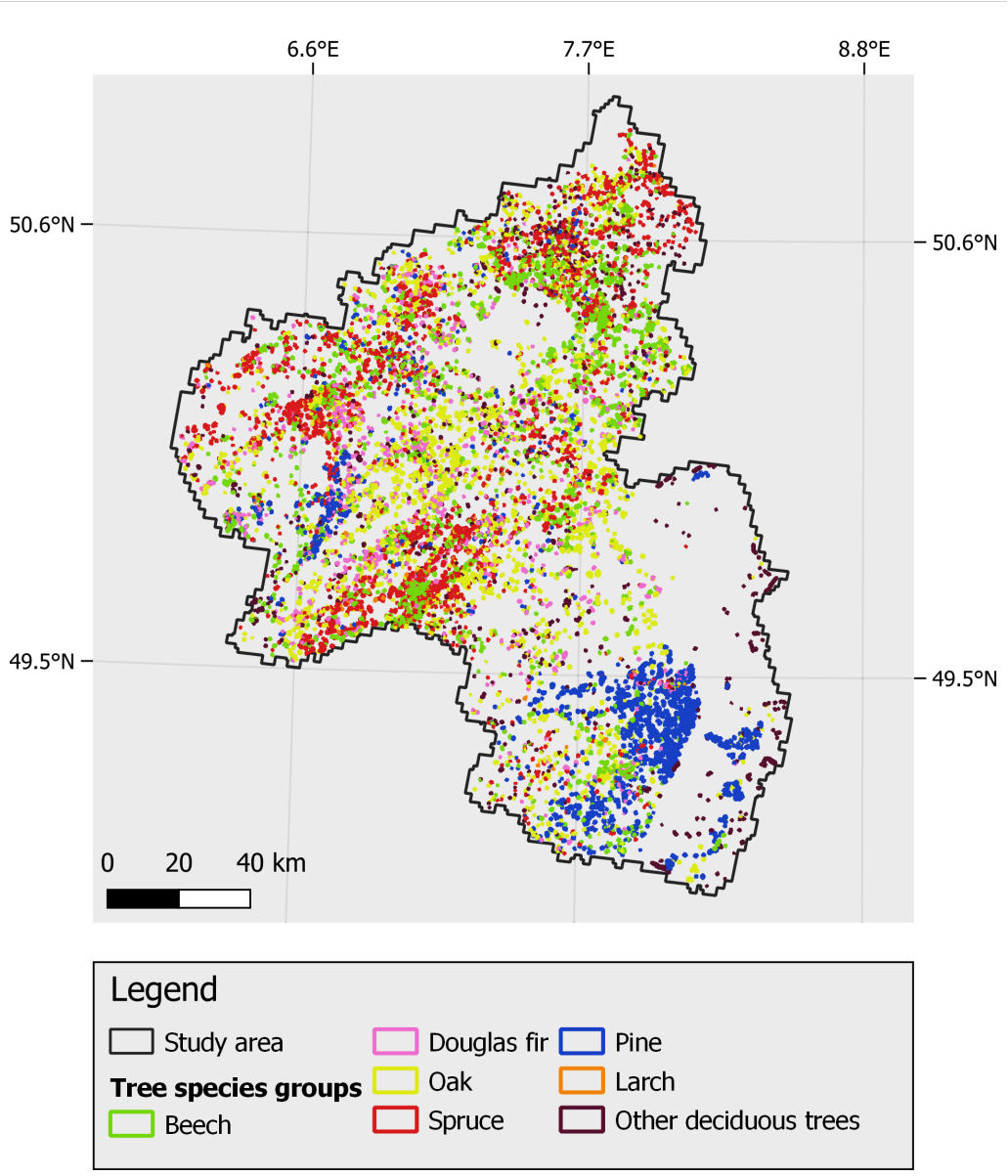


Figure 3.3: Spatial distribution of the tree species groups from the forest inventory. As the forest inventory only surveys state-owned forests the depicted polygons represent only a subset within the whole study area. Data: Landesforsten Rheinland-Pfalz (2014).

Sentinel-2 data

Multi-spectral optical data are proven to be adequate for tree species classifications (Grabska et al., 2019; Hemmerling et al., 2021; Hościło & Lewandowska, 2019; Immitzer et al., 2016; Wessel et al., 2018; Xi et al., 2021), making it 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 2019-2021 were downloaded at level 1C and further atmospherically as well as topographically corrected. Within FORCE near-infrared Landsat data was 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 three 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, June) and two images for leaf senescence (September, October) were created. In addition to the original bands, multiple spectral

indices reflecting vegetation properties were calculated. Table 3.1 shows the Sentinel-2 spectral bands and indices that were used in this study.

Table 3.1: 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.

Name	Description	Reference
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, Gritz, and Merzlyak (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	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)

LiDAR data

The LiDAR data are the key elements of this study, as its aim is to identify the potential of these heterogenous 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 nine years to only four years. As the transition is still ongoing, in this study data from a seven-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 3.4). Study areas consisting of different flight campaigns tend to also vary notably in technical properties like for example scan angle and flight altitude (Næsset, 2009; Ørka et al., 2010; Solberg et al., 2009). In total, 29 indices were computed from the LiDAR data using the remote sensing database (RSDB) of Wöllauer et al., 2020. The indices were computed for all areas identified as deciduous or coniferous forest 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., penetration rate of different vegetation layers), overall vegetation properties (e.g., aboveground biomass, vegetation coverage, leaf area index), and terrain features (e.g., elevation; see table 3.2).

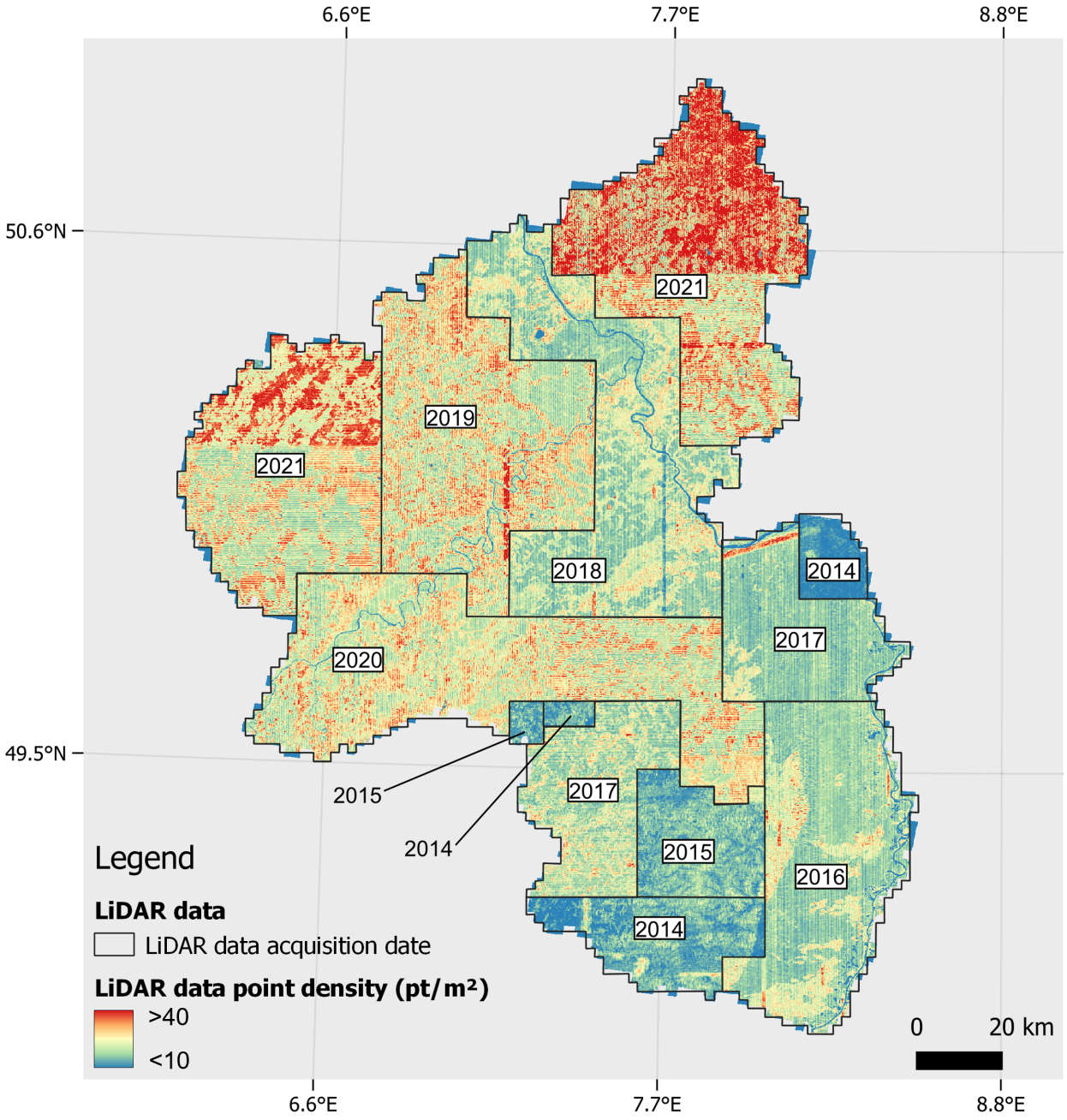


Figure 3.4: Properties of heterogeneous LiDAR data of Rhineland-Palatinate. The year in which the data were recorded as well as the calculated point density derived directly from the LiDAR data set, based on 100 m pixel are depicted. It is visible that there is a transition from lower point densities in earlier years to higher point densities in later years. Data: GeoBasis-DE / LVermGeoRP (2022).

Table 3.2: Overview of LiDAR indices characterizing the vegetation calculated with the remote sensing database (RSDB; Wöllauer et al., 2020; see appendix 3.A2 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 surface above ground - CHM raster based
CH sd	Standard deviation of canopy height
CHM sd	Standard deviation of surface above ground - CHM raster based
CH median	Median canopy height
CH skew	Skewness of canopy height distribution
CH curtosis	Excess kurtosis of 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 regeneration vegetation layer
PR understory	Penetration rate of 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) 1m	Vegetation coverage in 1 meter height
VC 2m	Vegetation coverage in 2 meter height
VC 5m	Vegetation coverage in 5 meter height
VC 10m	Vegetation coverage in 10 meter 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

3.2.3 Methods

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 different models (see figure 3.1). 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.

Balancing data and splitting into testing and training data

The 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 was balanced so 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 the number of pixels from within each polygon is chosen very small, many of the available polygons can be used, but only a few pixels of each, even of the very large polygons, will be used. On the other hand, if the number of pixels is chosen to be large, more polygons fall below this threshold and are therefore not considered further, but more pixels are considered 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 possible (for more details see appendix 3.A1 or the R code in the data availability statement). From the resulting data set, 20 % of the polygons from each class were retained for external testing, the rest was used for model training and validation (see tables 3.A1 and 3.A3 in the appendix). From all the training data sets for each of the successional stages model (one data set for each tree species group containing 80 % of the polygons of each class) all data was used to balance a data set for the tree species group model, which 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.

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 recognition and removal of variables that lead to overfitting (Meyer et al., 2018). 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 the 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 3.2.3) to evaluate the potential of LiDAR data for classifying successional stages.

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. Three models were trained independently 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 of all tree species groups across the entire area as baseline, and secondly, modeling of 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.

3.3 Results

This section presents the study findings of comparing the performances of different modeling approaches for successional stages of specific tree species groups and thereby, evaluates the potential of using heterogeneous LiDAR data. Additionally, the variable selection of the different models and the area-wide prediction of tree species groups specific successional stages throughout Rhineland-Palatinate were analysed.

3.3.1 Model performance

The structural models and hybrid models performed notably better than the spectral models 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. As the spectral information from Sentinel-2 data improved the performance of the hybrid models compared to the structural models, and Sentinel-2 data are freely and globally available, the utilization of hybrid models should be preferred. Therefore, and for a 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 appendix 3.A2. The two left columns of figure 3.5 show the results from the application to the test data sets of each successional stage and model, and the right column shows the difference of the proportion of correctly classified pixels between models. For the spectral models, an overall accuracy between 0.33 for the group 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 occured among adjacent successional stages. The only exception is the qualification stage of other deciduous trees, where only 5% of pixels were classified correctly and the most misclassifications were not even in the adjacent stage (dimensioning stage) but 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, 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 3.4.3). 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 3.5).

Figure 3.5: Model performances. The left column of plots shows the results of models applied to the test data sets only using Sentinel-2 variables (spectral models), 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.

3.3.2 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 3.2.3, 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 the hybrid model additionally used LiDAR-derived variables it is 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 3.1 and 3.2). Figure 3.6 shows a boxplot each for the spectral and the hybrid model 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 were selected. For all other variable groups, at least two models did not select any variables from the respective variable group (see table 3.3).
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 3.A4) 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 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 $= 0.48$).

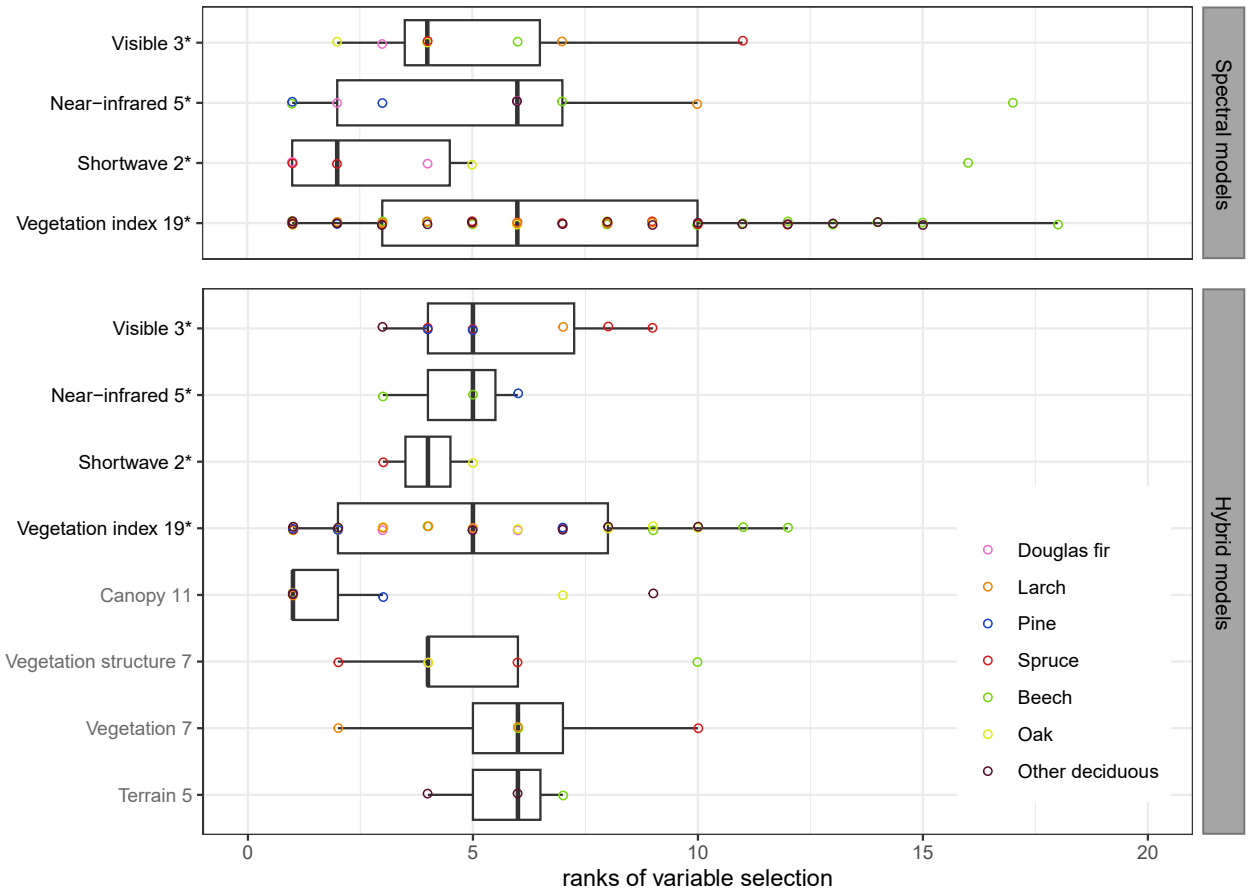


Figure 3.6: The ranks of variable groups from variable selection. Boxplots display the ranks of variables selected during feature selection for different variable groups. Colored dots show the rank separated by tree species group. The boxes in the plot show the interquartile range of the ranks, with 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 3.1 and 3.2), 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. *Each Sentinel-2 variable is available for seven months.

Table 3.3: 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. *Each Sentinel-2 variable is available for seven month.

Table 3.4

	Douglas fir	Larch	Pine	Spruce	Beech	Oak	Other decidicous 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	

Table 3.5

	Douglas fir	Larch	Pine	Spruce	Beech	Oak	Other decidicous 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	

3.3.3 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 0.81. Details of the model and its variable importance are provided in the appendix (see figure 3.A3 and table 3.A5). 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 when applied to the test data sets. For detailed confusion matrices see table 3.A5 and table 3.A6 in the appendix.

Figure 3.7 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 South-east (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, 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.

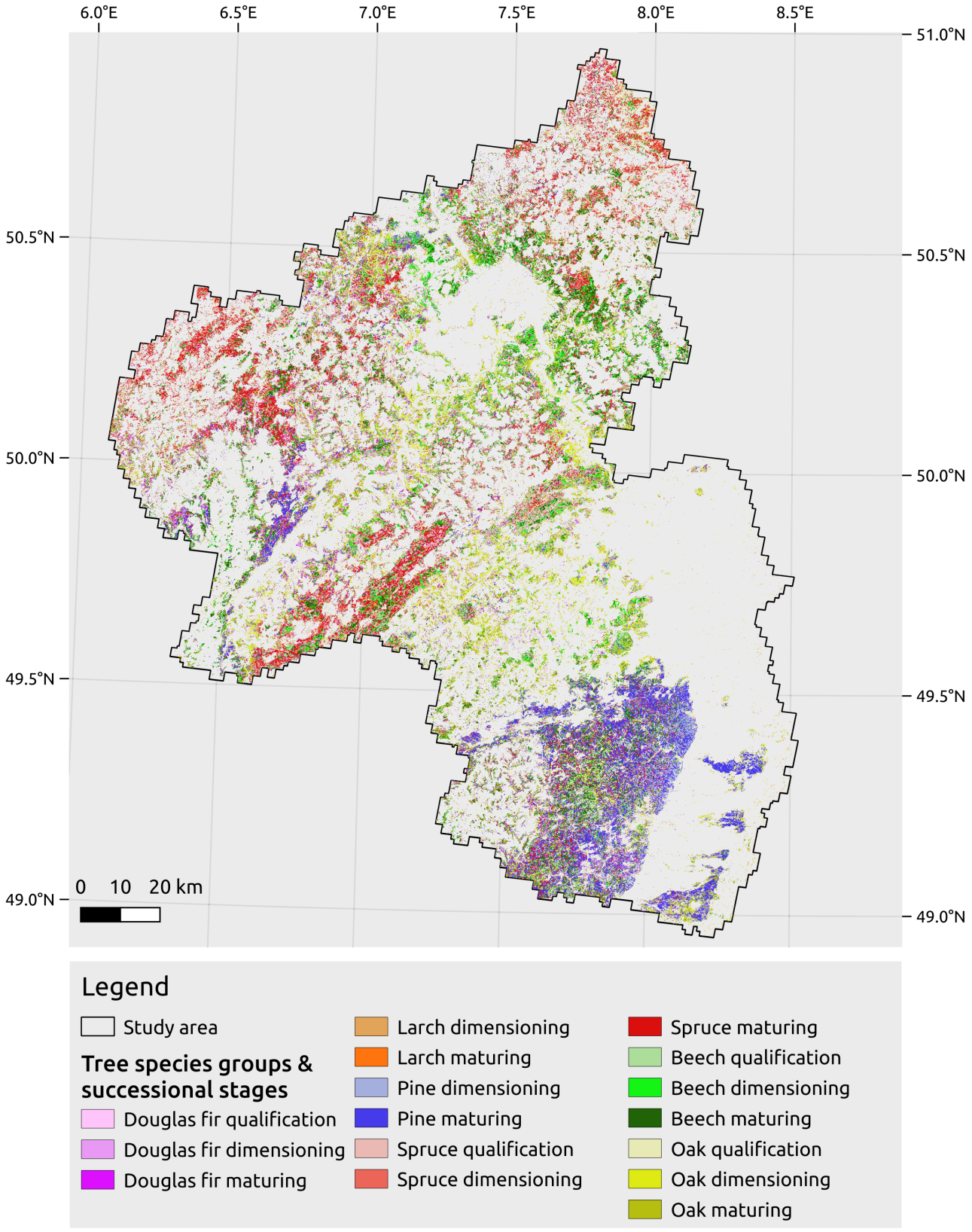


Figure 3.7: Area-wide map of tree species groups and its successional stages for Rhineland-Palatinate. Leaflet available at: https://envima.github.io/LidarForestModeling/l.

Figure 3.8 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 3.8b) with up to six years of temporal difference. Figure 3.8a shows an area dominated by deciduous species, while figure 3.8c 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.

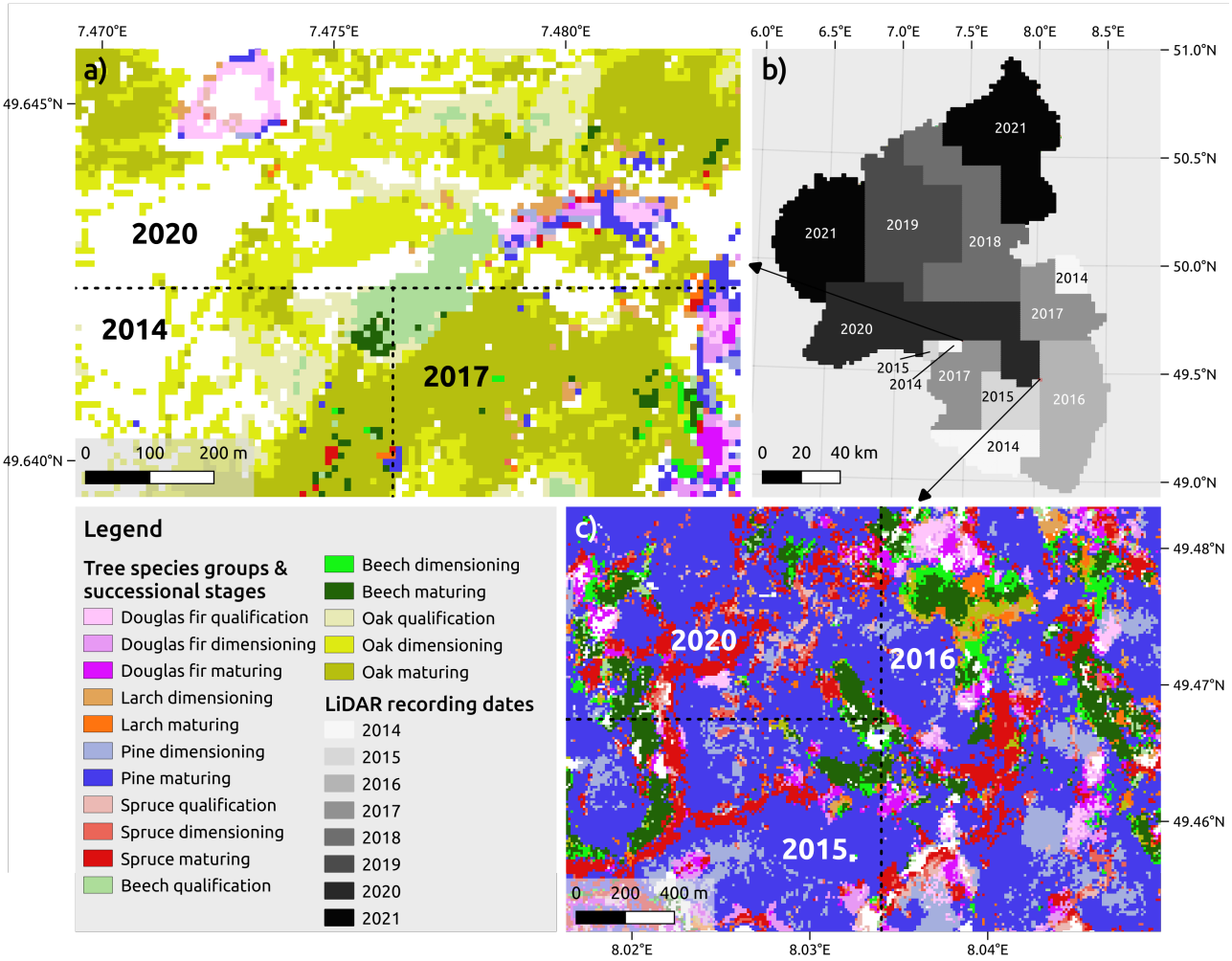


Figure 3.8: Detailed map of LiDAR survey borders. Variations of the tree species group specific successional stages are indicated with colors in two map sections in plot a) and c). These exemplary areas were chosen at boundaries of LiDAR scenes with large temporal differences. On plot b) the years of acquisition for the LiDAR data set can be identified with lighter colors for older data and darker colors for more recent data.

3.4 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 was 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.

3.4.1 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 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 multi-spectral 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 was grouped together in this class (see appendix table 3.A7). 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 3.3.3). 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 is essential due to their significant impact on biodiversity (Hilmers et al., 2018; Reif et al., 2013) and the preservation of forest ecosystem services.

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 will incorporate this information for modeling the habitat of endangered forest-dwelling bat species. This direct application underscores how such readily available but heterogeneous LiDAR data can significantly 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 is proven (Reddy, 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. At least more focus has to be put on the pre-processing of data and the adjustment of modeling techniques 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) to exploit the full potential for ecosystem monitoring, especially since they provide relevant information on the three-dimensional forest structure which cannot be substituted by passive sensors. Although recent high-resolution LiDAR data, 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 sci-

entific projects) and, when combined with publicly available Sentinel-2 data, 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.

