Geo-information Science and Remote Sensing

Thesis Report GIRS-2024-33

The use of LiDAR data in the acquisition of information about the habitat preferences of woodpeckers

Juriaan Rijsdijk

13-05-2024

WAGENINGEN
UNIVERSITY & RESEARCH

The use of LiDAR data in the acquisition of information about the habitat preferences of woodpeckers

Juriaan Rijsdijk

1298283

Supervisor:
Corné Vreugdenhil

External assistance
Juan Gallego-Zamorano

A thesis submitted in partial fulfilment of the degree of Master of Science

at Wageningen University and Research,

The Netherlands.

13-05-2024

Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2024-33
Wageningen University and Research
Laboratory of Geo-Information Science and Remote Sensing

Foreword

I would like to thank Corné Vreugdenhil for the valuable supervision during this thesis process. Corné was very much engaged during all the steps taken in this thesis, which I highly appreciate. Moreover, I would like to thank Juan Gallego-Zamorano and Henk Sierdsema. Both work at the foundation Sovon, the Dutch Centre for Field Ornithology. They helped me start up this research and provided me with woodpecker observation data for a study area in De Hoge Veluwe. Additionally, they provided me with information about the ecology of woodpeckers. Furthermore, after the proposal phase of this thesis, Juan provided me with valuable feedback during several visits at Sovon and during several online meetings.

Abstract

Variables used in ecological models usually concern aspects such as climate, soil and land cover. However, the structure of vegetation is a factor that is less commonly taken into account. Devices that have been used often so far to acquire vegetation data are multispectral imagery and synthetic aperture radar. Nevertheless, with the development of Light Detection and Ranging (LiDAR), vegetation could be mapped in more detail. In this thesis, it was researched how LiDAR data can be a valuable addition to ecological research into five different woodpecker species: the black woodpecker, great spotted woodpecker, middle spotted woodpecker, lesser spotted woodpecker and green woodpecker. In order to get more insight into these preferences, 14 different metrics representing aspects of vegetation structure were calculated from LiDAR point clouds. Three different types of metrics were calculated. At first, metrics were calculated that give information about the vertical complexity of vegetation, such as the standard deviation of vegetation height. Moreover, metrics were made that give information about the horizontal heterogeneity of vegetation, such as vegetation roughness. The third category is less commonly used in research. Namely, metrics were calculated with the use of polygons representing tree crowns, such as the distance to the forest edge. The metric values extracted at the locations of woodpecker observations were used in Maxent models in order to acquire information about the habitat preferences of woodpeckers. At first, one Maxent model was run with the observations of all five woodpecker species combined as input. The main results showed that the 95th percentile of vegetation height had the biggest importance and that the woodpecker family prefers trees that are taller than the average tree in the study area. Thereafter, a separate Maxent model was run for each woodpecker species in order to get more insight into the different habitat preferences of the species. The results showed several similarities, such as that the 95th percentile had the biggest variable importance for each species. The biggest difference that was found was that the results of the middle spotted woodpecker, lesser spotted woodpecker and green woodpecker showed that they have a preference for the edge of the forest, whereas no preferences were observed for the black woodpecker and great spotted woodpecker. Overall, this research has shown that LiDAR data can be a valuable addition to ecological research into woodpeckers. Moreover, the methodology used in this research could be an example of how LiDAR data can be used in ecological research into bird species in general. In future research, certain metrics could be improved in order to become more reliable for deriving information about vegetation structure that is relevant for woodpeckers. Moreover, Maxent models could be run with absence points instead of background points in order to increase the accuracy of the models. Namely, even though the accuracy values of the models used in this research were sufficient, there is still uncertainty about the reliability of the accuracy assessment, as the use of background points can lead to the model showing a lower accuracy than actually is the case.

Table of contents

1. Introduction

1.1 Vegetation structure

In ecological research, species distribution models (SDMs) are commonly used to find the relationship between the distribution of a species and environmental variables (Koma et al., 2022). In research on bird distributions, the inputs for SDMs are often variables concerning climate, topography, land cover and soil. These variables are important factors at a large scale. Nevertheless, at a smaller spatial extent (smaller than 100 km^2), vegetation structure can also be an important factor (Bakx et al., 2019). Vegetation structure is the three-dimensional arrangement of plants. Aspects such as plant height, distribution of vertical layers and density are part of the structure of a plant (Farwell et al., 2021). In this thesis, different trees, bushes and other singular plant units will be called vegetation units. The combination of these vegetation units in an area is called the vegetation composition of an area (Farwell et al., 2021). Especially vegetation heterogeneity, which is the diversity in vegetation in an area, has an influence on the spatial distribution of a species (Farwell et al., 2021). A higher variation of vegetation could lead to a higher variety of resources, which means that areas with a high vegetation heterogeneity could attract more species. Additionally, according to research of Kissling et al. (2023), important factors for habitat preferences such as nest site selection and the availability of food are strongly related to horizontal and vertical heterogeneity of vegetation. More specifically, the combination of different vegetation units in an area and the structure of these vegetation units play an important role in the habitat selection of birds. In this thesis, the terms habitat preferences and habitat selection are often used. Habitat preferences refer to certain versions of aspects of vegetation structure in a habitat. For example, the aspect tree height might be important for the habitat selection of a species and the specific preferred version of that aspect might be tall trees. Then, 'tall trees' is a habitat preference of this species. Thus, in this thesis a habitat preference is defined as the version of a vegetation structure aspect/characteristic. Additionally, there could be many more preferences that are more or less important for the species' choice for its habitat. Eventually, this trade-off between habitat preferences and their combination result in the habitat selection of a species.

1.2 LiDAR

In previous research, fine-scale vegetation variables were not included in SDMs due to that fine-scale vegetation was not yet measured at a large scale (Bakx et al., 2019). Nevertheless, remote sensing developments have resulted in that fine-scaled vegetation structures are now measured over larger areas and that resulting vegetation data are more detailed. Vegetation data were often measured by remote sensing products such as multispectral imagery and synthetic aperture radar (SAR) (Koma et al., 2022). However, the development of Light Detection and Ranging (LiDAR) data could improve the detail in which green structures are mapped. LiDAR is a technology in which a device sends out optical waves in order to measure the earth surface and its objects (Li et al., 2022). The sent out laser beams reflect against objects, the time until return is measured and subsequently the distance between the objects and the sensor can be derived. These measurements result in a point cloud of elevation data. This active remote sensing technique, in which laser pulses are emitted from a device on a plane or helicopter in order to measure the surface, is called airborne laser scanning (ALS) (Bakx et al., 2019). From the point cloud data, vegetation metrics can be calculated. Figure 1a shows the trajectory of LiDAR pulses sent out by a plane. Each pulse can reflect against multiple objects. Each time a wave returns to the sensor is counted, resulting in multiple returns (Esri, 2024b). Figure 1b shows an example of a point cloud created by LiDAR, as well as examples of which vegetation structure metrics can be calculated with this point cloud (Bakx et al., 2019). LiDAR metrics could be a more accurate representation of vegetation structure than radar and multispectral imagery derived metrics (Koma et al., 2022). Thus, as LiDAR data can be used to create more detailed vegetation structure layouts, there can be explored more how LiDAR data can be a used to map vegetation structure. Besides airborne laser scanning, also terrestrial laser scanning can be used to create vegetation metrics. The difference is that terrestrial laser scanning is performed by a person walking on the earth surface with a LiDAR device. This way of scanning could give a better view of the lower vegetation layers, whereas airborne laser scanning can provide more detail of the upper layers. In this thesis, airborne laser scanning data is used as it is available for bigger areas and is more useful for most metrics created during this thesis.