3.4.2 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 3.6 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 3.3). 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 four and six 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 appendix 3.A4) was often selected as one of the first and, therefore, most important variables.

3.4.3 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 3.3) 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 tree species group larch 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 compared to the usage of for example SPOT data in Stoffels et al. (2015) approach. 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 certain tree species groups (beech, 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 3.7 shows general spatial patterns of tree species that align with the actual forest patterns in Rhineland palatinate (see figure 3.3; 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 3.8 do not exhibit any distorting patterns. At the intersection of the 2014, 2017 and 2020 flight campaign boundaries in figure 3.8, 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.
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.

3.5 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 pre-processing 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. As this study showed, this could provide a valuable contribution to ecosystem research and, subsequently, to the preservation of forest ecosystem services.

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/.

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).

Competing Interests

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 10.11.2023.

Ethical Standards

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

Author Contributions

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). 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.

Appendix

Table 3.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	3 114	18
Larch	7 596	36	211	2 110	10
Pine	12 844	26	494	3 952	8
Spruce	30 810	237	130	7 410	57
Beech	61 152	168	364	14 560	40
Oak	18 573	123	151	4 983	33
Other deciduous trees	12 714	78	163	3 423	21

Table 3.A2: Overview of structural LiDAR indices with the names of the 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) 1m	vegetation_coverage_01m_CHM
VC 2m	vegetation_coverage_02m_CHM
VC 5m	vegetation_coverage_05m_CHM
VC 10m	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 3.A1: Pseudocode balancing.

Table 3.A3: 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

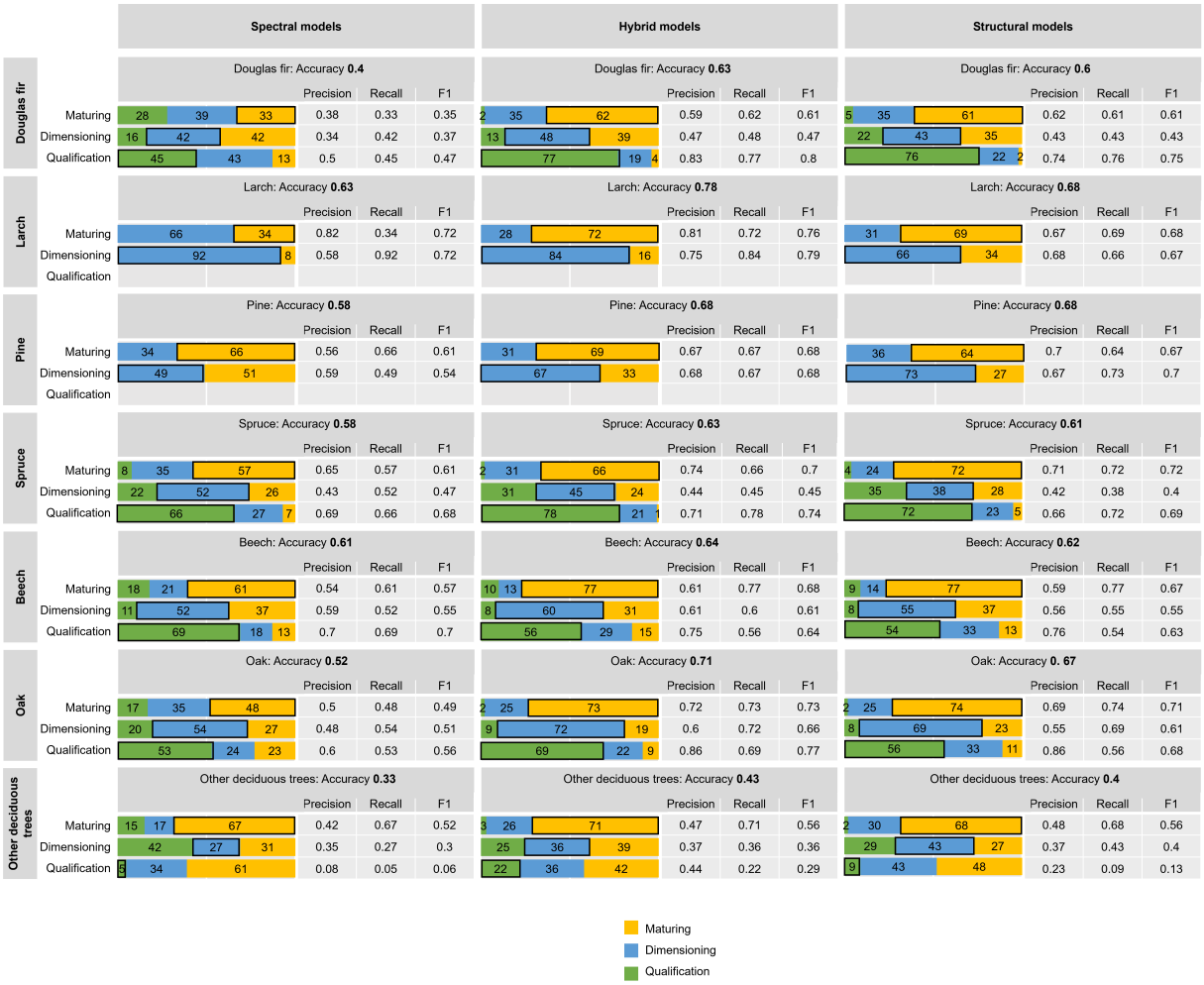


Figure 3.A2: Left Column of plots shows test results of models only using Sentinel-2 variables (spectral models), middle column shows test results using Sentinel-2 and LiDAR variables (hybrid models). The right column shows the test results of the models using only LiDAR variables (structural model). 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 3.A4: 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	

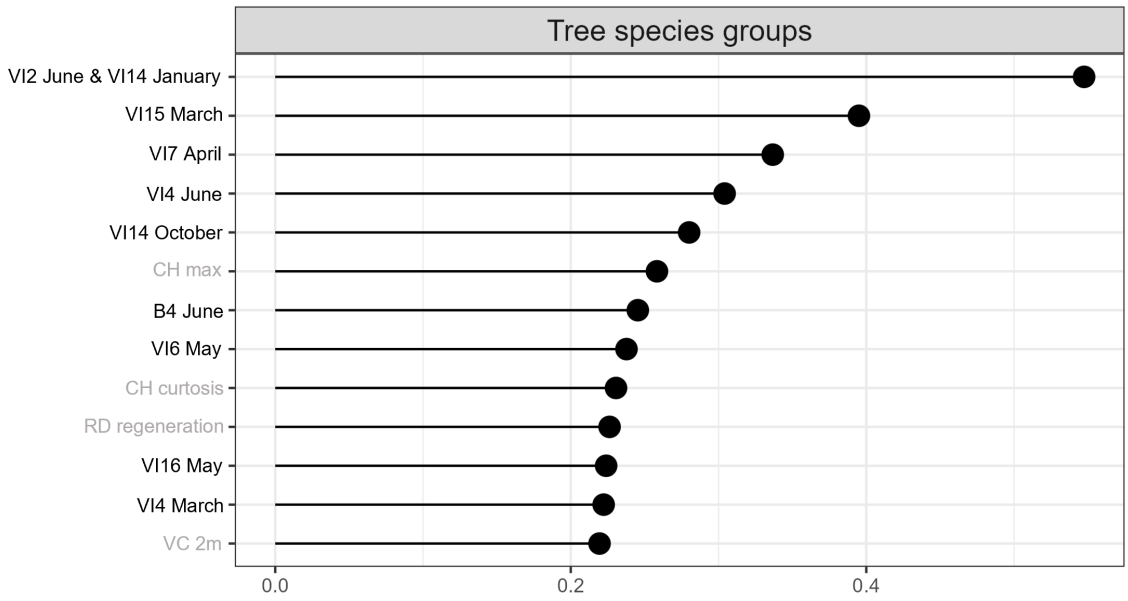


Figure 3.A3: Variable importance of the tree species groups model.

Table 3.A5: 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.

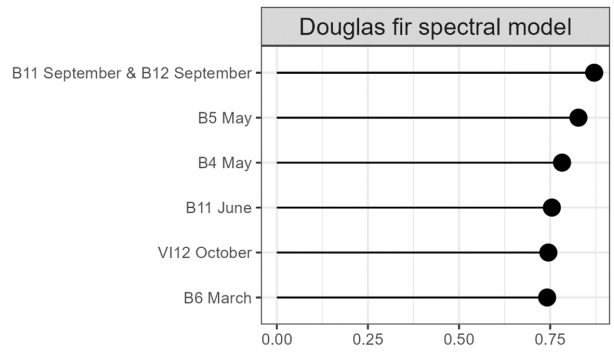
	Douglas fir	Larch	Pine	Spruce	Beech	Oak	Other deciduous trees
Douglas fir	85	0	4	2	0	1	2
Larch	1	86	2	2	1	2	4
Pine	5	3	90	2	5	1	2
Spruce	7	0	2	89	2	1	2
Beech	1	5	2	1	84	11	3
Oak	0	3	1	0	6	63	13
Other deciduous trees	1	2	0	3	1	23	74

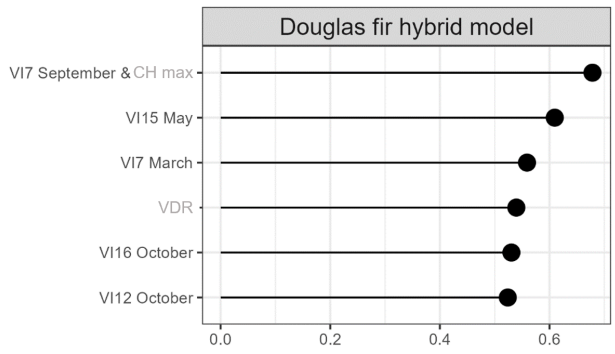
Table 3.A6: Confusion matrix for the hierarchical modeling of tree species groups and successional stages with letter indicating the successional stages (qualification: q, dimensioning: d, maturing: m). The values indicate the classified pixels as percentage.

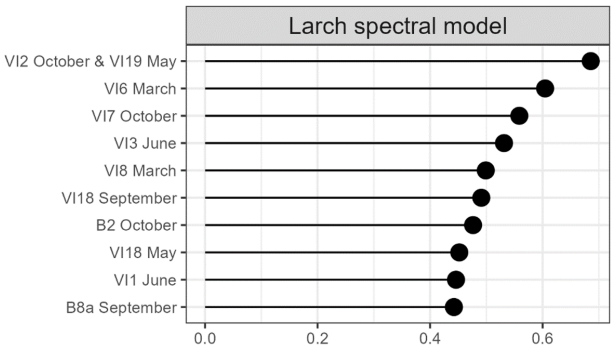
	Douglas fir q	d	m	Larch d	m	Pine d	m	Spruce q	d	m	Beech q	d	m	Oak q	d	m
Douglas fir q	65	12	2	0	0	0	0	2	0	0	0	0	0	1	0	0
douglas fir d	14	41	31	0	0	0	1	0	0	1	0	0	0	1	0	0
Douglas fir m	3	36	56	0	0	2	3	0	0	2	0	0	0	0	0	0
Larch d	0	1	1	76	22	0	0	1	0	0	1	2	0	2	1	0
Larch m	0	0	1	16	63	3	0	0	0	0	1	1	0	0	0	3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

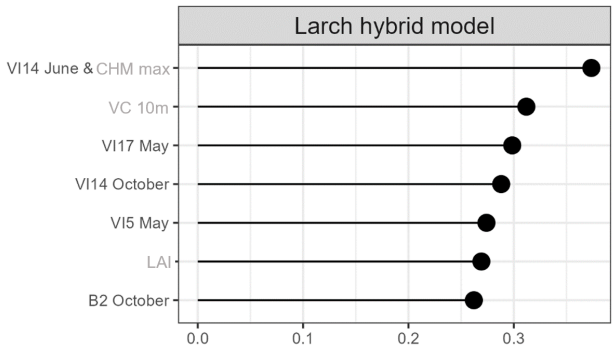
Table 3.A7: Most important species in the tree species groups in the forest inventory data.

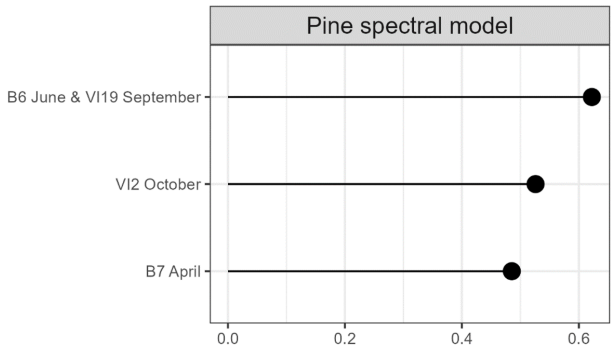
Douglas fir	
Douglas fir	Pseudotsuga menziesii
False cypress	Chamaecyparis
Juniper	Juniperus
Redwoods	Sequoioideae
Thuja	Thuja
Tsuga	Tsuga
Yew	Taxus
Larch	
European larch	Larix decidua
Japanese larch	Larix kaempferi
Pine	
Black pine	Pinus nigra
Pine	Pinus nigra
Ponderosa pine	Pindus ponderosa
Weymouth pine	Pinus strobus
Spruce	
Serbian spruce	Picea omorika
Sitka spruce	Picea sitchensis
Spruce	Picea
Beech	
Beech	Fagus
Oak	
Common oak	Quercus robur
Downy oak	Quercus pubescens
Sessile oak	Quercus petraea
Turkey oak	Quercus cerris
Other deciduous trees	
Short-lived deciduous trees	
Alder	Alnus
Aspen	Populus tremula
Balsam poplar	Populus balsamifera
Birch	Betula
Black poplar	Populus nigra
Cherry	Prunus
Downy birch	Betula pubescens
European crab apple	Malus sylvestris
European wild pear	Pyrus pyraster
Goat willow	Salix caprea
Poplar	Populus
Rowan	Sorbus aucuparia
Silver birch	Betula pendula
Long-lived deciduous trees	
Ash	Fraxinus excelsior
Black locust	Robinia pseudoacacia
Common hornbeam	Carpinus betulus
Common whitebeam	Sorbus aria
Eastern american black walnut	Juglans nigra
Elm	Ulmus
English walnut	Juglans regia
Field maple	Acer campestre
Large-leaved linden	Tilia platyphyllos
Linden	Tilia platyphyllos
Montpellier maple	Acer monspessulanum
Norway maple	Acer platanoides
Red oak	Quercus rubra
Service tree	Sorbus domestica
Shadbush	Amalanchier
Small-leaved linden	Tilia cordata
Sorbus species	Sorbus
Sweet chestnut	Castanea sativa
Sycamore maple	Acer pseudoplatanus
Wild service tree	Sorbus torminalis

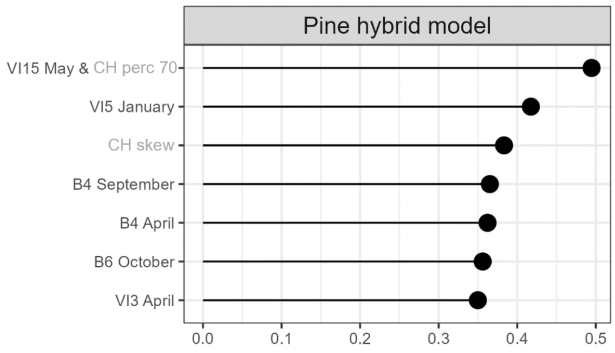

Douglas fir spectral model
B11 September & B12 September
B5 May
B4 May
B11 June
VI12 October
B6 March
0.00
0.25
0.50
0.75

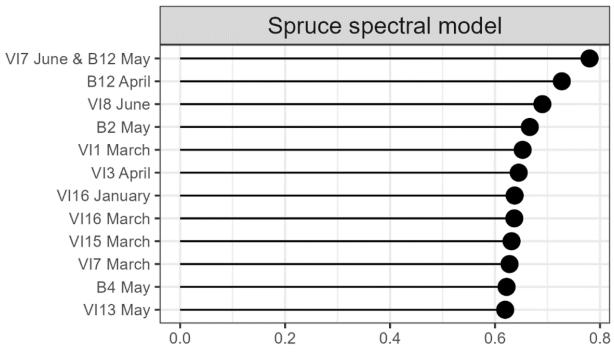

Douglas fir hybrid model
VI7 September & CH max
VI15 May
VI7 March
VDR
VI16 October
VI12 October
0.0
0.2
0.4
0.6

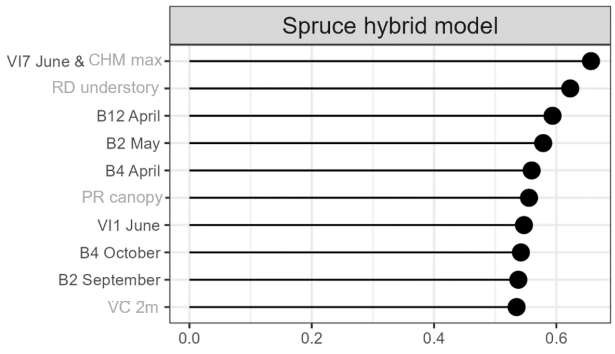

Larch spectral model
VI2 October & VI19 May
VI6 March
VI7 October
VI3 June
VI8 March
VI18 September
B2 October
VI18 May
VI1 June
B8a September
0.0
0.2
0.4
0.6


Larch hybrid model
VI14 June & CHM max
VC 10m
VI17 May
VI14 October
VI5 May
LAI
B2 October
0.0
0.1
0.2
0.3


Pine spectral model
B6 June & VI19 September
VI2 October
B7 April
0.0
0.2
0.4
0.6


Pine hybrid model
VI15 May & CH perc 70
VI5 January
CH skew
B4 September
B4 April
B6 October
VI3 April
0.0
0.1
0.2
0.3
0.4
0.5


Spruce spectral model
VI7 June & B12 May
B12 April
VI8 June
B2 May
VI1 March
VI3 April
VI16 January
VI16 March
VI15 March
VI7 March
B4 May
VI13 May
0.0
0.2
0.4
0.6
0.8


Spruce hybrid model
VI7 June & CHM max
RD understory
B12 April
B2 May
B4 April
PR canopy
VI1 June
B4 October
B2 September
VC 2m
0.0
0.2
0.4
0.6

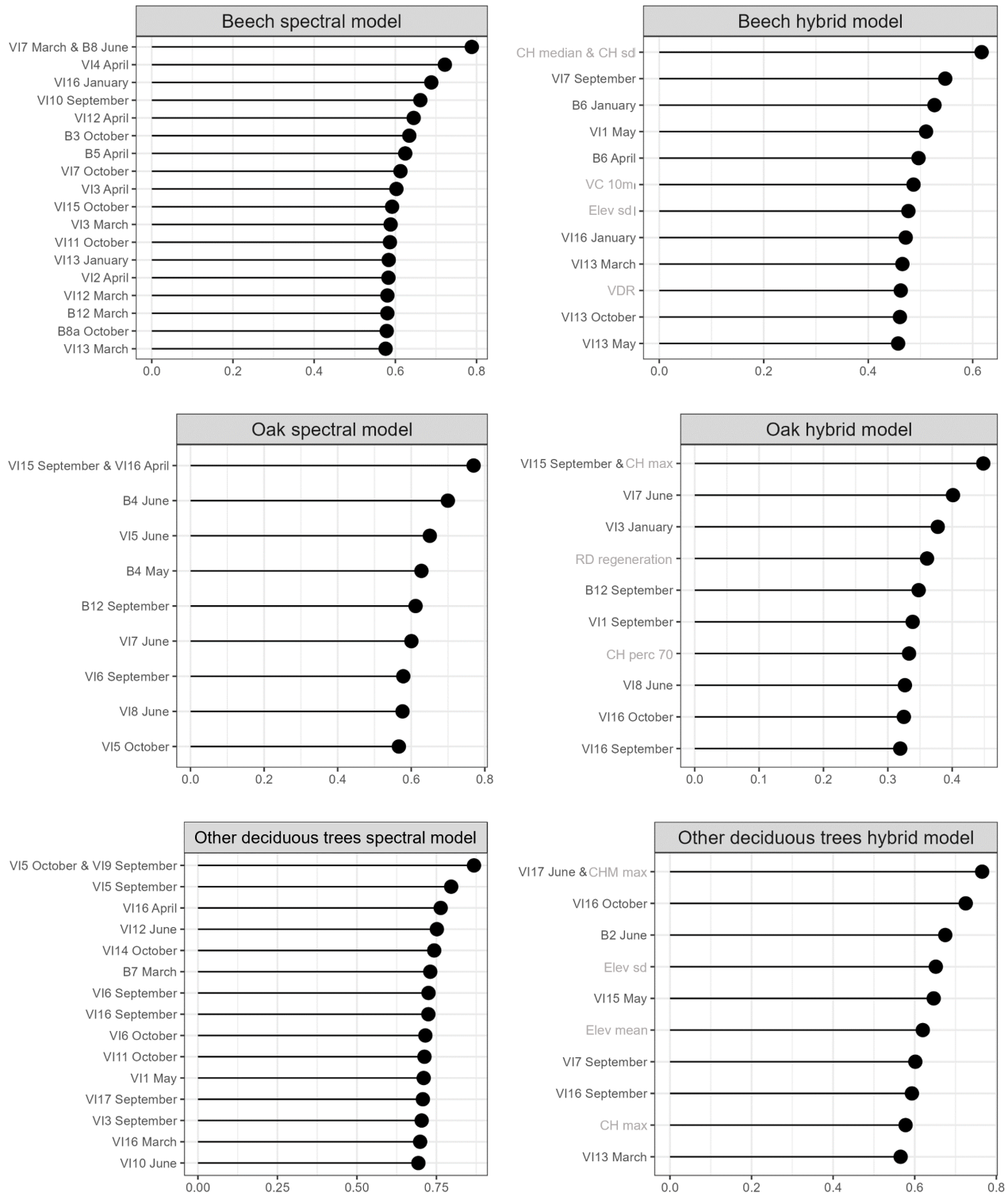


Figure 3.A4: Variable importance of hybrid and spectral tree species groups specific successional stages models. For each tree species group one model only using Sentinel-2 data (spectral on the left) and one using Sentinel-2 and LiDAR (hybrid on the right) is depicted. Sentinel-2 variables are labeled black and LiDAR variables are labeled gray at the y-axis.

Chapter 4

Using GEDI as training data for an ongoing mapping of landscape-scale dynamics of the plant area index

This chapter has been published as:

Ziegler, A., Heisig, J., Ludwig, M., Reudenbach, C., Meyer, H. and Nauss, T. **(2023)**, Using GEDI as training data for an ongoing mapping of landscape-scale dynamics of the plant area index, *Environmental Research Letters* 18(7). https://doi.org/10.1088/1748-9326/acde8f.

Abstract
Leaf or plant area index (LAI, PAI) information is frequently used to describe vegetation structure in environmental science. While field measurements are time-consuming and do not scale to landscapes, model-based air- or spaceborne remote-sensing methods have been used for many years for area-wide monitoring. As of 2019, NASA's Global Ecosystem Dynamics Investigation (GEDI) mission delivers a point-based LAI product with 25 m footprints and periodical repetition. This opens up new possibilities in integrating GEDI as frequently generated training samples with high resolution (spectral) sensors. However, the foreseeable duration of the system installed on the ISS is limited. In this study we want to test the potential of GEDI for regional comprehensive LAI estimations throughout the year with a focus on its usability beyond the lifespan of the GEDI mission. We study the landscape of Hesse, Germany, with its pronounced seasonal changes. Assuming a relationship between GEDI's PAI and Sentinel-1 and -2 data, we used a Random Forest approach together with spatial variable selection to make predictions for new Sentinel scenes. The model was trained with two years of GEDI PAI data and validated against a third year to provide a robust and temporally independent model validation. This ensures the applicability of the validation for years outside the training period, reaching a total RMSE of 1.12. Predictions for the test year showed the expected seasonal and spatial patterns indicated by RMSE values ranging between 0.75 and 1.44, depending on the land cover class. The overall prediction performance shows good agreement with the test data set of the independent year which supports our assumption that the usage of GEDI's PAI beyond the mission lifespan is feasible for regional studies.