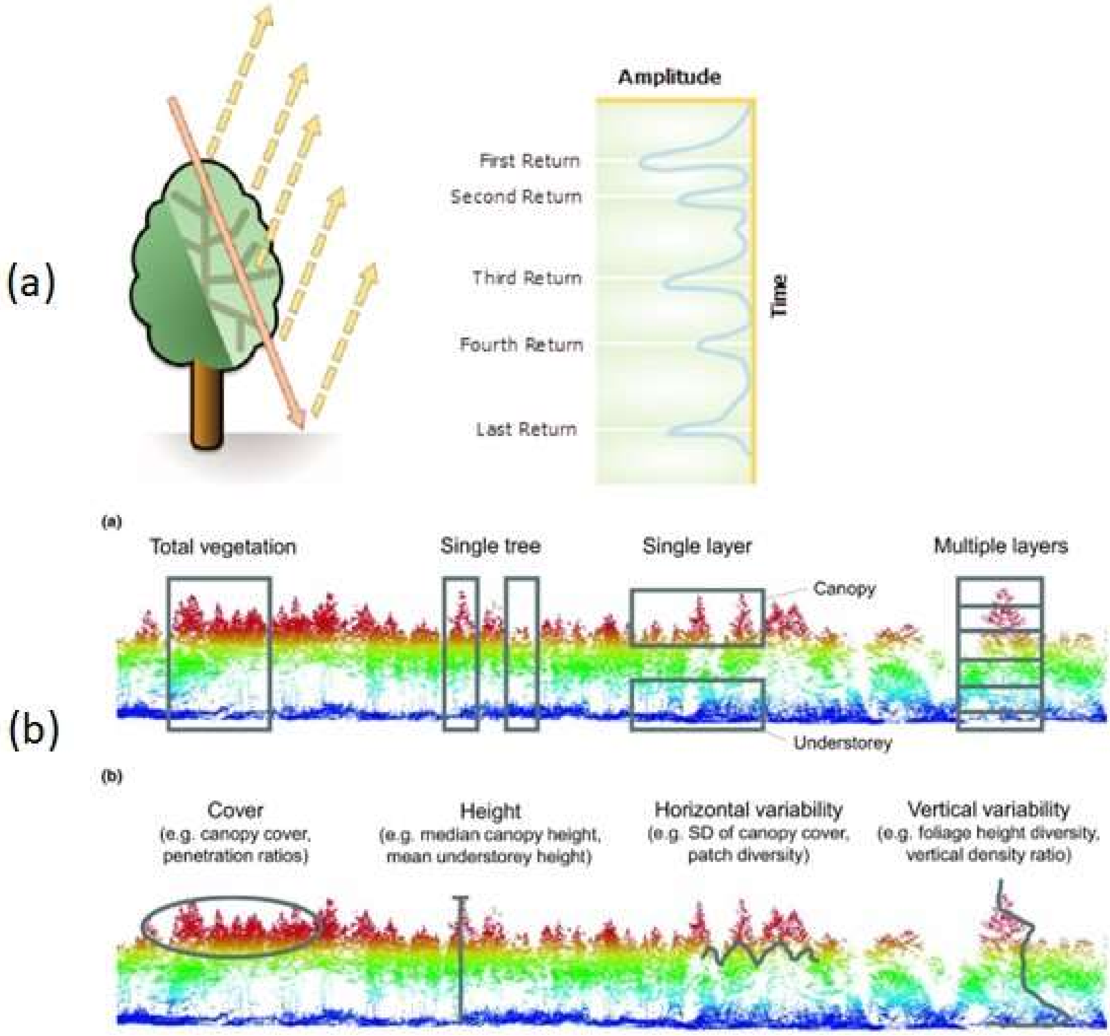


Figure 1: (a) Trajectory of a LiDAR pulse and its multiple returns (Esri, 2024b). (b) Point cloud of a forest (Bakx et al., 2019). 'a' shows different vegetation layers of a forest. 'b' shows different metrics that can be calculated for these different vegetation layers.