4.1 Introduction

Leaf area index (LAI) describes the structure of vegetation as the ratio of leaf area to ground area. LAI is a highly relevant variable for interactions between vegetation and atmosphere and is therefore one of the proclaimed Essential Climate Variables (ECV) (GCOS, 2021). The plant area index (PAI) is defined as half of the total plant area per unit ground surface. Compared to the LAI, not only leaves, but all above-ground plant components such as branches and trunks are considered. However, the PAI is closely related to the LAI (Feret et al., 2008; Myneni et al., 2001; Tang et al., 2012; Weiss et al., 2007). On a global scale, carbon flux and evapotranspiration are massively influenced by LAI. At the regional scale, in addition to its use in models for carbon assessment e.g. (Tharammal et al., 2019), the importance of LAI is, for example, evident in runoff models as it affects not only evapotranspiration rates but also immediate water retention (Huang et al., 2022; Seo & Kim, 2021; Tesemma et al., 2015). Since LAI varies greatly due to seasonality, the use of a temporally as well as spatially (Huang et al., 2022) detailed data set could help improve dynamic modeling of such variables.
Various field methods and techniques exist for the direct (destructive) measurement and the indirect estimation or inverse modeling of the LAI (Fang et al., 2019; Zheng & Moskal, 2009). However, direct or field-based methods in general are time-consuming and therefore do not scale over time and space. In contrast, air-borne light detection and ranging (LiDAR) missions provide high-resolution wall-to-wall data on vegetation structure at landscape level and can be used for spatial mapping of LAI (Wang & Fang, 2020; Yan et al., 2019). Nonetheless, they have limitations with respect to temporal repetition, which usually does not allow for spatial time series that are dense enough to derive phenological information. Global satellite-based products, e.g. from MODIS (Myneni et al., 2001; Qiao et al., 2019) or Sentinel-3 (Fuster et al., 2020) fill this gap by providing regular repetitions, but the resolutions of typically 300 m to 1 km are too coarse for small-scale differentiated studies. To bridge this gap between current global space-borne products and air-borne or field-based missions, the utilization of higher resolution radar and optical sensor satellites, e.g. from Sentinel-1 or -2 data, is promising (Baghdadi et al., 2016; Frampton et al., 2013; Kganyago et al., 2020; Luo et al., 2020; Padalia et al., 2020; Pasqualotto et al., 2019; Wang et al., 2019). Yet, this requires extensive ground truth data and sophisticated modeling strategies, to link optical and radar data to the response variable - the LAI. This leads back to the point that field observations are not sufficiently comprehensive in the spatial and temporal domain and hence do not provide a sufficient baseline for model training. Bringing LiDAR into space with NASA's new Global Ecosystem Dynamics Investigation (GEDI) mission in December 2018 was a big step towards almost global, space-borne, and direct observation of vegetation structure which may provide training and testing samples with a much higher temporal repetition rate. Since January 2019 data sets are available and studies confirm the high potential for ecosystem monitoring (Boucher et al., 2020; Di Tommaso et al., 2021; Healey et al., 2020; Kacic et al., 2021; Marselis et al., 2020; Potapov et al., 2021; Rishmawi et al., 2021, 2022; Wang et al., 2022; Xi et al., 2022).
A Level 2B standard product of GEDI is the plant area index (PAI) that is closely related to the LAI (see section 4.2.2). Version 2 of the PAI product provides information for footprints with 25 m in diameter with a spacing of 60 m along track and 600 m across track (Dubayah et al., 2020). To derive wall-to-wall products from GEDI's PAI, ESA's Sentinel-1 and Sentinel-2 systems are promising candidates, as vegetation structure interferes with radar and optical wavelengths, and since both sensors come with a high spatial and temporal resolution. The radar observations of Sentinel-1 capture vertical vegetation heterogeneity similar to LiDAR observations (Bae et al., 2019) and the multi-spectral scanner observations of Sentinel-2 are indicators for plant physiology, vegetation type and biomass. Previous studies have shown a high potential to estimate LAI from a combination of different spectral bands using non-linear models (Baghdadi et al., 2016; Jiang et al., 2020; Korhonen et al., 2017; Luo et al., 2020; Verrelst et al., 2015; Wang et al., 2019). The aim of this study is to use the PAI as estimated by GEDI to produce wall-to-wall maps by an integration of Sentinel optical and radar data. The integration of GEDI's PAI and Sentinel-1/-2 data was already proven to be feasible (Di Tommaso et al., 2021; Kacic et al., 2021; Rishmawi et al., 2021, 2022). However, previous studies that aimed at wall-to-wall mapping of GEDI-derived variables used temporally aggregation, i.e. long-term means of such variables (Chen et al., 2021; Dorado-Roda et al., 2021; Francini et al., 2022; Healey et al., 2020; Khati et al., 2021; Potapov et al., 2020; Shendryk, 2022; Verhelst et al., 2021). This

is certainly suitable for a detection of large-scale spatial patterns, however, does not support seasonality and is hence not suitable for studies that require the consideration of phenology. The potential of learning the seasonal dynamics by using the different overpasses of the GEDI with the corresponding optical and radar data from Sentinel is, to our best knowledge, not analyzed yet. In this study we test the potential of matching GEDI point data to the temporally closest Sentinel scene in a machine learning approach, with the aim to derive wall-to-wall predictions of the PAI with a temporal resolution that is in accordance with the availability of the frequent Sentinel scenes. Motivation for this study comes from the limited duration of the GEDI mission onboard the ISS. Therefore it is of particular importance to apply and test the trained model beyond its training period. Hence, in this study, we use two of the three years of GEDI data to train a model and we validate the performance with the remaining year. The German state Hesse was selected as a study area because of its heterogeneous landscape of forests, pastures and cultivated land. Since we expect land cover to cause differences in model performance, results are interpreted by taking the different types into account.

4.2 Methods

4.2.1 Study area

The state of Hesse, Germany (about 21 000 km^2) was used as the study area (see figure 4.1). The area with low mountain ranges and a temperate climate with pronounced seasonal changes is composed of 26% non-irrigated arable land, 25% broad-leaved forest, 20% pastures, 13% coniferous forest and 3% mixed forest according to the Coordination of information on the environment (CORINE) land cover inventory 2018 (European Union, Copernicus Land Monitoring System, 2018). The remaining area is mostly covered by urban areas and some smaller patches of other land cover classes that will not be considered in this study.

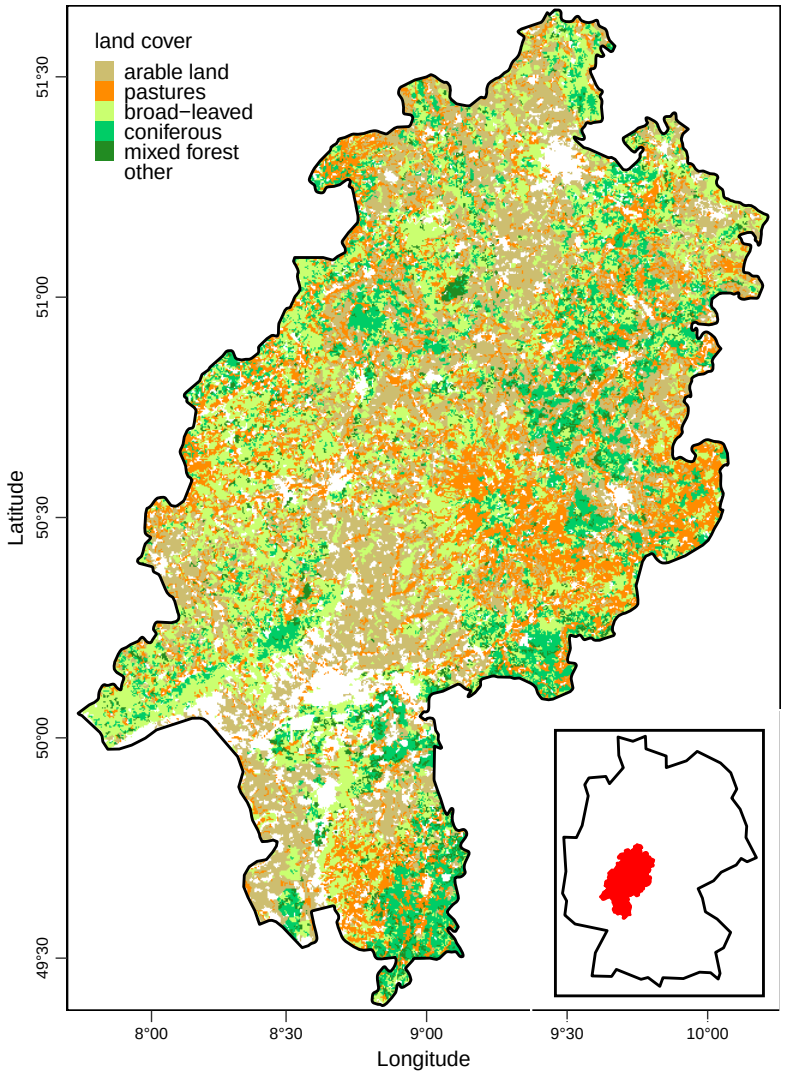


Figure 4.1: CORINE land cover classes for the study area Hesse. Small sketch in the bottom-right corner shows the location of Hesse within Germany.

4.2.2 Data sources

Data pre-processing was mainly executed in Google Earth Engine (GEE) to handle the large data volume (Gorelick et al., 2017). All steps of model training and evaluation were performed in R (R Core Team, 2022). For the availability of scripts and data sets see the data availability statement (see

Plant Area Index GEDI data

The GEDI mission was operational from March 2019 to March 2023 and is tentatively scheduled to provide additional data beginning in the fall of 2024. All orbits that intersected the study area within the study period were identified using the rGEDI package in R (Silva et al., 2021). The level 2B version 2 PAI product was obtained within GEE and contains footprints with a diameter of 25 m for eight parallel tracks spaced 600 m across track and 60 m along track (Dubayah et al., 2020). For our study area it offers about one daily overpass but no complete spatial coverage or overlap (Dubayah et al., 2020). Of the available 1 441 GEDI overflights from April 2019 until December 2021, 1 347 potential overflights remained after excluding

footprints with a low quality flag and GEDI's sensitivity value below 0.9 in the Level 2B PAI product (Tang et al., 2019).

Sentinel-1 data

To provide spatially continuous predictors for PAI, the ground-range-detected high-resolution product of Sentinel-1's C-Band radar, retrieved directly within GEE, was used for the study. Both polarization modes (VV: vertically transmitted and received, VH: vertically transmitted and horizontally received) with a resolution of 10 m and their difference and ratio (VV-VH, VV/VH) were used as potential predictor variables since these metrics showed to be sensitive to seasonal changes in forested areas (Frison et al., 2018). For the selection process of appropriate scenes and the matching with GEDI data, see section 4.2.3.

Sentinel-2 data

Sentinel-2 Level 2A multi-spectral data, retrieved directly within GEE, provided the second source of potential predictor variables for the PAI. All available spectral bands except band 10, which lies within the atmospheric absorption bands of water vapour and carbon dioxide, were selected as potential predictors in their original resolution. The normalized difference vegetation index (NDVI), enhanced vegetation index (EVI) and the inverted red-edge chlorophyll index (IRECI), that copes well with the problem of oversaturation in LAI products, were computed (Frampton et al., 2013). For the selection process of appropriate scenes and the matching with GEDI data, see section 4.2.3.

4.2.3 Processing Methods

Temporal and spatial matching of GEDI data with Sentinel-1/-2

Temporally matching Sentinel data were queried for each of the 1 441 available GEDI overpasses. We allowed a temporal difference of +/- 3 days for Sentinel-1 and +/- 5 days for Sentinel-2 and a maximum cloud cover of 50 % per scene. If more than one scene was available within this window, we chose the temporally closest cloud-free pixel to the GEDI acquisition date. All pixels selected according to this procedure were combined into a mosaic and used in the following for the corresponding GEDI recording. If no adequate Sentinel data were available within this time window, the GEDI overpass was not considered for further analysis. For each of the 25 m footprints, all intersecting pixels were extracted from the mosaic and used as predictor variables for the model.

Data quality and sampling strategy

The availability of sufficiently large data sets allowed generous filtering to use only high quality data points, especially because the GEDI location error adds additional uncertainty. Therefore, GEDI footprints were excluded if located within a 1000 m buffer around pixels marked as clouds or cloud shadows or flagged as defective by Sentinel's scene classification band (SCL). If footprints were flagged in the GEDI quality band or covered mixed land cover classes according to the CORINE inventory they were excluded from the study. Additionally, only the five main land cover classes were taken into consideration (see section 4.2.1). Outliers were rigorously eliminated from Sentinel and PAI data by excluding the upper and lower 0.1 % of the data points. This temporal and spatial matching as well as sub-sampling resulted in 7 132 148 valid observations from 1 327 GEDI overpasses between April 2019 and December 2021. Due to the massive amount of data points we randomly sampled 150 000 of those high quality points for model training (127 449 points) and testing (22 551 points, see figure 4.2). The number of samples for the training data set across land cover classes and month is visible in figure 4.2 and roughly follows the distribution corresponding to the proportions of land cover classes (see section 4.2.1) and prevailing weather conditions of the seasons.

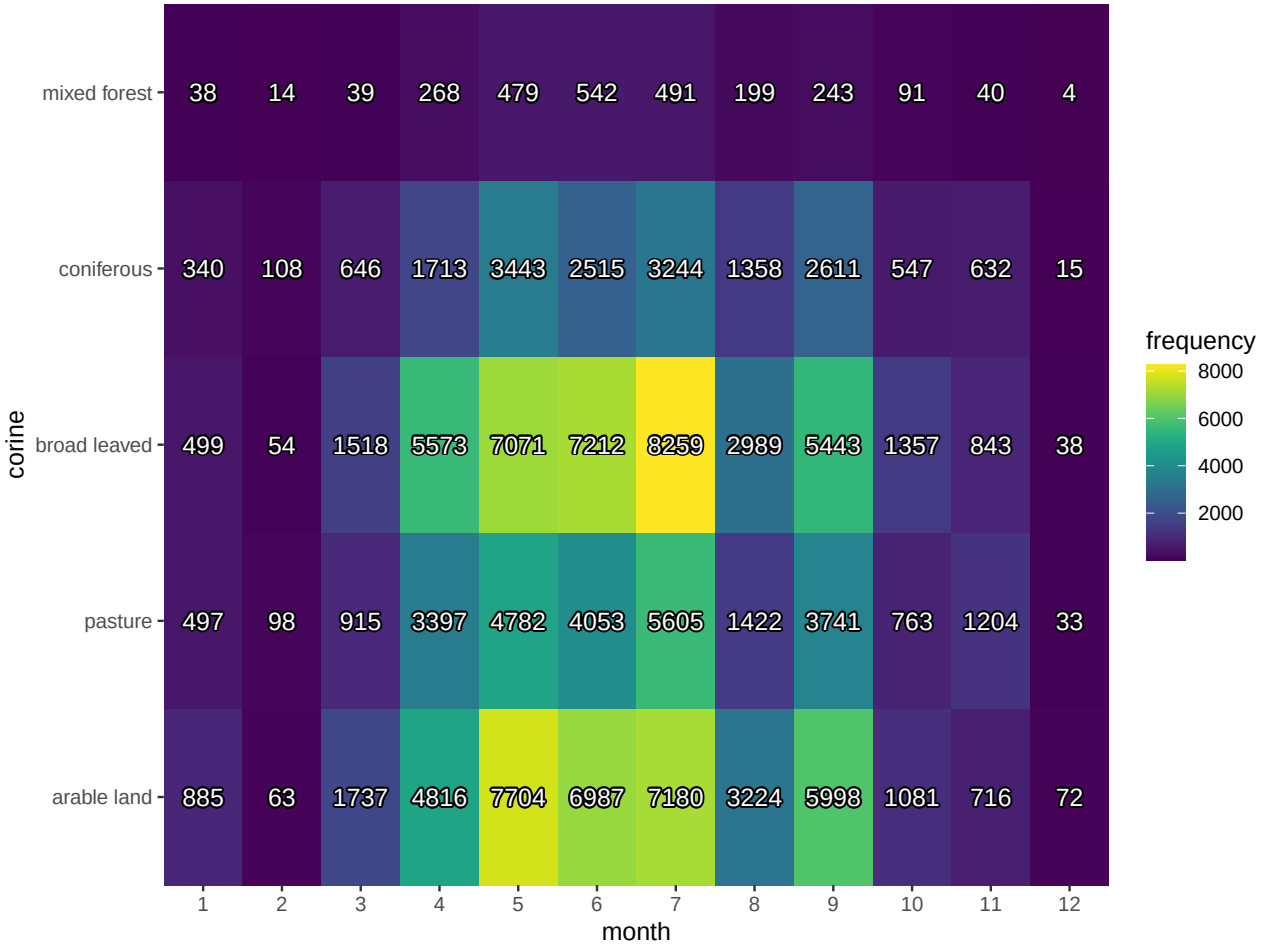


Figure 4.2: Number of available PAI observations in training data by month and land cover class.

Model training

In sum 16 predictors formed the set of training data: two from Sentinel-1 (see section 4.2.2); nine Sentinel-2 channels and five derived indices. A random forest machine learning approach was used to model PAI with the R packages caret (Kuhn, 2008), ranger (Wright & Ziegler, 2017) and CAST (Meyer, 2018). During model tuning a spatial cross-validation (CV) was applied to evaluate the potential of models to predict GEDI PAI for new spatial areas. Therefore, we divided the data into 20 folds by keeping footprints from one orbit placed in the same fold, to ensure spatial and temporal independence between folds. For training, we only took data from 2019 and 2020 and used the 2021 data to assess the ability of the trained model to make predictions beyond the training phase. During model tuning, a spatial forward feature selection as explained in Meyer et al. (2018) and Meyer, Reudenbach, et al. (2019) was applied: from all 18 potential predictor variables, only those that led to the lowest spatial CV error (RMSE) were selected for the model training. The models were trained with 50 trees and the hyperparameter mtry was tuned with three different numbers of variables included in the respective training iteration step (2,6,11). After tuning, model performance was quantified with the remaining 2021 data by computing R^2 and RMSE. Since the land use classes differ strongly in their variance, the RMSE was additionally normalized with the standard deviation (RMSE/sd) to allow a direct comparison of the performances between the different classes. To visually interpret the spatial patterns of the predictions, we applied the model to monthly composites of the predictors for the year 2021.

4.3 Results

4.3.1 Comparison of the temporal phenology dynamics of GEDI PAI and Sentinel-2 NDVI

To assess the plausibility of the GEDI PAI dynamics compared to the well-established NDVI, we first assessed the time series of both data sets. The temporal development of PAI from GEDI and NDVI from Sentinel-2 show clear patterns of vegetation phenology. While the NDVI reflects the phenology of vegetation by spectral properties responding to green biomass, the PAI reflects phenology by a change in the structure. The temporal dynamics in both, PAI and NDVI, reflect general differences especially during the summer months between arable land and pastures (lower values) and different forest types (higher values), but feature a large within-class variability (figure 4.3). The forest bud burst and leaf growth is well represented in the NDVI between February and May and also clearly visible in the PAI. The three forest types show different seasonal patterns in the NDVI, which also corresponds to the patterns in PAI. For pastures, the similarity between NDVI and PAI dynamics is less distinctly compared to the forest classes, as expected. However, the same tendencies can be observed considering an increasing variability during the summer months. The most obvious difference in PAI compared to what is reflected by the NDVI is observed for arable land. Here the NDVI shows clear seasonal patterns, which are not reflected by the median monthly PAI.

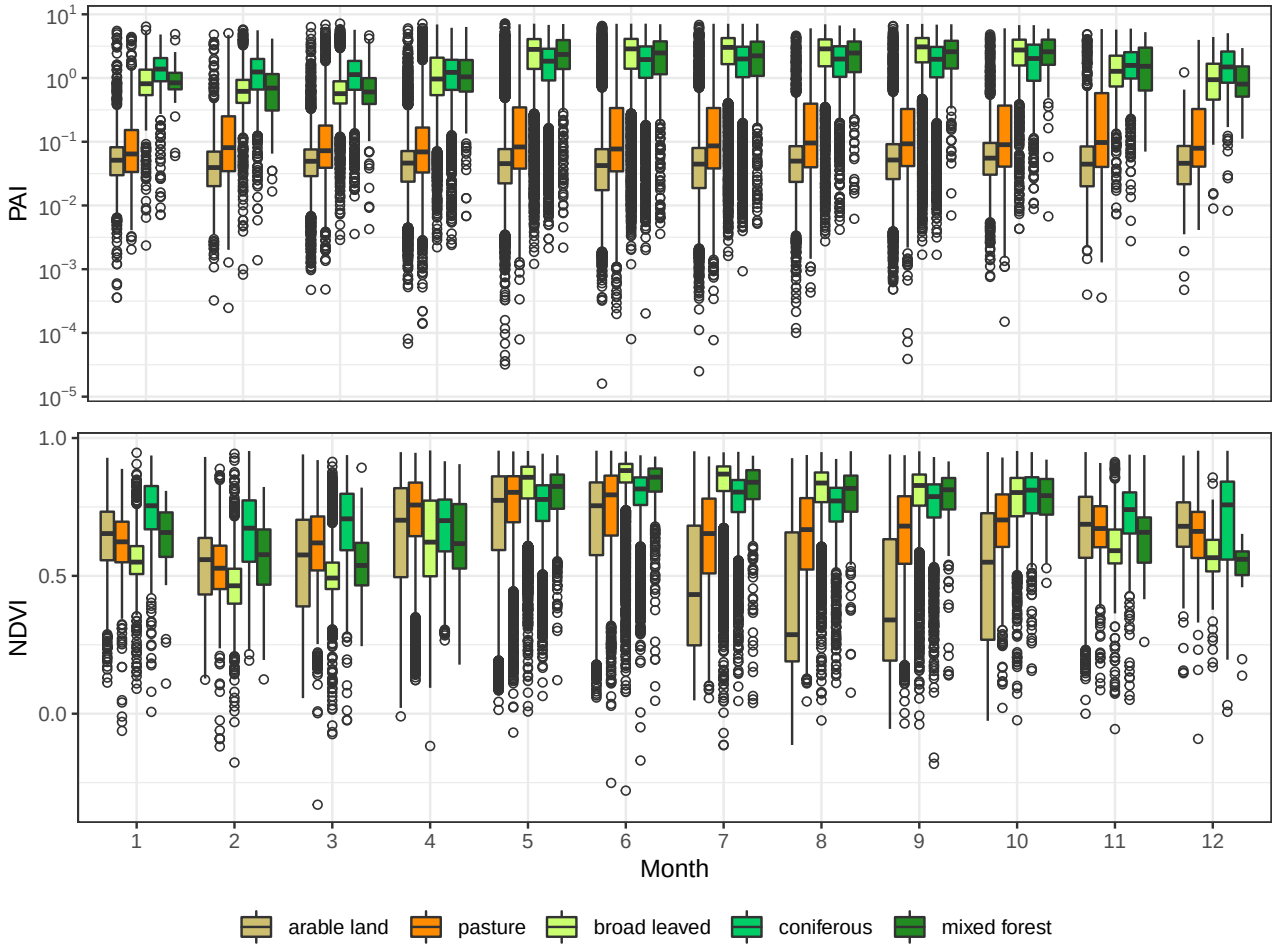


Figure 4.3: Temporal variability of the GEDI's PAI (top) and Sentinel's NDVI (bottom) per land cover class. The NDVI values are extracted from the same locations as the PAI and include a total of 150 000 points across the years 2019-2021.

4.3.2 Assessment of the trained model

The spatial forward feature selection revealed that eleven out of 16 predictors were useful for spatial predictions of the PAI (figure 4.4). The best seven predictors include two visible bands (bands 3 and 4), two Sentinel-1 bands (VV and VH), two near-infrared bands from the red edge spectrum (bands 5 and 7) and one vegetation index (EVI). Both Sentinel-1 indices, and the two vegetation indices from Sentinel-2 data as well as the short wave infrared band (band 9) could not improve the model further.

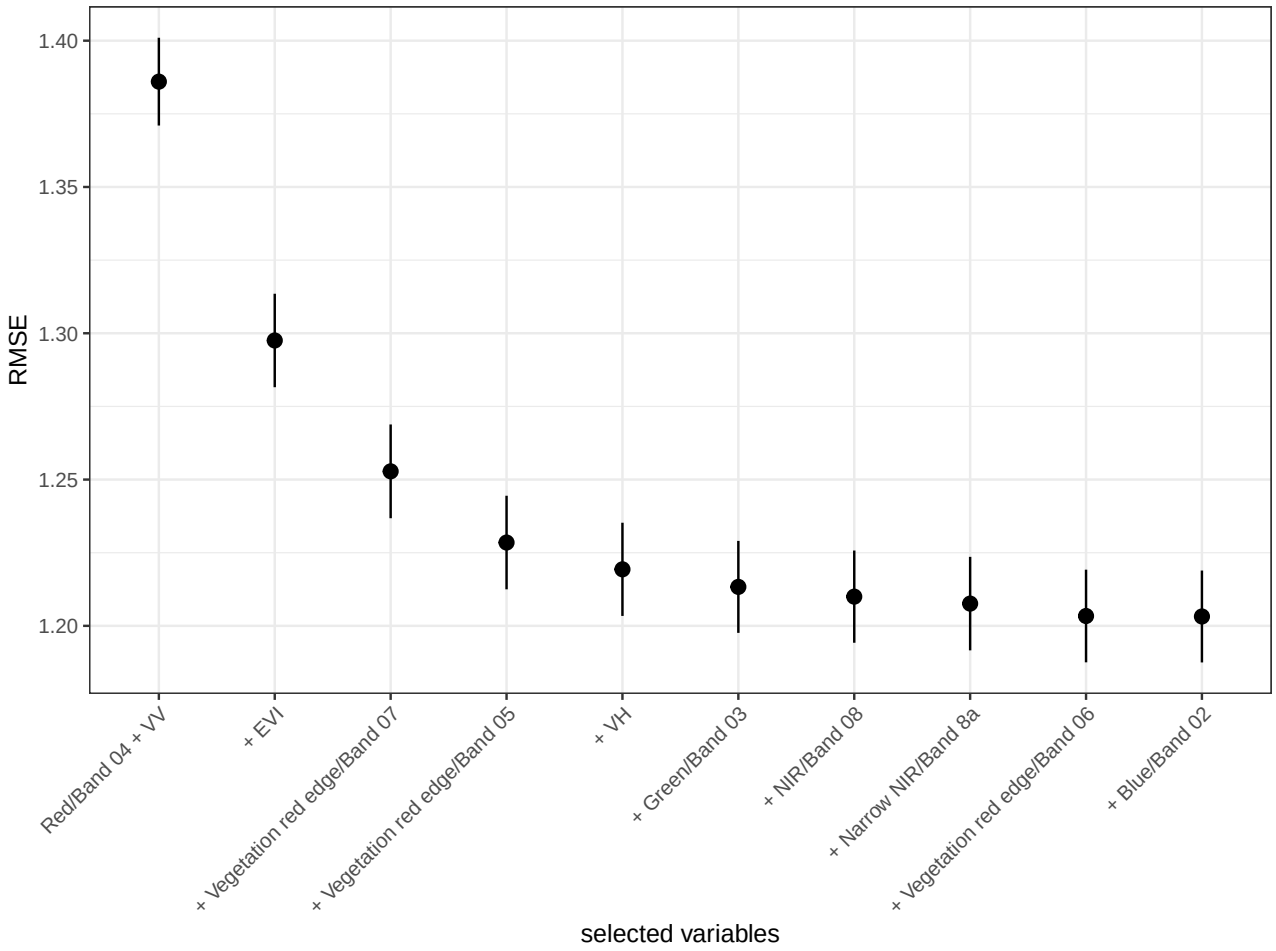


Figure 4.4: Ranking by RMSE of the predictor variables that were chosen in the feature selection to be included in the model.

The cross validation error for the 2019 and 2020 data set reached a R^2 value of 0.45 (RMSE: 1.20, RMSE/sd: 0.83). Based on the independent validation data set from 2021, the model reached an overall R^2 performance of 0.40 (RMSE: 1.12, RMSE/sd: 0.78). The difference between the cross-validation performance and the external validation statistics can be regarded as an indicator of the models potential to be applied beyond the training period. A large difference would indicate that the model performs significantly better within the training period. In our case, the cross-validation error (gray) agrees well with the external test data set when validated across the land cover classes.
Figure 4.5 shows a heat map of the RMSE/sd for the independent validation by land cover class and month. Arable land performed worse than forest classes, overall and for each individual month except for September. In the winter months they fall far behind other classes with RMSE/sd above 5 (see figure 4.5). Forested classes follow a seasonal pattern with better performance in the summer months. The ranking of the performances of the forested classes alternates (see figure 4.5). In most months, coniferous forest has the best performance, despite an overall worse performance than mixed or broad-leaved forest (see table 4.1). The monthly breakdown of the external validation (see figure 4.6) demonstrates that all RMSE/sd values are below 1 and reveals a seasonal pattern with a better performance during the summer months.

Regarding the error metrics throughout the whole year split up by land cover class as shown in table 4.1 RMSE/sd values stay below 1 for all forested classes and above 1 for arable land and pastures.

Table 4.1: Prediction errors that were externally validated against a testing data set from a different year, differentiated by land cover class (white rows). The gray row indicates the error from the cross validation within the model training.

corine	Rsquared	RMSE	RMSE/sd
mixed forest	0.29	1.25	0.84
coniferous forest	0.20	1.21	0.90
broad-leaved forest	0.25	1.44	0.88
pastures	0.14	0.91	1.06
arable land	0.05	0.75	1.46
all	0.40	1.12	0.78
all (cross validation error)	0.45	1.20	0.83

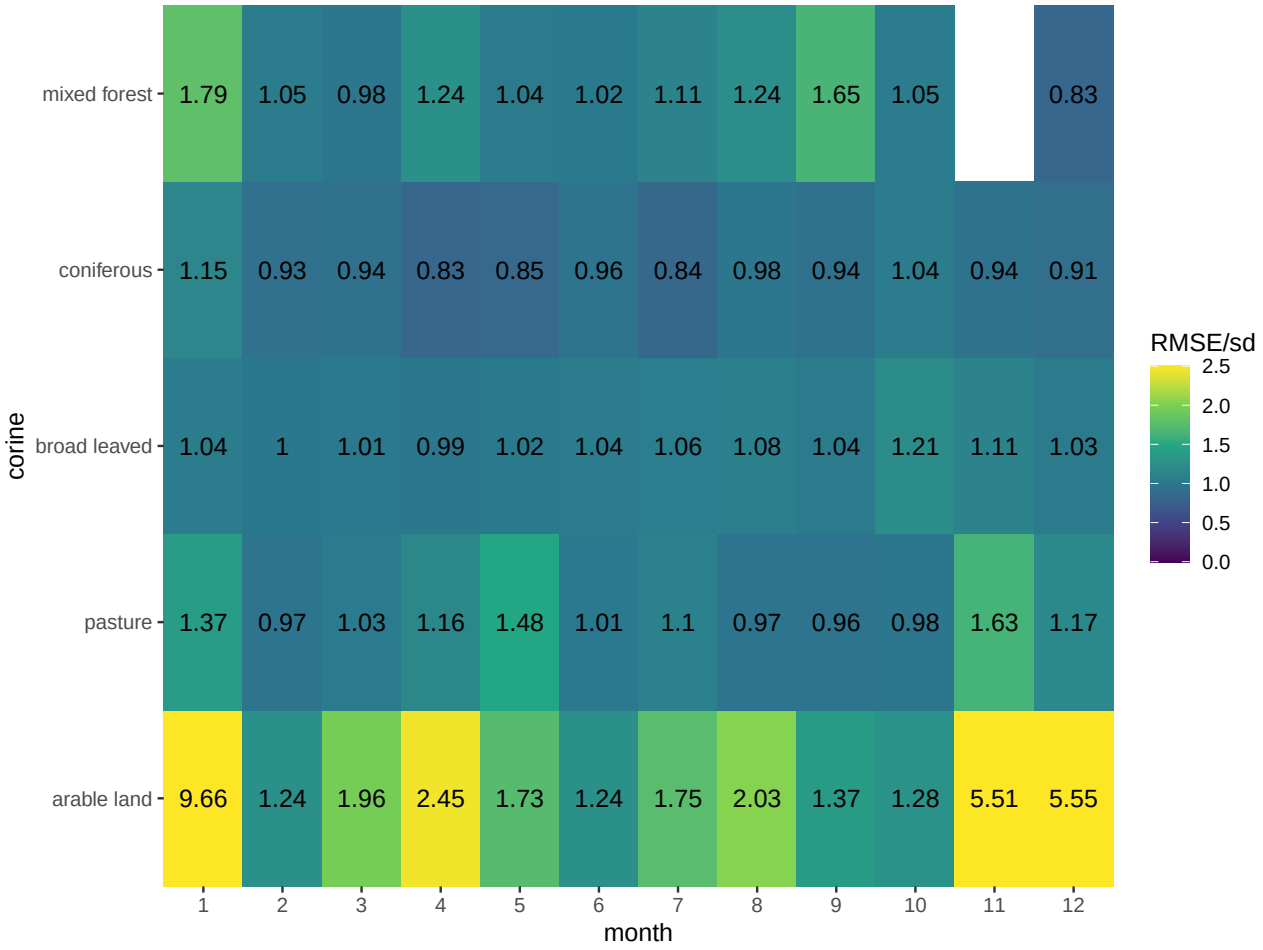


Figure 4.5: Assessment of prediction errors (RMSE/sd) for the 2021 validation data by month and land cover class.

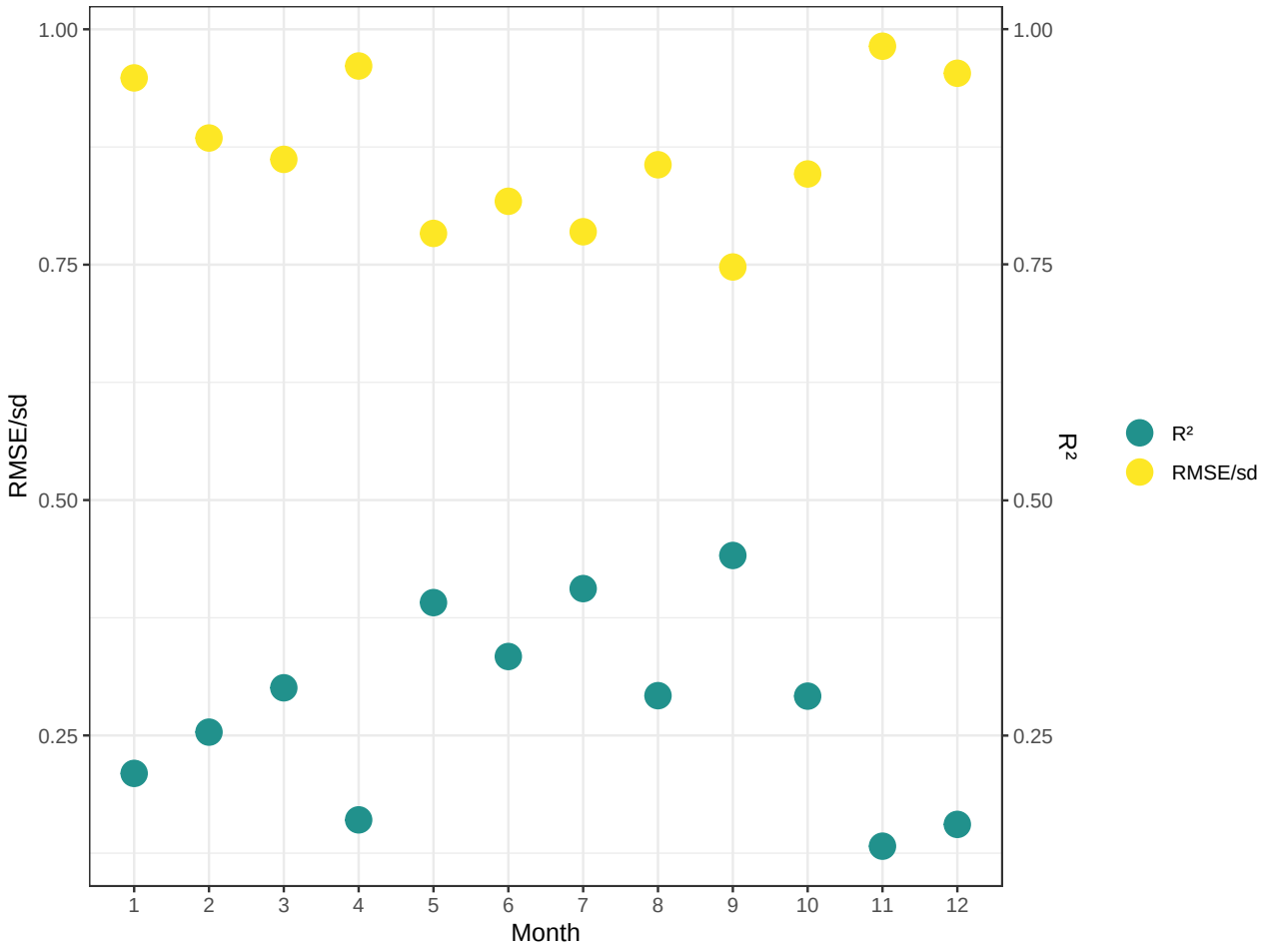


Figure 4.6: R2 and RMSE/sd error metrics for the 2021 validation data by month.

4.3.3 Spatio-temporal prediction

Spatial wall-to-wall predictions for the study area were computed based on monthly median values of Sentinel-1 and Sentinel-2 data from 2021. Figure 4.7 shows a sequence of four PAI predictions covering the growing season 2021. For a comparison of all 12 months see figure 4.A1 in the appendix. The general expectation that PAI increases significantly during the growing season and that there are significant differences between forest and non-forest are evident in figure 4.7. This impression is also confirmed by analyzing the predicted PAI dynamics separately for each land cover class (boxplots in figure 4.8, analogue to figure 4.3). Predictions during January and November, however, are based on considerably less available pixels than during the rest of the year (figure 4.9). To check the agreement of the annual dynamics, Pearson's correlation coefficient was calculated between the monthly medians for the 2021 observations and the 2021 predictions. As detailed in table 4.2, this reveals a significant positive correlation for all forest classes but can not find a significant correlation for arable land and pastures.

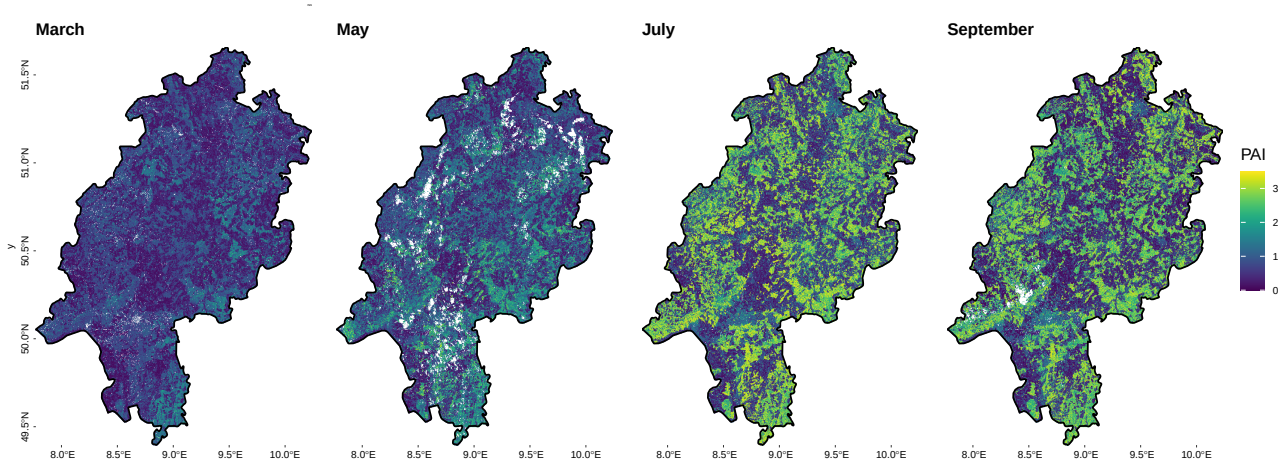


Figure 4.7: Spatial prediction of PAI for March, May, July and September for Hesse, Germany.

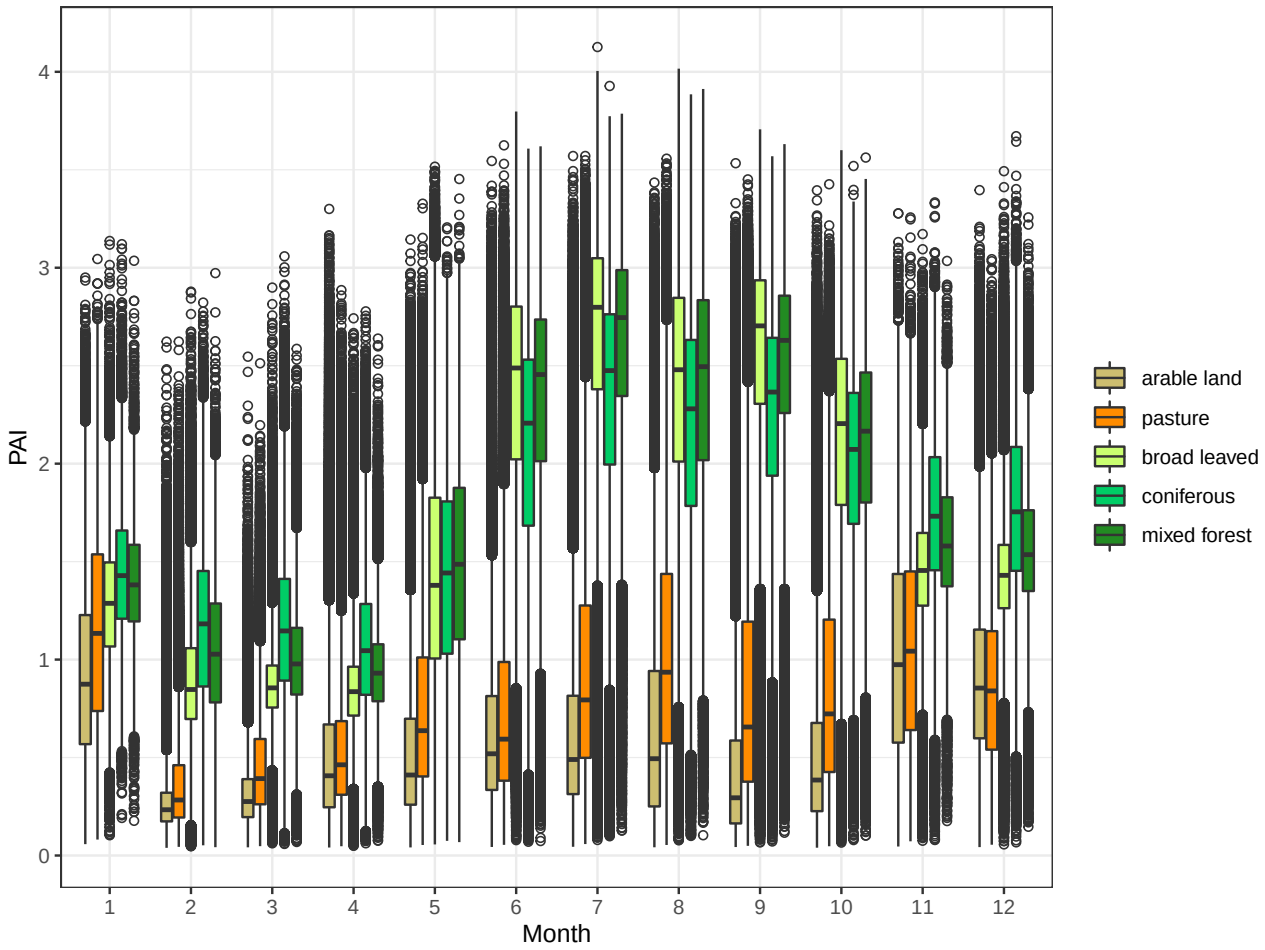


Figure 4.8: Temporal variability of the GEDI's PAI per land cover class including all pixels of monthly spatial predictions for 2021.

Table 4.2: Pearson correlation of median monthly values between observations of testing data set and pixel values of area-wide spatial prediction. Significant results, in terms of the p-value, are marked bold. *Due to missing values in November for mixed forest, this value is only based on 11 median monthly values.

corine	correlation	p-value
arable land 12	0.018	0.956
pastures 18	-0.004	0.990
broad-leave forest 23	**0.910**	**0.000**
coniferous forest 24	**0.729**	**0.007**
mixed forest 25 *	**0.821**	**0.002**

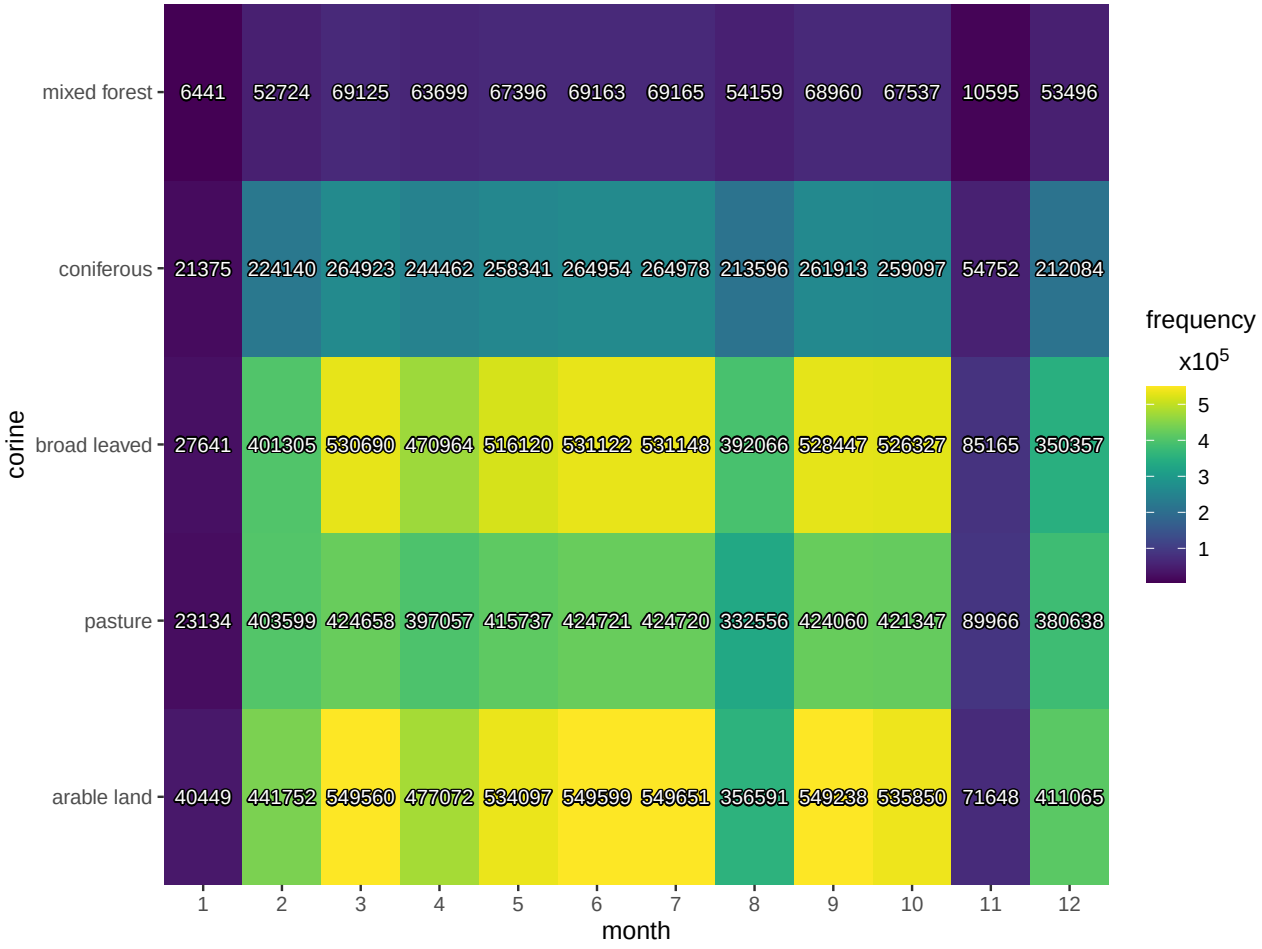


Figure 4.9: Number of available pixels in the monthly spatial predictions by land cover class.

4.4 Discussion

GEDI's PAI product generally reflects the expected temporal variability in the study area. In forested areas the increase in PAI during the early growing season corresponds to bud burst and leaf growth. On pastures, a slight gain together with an extended variability marks the start of the growing season. Seasonal patterns are more distinct for forested areas than for pastures and arable land. High variability of PAI on pastures during the summer months can be explained by irregular management activities such as mowing or grazing as well as isolated trees.

The spatial prediction of PAI also generally reflects the expected seasonal dynamics described in section 4.3.1 (see figure 4.7 and appendix 4.A1). PAI increases in forests (see figure 4.1 and figure 4.8) during

spring and fluctuates in summer. The latter is probably the result of alternating extreme weather conditions in Germany. Related effects add to existing tree conditions induced by recent droughts (DWD, 2019, 2020). For arable land and pastures, the temporal dynamic in figure 4.8 looks similar to the observed seasonal course (figure 4.3). For January and November the number of valid pixels is extremely low compared to the other months due to persistent and area-wide cloud coverage (see figure 4.9 and appendix 4.A1) and do therefore not allow for meaningful interpretation.

The PAI prediction was generally more accurate for the forest classes with broad-leaved forest and mixed forest performing even better than coniferous forest. This can likely be explained by the larger vertical variability of forests compared to other land cover types like arable land or pastures, and different moisture properties. Both GEDI LiDAR and the Sentinel sensors can capture differences in PAI more easily if vegetation height and structural variability are large. However, mixed forest shows inconsistency in performance throughout the year. This may either originate from its diverse ecological and structural composition and related spatio-temporal anomalies throughout the seasons, or from the relatively small number of reference footprints compared to the other land cover classes. Uncertainties in class assignment can be expected, especially from the delineation of pure deciduous or coniferous forest and mixed forest. This may be an additional factor causing lower model performance in the mixed forest class. In principle, and despite the cross-check with the CORINE land cover classes, heterogeneous training footprints cannot be excluded from the analysis with absolute certainty. Since pastures and arable land in the study area are typically embedded in more heterogeneous structures (small settlements, roads, edges) that are not captured by CORINE, it is likely that these classes are more affected by impure footprints compared to larger homogeneous forest areas.

The temporal dynamics between observations and area-wide predicted pixel values for the test year 2021 agree well for forested areas. A significant positive correlation was found for the forested classes but not for arable land and pastures (see table 4.2). This may partly be explained by the overall smaller variance in PAI as well as individual growth and harvesting or mowing events of different crop types. Even though GEDI and Sentinel acquisition times can lie no further apart than three days, these areas can change significantly in the meantime, resulting in a mismatch between the data sets and, hence, an unfavorable effect on prediction quality. The monthly performance across all CORINE classes (see figure 4.6) shows a clear seasonal pattern with better performances during the summer months. The monthly RMSE/sd across all land cover classes scores below 1 and therefore lower than the standard deviation in the original data set. This proves that our model is generally usable throughout the year. January, April, November and December are the months with the weakest performance. This behavior may partially be attributed to sparse data availability and therefore extremely low RMSE/sd values for those months for arable land (see figure 4.5).

A performance comparison with other studies can only provide limited indications in this case. Most other studies used field-based measurements as reference and observed the related yet not identical LAI as the response variable. The accuracy of GEDI's PAI for different landscapes has only been explored in few studies. Dhargay et al. (2022) tested the accuracy of GEDI's PAI in Australia and found rather poor agreement and a significant underestimation compared to estimates derived from air-borne LiDAR data. According to their assessment, however, both the complex study area and the time lag between air-borne and space-borne observations can be possible sources of error. They further used only about one month of GEDI data, which does not cover a time frame large enough to study temporal dynamics. Kacic et al. (2021) examined the integration of GEDI's PAI and sentinel data in Paraguay's forests and aggregated the data across one entire dry season (RMSE = 0.3, R^2 around 0.5). Rishmawi et al. (2021) produced contiguous PAI maps at 1 km resolution over the United States by integrating GEDI and VIIRS data (RMSE = 0.09, R^2 = 0.76). Since we use data with high spatial and temporal resolution, the performance of our models is worse, as would be expected, but difficult to relate directly. Few studies, including Miranda et al. (2020), use PAI winter observations to calculate the woody proportion of forest areas. With this information they then calculate effective LAI during the growing season. This approach could potentially also be applied to GEDI studies and should be investigated further.