1.3 Woodpeckers

Over the past three decades, the conservation status of avifauna has decreased globally (Lees et al., 2022). As in 2022, 13.5% of the 10994 official birds species are labelled as under threat of global extinction on the Red List of the International Union for Conservation of Nature (IUCN). Important factors in bird decline are habitat loss and climate change. The factor habitat loss is often caused by human transformations of landscapes, such as agricultural expansion and urbanization. Agricultural expansion has a homogenizing effect on a landscape, meaning that landscapes become more similar to each other (Endenburg et al., 2019). This could lead to more similar forest bird communities in these landscapes. Also the urbanization of landscapes has a homogenizing effect on bird habitats neighbouring these areas (Sidemo-Holm et al., 2022). Namely, urbanization can lead to habitat fragmentation, noise pollution and light pollution, contributing to a lower bird species richness in these habitats. In the Netherlands, woodpeckers (*Picidae*) are in contrast to many bird species not endangered. Woodpeckers are resident birds and breed in the Netherlands. Moreover, as they have a big presence in the Netherlands, they are protected birds (Vogelbescherming Nederland, n.d.a). The most common woodpecker species in the Netherlands are the green woodpecker (*Picus viridis*), lesser spotted woodpecker (*Dryobates minor*), middle spotted woodpecker (*Dendrocoptes medius*), great spotted woodpecker (*Dendrocopos major*) and black woodpecker (*Dryocopus martius*). Although they are different species, they all have in common that they excavate cavities in trees. This has two main purposes. At first, the cavities could be used as a roost, a place for woodpeckers to rest (Jackson & Jackson, 2004). Secondly, the cavities can be used as nests for them to breed and protect their hatchlings from predators such as pine marten and owls (Puverel et al., 2019). By creating cavities, woodpeckers fulfil and important function in ecosystems (Zawadzki & Sławski, 2023). Namely, they provide ecosystem functions for other species. For example, microhabitat formation occurs in the created cavities and other species make use of the cavities when they are abandoned (Puverel et al., 2019). Thus, woodpeckers are important for conserving cavity dependent species.

1.4 LiDAR in ecological research

Even though the five woodpecker species mentioned are not endangered in the Netherlands, the observed trend of homogenization due to factors such as agricultural expansion and urbanization could also put pressure on the existence of these species. Especially because woodpeckers have an important role in ecosystems, it is important to preserve this family. In order to preserve woodpeckers, more information could be obtained about their habitat preferences. Furthermore, it was earlier mentioned that more research could be done in the creation of metrics that represent aspects of vegetation structure with LiDAR data. As woodpeckers are dependent on trees to breed in, aspects of vegetation structure, such as vegetation height, could be a big factor in the habitat selection of woodpeckers. This makes the woodpecker family a suitable family to research regarding their preferences for different vegetation structures. Thus, it can be explored more how LiDAR data can be used in ecological research in order to acquire more information about the habitat preferences of the woodpecker family in the Netherlands. Besides that research on woodpeckers specifically would provide more information about this family, the results of this research would also show how LiDAR data can improve ecological research into bird distributions in general.

1.5 Previous research

In prior research into the spatial distribution of birds, different metrics representing vegetation structure were calculated with the use of LiDAR data. Bakx et al. (2019) did research on which LiDAR metrics were most commonly used in 50 different papers. All papers had in common that LiDAR-derived vegetation metrics were calculated in order to research the distribution of species (mostly bird species). 77 unique metrics were found and Bakx et al. (2019) divided these metrics into several categories: vegetation cover, vegetation height, horizontal variability and vertical variability (see Figure 1b). Additionally, the research showed that these metrics could be calculated for different vegetation layers. As metrics that are part of these categories are important in many previous research, these metrics could be used as a good starting point for this thesis research. Furthermore, in research of Burns et al. (2020), LiDAR data were used to create canopy structure variables. These variables, along with other variables such as climatic variables, were used as predictor variables in machine learning models. These models were combined in order to calculate the variable importance of the variables. The results showed that canopy structure variables were the most important when predicting bird distribution in coniferous forests. As canopy structure variables are important for predictive models, it could be useful to assess their importance in descriptive models as well, using already existing bird distribution data. Nevertheless, generally less focus is put on the vegetation surrounding the nesting trees. For example, there could be explored more if these nesting trees are located in a dense forest or that they are more isolated. Also the influence of shrubs and bushes on the woodpeckers' choice for nesting trees could be researched.

1.6 Vector-based vegetation metrics

In the earlier described research of Bakx et al. (2019), most of the 50 researched papers used an area-based approach to calculate the different metrics. In this approach, the point cloud is rasterized, meaning that raster layers are created by calculating a value using the LiDAR points located in the area of each raster cell and assigning this value to the cell. The disadvantage of this approach is that some detail is lost when point clouds are rasterized to cells with a certain cell size. As a solution, an object-based approach can be used. In this method, the point cloud is segmented directly into objects, such as trees and hedges. This results into layouts with vectors such as polygons representing trees. Nevertheless, in the 50 studied papers, this approach was rarely used. That is why there is much that can be explored when using an object-based approach (besides an area-based approach) to create vegetation metrics. For example, vector-layers in which polygons represent vegetation units can be used as the input for different new vegetation rasters. For example, rasters could be created based on tree polygons.

1.7 Habitat preferences of different woodpecker species

Besides that there is much to explore in what LiDAR data can add to research into the habitat preferences of the woodpecker family in general, there is also a knowledge gap on the differences in habitat preferences between the five woodpecker species. There are several preferences of the five species that could be explored more during this thesis. Some of these preferences will be described below.

At first, the green woodpecker prefers to breed in trees close to open grass areas. In these open grass areas live ants, which are prey for this species (Villanúa et al., 2023). Moreover, snags are used by lesser spotted woodpeckers as both nesting and foraging site (Olsson et al., 1992). These are standing dead trees (Stitt et al., 2022). This species prefers areas with a higher density of snags (Olsson et al., 1992). The lesser spotted woodpecker is a small species, thus the wood of the trees needs to be soft in order to excavate holes (Smith & Charman, 2012). The dead wood of snags and dead branches located high in old trees are easier to excavate. For the middle spotted woodpecker, a deciduous forest with open canopies is preferred (Kosinski & Winiecki, 2004). There should be relatively isolated trees, preferably oak trees with a medium-sized crown. They prefer these types of habitats as open crowns increases the accessibility for birds. Moreover, a more open forest facilitates species with more sunlight. This attracts certain arthropods, which are hunted for by the middle spotted woodpeckers. Thus, the suitability for foraging is an important factor in the habitat selection of this species. The great spotted woodpecker forages in the trunks of trees, usually trees with a fissured bark in which arthropods can be found (Kosiński, 2006). The bird prefers older trees, as these have a bigger diameter and are therefore big enough for this bird to excavate nesting holes. Finally, the black woodpecker prefers trees that are more isolated, making the bird less prone to predators than when trees and the branches of trees would be located nearby (Puverel et al., 2019).

1.8 Overall research aim and research questions

This research has several aims. At first, the three sub-questions (RQ1, RQ2 and RQ3) will be explained. Thereafter, the general aim of this research will be discussed.

RQ1: Which metrics that provide information about vegetation structure and are potentially relevant for woodpeckers can be derived from LiDAR data?

The first goal of this research is to calculate LiDAR-derived vegetation metrics, also called LiDAR metrics, that are relevant for acquiring information about aspects of vegetation structure. These LiDAR-derived vegetation metrics could be important for the habitat selection of the woodpeckers. That is why the potential relevance of the LiDAR metrics, when used in the next steps of the research, for giving information about the habitat preferences of woodpeckers is also taken into account during the creation of the metrics.

RQ2: Which aspects of vegetation structure represented by LiDAR metrics have the biggest influence on the habitat selection of the woodpecker family?

The second goal of this research is to determine which metrics representing aspects of vegetation structure are the most important in the habitat selection of the woodpecker family. In this sub-question, several topics will be discussed. At first, it is described how the importance of certain metrics can be derived from the results and then ecological explanations for certain metrics being important will be discussed. Additionally, also the specific habitat preferences, the versions (values) of aspects (metrics) of vegetation structure, will be explained. For example, an aspect of vegetation might be vegetation height, but the version of this aspect could be a small vegetation height.