Other study design components, that complicate a direct comparison include, study areas, vegetation types, study seasons, variance in the data, validation strategies and other aspects of the study design

that vary considerably between publications. In addition, RMSE/sd values which facilitate a comparison across studies are rarely communicated and many previous studies restricted data analysis to plots that were clearly dominated by a single species (Brown et al., 2019), or featured homogeneous cover (Cohrs et al., 2020; Korhonen et al., 2017). Nevertheless, it is helpful to put the results of this study in the context of previous research on LAI estimation. The GEDI models in our study generally reached lower R^2 values for arable land and pastures than models of previous studies. Frampton et al. (2013) and Gitelson, Viña, et al. (2003) report a training error R^2 value range of 0.36 to 0.88 in their LAI models, differing drastically from the performance of our corresponding model (independently validated $R^2 = 0.05$). For pastures, Baghdadi et al. (2016) presented R^2 values between 0.65 and 0.89 in contrast to 0.14 in our study. For agricultural areas, RMSE values of previous studies ranged from 0.44 to 0.68 (Delegido et al., 2011; Luo et al., 2020; Verrelst et al., 2015) which is slightly better than our model (RMSE = 0.75). In the case of pastures we were able to obtain a lower error (RMSE = 0.91) compared to a study by Wang et al. (2019) (RMSE = 1.09). For the forest classes, the GEDI-based approach achieved a slightly lower performance compared to other studies. While our RMSE scores are above 1.21, previous work by Korhonen et al. (2017) and Meyer, Heurich, et al. (2019) reached values between 0.8 and 0.9 in models based on field-observations. The study of Cohrs et al. (2020) in coniferous forest reached RMSE values of 0.63 to 0.89 with linear models while our validation reached an RMSE of 1.21. For deciduous forest, Brown et al. (2019) achieved RMSE values of 1.55 using the Sentinel Application Platform (SNAP) algorithm, and up to 0.47 using an optimized algorithm, compared to an RMSE of 1.44 achieved in this study. In addition to previously described limitations in comparability, it is likely that the systematic, but spatio-temporally irregular coverage of GEDI footprints bares more challenges for the statistical modeling process compared to studies which do not use GEDI data. Luo et al. (2020) for example used time series from a fixed set of plots and analyzed more homogeneous data than expected from the shifting GEDI footprints.
The different validation methods also have an impact on the comparability of the results. The spatial and temporal cross-validation used in this study tests the applicability to new data, which in principle leads to poorer validation measures, but provides a realistic picture for prediction on new data. However, regardless of the absolute performance, the good agreement of error measures between the model-internal cross validation and the validation with the external test data set (see table 4.1) shows that modeling of LAI in regional studies is feasible for applications beyond the lifetime of the GEDI mission.

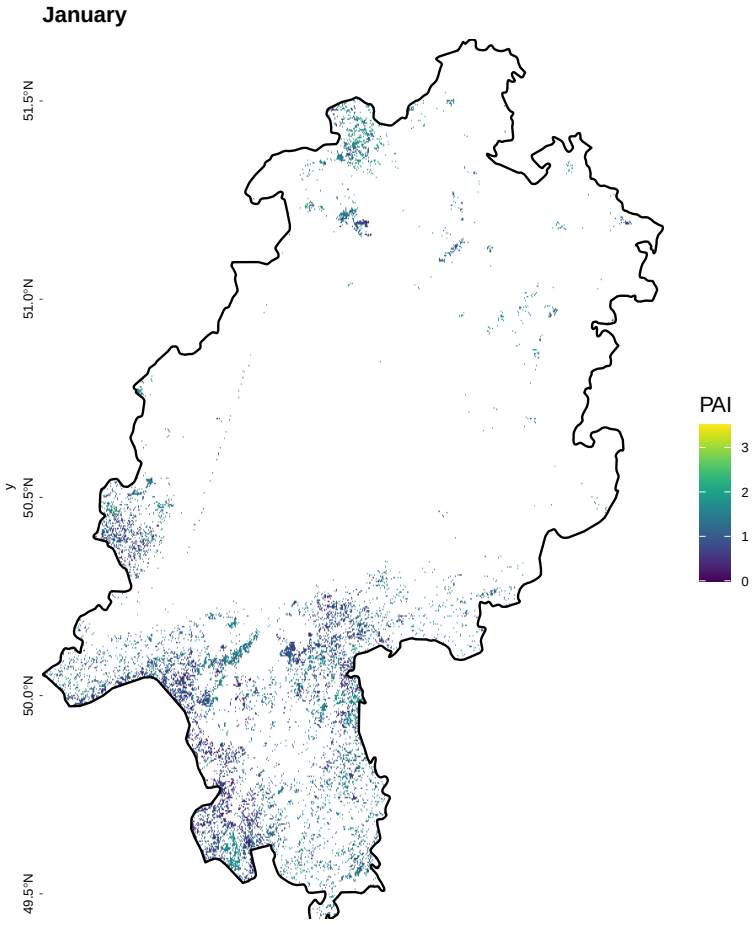
4.5 Conclusion

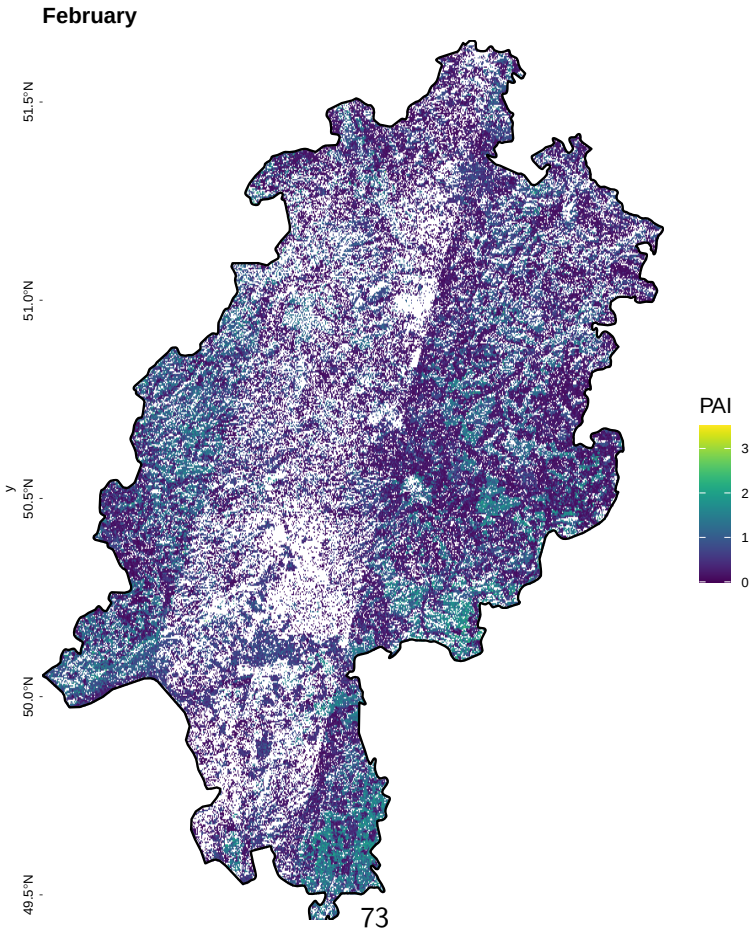
This study presents a first approach to match GEDI's PAI observations with their closest Sentinel-1 and -2 pixels in space and time to compile monthly wall-to-wall maps of PAI in heterogeneous landscapes. The high spatial resolution of the predictions and regular repetition rates improve the availability of information compared to current operational global LAI products. This study demonstrates that a stable year-round monitoring of the key vegetation variable PAI in a heterogeneous landscape is possible. However, our findings reveal that there are great differences in predictive power across land cover classes, the use of multi-temporal variables might be an option to optimize the model further. We further found the prediction of PAI within forests to be more stable possibly due to its variability in vertical structure. Overall our results show good agreement for predictions within the time range of our training data and one year beyond. This leads to the conclusion that our approach can even be applied to time periods outside GEDI's life-span.

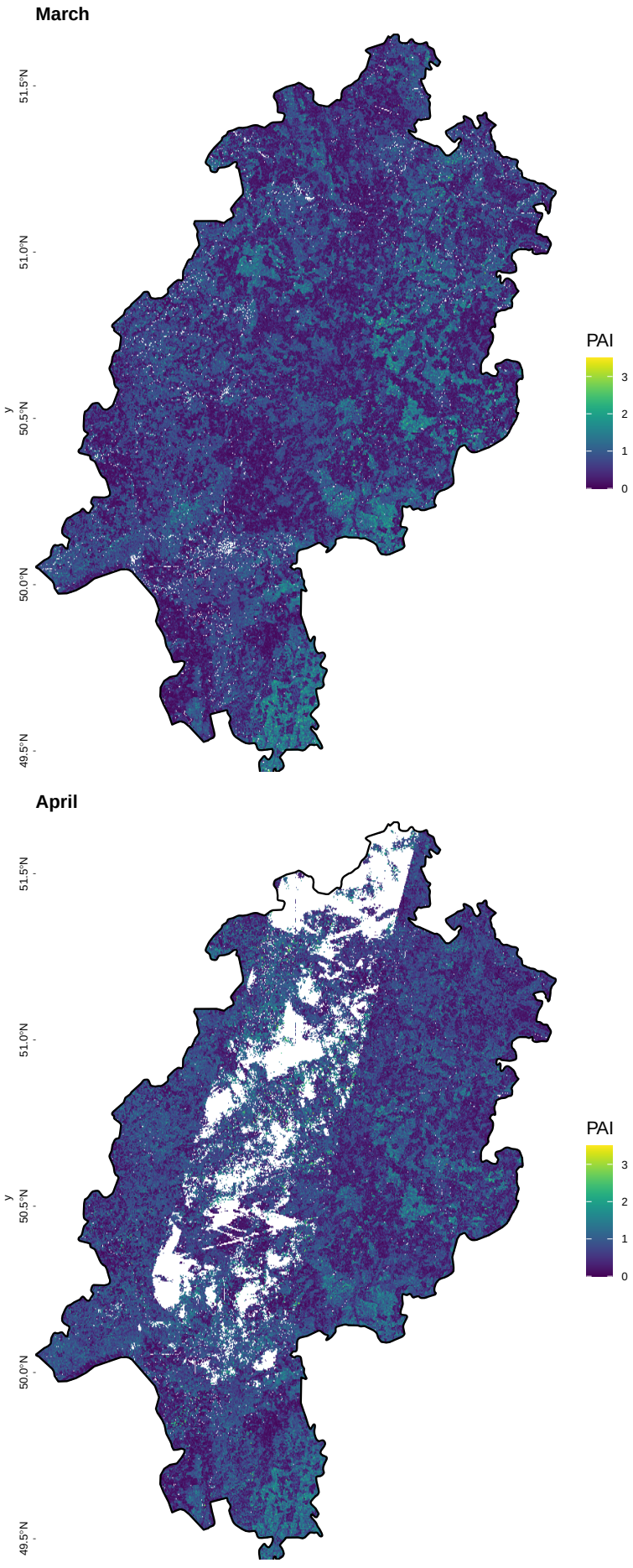
Data Availability Statement

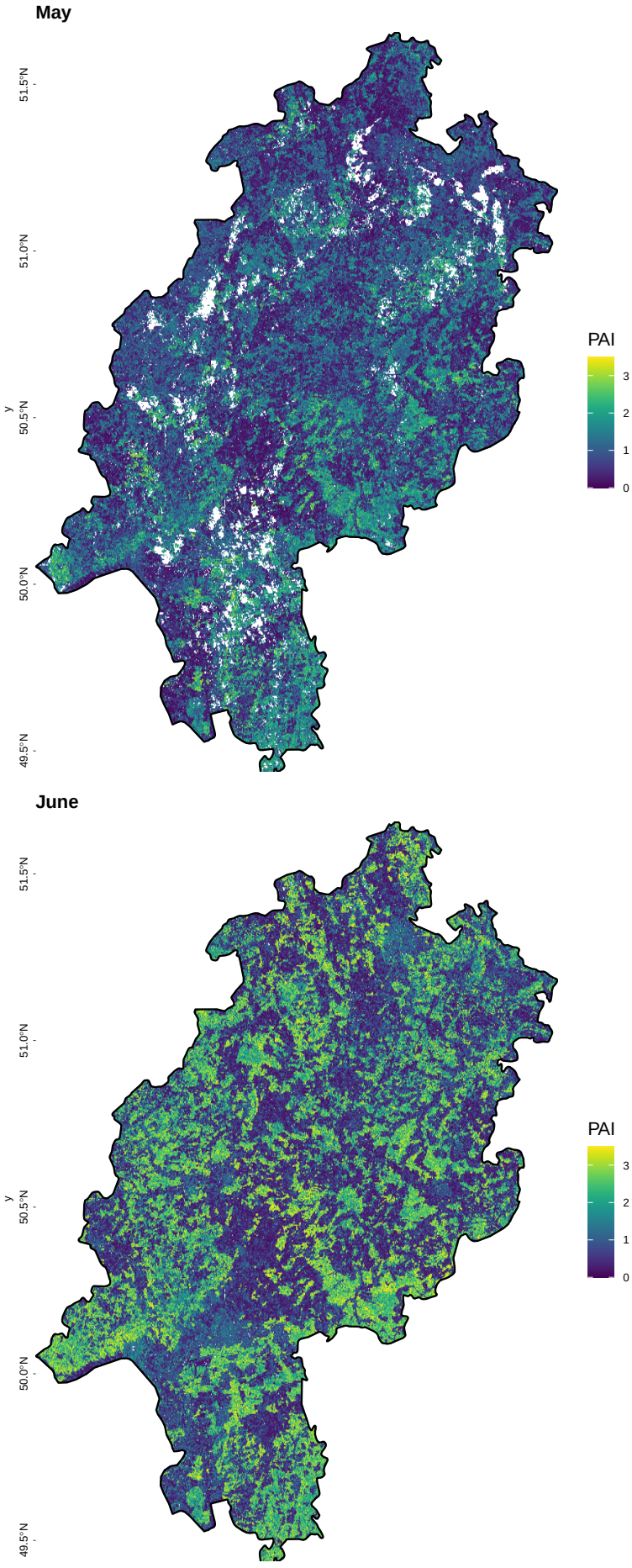
R scripts used for this study are available under a GPL 3.0 license as Git repository at github.com. A release of the Git repository to reproduce the results of the study is available at https://github.com/aliceziegler/GediEngineR, accessed on 14.3.2023. Data supporting the findings of this study are freely available and can be accessed using the routines on Github.

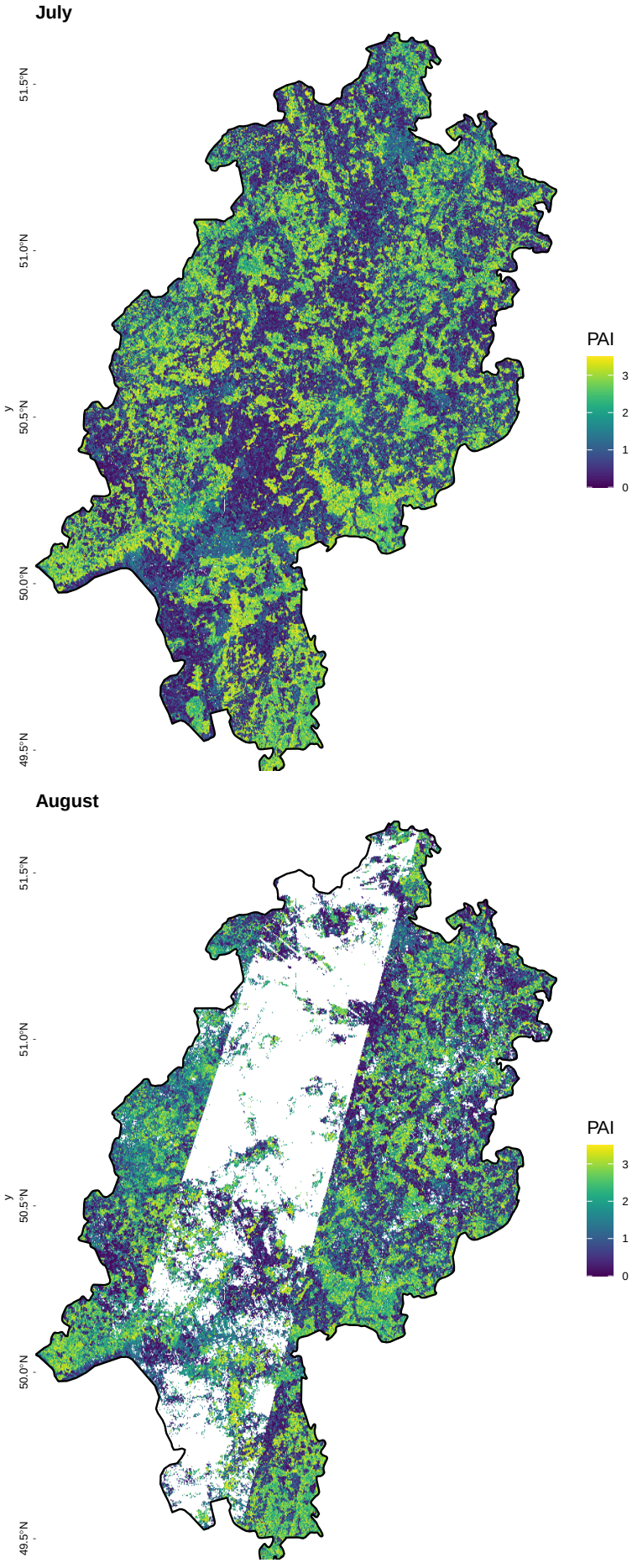
Appendix

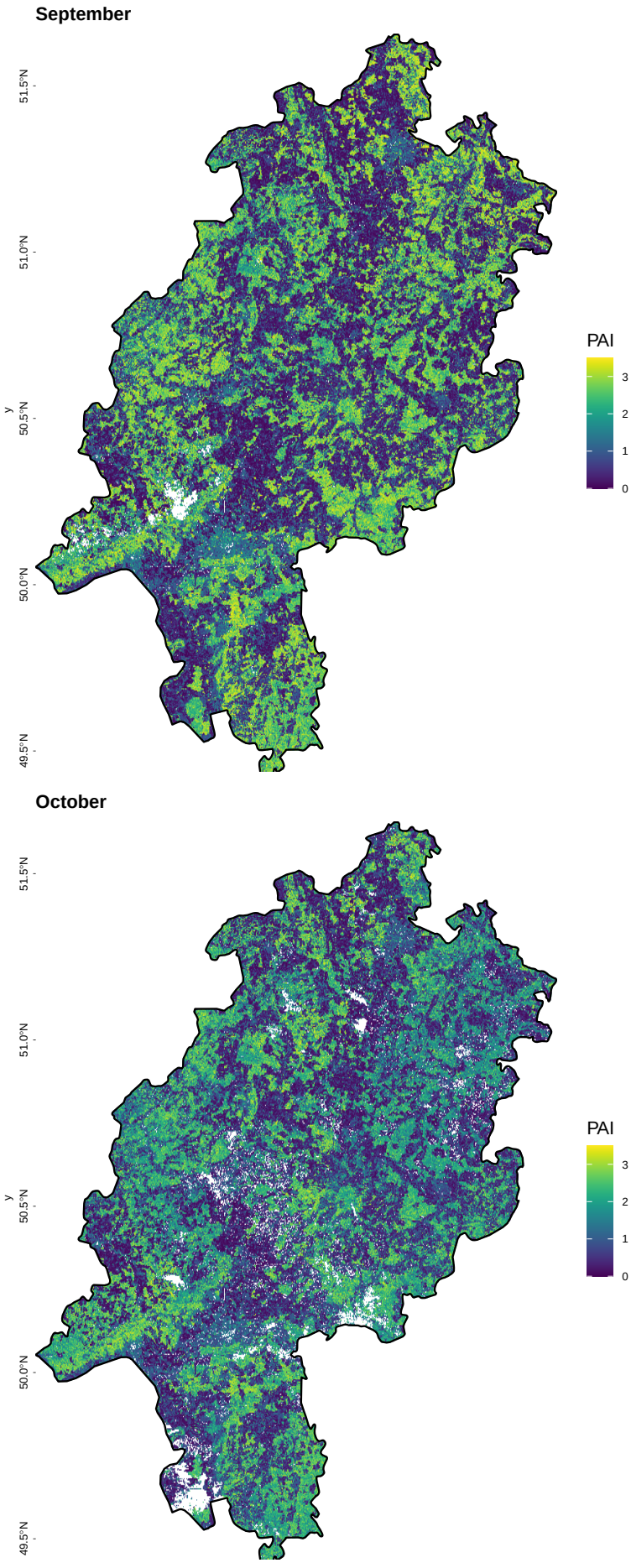

January
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0


February
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0


March
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0
April
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0


May
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0
June
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0


July
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0
August
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0


September
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0
October
51.5°N
51.0°N
50.5°N
50.0°N
49.5°N
y
PAI
3
2
1
0

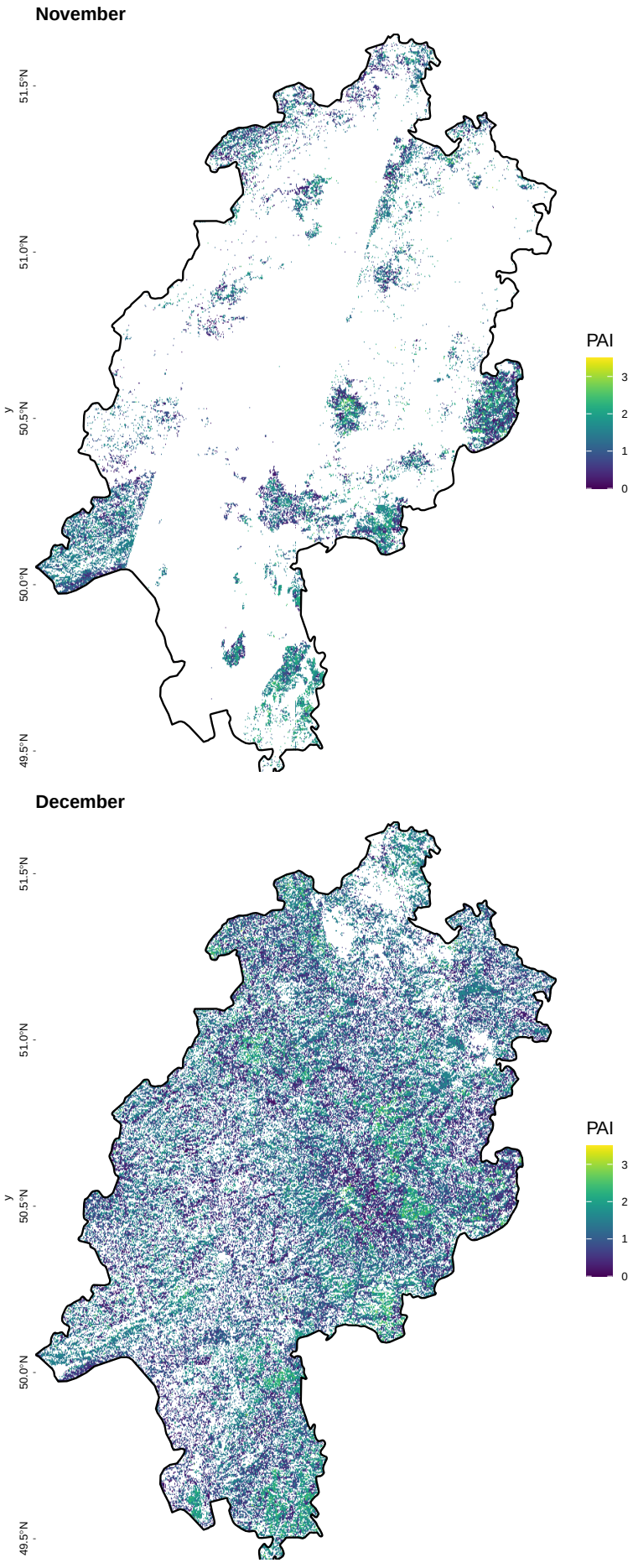


Figure 4.A1: Monthly predictions of PAI in Hesse (Germany).

Chapter 5

Synthesis

With this thesis I aimed at deepening the understanding of the utility of LiDAR data for the quantification of different biodiversity measures depending on the represented ecological question. LiDAR data offer unprecedented possibilities to quantify structure via remote sensing approaches and therefore I assumed a great potential for biodiversity mapping, nonetheless with challenges. In the individual chapters of my thesis, I aimed at addressing several challenges to contribute to a better applicability of LiDAR for biodiversity research. In the following paragraphs I will draw conclusions about the contributions of the three case studies to the challenges outlined in chapter 1.1.4.

5.1 Disentangling of structural LiDAR information from co-factors

In the context of biodiversity, LiDAR data offer diverse opportunities for meaningful applications that use vegetation structure as a direct or indirect proxy. To understand the importance of LiDAR data for biodiversity mapping, it is important to disentangle the mixed effects that impact the biodiversity measure in the modeling approaches. I confronted this challenge with the example of modeling animal species richness at Mount Kilimanjaro (chapter 2). I conclude, that in the context of predicting biodiversity by using LiDAR-derived vegetation structure as a potential driver, interpretation of the results have to take the relief of the study area into account, as it influences multiple environmental factors and thereby overshadows the effect of structure and hence the relevance of LiDAR data. This comparison between LiDAR-derived vegetation structure and elevation as predictors shows the great necessity of a holistic understanding of the local ecosystem with all its components and influences and not just the examination of a specific aspect of interest. Additionally, the obvious change of vegetation structure with vegetation zones along the elevation gradient enhances the perpetual problem of sufficient training data. Due to the cost of field campaigns, obtaining a sufficient number of in-situ measurements is almost always a considerable challenge. Generally, efforts towards standardized or comparable data acquisition in the field could provide added value to the effort of transferable knowledge. For instance, the enhancement of inventories with matching LiDAR variables would complement field work for spatially comprehensive data sets. However, in this study, with a changing vegetation structure driven by elevation and the corresponding abiotic factors, the available data were not sufficient to fully disentangle the effects of LiDAR data and co-factors (elevation, land use). For follow up projects, there is a need for more training data per vegetation zone to account for intra-zonal changes and relations within. This would probably assign more value to the LiDAR data, as otherwise finer differences are concealed within the larger patterns.

5.2 Systematic comparison across taxa

Another raised challenge was the systematic comparability between modeled taxa. Multi-taxa approaches for the prediction of animal species richness need intensive field campaigns to collect data used for model training and therefore are rather rare. The study presented in chapter 2 followed an approach for modeling

species richness to provide systematically comparable results across 17 taxa and four feeding guilds. Results of this study are of particular value, as the methodology, which is consistent across all considered taxa, permits more widely applicable statements. As results are not limited to single animal assemblages, but available for an abundant amount of taxa, findings have greater relevance in the research context. Only the multi-taxa approach in this study allowed a general cross-taxa statement about the significant difference in the prediction performance of flying taxa compared to non-flying taxa. Consequently, this multi-taxa approach might provide a more robust base to transfer the results for example to different study designs. This comparability is the groundwork for bridging the gap between local and global scales and eventually draw conclusions for global mechanisms. Hence, it represents a crucial component in the pursuit for global ecosystem monitoring.