RQ3: How do the habitat preferences of the five separate woodpecker species differ?

In De Hoge Veluwe, different woodpecker species with potentially different habitat preferences are present. That is why besides the general habitat preferences of the woodpecker family (RQ2), also the habitat preferences of the five separate woodpecker species will be derived and compared. In RQ3, there will be focused more on describing the different habitat preferences of the woodpecker species and the importance of the different metrics will be explained less elaborately. This is done, as the general importance of the different metrics will already be explained in RQ2. Moreover, the specific habitat preferences of the separate woodpecker species can be supported by more in depth ecological literature than the habitat preferences of the woodpecker family in general.

Main question: How can the use of airborne LiDAR data improve the understanding of the habitat preferences of five woodpecker species in De Hoge Veluwe, the Netherlands?

The main aim of this research is to get more insight into the additional value of LiDAR data in ecological research into woodpeckers. It is researched how LiDAR data can be used to acquire information about aspects of vegetation structure. This information about vegetation structure can subsequently be used to derive information about the preferred versions of vegetation aspects of woodpeckers. In this way, the information derived from LiDAR could support already existing literature about the ecology of woodpeckers. Besides that this research could provide more information about the habitat preferences of woodpeckers in De Hoge Veluwe, the results of this research could also be an example of how LiDAR data can be applied in ecological research into bird species in general.

2. Methodology

In this methodology section, at first, the study area and data used in this research will be described. Thereafter, the methodology for answering RQ1, RQ2 and RQ3 will be described.

2.1 Study area

Figure 2 shows the study area of this research: De Hoge Veluwe. There is distinguished between deciduous and coniferous trees using the LGN2022 dataset (4TU, 2022). This dataset shows the different land use classes in the Netherlands. In this case, the raster cells with only deciduous or coniferous trees are selected. This area has a relatively young soil and many areas with coniferous trees can be found.