5.3 Potential of irregular LiDAR data

Generally, a comprehensive monitoring and consequently the knowledge about the current state of biodiversity, is the inevitable foundation for any targeted conservation strategy. Nevertheless, the inherent discontinuity of some data sets poses another important ongoing challenge in the applicability of LiDAR data. The trade-off between spatial resolution and temporal repetition is very prominent in LiDAR data sets. Typically, project-based flight campaigns as in chapter 2 are singular events with a high spatial resolution for a limited area. Governmental flight campaigns offer data sets for whole countries or federal states with medium resolution and a repetition rate of a couple of years. In contrast, the global spaceborne GEDI mission has a repetition rate of several times a month but provides 25 m footprints in a scattered pattern. Yet, especially when integrated with other data sets, good advances were made to mitigate the discontinuities and exploit the advantages of the vertical information. While the recent GEDI spaceborne LiDAR mission surveyed the surface in regularly spaced footprints and provided regular overflights of the same area, the coordinates of the footprints might be shifted.
The temporal resolution of GEDI offers highly valuable continuous structure-based information and the idea of chapter 4 was that, with the integration of additional area-wide data sets, the spatial gaps between footprints could be filled to provide a monitoring that is continuous in space and time. It was shown that the integration of spaceborne leaf area index measures derived from structural features in combination with Sentinel-1 and Sentinel-2 data worked satisfactory for the purpose of seasonally explicit area-wide prediction of the leaf area index as an example of plant traits.
Even the successful temporal prediction beyond the lifespan of the GEDI sensor was demonstrated and was promising for the use of spaceborne missions with a limited lifespan for ongoing biodiversity monitoring. This allows me to conclude that the approach was adequately fitted to both the temporal and spatial aspects of the depicted phenomenon. The results show that within the right context and scale, even spatially discontinuous observations can be beneficial to regional studies if processed adequately. This means that even low-resolution spaceborne structural data in a scattered footprint pattern can add valuable information and bridge gaps of certain comprehensive regional ecological questions by integrating appropriate data sets. Generally, in contrast to that, area-wide airborne LiDAR data can be highly relevant for filling spatial and temporal observational gaps in point-based in-situ biodiversity measurements. With the growing availability of LiDAR data from different sources, the achievement in using heterogeneous data for the complex prediction of tree species-specific successional stages in chapter 3 provided a positive example for a value chain derived from readily available but less aligned data in terms of spatial resolution and time of data acquisition. It was shown that, in combination with spectral Sentinel-2 data, those heterogeneous data had a high potential for classifying vegetation processes such as successional stages in temperate forests. In the context of successional stages, even the rather low-resolution data of the governmental LiDAR campaigns seemed to adequately capture and represent structures and patterns within these forest sections. The matching of the temporal scale of the rather slowly developing successional stages in forests were within the same magnitude and therefore matching to the temporal differences in the LiDAR data. It was demonstrated that on the scale of a whole federal state, with appropriate data balancing and an adapted modeling approach, the prediction of slowly developing forest succession could benefit even from those

heterogeneous data sets. This might initially be surprising but circles back to the overall question and the assumption that to a large extent the proper matching of the ecological question to the appropriate data set is responsible for the utility of data for biodiversity processes. The results highly recommend to evaluate the area and process in space and time and to take even data into consideration that look unfavorable initially. The cost-benefit-ratio for a specialized flight campaign for this endeavor appears disproportionately high.

5.4 Linking ecological process understanding to appropriate remote sensing data sets

In general, LiDAR's great potential for biodiversity monitoring and modeling always needs to be analyzed and utilized in the context of other data sets, particularly high-quality field data for ground truthing. Due to the capability of LiDAR to sense vegetation structure below the surface, it allows for major improvements for biodiversity monitoring. However, each research question has different demands regarding the temporal and spatial scales. Therefore, it is of extreme importance to know and understand the ecological processes behind environmental patterns to support a method-centered advancement of data integration. The three individual case studies were conducted illuminating different aspects of biodiversity using LiDAR data sets different in quality, acquisition mode and scale. Each of the three studies provided insights at the interface between biodiversity and data science. Chapter 2 highlighted the importance to consider the characteristics of the study area to select appropriate data sets, chapter 3 provided a positive example for the usage of readily available heterogeneous data for a slow evolving process and chapter 4 provided insights about the integration of scattered LiDAR footprints with comprehensive spectral data for seasonally explicit mapping of leaf area index. All studies highlighted the importance of the understanding of the ecological processes for the utility of LiDAR data.

5.5 Conclusion

LiDAR data can be beneficial, for multiple ecological purposes, as shown in the studies included in this thesis, due to the general relevance of vegetation structure for biodiversity. LiDAR data offer great chances to supplement both traditional field observations as well as optical remote sensing data sets. The call for interdisciplinary collaborations between ecologists and the remote sensing community is highly relevant not only for the selection of appropriate variables but also in an even more general manner of choosing appropriate data that match all the requirements essential for the respective underlying ecological process. Future research should continue to deepen the understanding at the interface of technology and nature to systematically advance biodiversity monitoring as a foundation for targeted nature conservation strategies.

Bibliography

Acebes, P., Lillo, P., & Jaime-Gonzalez, C. (2021). Disentangling LiDAR Contribution in Modelling Species-Habitat Structure Relationships in Terrestrial Ecosystems Worldwide. A Systematic Review and Future Directions. *Remote Sensing*, *13*(17), 3447. https://doi.org/10.3390/rs13173447

Ahmed, N., Atzberger, C., & Zewdie, W. (2021). Species Distribution Modelling performance and its implication for Sentinel-2-based prediction of invasive Prosopis juliflora in lower Awash River basin, Ethiopia. *Ecological Processes*, *10*(1), 18. https://doi.org/10.1186/s13717-021-00285-6

Asner, G. P., & Martin, R. E. (2016). Spectranomics: Emerging science and conservation opportunities at the interface of biodiversity and remote sensing. *Global Ecology and Conservation*, *8*, 212–219. https://doi.org/https://doi.org/10.1016/j.gecco.2016.09.010

Bae, S., Levick, S. R., Heidrich, L., Magdon, P., Leutner, B. F., Wöllauer, S., Serebryanyk, A., Nauss, T., Krzystek, P., Gossner, M. M., Schall, P., Heibl, C., Bässler, C., Doerfler, I., Schulze, E.-D., Krah, F.-S., Culmsee, H., Jung, K., Heurich, M., … Müller, J. (2019). Radar vision in the mapping of forest biodiversity from space. *Nature Communications*, *10*(1), 4757. https://doi.org/10.1038/s41467-019-12737-x

Baghdadi, N. N., El Hajj, M., Zribi, M., & Fayad, I. (2016). Coupling SAR C-Band and Optical Data for Soil Moisture and Leaf Area Index Retrieval Over Irrigated Grasslands. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(3, SI), 1229–1243. https://doi.org/10.1109/JSTARS.2015.2464698

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

Barnes, E., Clarke, T., Richards, S., Colaizzi, P., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., & 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*.

Basset, Y., Cizek, L., Cuenoud, P., Didham, R., Guilhaumon, F., Missa, O., Novotny, V., Ødegaard, F., Roslin, T., Schmidl, J., K. Tishechkin, A., N. Winchester, N., Roubik, D., Aberlenc, H.-P., Bail, J., Barrios, H., Bridle, J., Castaño, G., Corbara, B., & Leponce, M. (2012). Arthropod Diversity in a Tropical Forest. *Science*, *338*, 1481–1484. https://doi.org/10.1126/science.1226727

Berg, Å. (1997). Diversity and abundance of birds in relation to forest fragmentation, habitat quality and heterogeneity. *Bird Study*, *44*(3), 355–366. https://doi.org/10.1080/00063659709461071

Bergen, K. M., Goetz, S. J., Dubayah, R. O., Henebry, G. M., Hunsaker, C. T., Imhoff, M. L., Nelson, R. F., Parker, G. G., & Radeloff, V. C. (2009). Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences*, *114*(G2). https://doi.org/10.1029/2008JG000883

Berveglieri, A., Imai, N. N., Christovam, L. E., Galo, M. L. B. T., Tommaselli, A. M. G., & Honkavaara, E. (2021). Analysis of trends and changes in the successional trajectories of tropical forest using the Landsat NDVI time series. *Remote Sensing Applications: Society and Environment*, *24*, 100622. https://doi.org/10.1016/j.rsase.2021.100622

Berveglieri, A., Imai, N. N., Tommaselli, A. M., Casagrande, B., & 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*(2018), 548–558. https://doi.org/10.1016/j.isprsjprs.2018.11.002

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;jsessionid=1A403B5E29DCE9F51A42A4DC2D4366AD.intranet922?__blob=publicationFile&v=3

Borthagaray, A. I., Arim, M., & Marquet, P. A. (2012). Connecting landscape structure and patterns in body size distributions. *OIKOS*, *121*(5), 697–710. https://doi.org/10.1111/j.1600-0706.2011.19548.x

Boucher, P. B., Hancock, S., Orwig, D. A., Duncanson, L., Armston, J., Tang, H., Krause, K., Cook, B., Paynter, I., Li, Z., Elmes, A., & Schaaf, C. (2020). Detecting Change in Forest Structure with Simulated GEDI Lidar Waveforms: A Case Study of the Hemlock Woolly Adelgid (HWA; Adelges tsugae) Infestation. *REMOTE SENSING*, *12*(8). https://doi.org/10.3390/rs12081304

Breidenbach, J., Waser, L. T., Debella-Gilo, M., Schumacher, J., Rahlf, J., Hauglin, M., Puliti, S., & Astrup, R. (2021). National mapping and estimation of forest area by dominant tree species using Sentinel-2 data. *Canadian Journal of Forest Research*, *51*(3), 365–379. https://doi.org/10.1139/cjfr-2020-0170

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brokaw, N. V. L., & Lent, R. A. (1999). Vertical structure. In M. L. Hunter (Ed.), *Maintaining Biodiversity in Forest Ecosystems* (pp. 373–399). Cambridge University Press. https://doi.org/10.1017/CBO9780511613029.013

Brown, L. A., Ogutu, B. O., & Dash, J. (2019). Estimating Forest Leaf Area Index and Canopy Chlorophyll Content with Sentinel-2: An Evaluation of Two Hybrid Retrieval Algorithms. *Remote Sensing*, (15), 1752. https://dx.doi.org/10.3390/rs11151752

Burns, P., Clark, M., Salas, L., Hancock, S., Leland, D., Jantz, P., Dubayah, R., & Goetz, S. J. (2020). Incorporating canopy structure from simulated GEDI lidar into bird species distribution models. *Environmental Research Letters*, *15*(9), 095002. https://doi.org/10.1088/1748-9326/ab80ee

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L., Jung, M., Gamon, J. A., Reichstein, M., & Running, S. W. (2021). A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, *7*(9), eabc7447. https://doi.org/10.1126/sciadv.abc7447

Cao, S., Yu, Q., Sanchez-Azofeifa, A., Feng, J., Rivard, B., & Gu, Z. (2015). Mapping tropical dry forest succession using multiple criteria spectral mixture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *109*(2015), 17–29. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2015.08.009

Carrascal, L. M., Galván, I., & Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, *118*(5), 681–690. https://doi.org/10.1111/j.1600-0706.2008.16881.x

Cavard, X., Macdonald, S. E., Bergeron, Y., & Chen, H. Y. (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

Cavender-Bares, J., Schneider, F. D., Santos, M. J., Armstrong, A., Carnaval, A., Dahlin, K. M., Fatoyinbo, L., Hurtt, G. C., Schimel, D., Townsend, P. A., Ustin, S. L., Wang, Z., & Wilson, A. M. (2022). Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nature Ecology & Evolution*, *6*(5), 506–519. https://doi.org/10.1038/s41559-022-01702-5

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, *114*(30), E6089–E6096. https://doi.org/10.1073/pnas.1704949114

Chen, J. M. (1996). Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Canadian Journal of Remote Sensing*, *22*(3), 229–242. https://doi.org/10.1080/07038992.1996.10855178

Chen, L., Ren, C., Zhang, B., Wang, Z., Liu, M., Man, W., & Liu, J. (2021). Improved estimation of forest stand volume by the integration of GEDI LiDAR data and multi-sensor imagery in the Changbai

Mountains Mixed forests Ecoregion (CMMFE), northeast China. *International Journal of Applied Earth Observation and Geoinformation*, *100*. https://doi.org/10.1016/j.jag.2021.102326

Clawges, R., Vierling, K. T., Vierling, L. A., & Rowell, E. (2008). The use of airborne LiDAR to assess avian species diversity, density, and occurrence in a pine/aspen forest. *Remote Sensing of Environment*.

Cohrs, C. W., Cook, R. L., Gray, J. M., & Albaugh, T. J. (2020). Sentinel-2 Leaf Area Index Estimation for Pine Plantations in the Southeastern United States. *Remote Sensing*, (9), 1406. https://dx.doi.org/10.3390/rs12091406

Corona, P., Chirici, G., McRoberts, R. E., Winter, S., & Barbati, A. (2011). Contribution of large-scale forest inventories to biodiversity assessment and monitoring. *Forest Ecology and Management*, *262*(11), 2061–2069. https://doi.org/10.1016/j.foreco.2011.08.044

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

Delegido, J., Verrelst, J., Alonso, L., & Moreno, J. (2011). Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors*, (7), 7063–7081. https://dx.doi.org/10.3390/s110707063

Dhargay, S., Lyell, C. S., Brown, T. P., Inbar, A., Sheridan, G. J., & Lane, P. N. J. (2022). Performance of GEDI Space-Borne LiDAR for Quantifying Structural Variation in the Temperate Forests of South-Eastern Australia. *Remote Sensing*, *14*(15). https://doi.org/10.3390/rs14153615

Di Tommaso, S., Wang, S., & Lobell, D. B. (2021). Combining GEDI and Sentinel-2 for wall-to-wall mapping of tall and short crops. *Environmental Research Letters*, *16*(12). https://doi.org/10.1088/1748-9326/ac358c

Dorado-Roda, I., Pascual, A., Godinho, S., Silva, C. A., Botequim, B., Rodriguez-Gonzalvez, P., Gonzalez-Ferreiro, E., & Guerra-Hernandez, J. (2021). Assessing the Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests. *Remote Sensing*, *13*(12). https://doi.org/10.3390/rs13122279

Duan, M., Bax, C., Laakso, K., Mashhadi, N., Mattie, N., & Sanchez-Azofeifa, A. (2023). Characterizing Transitions between Successional Stages in a Tropical Dry Forest Using LiDAR Techniques. *Remote Sensing*, *15*(2), 479. https://doi.org/10.3390/rs15020479

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, *1*, 100002. https://doi.org/10.1016/j.srs.2020.100002

Duncanson, L., Kellner, J. R., Armston, J., Dubayah, R., Minor, D. M., Hancock, S., Healey, S. P., Patterson, P. L., Saarela, S., Marselis, S., Silva, C. E., Bruening, J., Goetz, S. J., Tang, H., Hofton, M., Blair, B., Luthcke, S., Fatoyinbo, L., Abernethy, K., ... Zgraggen, C. (2022). Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sensing of Environment*, *270*, 112845. https://doi.org/10.1016/j.rse.2021.112845

DWD. (2019). *Jahrbuch 2019*. https://www.dwd.de/DE/leistungen/jahresberichte_dwd/jahresberichte_pdf/jahresbericht_2019.pdf?__blob=publicationFile&v=8

DWD. (2020). *Jahrbuch 2020*. https://www.dwd.de/DE/leistungen/jahresberichte_dwd/jahresberichte/2020.html

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

Estes, L., Elsen, P. R., Treuer, T., Ahmed, L., Caylor, K., Chang, J., Choi, J. J., & Ellis, E. C. (2018). The spatial and temporal domains of modern ecology. *Nature Ecology & Evolution*, *2*(5), 819–826. https://doi.org/10.1038/s41559-018-0524-4

Falkowski, M. J., Evans, J. S., Martinuzzi, S., Gessler, P. E., & Hudak, A. T. (2009). Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment*, *113*(5), 946–956. https://doi.org/10.1016/j.rse.2009.01.003

Fang, H., Baret, F., Plummer, S., & Schaepman-Strub, G. (2019). An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. *Reviews of Geophysics*, *57*(3), 739–799. https://doi.org/10.1029/2018RG000608

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, *186*(2016), 64–87. https://doi.org/10.1016/j.rse.2016.08.013

Felix, A. B., Campa, H., Millenbah, K. F., Winterstein, S. R., & Moritz, W. E. (2004). Development of landscape-scale habitat-potential models for forest wildlife planning and management. *Wildlife Society Bulletin*, *32*(3), 795–806. 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, A. M., Björkman, C., Sang, Å. O., Jonsell, M., Holmström, E., Nilsson, U., Rönnberg, J., Kalén, C., & Lindbladh, M. (2020). The tree species matters: Biodiversity and ecosystem service implications of replacing Scots pine production stands with Norway spruce. *Ambio*, *49*(5), 1035–1049. https://doi.org/10.1007/s13280-019-01259-x

Feret, J.-B., François, C., Asner, G. P., Gitelson, A. A., Martin, R. E., Bidel, L. P. R., Ustin, S. L., Maire, G. l., & Jacquemoud, S. (2008). PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, *112*(6), 3030–3043. https://doi.org/https://doi.org/10.1016/j.rse.2008.02.012

Fernández-Manso, A., Fernández-Manso, O., & Quintano, C. (2016). Sentinel-2A red-edge spectral indices suitability for discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation*, *50*(2016), 170–175. https://doi.org/10.1016/j.jag.2016.03.005

Flashpohler, D. J., Giardina, C. P., Asner, G. P., Hart, P., Price, J., & Lyons, C. K. (2010). Long-term effects of fragmentation and fragment properties on bird species richness in Hawaiian forests. *Biological Conservation*, *143*(2), 280–288. https://doi.org/https://doi.org/10.1016/j.biocon.2009.10.009

Fouladinejad, F., Matkan, A., Hajeb, M., & Brakhasi, F. (2019). History and Applications of Space-Borne LiDARs. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-4-W18*, 407–414. https://doi.org/10.5194/isprs-archives-XLII-4-W18-407-2019

Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 83–92. https://dx.doi.org/10.1016/j.isprsjprs.2013.04.007

Francini, S., D'Amico, G., Vangi, E., Borghi, C., & Chirici, G. (2022). Integrating GEDI and Landsat: Spaceborne Lidar and Four Decades of Optical Imagery for the Analysis of Forest Disturbances and Biomass Changes in Italy. *SENSORS*, *22*(5). https://doi.org/10.3390/s22052015

Frantz, D. (2019). FORCE - Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sensing*, *11*(9), 1124. https://doi.org/10.3390/rs11091124

Frison, P.-L., Fruneau, B., Kmiha, S., Soudani, K., Dufrêne, E., Toan, T. L., Koleck, T., Villard, L., Mougin, E., & Rudant, J.-P. (2018). Potential of Sentinel-1 Data for Monitoring Temperate Mixed Forest Phenology. *Remote Sensing*, (12), 2049. https://dx.doi.org/10.3390/rs10122049

Froidevaux, J. S. P., Zellweger, F., Bollmann, K., Jones, G., & Obrist, M. K. (2016). From field surveys to LiDAR: Shining a light on how bats respond to forest structure. *Remote Sensing of Environment*, *175*, 242–250. https://doi.org/10.1016/j.rse.2015.12.038

Fujiki, S., Okada, K.-i., Nishio, S., & 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*(2016), 280–293. https://doi.org/10.1016/j.isprsjprs.2016.06.008

Fuster, B., Sánchez-Zapero, J., Camacho, F., García-Santos, V., Verger, A., Lacaze, R., Weiss, M., Baret, F., & Smets, B. (2020). Quality Assessment of PROBA-V LAI, fAPAR and fCOVER Collection 300 m Products of Copernicus Global Land Service. *Remote Sensing*, *12*(6). https://doi.org/10.3390/rs12061017

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M. C., Fröberg, M., Stendahl, J., Philipson, C. D., Mikusiński, G., Andersson, E., Westerlund, B., Andrén, H., Moberg, F., Moen, J., & Bengtsson, J. (2013). Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications*, *4*(1), 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*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gao, T., Hedblom, M., Emilsson, T., & Nielsen, A. B. (2014). The role of forest stand structure as biodiversity indicator. *Forest Ecology and Management*, *330*, 82–93. https://doi.org/10.1016/j.foreco.2014.07.007

GCOS. (2021). *The Status of the Global Climate Observing System 2021: The GCOS Status Report (GCOS-240)* (Status Report). GCOS. Geneva. https://ane4bf-datap1.s3.eu-west-1.amazonaws.com/wmod8_gcos/s3fs-public/gcos-status_report_full_text-240_lr_compressed.pdf?FDdn12yqICplxugb2V7hTQ9lTIcMRQFd=

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 Lasercsanner-Messaufnahme einschließlich Klassifizierung - Freistaat Sachsen. Retrieved July 31, 2023, from https://geomis.sachsen.de/geomis-client/?lang=de#/datasets/iso/99815cb4-f6de-4bef-ac7f-e8e005e3f1d9

Getzin, S., Fischer, R., Knapp, N., & Huth, A. (2017). Using airborne LiDAR to assess spatial heterogeneity in forest structure on Mount Kilimanjaro. *Landscape Ecology*. https://doi.org/10.1007/s10980-017-0550-7

Gitelson, A., & Merzlyak, M. N. (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*(3), 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (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*(3), 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters*, (5), n/a–n/a. https://dx.doi.org/10.1029/2002gl016450

Goetz, S., Steinberg, D., Dubayah, R., & Blair, B. (2007). Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sensing of Environment*, *108*, 254–263.

Google Maps. (2020). *Mt Kilimanjaro, Maxar Technologies (2020), CNES/Airbus (2020)*. https://www.google.de/maps

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2017.06.031

Grabska, E., Hostert, P., Pflugmacher, D., & Ostapowicz, K. (2019). Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing*, *11*(10), 1197. https://doi.org/10.3390/rs11101197

Graf, W., Kleinn, C., Schall, P., Nauss, T., Detsch, F., & Magdon, P. (2019). Analyzing the relationship between historic canopy dynamics and current plant species diversity in the herb layer of temperate forests using long-term Landsat time series. *Remote Sensing of Environment*, *232*, 111305. https://doi.org/10.1016/j.rse.2019.111305

Guo, Q., Su, Y., Hu, T., Guan, H., Jin, S., Zhang, J., Zhao, X., Xu, K., Wei, D., Kelly, M., & Coops, N. C. (2021). Lidar Boosts 3D Ecological Observations and Modelings: A Review and Perspective. *IEEE Geoscience and Remote Sensing Magazine*, *9*(1), 232–257. https://doi.org/10.1109/MGRS.2020.3032713

Haase, P., Tonkin, J. D., Stoll, S., Burkhard, B., Frenzel, M., Geijzendorffer, I. R., Häuser, C., Klotz, S., Kühn, I., McDowell, W. H., Mirtl, M., Müller, F., Musche, M., Penner, J., Zacharias, S., & Schmeller, D. S. (2018). The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. *Science of the Total Environment*, *613-614*, 1376–1384. https://doi.org/10.1016/j.scitotenv.2017.08.111

Healey, S. P., Yang, Z., Gorelick, N., & Ilyushchenko, S. (2020). Highly Local Model Calibration with a New GEDI LiDAR Asset on Google Earth Engine Reduces Landsat Forest Height Signal Saturation. *Remote Sensing*, *12*(17), 2840. https://doi.org/10.3390/rs12172840

Heidrich, L., Brandl, R., Ammer, C., Bae, S., Bässler, C., Doerflerh, I., Fischer, M., Gossner, M. M., Heurich, M., Heibl, C., Jung, K., Krzystek, P., Levick, S., Magdon, P., Schall, P., Schulze, E.-D., Seibold, S., Simons, N. K., Thorn, S., … Müller, J. (2023). Effects of heterogeneity on the ecological diversity and redundancy of forest fauna. *Basic and Applied Ecology*, S1439179123000610. https://doi.org/10.1016/j.baae.2023.10.005

Hemmerling, J., Pflugmacher, D., & 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., & Müller, J. (2018). Biodiversity along temperate forest succession. *Journal of Applied Ecology*, *55*(6), 2756–2766. https://doi.org/10.1111/1365-2664.13238

Hisano, M., Searle, E. B., & Chen, H. Y. H. (2018). Biodiversity as a solution to mitigate climate change impacts on the functioning of forest ecosystems. *Biological Reviews of the Cambridge Philosophical Society*, *93*(1), 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., & Kuenzer, C. (2020). Earth Observation Based Monitoring of Forests in Germany: A Review. *Remote Sensing*, *12*(21), 3570. https://doi.org/10.3390/rs12213570

Hooper, D., Chapin, F., Ewel, J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J., Lodge, D., Loreau, M., Naeem, S., Schmid, B., Setala, H., Symstad, A., Vandermeer, J., & Wardle, D. (2005). Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs*, *75*(1), 3–35. https://doi.org/10.1890/04-0922

Hościło, A., & Lewandowska, A. (2019). Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data. *Remote Sensing*, *11*(8), 929. https://doi.org/10.3390/rs11080929

Huang, B., Yang, Y., Li, R., Zheng, H., Wang, X., Wang, X., & Zhang, Y. (2022). Integrating Remotely Sensed Leaf Area Index with Biome-BGC to Quantify the Impact of Land Use/Land Cover Change on Water Retention in Beijing. *Remote Sensing*, *14*(3), 743. https://doi.org/10.3390/rs14030743

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., & Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1-2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Humphrey, J. W., Hawes, C., Peace, A. J., Ferris-Kaan, R., & Jukes, M. R. (1999). Relationships between insect diversity and habitat characteristics in plantation forests. *Forest Ecology and Management*, *113*(1), 11–21. https://doi.org/10.1016/S0378-1127(98)00413-7

HVBG - Hessische Verwaltung für Bodenmanagement und Geoinformation. (2023). Airborne Laserscanning. Luftgestütze Messverfahren. Retrieved July 31, 2023, from https://hvbg.hessen.de/landesvermessung/geotopographie/3d-daten/airborne-laserscanning

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing*, *8*(3), 166. https://doi.org/10.3390/rs8030166

Isenburg, M. (n.d.). *LAStools - efficient LiDAR processing software*. http://rapidlasso.com/LAStools.

Jenness, J. (2004). Calculating Landscape Surface Area from Digital Elevation Models. *Wildlife Society Bulletin*, *32*, 829–839. https://doi.org/10.2193/0091-7648(2004)032[0829:CLSAFD]2.0.CO;2

Jetz, W., McGeoch, M., Robert, G., Ferrier, S., Beck, J., Costello, M., Fernández, M., Geller, G., Keil, P., Merow, C., Meyer, C., Muller-Karger, F., Regan, E., Schmeller, D., & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-019-0826-1

Jiang, F., Smith, A. R., Kutia, M., Wang, G., Liu, H., & Sun, H. (2020). A Modified KNN Method for Mapping the Leaf Area Index in Arid and Semi-Arid Areas of China. *Remote Sensing*, (11), 1884. https://dx.doi.org/10.3390/rs12111884

Jung, K., Kaiser, S., Böhm, S., Nieschulze, J., & Kalko, E. K. V. (2012). Moving in three dimensions: Effects of structural complexity on occurrence and activity of insectivorous bats in managed forest stands. *Journal of Applied Ecology*, *49*, 523–531.

Kacic, P., Hirner, A., & Da Ponte, E. (2021). Fusing Sentinel-1 and-2 to Model GEDI-Derived Vegetation Structure Characteristics in GEE for the Paraguayan Chaco. *Remote Sensing*, *13*(24). https://doi.org/10.3390/rs13245105

Kalinicheva, E., Landrieu, L., Mallet, C., & Chehata, N. (2022, April). Multi-Layer Modeling of Dense Vegetation from Aerial LiDAR Scans. https://doi.org/10.48550/arXiv.2204.11620

Kaspari, M., & Weiser, M. D. (1999). The size–grain hypothesis and interspecific scaling in ants. *Functional Ecology*, *13*(4), 530–538. https://doi.org/10.1046/j.1365-2435.1999.00343.x

Kaufman, Y., & Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(2), 261–270. https://doi.org/10.1109/36.134076

Kganyago, M., Mhangara, P., Alexandridis, T., Laneve, G., Ovakoglou, G., & Mashiyi, N. (2020). Validation of sentinel-2 leaf area index (LAI) product derived from SNAP toolbox and its comparison with global LAI products in an African semi-arid agricultural landscape. *Remote Sensing Letters*, *11*(10), 883–892. https://doi.org/10.1080/2150704X.2020.1767823

Khati, U., Lavalle, M., & Singh, G. (2021). The Role of Time-Series L-Band SAR and GEDI in Mapping Sub-Tropical Above-Ground Biomass. *FRONTIERS IN EARTH SCIENCE*, *9*. https://doi.org/10.3389/feart.2021.752254

Khosravipour, A., Skidmore, A. K., Isenburg, M., Wang, T., & Hussin, Y. A. (2014). Generating Pit-free Canopy Height Models from Airborne Lidar. *Photogrammetric Engineering & Remote Sensing*, *80*(9), 863–872. https://doi.org/10.14358/PERS.80.9.863

Kissling, W. D., Ahumada, J. A., Bowser, A., Fernandez, M., Fernández, N., García, E. A., Guralnick, R. P., Isaac, N. J. B., Kelling, S., Los, W., McRae, L., Mihoub, J.-B., Obst, M., Santamaria, M., Skidmore, A. K., Williams, K. J., Agosti, D., Amariles, D., Arvanitidis, C., ... Hardisty, A. R. (2018). Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biological Reviews*, *93*(1), 600–625. https://doi.org/10.1111/brv.12359

Korhonen, L., Hadi, Packalen, P., & Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sensing of Environment*, 259–274. https://dx.doi.org/10.1016/j.rse.2017.03.021

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, Articles*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M. (2018). *Caret: Classification and Regression Training*. https://CRAN.R-project.org/package=caret

Labenski, P., Ewald, M., Schmidtlein, S., Heinsch, F. A., & Fassnacht, F. E. (2023). Quantifying surface fuels for fire modelling in temperate forests using airborne lidar and Sentinel-2: Potential and limitations. *Remote Sensing of Environment*, *295*, 113711. https://doi.org/10.1016/j.rse.2023.113711

Landesamt für Vermessungs- und Katasterverwaltung Rheinland-Pflaz. (2021). Laserpunkte Produktbeschreibung. Retrieved November 1, 2023, from https://lvermgeo.rlp.de/fileadmin/lvermgeo/pdf/produktblaetter/ProduktbeschreibungRP_Laserpunkte.pdf

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/

Lange, M., Feilhauer, H., Kühn, I., & Doktor, D. (2022). Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series. *Remote Sensing of Environment*, *277*, 112888. https://doi.org/10.1016/j.rse.2022.112888

Lassau, S. A., & Hochuli, D. F. (2004). Effects of habitat complexity on ant assemblages. *Ecography*, *27*(2), 157–164. https://doi.org/10.1111/j.0906-7590.2004.03675.x

Laurin, G. V., Chen, Q., Lindsell, J. A., Coomes, D. A., Frate, F. D., Guerriero, L., Pirotti, F., & Valentini, R. (2014). Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. *Isprs Journal of Photogrammetry and Remote Sensing*, *89*, 49–58. https://doi.org/10.1016/j.isprsjprs.2014.01.001

Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). Lidar Remote Sensing for Ecosystem Studies. *BioScience*, *52*(1), 19. https://doi.org/https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2

Lesak, A. A., Radeloff, V., Hawbaker, T., Pidgeon, A. M., Gobakken, T., & Contrucci, K. (2011). Follow publication Modeling forest song bird species richness using LiDAR-derived measures of forest structure. *Remote Sensing of Environment*, *115*(11), 2823–2835.

Li, J., Hu, B., & Noland, T. L. (2013). Classification of tree species based on structural features derived from high density LiDAR data. *Agricultural and Forest Meteorology*, *171-172*, 104–114. https://doi.org/10.1016/j.agrformet.2012.11.012

Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., & Chen, H. (2020). High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation*, *92*, 102163. https://doi.org/10.1016/j.jag.2020.102163

Liu, A., Cheng, X., & Chen, Z. (2021). Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sensing of Environment*, *264*. https://doi.org/10.1016/j.rse.2021.112571

Lohani, B., & Ghosh, S. (2017). Airborne LiDAR Technology: A Review of Data Collection and Processing Systems. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, *87*(4), 567–579. https://doi.org/10.1007/s40010-017-0435-9

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J., Hector, A., Hooper, D., Huston, M., Raffaelli, D., Schmid, B., Tilman, D., & Wardle, D. (2001). Ecology - Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science*, *294*(5543), 804–808. https://doi.org/10.1126/science.1064088

Luo, L., Zhai, Q., Su, Y., Ma, Q., Kelly, M., & Guo, Q. (2018). Simple method for direct crown base height estimation of individual conifer trees using airborne LiDAR data. *Optics Express*, *26*(10), A562–A578. https://doi.org/10.1364/OE.26.00A562

Luo, P., Liao, J., & Shen, G. (2020). Combining Spectral and Texture Features for Estimating Leaf Area Index and Biomass of Maize Using Sentinel-1/2, and Landsat-8 Data. *IEEE ACCESS*, *8*, 53614–53626. https://doi.org/10.1109/ACCESS.2020.2981492

Luo, S., Wang, C., Xi, X., Pan, F., Peng, D., Zou, J., Nie, S., & Qin, H. (2017). Fusion of airborne LiDAR data and hyperspectral imagery for aboveground and belowground forest biomass estimation. *Ecological Indicators*, *73*, 378–387. https://doi.org/10.1016/j.ecolind.2016.10.001

LVermGeo - 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 Geoportal. (2023). Elevation data. Retrieved July 31, 2023, from https://geoportaal.maaamet.ee/eng/Spatial-Data/Elevation-data-p308.html

Macarthur, R., & Macarthur, J. W. (1961). On Bird Species Diversity. *Ecology*, *42*(3), 594–598. https://doi.org/10.2307/1932254

Mairota, P., Cafarelli, B., Labadessa, R., Lovergine, F., Tarantino, C., Lucas, R., Nagendra, H., & Didham, R. (2015). Very high resolution Earth Observation features for monitoring plant and animal community structure across multiple spatial scales in protected areas. *International Journal of Applied Earth Observations and Geoinformation*, *online*. https://doi.org/10.1016/j.jag.2014.09.015

Maltman, J. C., Hermosilla, T., Wulder, M. A., Coops, N. C., & White, J. C. (2023). Estimating and mapping forest age across Canada's forested ecosystems. *Remote Sensing of Environment*, *290*(2023), 113529. https://doi.org/10.1016/j.rse.2023.113529

Mandl, L., Stritih, A., Seidl, R., Ginzler, C., & Senf, C. (2023). Spaceborne LiDAR for characterizing forest structure across scales in the European Alps. *Remote Sensing in Ecology and Conservation*, *n/a*(n/a). https://doi.org/10.1002/rse2.330

Margono, B. A., Turubanova, S., Zhuravleva, I., Potapov, P., Tyukavina, A., Baccini, A., Goetz, S., & Hansen, M. C. (2012). Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010. *Environmental Research Letters*, *7*(3), 034010. https://doi.org/10.1088/1748-9326/7/3/034010

Marselis, S. M., Abernethy, K., Alonso, A., Armston, J., Baker, T. R., Bastin, J.-F., Bogaert, J., Boyd, D. S., Boeckx, P., Burslem, D. F. R. P., Chazdon, R., Clark, D. B., Coomes, D., Duncanson, L., Hancock, S., Hill, R., Hopkinson, C., Kearsley, E., Kellner, J. R., ... Dubayah, R. (2020). Evaluating the potential of full-waveform lidar for mapping pan-tropical tree species richness. *Global Ecology and Biogeography*, *29*(10), 1799–1816. https://doi.org/https://doi.org/10.1111/geb.13158

Martins, A. C. M., Willig, M. R., Presley, S. J., & Marinho-Filho, J. (2017). Effects of forest height and vertical complexity on abundance and biodiversity of bats in Amazonia. *Forest Ecology and Management*, *391*, 427–435. https://doi.org/10.1016/j.foreco.2017.02.039

Masemola, C., Cho, M. A., & Ramoelo, A. (2020). Sentinel-2 time series based optimal features and time window for mapping invasive Australian native Acacia species in KwaZulu Natal, South Africa. *International Journal of Applied Earth Observation and Geoinformation*, *93*, 102207. https://doi.org/10.1016/j.jag.2020.102207

McCain, C. M., & Grytnes, J.-A. (2010). Elevational Gradients in Species Richness. In *eLS*. American Cancer Society. https://doi.org/10.1002/9780470015902.a0022548

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Melin, M., Hinsley, S. A., Broughton, R. K., Bellamy, P., & Hill, R. A. (2018). Living on the edge: Utilising lidar data to assess the importance of vegetation structure for avian diversity in fragmented woodlands and their edges. *Landscape Ecology*, *33*(6), 895–910. https://doi.org/10.1007/s10980-018-0639-7

Meyer, H. (2018). *CAST: 'caret' Applications for Spatial-Temporal Models*. https://CRAN.R-project.org/package=CAST

Meyer, H., Milà, C., Ludwig, M., & 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., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, *101*(2018), 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, *411*(2019), 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Meyer, L. H., Heurich, M., Beudert, B., Premier, J., & Pflugmacher, D. (2019). Comparison of Landsat-8 and Sentinel-2 Data for Estimation of Leaf Area Index in Temperate Forests. *Remote Sensing*, (10), 1160. https://dx.doi.org/10.3390/rs11101160

Michałowska, M., & Rapiński, J. (2021). A Review of Tree Species Classification Based on Airborne LiDAR Data and Applied Classifiers. *Remote Sensing*, *13*(3), 353. https://doi.org/10.3390/rs13030353

Miranda, R. D. Q., Nóbrega, R. L. B., Moura, M. S. B. D., Raghavan, S., & Galvíncio, J. D. (2020). Realistic and simplified models of plant and leaf area indices for a seasonally dry tropical forest. *International Journal of Applied Earth Observation and Geoinformation*, 101992. https://dx.doi.org/10.1016/j.jag.2019.101992

MITMA - Ministry of Transport, Mobility and Urban Agenda National Geographic Institute. (2023). National Plan for aerial Orthophotography. National Plan fpr Territory Obervation. Retrieved July 31, 2023, from https://pnoa.ign.es/web/portal/pnoa-lidar/estado-del-proyecto

Morse, D., Lawton, J., Dodson, M., & Williamson, M. (1985). Fractal Dimension of Vegetation and the Distribution of Arthropod Body Length. *Nature*, *314*(6013), 731–733. https://doi.org/10.1038/314731a0

Müller, J., Bae, S., Röder, J., Chao, A., & Didham, R. K. (2014). Airborne LiDAR reveals context dependence in the effects of canopy architecture on arthropod diversity. *Forest Ecology and Management*, *312*, 129–137. https://doi.org/10.1016/j.foreco.2013.10.014

Müller, J., & Brandl, R. (2009). Assessing biodiversity by remote sensing in mountainous terrain: The potential of LiDAR to predict forest beetle assemblages. *Journal of Applied Ecology*, *46*(4), 897–905. https://doi.org/10.1111/j.1365-2664.2009.01677.x

Müller, J., Brandl, R., Brändle, M., Förster, B., de Araujo, B. C., Gossner, M. M., Ladas, A., Wagner, M., Maraun, M., Schall, P., Schmidt, S., Heurich, M., Thorn, S., & Seibold, S. (2018). LiDAR-derived canopy structure supports the more-individuals hypothesis for arthropod diversity in temperate forests. *Oikos*, *127*(6), 814–824. https://doi.org/10.1111/oik.04972

Müller, J., Moning, C., Bässler, C., Heurich, M., & Brandl, R. (2009). Using airborne laser scanning to model potential abundance and assemblages of forest passerines. *Basic and Applied Ecology*, *10*(7), 671–681.

Müller, J., Stadler, J., & Brandl, R. (2010). Composition versus physiognomy of vegetation as predictors of bird assemblages: The role of LiDAR. *Remote Sensing of Environment*, *114*(3), 490–495.

Müller, J., & Vierling, K. (2014). Assessing Biodiversity by Airborne Laser Scanning. In M. Maltamo, E. Næsset, & J. Vauhkonen (Eds.), *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies* (pp. 357–374). Springer Netherlands. https://doi.org/10.1007/978-94-017-8663-8_18

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, (6772), 853–858. https://dx.doi.org/10.1038/35002501

Myneni, R., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G., Lotsch, A., Friedl, M., Morisette, J., Votava, P., Nemani, R., & Running, S. (2001). Global Products Of Vegetation Leaf Area And Fraction Absorbed Par From Year One Of Modis Data. *Remote Sensing of Environment*, *83*, 214–231. https://doi.org/10.1016/S0034-4257(02)00074-3

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*(1), 148–159. https://doi.org/10.1016/j.rse.2008.09.001

Næsset, E., Bollandsås, O. M., & Gobakken, T. (2005). Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment*, *94*(4), 541–553. https://doi.org/10.1016/j.rse.2004.11.010

Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H. G., Tarantino, C., Adamo, M., & Mairota, P. (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators*, *33*(SI), 45–59. https://doi.org/10.1016/j.ecolind.2012.09.014

Nagendra, H., & Rocchini, D. (2008). High resolution satellite imagery for tropical biodiversity studies: The devil is in the detail. *Biodiversity and Conservation*, *17*, 3431–3442. https://doi.org/10.1007/s10531-008-9479-0

Newbold, T., Hudson, L., Hill, S., Contu, S., Lysenko, I., Senior, R., Börger, L., Bennett, D., Choimes, A., Collen, B., Day, J., De Palma, A., Diaz, S., Echeverria-Londono, S., Edgar, M., Feldman, A., Garon, M., Harrison, M., Alhusseini, T., & Purvis, A. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, *520*, 45–50. https://doi.org/10.1038/nature14324

NLS - National Land Survey of Finland. (2023). Laser scanning and aerial photography. Retrieved July 31, 2023, from https://www.maanmittauslaitos.fi/laserkeilaus-ja-ilmakuvaus

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

Novotny, V., Drozd, P., Miller, S., Kulfan, M., Janda, M., Basset, Y., & D. Weiblen, G. (2006). Why are there so many species of herbivorous insects in tropcial rainforests? *Nature*, *313*, 1115–1118.

OpenStreetMap. (2023). Base map from OpenStreetMap. https://www.openstreetmap.org/copyright.

Ørka, H. O., Næsset, E., & Bollandsås, O. M. (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*(7), 1445–1461. https://doi.org/10.1016/j.rse.2010.01.024

Ozanne, C. M. P., Anhuf, D., Boulter, S. L., Keller, M., Kitching, R. L., Körner, C., Meinzer, F. C., Mitchell, A. W., Nakashizuka, T., Dias, P. L. S., Stork, N. E., Wright, S. J., & Yoshimura, M. (2003). Biodiversity Meets the Atmosphere: A Global View of Forest Canopies. *Science*, *301*(5630), 183–186. https://doi.org/10.1126/science.1084507

Padalia, H., Sinha, S. K., Bhave, V., Trivedi, N. K., & Senthil Kumar, A. (2020). Estimating canopy LAI and chlorophyll of tropical forest plantation (North India) using Sentinel-2 data. *Advances in Space Research*, (1), 458–469. https://dx.doi.org/10.1016/j.asr.2019.09.023

Parisi, F., Vangi, E., Francini, S., Chirici, G., Travaglini, D., Marchetti, M., & Tognetti, R. (2022). Monitoring the abundance of saproxylic red-listed species in a managed beech forest by landsat temporal metrics. *Forest Ecosystems*, *9*, 100050. https://doi.org/10.1016/j.fecs.2022.100050

Pasqualotto, N., Delegido, J., Van Wittenberghe, S., Rinaldi, M., & Moreno, J. (2019). Multi-Crop Green LAI Estimation with a New Simple Sentinel-2 LAI Index (SeLI). *Sensors*, *19*(4). https://doi.org/10.3390/s19040904

PEFC - Arbeitsgruppe Rheinland-Pfalz. (2010). 3. Regionaler Waldbericht Rheinland-Pfalz, 49.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Hoft, R., Hurtt, G., Jetz, W., ... Wegmann, M. (2013). Essential Biodiversity Variables. *Science*, *339*(6117), 277–278. https://doi.org/10.1126/science.1229931

Pereira, H. M., Navarro, L. M., & Martins, I. S. (2012). Global Biodiversity Change: The Bad, the Good, and the Unknown. In Gadgil, A and Liverman, DM (Ed.), *Annual Review of Environment and Resources* (Vol. 37). Annual Reviews.

Peters, M., Andreas, H., Appelhans, T., Becker, J., Behler, C., Classen, A., Detsch, F., Ensslin, A., Ferger, S., B. Frederiksen, S., Gebert, F., Gerschlauer, F., Gütlein, A., Helbig-Bonitz, M., Hemp, C., Kindeketa, W., Kühnel, A., Mayr, A., Mwangomo, E., & Steffan-Dewenter, I. (2019). Climate–land-use interactions shape tropical mountain biodiversity and ecosystem functions. *Nature*, *568*, 1–5. https://doi.org/10.1038/s41586-019-1048-z

Peters, M. K., Hemp, A., Appelhans, T., Behler, C., Classen, A., Detsch, F., Ensslin, A., Ferger, S. W., Frederiksen, S. B., Gebert, F., Haas, M., Helbig-Bonitz, M., Hemp, C., Kindeketa, W. J., Mwangomo, E., Ngereza, C., Otte, I., Roeder, J., Rutten, G., ... Steffan-Dewenter, I. (2016). Predictors of elevational biodiversity gradients change from single taxa to the multi-taxa community level. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms13736

Pettorelli, N., Laurance, W., O'Brien, T., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: Opportunities and challenges. *Journal of Applied Ecology*, *51*. https://doi.org/10.1111/1365-2664.12261

Pettorelli, N., Wegmann, M., Skidmore, A., Mücher, S., Dawson, T. P., Fernandez, M., Lucas, R., Schaepman, M. E., Wang, T., O'Connor, B., Jongman, R. H., Kempeneers, P., Sonnenschein, R., Leidner, A. K., Böhm, M., He, K. S., Nagendra, H., Dubois, G., Fatoyinbo, T., ... Geller, G. N. (2016). Framing the concept of satellite remote sensing essential biodiversity variables: Challenges and future directions (D. Boyd, Ed.). *Remote Sensing in Ecology and Conservation*, *2*(3), 122–131. https://doi.org/10.1002/rse2.15

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., & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, *11*(1), 4540. https://doi.org/10.1038/s41467-020-18321-y

Polechová, J., & Storch, D. (2019, January). Ecological Niche. In B. Fath (Ed.), *Encyclopedia of Ecology (Second Edition)* (pp. 72–80). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.11113-3

Poorter, L., Amissah, L., Bongers, F., Hordijk, I., Kok, J., Laurance, S. G. W., Lohbeck, M., Martínez-Ramos, M., Matsuo, T., Meave, J. A., Muñoz, R., Peña-Claros, M., & van der Sande, M. T. (2023). Successional theories. *Biological Reviews*, *n/a*(n/a). https://doi.org/10.1111/brv.12995

Potapov, P., Hansen, M. C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina, A., & Ying, Q. (2020). Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping. *Remote Sensing*, *12*(3), 426. https://doi.org/10.3390/rs12030426

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/j.rse.2020.112165

Proença, V., Martin, L. J., Pereira, H. M., Fernandez, M., McRae, L., Belnap, J., Böhm, M., Brummitt, N., García-Moreno, J., Gregory, R. D., Honrado, J. P., Jürgens, N., Opige, M., Schmeller, D. S., Tiago, P., & Swaay, C. A. M. v. (2017). Global biodiversity monitoring: From data sources to Essential Biodiversity Variables. *Biological Conservation*, *213*, 256–263. https://doi.org/10.1016/j.biocon.2016.07.014

Qiao, K., Zhu, W., Xie, Z., & Li, P. (2019). Estimating the Seasonal Dynamics of the Leaf Area Index Using Piecewise LAI-VI Relationships Based on Phenophases. *Remote Sensing*, *11*(6). https://doi.org/10.3390/rs11060689

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Rapinel, S., Mony, C., Lecoq, L., Clément, B., Thomas, A., & Hubert-Moy, L. (2019). Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities. *Remote Sensing of Environment*, *223*, 115–129. https://doi.org/10.1016/j.rse.2019.01.018

Rechsteiner, C., Zellweger, F., Gerber, A., Breiner, F. T., & Bollmann, K. (2017). Remotely sensed forest habitat structures improve regional species conservation. *Remote Sensing in Ecology and Conservation*, *3*(4), 247–258. https://doi.org/10.1002/rse2.46

Reddy, C. S. (2021). Remote sensing of biodiversity: What to measure and monitor from space to species? *Biodiversity and Conservation*, *30*(10), 2617–2631. https://doi.org/10.1007/s10531-021-02216-5

Reddy, C. S., Kurian, A., Srivastava, G., Singhal, J., Varghese, A. O., Padalia, H., Ayyappan, N., Rajashekar, G., Jha, C. S., & Rao, P. V. N. (2021). Remote sensing enabled essential biodiversity variables for biodiversity assessment and monitoring: Technological advancement and potentials. *Biodiversity and Conservation*, *30*(1), 1–14. https://doi.org/10.1007/s10531-020-02073-8

Reif, J., Marhoul, P., & Koptík, J. (2013). Bird communities in habitats along a successional gradient: Divergent patterns of species richness, specialization and threat. *Basic and Applied Ecology*, *14*(5), 423–431. https://doi.org/10.1016/j.baae.2013.05.007

Richardson, J. J., Moskal, L. M., & Kim, S.-H. (2009). Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR. *Agricultural and Forest Meteorology*, *149*(6), 1152–1160. https://doi.org/10.1016/j.agrformet.2009.02.007

Rishmawi, K., Huang, C., Schleeweis, K., & Zhan, X. (2022). Integration of VIIRS Observations with GEDI-Lidar Measurements to Monitor Forest Structure Dynamics from 2013 to 2020 across the Conterminous United States. *Remote Sensing*, *14*(10), 2320. https://doi.org/10.3390/rs14102320

Rishmawi, K., Huang, C., & Zhan, X. (2021). Monitoring Key Forest Structure Attributes across the Conterminous United States by Integrating GEDI LiDAR Measurements and VIIRS Data. *Remote Sensing*, *13*(3), 442. https://doi.org/10.3390/rs13030442

Rocchini, D., Balkenhol, N., Carter, G. A., Foody, G. M., Gillepsie, T. W., He, K. S., Kark, S., Levin, N., Lucas, K., Luoto, M., Nafendra, H., Oldeland, J., Ricotta, C., Southworth, J., & Neteler, M. (2010). Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecological Informatics*, *5*, 318–329. https://doi.org/https://doi.org/10.1016/j.ecoinf.2010.06.001

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

Rufin, P., Frantz, D., Yan, L., & Hostert, P. (2021). Operational Coregistration of the Sentinel-2A/B Image Archive Using Multitemporal Landsat Spectral Averages. *IEEE Geoscience and Remote Sensing Letters*, *18*(4), 712–716. https://doi.org/10.1109/LGRS.2020.2982245

Saarela, S., Wästlund, A., Holmström, E., Mensah, A. A., Holm, S., Nilsson, M., Fridman, J., & Ståhl, G. (2020). Mapping aboveground biomass and its prediction uncertainty using LiDAR and field data, accounting for tree-level allometric and LiDAR model errors. *Forest Ecosystems*, *7*(1), 43. https://doi.org/10.1186/s40663-020-00245-0

Sandel, B., & Smith, A. B. (2009). Scale as a lurking factor: Incorporating scale-dependence in experimental ecology. *Oikos*, *118*(9), 1284–1291. https://doi.org/10.1111/j.1600-0706.2009.17421.x

Sarty, M., Abbott, K. L., & Lester, P. J. (2006). Habitat complexity facilitates coexistence in a tropical ant community. *Oecologia*, *149*(3), 465–473. https://doi.org/10.1007/s00442-006-0453-9

Schooler, S. L., & Zald, H. S. J. (2019). Lidar Prediction of Small Mammal Diversity in Wisconsin, USA. *Remote Sensing*, *11*(19). https://doi.org/10.3390/rs11192222

Seidl, R., Schelhaas, M.-J., & Lexer, M. J. (2011). Unraveling the drivers of intensifying forest disturbance regimes in Europe: Drivers of forest disturbance intensification. *Global Change Biology*, *17*(9), 2842–2852. https://doi.org/10.1111/j.1365-2486.2011.02452.x

Senf, C. (2022). Seeing the System from Above: The Use and Potential of Remote Sensing for Studying Ecosystem Dynamics. *Ecosystems*, *25*(8), 1719–1737. https://doi.org/10.1007/s10021-022-00777-2

Seo, H., & Kim, Y. (2021). Role of remotely sensed leaf area index assimilation in eco-hydrologic processes in different ecosystems over East Asia with Community Land Model version 4.5 – Biogeochemistry. *Journal of Hydrology*, *594*, 125957. https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.125957

Shendryk, Y. (2022). Fusing GEDI with earth observation data for large area aboveground biomass mapping. *International Journal of Applied Earth Observation and Geoinformation*, *115*, 103108. https://doi.org/https://doi.org/10.1016/j.jag.2022.103108

Siemann, E., Haarstad, J., & Tilman, D. (1999). Dynamics of plant and arthropod diversity during old field succession. *Ecography*, *22*(4), 406–414. https://doi.org/10.1111/j.1600-0587.1999.tb00577.x

Silva, C. A., Hamamura, C., Valbuena, R., Hancock, S., Cardil, A., Broadbent, E. N., Almeida, D. R. A. d., Junior, C. H. L. S., & Klauberg, C. (2021). *rGEDI: NASA's Global Ecosystem Dynamics Investigation (GEDI) Data Visualization and Processing*. https://CRAN.R-project.org/package=rGEDI

Simonson, W. D., Allen, H. D., & Coomes, D. A. (2014). Applications of airborne lidar for the assessment of animal species diversity. *Methods in Ecology and Evolution*, *5*(8), 719–729. https://doi.org/10.1111/2041-210X.12219

Smart, L. S., Swenson, J. J., Christensen, N. L., & Sexton, J. O. (2012). Three-dimensional characterization of pine forest type and red-cockaded woodpecker habitat by small-footprint, discrete-return lidar. *Forest Ecology and Management*, *281*, 100–110. https://doi.org/10.1016/j.foreco.2012.06.020

Solberg, S., Brunner, A., Hanssen, K. H., Lange, H., Næsset, E., Rautiainen, M., & Stenberg, P. (2009). Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sensing of Environment*, *113*(11), 2317–2327. https://doi.org/10.1016/j.rse.2009.06.010

Soliveres, S., van der Plas, F., Manning, P., Prati, D., Gossner, M. M., Renner, S. C., Alt, F., Arndt, H., Baumgartner, V., Binkenstein, J., Birkhofer, K., Blaser, S., Bluethgen, N., Boch, S., Boehm, S., Boerschig, C., Buscot, F., Diekoetter, T., Heinze, J., … Allan, E. (2016). Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature*, *536*(7617), 456+. https://doi.org/10.1038/nature19092

Stanska, M., & Stanski, T. (2017). Body size distribution of spider species in various forest habitats. *POLISH JOURNAL OF ECOLOGY*, *65*(4), 359–370. https://doi.org/10.3161/15052249PJE2017.65.4.005

Stein, A., Gerstner, K., & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales (H. Arita, Ed.). *Ecology Letters*, *17*(7), 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., & Ontrup, G. (2015). Satellite-Based Derivation of High-Resolution Forest Information Layers for Operational Forest Management. *Forests*, *6*(6), 1982–2013. https://doi.org/10.3390/f6061982

Sumnall, M. J., Hill, R. A., & Hinsley, S. A. (2016). Comparison of small-footprint discrete return and full waveform airborne lidar data for estimating multiple forest variables. *Remote Sensing of Environment*, *173*, 214–223. https://doi.org/10.1016/j.rse.2015.07.027

Tang, H., Armston, J., Hancock, S., Marselis, S., Goetz, S., & Dubayah, R. (2019). Characterizing global forest canopy cover distribution using spaceborne lidar. *Remote Sensing of Environment*, *231*. https://doi.org/10.1016/j.rse.2019.111262

Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D. B., & Blair, B. (2012). Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sensing of Environment*, *124*, 242–250. https://doi.org/https://doi.org/10.1016/j.rse.2012.05.005

TEEB - The Economics of Ecosystems and Biodiversity. (2010). *Ecological and Economic Foundations* (P. Kumar, Ed.). Earthscan.

Tesemma, Z. K., Wei, Y., Peel, M. C., & Western, A. W. (2015). The effect of year-to-year variability of leaf area index on variable infiltration capacity model performance and simulation of runoff. *Advances in Water Resources*, *83*, 310–322. https://doi.org/10.1016/j.advwatres.2015.07.002

Tew, E. R., Conway, G. J., Henderson, I. G., Milodowski, D. T., Swinfield, T., & Sutherland, W. J. (2022). Recommendations to enhance breeding bird diversity in managed plantation forests determined using LiDAR. *Ecological Applications*, *32*(7), e2678. https://doi.org/10.1002/eap.2678

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M. C., Schwager, M., & Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: The importance of keystone structures. *Journal of Biogeography*, *31*(1), 79–92. https://doi.org/10.1046/j.0305-0270.2003.00994.x

Tharammal, T., Bala, G., Devaraju, N., & Nemani, R. (2019). A review of the major drivers of the terrestrial carbon uptake: Model-based assessments, consensus, and uncertainties. *Environmental Research Letters*, *14*(9), 093005. https://doi.org/10.1088/1748-9326/ab3012

Thünen-Institut. (2012a). Order code: 69Z1JI_l321of_2012_l322, 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. (2012b). Order code: 69Z1JI_l321of_2012_l322, 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. (2012c). Order code: 69Z1JI_l321of_2012_l322, 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: 03.12.2021. Dritte Bundeswaldinventur - Ergeb.

Toivonen, J., Kangas, A., Maltamo, M., Kukkonen, M., & 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

Torre-Tojal, L., Bastarrika, A., Boyano, A., Lopez-Guede, J. M., & Graña, M. (2022). Above-ground biomass estimation from LiDAR data using random forest algorithms. *Journal of Computational Science*, *58*, 101517. https://doi.org/10.1016/j.jocs.2021.101517

Trumbore, S., Brando, P., & Hartmann, H. (2015). Forest health and global change. *Science*, *349*(6250), 814–818. https://doi.org/10.1126/science.aac6759

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. Retrieved October 25, 2023, from https://wedocs.unep.org/xmlui/handle/20.500.11822/9814

United Nations Department of Economic and Social Affairs. (2023, July). The Sustainable Development Goals Report 2023: Special Edition. Retrieved October 25, 2023, from http://desapublications.un.org/publications/sustainable-development-goals-report-2023-special-edition

United Nations Environmental Programme. (1992). Convention on Biological Diversity. https://wedocs.unep.org/20.500.11822/29146

Vanden Borre, J., Paelinckx, D., Mücher, C. A., Kooistra, L., Haest, B., De Blust, G., & Schmidt, A. M. (2011). Integrating remote sensing in Natura 2000 habitat monitoring: Prospects on the way forward. *Journal for Nature Conservation*, *19*(2), 116–125. https://doi.org/10.1016/j.jnc.2010.07.003

Vehmas, M., Eerikäinen, K., Peuhkurinen, J., Packalén, P., & Maltamo, M. (2009). Identification of boreal forest stands with high herbaceous plant diversity using airborne laser scanning. *Forest Ecology and Management*, *257*(1), 46–53. https://doi.org/10.1016/j.foreco.2008.08.016

Verhelst, K., Gou, Y., Herold, M., & Reiche, J. (2021). Improving Forest Baseline Maps in Tropical Wetlands Using GEDI-Based Forest Height Information and Sentinel-1. *Forests*, *12*(10). https://doi.org/10.3390/f12101374

Verrelst, J., Rivera, J. P., Veroustraete, F., Muñoz-Marí, J., Clevers, J. G. P. W., Camps-Valls, G., & Moreno, J. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods - A comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, 260–272. https://dx.doi.org/10.1016/j.isprsjprs.2015.04.013

Vidal, C., Alberdi, I., Redmond, J., Vestman, M., Lanz, A., & Schadauer, K. (2016). The role of European National Forest Inventories for international forestry reporting. *Annals of Forest Science*, *73*(4), 793–806. https://doi.org/10.1007/s13595-016-0545-6

Vierling, K. T., Bässler, C., Brandl, R., Vierling, L. A., Weiss, I., & Müler, J. (2011). Spinning a laser web: Predicting spider distributions using LiDAR. *Ecological Applications*, *21*(2), 577–588. https://doi.org/10.1890/09-2155.1

Vogeler, J. C., Hudak, A. T., Vierling, L. A., Evans, J., Green, P., & Vierling, K. T. (2014). Terrain and vegetation structural influences on local avian species richness in two mixed-conifer forests. *Remote Sensing of Environment*, *147*, 13–22.

Vogeler, J. C., Hudak, A. T., Vierling, L. A., & Vierling, K. T. (2013). Lidar-derived Canopy Architecture Predicts Brown Creeper Occupancy of Two Western Coniferous Forests. *Condor*, *115*(3), 614–622. https://doi.org/10.1525/cond.2013.110082

Wallis, C. I. B., Brehm, G., Donoso, D. A., Fiedler, K., Homeier, J., Paulsch, D., Suessenbach, D., Tiede, Y., Brandl, R., Farwig, N., & Bendix, J. (2017). Remote sensing improves prediction of tropical montane species diversity but performance differs among taxa. *Ecological Indicators*, *83*, 538–549. https://doi.org/10.1016/j.ecolind.2017.01.022

Wang, C., Elmore, A. J., Numata, I., Cochrane, M. A., Lei, S., Hakkenberg, C. R., Li, Y., Zhao, Y., & Tian, Y. (2022). A Framework for Improving Wall-to-Wall Canopy Height Mapping by Integrating GEDI LiDAR. *Remote Sensing*, *14*(15). https://doi.org/10.3390/rs14153618

Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R. B., & Chang, Q. (2019). Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *154*, 189–201. https://doi.org/10.1016/j.isprsjprs.2019.06.007

Wang, Y., & Fang, H. (2020). Estimation of LAI with the LiDAR Technology: A Review. *Remote Sensing*, *12*(20). https://doi.org/10.3390/rs12203457

Warren, R., Price, J., Graham, E., Forstenhaeusler, N., & VanDerWal, J. (2018). The projected effect on insects, vertebrates, and plants of limiting global warming to 1.5°C rather than 2°C. *Science*, *360*(6390), 791–795. https://doi.org/10.1126/science.aar3646

Weiss, M., Frederic, B., Garrigues, S., & Lacaze, R. (2007). LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: Validation and comparison with MODIS collection 4 products. *Remote Sensing of Environment*, *110*, 317–331. https://doi.org/10.1016/j.rse.2007.03.001

Welle, T., Aschenbrenner, L., Kuonath, K., Kirmaier, S., & Franke, J. (2022). Mapping Dominant Tree Species of German Forests. *Remote Sensing*, *14*(14), 3330. https://doi.org/10.3390/rs14143330

Wessel, M., Brandmeier, M., & 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*(9), 1419. https://doi.org/10.3390/rs10091419

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., & Tompalski, P. (2016). Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Canadian Journal of Remote Sensing*, *42*(5), 619–641. https://doi.org/10.1080/07038992.2016.1207484

Whitehurst, A. S., Swatantran, A., Blair, J. B., Hofton, M. A., & Dubayah, R. (2013). Characterization of Canopy Layering in Forested Ecosystems Using Full Waveform Lidar. *Remote Sensing*, *5*(4), 2014–2036. https://doi.org/10.3390/rs5042014

Wiens, J. A. (2009). Landscape Ecology as a Foundation for Sustainable Conservation. *Landscape Ecology*, (24), 1053–1065.

Wilson, E. O., & Peter, F. M. (1988). Structural and Functional Diversity in Temperate Forests. In *Biodiversity*. National Academies Press (US). Retrieved October 26, 2023, from https://www.ncbi.nlm.nih.gov/books/NBK219319/

Wöllauer, S., Zeuss, D., Magdon, P., & 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*(3), 414–426. https://doi.org/10.1111/ecog.05266

Wright, M. N., & Ziegler, A. (2017). Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*(1), 1–17. https://doi.org/10.18637/jss.v077.i01

Wu, H., & Li, Z.-L. (2009). Scale Issues in Remote Sensing: A Review on Analysis, Processing and Modeling. *Sensors*, *9*(3), 1768–1793. https://doi.org/10.3390/s90301768

Xi, Y., Ren, C., Tian, Q., Ren, Y., Dong, X., & 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

Xi, Y., Tian, Q., Zhang, W., Zhang, Z., Tong, X., Brandt, M., & Fensholt, R. (2022). Quantifying understory vegetation density using multi-temporal Sentinel-2 and GEDI LiDAR data. *GIScience & Remote Sensing*, *59*(1), 2068–2083. https://doi.org/10.1080/15481603.2022.2148338

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*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

Yan, G., Hu, R., Luo, J., Weiss, M., Jiang, H., Mu, X., Xie, D., & Zhang, W. (2019). Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. *Agricultural and Forest Meteorology*, *265*, 390–411. https://doi.org/10.1016/j.agrformet.2018.11.033

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. (2018). Out-

standing Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*, *33*(10), 790–802. https://doi.org/10.1016/j.tree.2018.08.001

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

Zellweger, F., Roth, T., Bugmann, H., & Bollmann, K. (2017). Beta diversity of plants, birds and butterflies is closely associated with climate and habitat structure. *Global Ecology and Biogeography*, *26*(8), 898–906. https://doi.org/10.1111/geb.12598

Zhao, G., Sanchez-Azofeifa, A., Laakso, K., Sun, C., & Fei, L. (2021). Hyperspectral and Full-Waveform LiDAR Improve Mapping of Tropical Dry Forest's Successional Stages. *Remote Sensing*, *13*(19), 3830. https://doi.org/10.3390/rs13193830

Zheng, G., & Moskal, L. M. (2009). Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors*, (4), 2719–2745. https://dx.doi.org/10.3390/s90402719

Zhu, X., Liu, J., Skidmore, A. K., Premier, J., & Heurich, M. (2020). A voxel matching method for effective leaf area index estimation in temperate deciduous forests from leaf-on and leaf-off airborne LiDAR data. *Remote Sensing of Environment*, *240*, 111696. https://doi.org/10.1016/j.rse.2020.111696

Ziegler, A., Meyer, H., Otte, I., Peters, M. K., Appelhans, T., Behler, C., Böhning-Gaese, K., Classen, A., Detsch, F., Deckert, J., Eardley, C. D., Ferger, S. W., Fischer, M., Gebert, F., Haas, M., Helbig-Bonitz, M., Hemp, A., Hemp, C., Kakengi, V., ... Nauss, T. (2022). Potential of Airborne LiDAR Derived Vegetation Structure for the Prediction of Animal Species Richness at Mount Kilimanjaro. *Remote Sensing*, *14*(3), 786. https://doi.org/10.3390/rs14030786

Chapter 6

Appendix

6.1 Summary

6.1.1 Summary - English

For the containment of the ongoing biodiversity loss, a targeted documentation of the status and dynamics of ecosystems is necessary. With LiDAR remote sensing, it is possible to quantify the vegetation structure, which plays a vital role for biodiversity. Technical developments in remote sensing and especially LiDAR hold great potential to implement comprehensive monitoring strategies but still pose some challenges. In this thesis, the ability of LiDAR data for predicting animal species across taxa, mapping tree species group specific successional stages as well as seasonal mapping of the leaf area index was systematically analyzed. With three case studies that vary by scale, data and ecological process, I aimed at analyzing the utility of LiDAR data for the mapping of biodiversity in the context of different data sources.

In case study one, high-resolution airborne LiDAR data, were especially obtained at Mount Kilimanjaro for the purpose of modeling animal species richness. Taxonomic data from an extensive multi-taxa field campaign were used to study the contribution of LiDAR data compared to environmental co-variables for species richness models. For most animal groups, the superior influence of elevation compared to vegetation structure was demonstrated. As the natural processes in the study area at the slopes of Mount Kilimanjaro are highly influenced by the elevation gradient the especially collected structural LiDAR data could not significantly add additional value. The study was designed as a multi-taxa approach, therefore results were compared systematically across taxa which is a valuable component for eventually bridging the gap between small scale local studies and global mechanisms.

Those specifically commissioned high-resolution LiDAR flight campaigns are resource intensive and not always within the possibilities of ecological projects. In many countries, however, governmental LiDAR data with rather low spatial resolution and a repetition rate of a couple of years exist. At the scale of federal states, wall-to-wall data consist of several observation dates and have varying quality. With that in mind, the benefit of these data initially appears questionable, which is probably why they are rarely used for the classification of successional stages.
In the second case study, the potential of such data for modeling forest successional stages for Rhineland-Palatinate was analyzed. By properly adapting data processing and modeling strategies, it was shown that for mapping of forest successional stages, even highly heterogeneous LiDAR data, in addition to the commonly used spectral data, could make a valuable contribution to biodiversity monitoring.

In the third case study included in this thesis, the seasonal changes of leaf area index with spaceborne LiDAR data were investigated. This study was based on the rather new LiDAR data from the spaceborne GEDI (Global Ecosystem Dynamics Investigation) mission. However, with its scattered collection of 25 m footprints and with overflights of the same area every couple of weeks the GEDI sampling design is neither comprehensive nor repetitive for the exact same points. This initially presented novel challenges since wall-to-wall monitoring is desirable but cannot be directly derived from GEDI data. The regular

overflights of the same areas, however, open up new possibilities for the integration with area-wide sensors. In this study the GEDI data were integrated with radar and optical data (Sentinel-1 and -2) to provide a spatio-temporally continuous mapping of the leaf area index. This approach was the first to analyze the monthly dynamics of the leaf area index derived from structural GEDI data comprehensively for a regional study. The satisfactory results of this integration method for different types of land use seemed promising in regards to further approaches in the direction of using multi-temporal data sets for regional ecological monitoring. Hence, even though planned to monitor global ecosystem dynamics, it was shown that the recorded GEDI data hold the potential of delivering valuable insights even for regional studies.
In summary, this thesis provides novel insights into the applicability of LiDAR data for biodiversity research. The general relevance of vegetation structure for biodiversity and LiDAR's ability to quantify three-dimensional vegetation structure offer great chances to support various ecological questions on different scales.

6.1.2 Zusammenfassung - deutsch

Um den fortschreitenden Biodiversitätsverlust einzudämmen, ist eine gezielte Dokumentation des Ist-Zustands und der Entwicklung von Ökosystemen notwendig. Mit LiDAR-Daten ist es möglich, die Vegetationsstruktur, die für die biologische Vielfalt eine entscheidende Rolle spielt, zu quantifizieren. Die technischen Entwicklungen in der Fernerkundung und insbesondere die wachsende Verfügbarkeit von LiDAR-Daten bergen ein großes Potenzial für die Umsetzung umfassender Monitoring-Strategien, stellen aber auch einige Herausforderungen dar. In dieser Arbeit wurde das Potenzial von LiDAR-Daten zur Vorhersage von Tierarten über mehrere Taxa hinweg zur Kartierung von baumartengruppenspezifischen Sukzessionsstadien sowie zur saisonalen räumlichen Vorhersage des Blattflächenindexes systematisch analysiert. Anhand von drei Fallstudien, die sich nach Skala, Datengrundlage und ökologischem Prozess unterscheiden, habe ich den Nutzen von LiDAR-Daten für die Erfassung der biologischen Vielfalt im Kontext unterschiedlicher Datenquellen analysiert.

In Fallstudie eins wurden hochaufgelöste Daten aus LiDAR-Befliegungen verwendet, die speziell dafür am Kilimandscharo erhoben wurden. Hiermit sollte die Artenvielfalt der Fauna aus einer umfangreichen Multitaxa-Feldkampagne modelliert werden, um das Potenzial der LiDAR-Daten im Vergleich zu Umwelt-Kovariablen zu untersuchen. Für den Großteil der Tiergruppen konnte der überlegene Einfluss der Geländehöhe im Vergleich zur Vegetationsstruktur nachgewiesen werden. Da die natürlichen Prozesse im Untersuchungsgebiet an den Hängen des Kilimandscharo stark durch den Höhengradienten beeinflusst werden, konnten die speziell erhobenen strukturellen LiDAR-Daten keinen signifikanten zusätzlichen Mehrwert liefern. Die Studie wurde als Multitaxa-Ansatz konzipiert, so dass die Ergebnisse systematisch zwischen verschiedenen Taxa verglichen werden konnten, was einen wertvollen Beitrag leistet, um letztendlich die Lücke zwischen kleinräumigen lokalen Studien und globalen Mechanismen zu schließen.

Speziell in Auftrag gegebene hochaufgelösten LiDAR-Befliegungsdaten sind ressourcenintensiv und liegen nicht immer im Rahmen der Möglichkeiten ökologischer Projekte. In vielen Ländern gibt es jedoch amtliche LiDAR-Daten mit eher geringer räumlicher Auflösung und einer Wiederholungsrate von einigen Jahren. Auf der Ebene der Bundesländer setzen sich die flächendeckenden Daten typischerweise aus mehreren Teilen zusamen, die zu verschiedenen Beobachtungszeitpunkten und in unterschiedlichen Qualitäten vorliegen. Vor diesem Hintergrund erscheint der Nutzen dieser Daten zunächst fraglich, weshalb sie wahrscheinlich nur selten für die Klassifizierung von Sukzessionsstadien verwendet werden.
In der zweiten Fallstudie wurde das Potenzial solcher Daten für die Modellierung von Waldentwicklungsstufen in Rheinland-Pfalz analysiert. Durch eine geeignete Anpassung der Datenverarbeitungs- und Modellierungsstrategien konnte deutlich gezeigt werden, dass für die Kartierung von Waldentwicklungsstufen zusätzlich zu den üblicherweise verwendeten Spektraldaten auch sehr heterogene LiDAR-Daten einen wertvollen Beitrag zum Biodiversitätsmonitoring leisten können.

In der dritten Fallstudie in dieser Arbeit wurden die saisonalen Veränderungen des Blattflächenindexes mit satellitengestützten LiDAR-Daten untersucht. Diese Studie basiert auf den relativ neuen LiDAR-Daten der

GEDI-Mission (Global Ecosystem Dynamics Investigation). Mit seiner stichprobenartigen Erfassung von runden Aufzeichnungsflächen mit 25 m Durchmesser und den Überflügen desselben Gebiets im Rhythmus einiger Wochen ist das Aufnahmedesign jedoch weder flächendeckend noch wird die Erfassung für genau dieselben Punkte wiederholt. Dies stellt zunächst eine neue Herausforderung dar, da ein flächendeckendes Monitoring wünschenswert ist, aber nicht direkt aus den GEDI-Daten abgeleitet werden kann. Die regelmäßigen Überflüge der gleichen Gebiete eröffnen jedoch neue Möglichkeiten für die Integration mit anderen Sensoren. In dieser Studie wurden die GEDI-Daten mit Radar- und optischen Daten (Sentinel-1 und -2) integriert, um eine raum-zeitlich kontinuierliche Vorhersage des Blattflächenindexes zu erhalten. Dieser Ansatz war der erste, der die monatliche Dynamik des aus strukturellen GEDI-Daten abgeleiteten Blattflächenindexes für eine regionale Studie umfassend analysiert. Die zufriedenstellenden Ergebnisse dieser Integrationsmethode für verschiedene Landnutzungsarten erschienen vielversprechend im Hinblick auf weitere Ansätze welche die Nutzung multitemporaler Datensätze für das regionale ökologische Monitoring anstreben. Somit konnte gezeigt werden, dass die GEDI-Daten, auch wenn sie für die Überwachung der globalen Ökosystemdynamik geplant waren, das Potenzial haben, auch für regionale Studien wertvolle Erkenntnisse zu liefern.

Zusammenfassend lässt sich sagen, dass die vorliegende Arbeit neue Erkenntnisse über die Nutzung verschiedener LiDAR-Datensätze für die Biodiversitätsforschung liefert. Die allgemeine Bedeutung der Vegetationsstruktur für die Biodiversität und die Fähigkeit von LiDAR-Daten, die dreidimensionale Vegetationsstruktur zu quantifizieren, bieten große Chancen, verschiedene ökologische Fragestellungen auf unterschiedlichen Skalen zu bereichern.

6.2 Acknowledgements

With the completion of my PhD, I reflect on an exciting, defining, and challenging time. I would like to express my gratitude to the people who accompanied and supported me during this time.

First of all, I want to thank Thomas, Chris and Hanna, who, from the beginning, have shaped my academic development in Marburg in very different but equally intense ways. Your special ability to create an honest environment with genuine exchange on an equal footing is something very special and has profoundly influenced me.

I also want to express thanks to my co-authors, working group members and colleagues who have been with me throughout this journey. In particular, I want to thank Jürgen, Katinka, Lisa, Marvin and Rieke. The exchange and support from all of you have enriched my life both professionally and personally over the past years. This journey wouldn't even have started without Lukas and Tim. You have both ignited my interest for working on research projects, thank you for that.

Special thanks go to my family and friends. Thank you for always supporting me, enduring with me, and turning everything into a truly good time together. I am grateful for my grandparents for always being enthusiastic about science and for consistently showing interest in my endeavors at university. I am thankful for all the support from my parents and my brother, and for the time we spent together during which curiosity meant fun and was always nurtured. In particular, I also want to thank Fabian. Thank you for your calm and unconditional support, always being there for me, having my back, and making all the highs and lows a whole lot better.

Thanks to all of you, your support is truly inspiring and incredibly meaningful to me. Thank you!