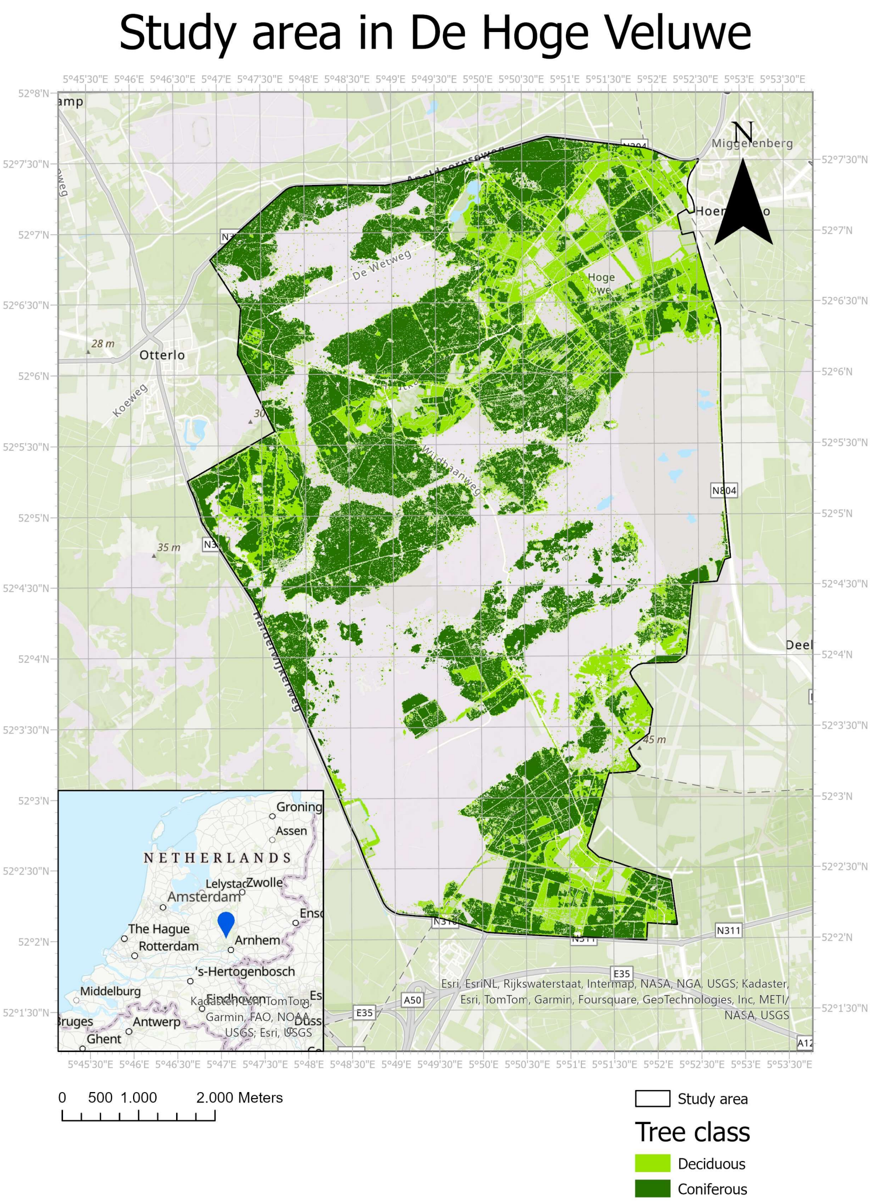


Figure 2: Study area in De Hoge Veluwe and its tree types. Dark green areas represent areas containing coniferous trees. Light green areas represent areas containing deciduous trees.

2.2 Data

2.2.1 AHN4

The ‘Actueel Hoogtebestand Nederland 4’ (AHN4) data were used in this research. This dataset is the most recent dataset showing elevation data of the Netherlands, measured between 2020 and 2022 (AHN, n.d.a). In the study area in De Hoge Veluwe specifically, LiDAR data were obtained in February of 2022 (ArcGIS, 2024). AHN4 point clouds were downloaded from Geotiles (Geotiles, 2024). These data were provided in smaller tiles of 1 by 1.25 kilometer. In total, 60 of these smaller tiles were downloaded. A table of the used tiles can be found in appendix L.

2.2.2 Woodpecker observation data

The foundation Sovon, which is the Dutch Centre for Field Ornithology, has provided woodpecker observation data of the study area. Different types of observations were made in the field and these observation points were each given a label. If an observation point is indicated as ‘true’, the bird breeds or lives at that location. If a point is indicated as ‘false’, the bird was seen at that particular location but the bird did not breed or live there. Only observation points with the label ‘true’ were selected, as for these points can be acknowledged with the most certainty that they live at that location. The ‘false’ points could be the same birds as the birds at the locations of the ‘true’ points, but then these birds were spotted at a location where the bird does not live. Namely, a bird could be flying around and be spotted multiple times. In total, for the five woodpecker species, 2658 observations were done in De Veluwe. When filtered, 712 observations remained.

2.3 RQ1: LiDAR metrics

In this part, the methodology of sub-question 1 will be described. In order to derive information about aspects of vegetation structure, 14 LiDAR metrics were eventually created. Table 1 shows these metrics and their division into three categories: vertical complexity metrics, horizontal heterogeneity metrics and vector-based metrics. For each metric, a short description is given. These metrics were selected with the use of literature (see ‘References’ column). In general, metrics were selected based on if they can give information about certain aspects of vegetation structure. Each metric represents a certain aspect of vegetation structure, such as the mean height of vegetation. While selecting these metrics, there was taken into account if their aspects of vegetation structure could potentially be useful to derive information about the habitat preferences of woodpeckers. Nevertheless, if the created metrics are useful for acquiring information about woodpeckers will be tested in RQ2 and RQ3. In RQ1, there will be solely researched which metrics are suitable/reliable for representing vegetation structure.

In the methodology of RQ1, there will be described why these metrics were chosen and why they are relevant. There will be described which aspects of vegetation structure the created metrics are supposed to represent and why these metrics could be potentially important for acquiring information about the habitat preferences of woodpeckers. Moreover, the creation of the metrics will be discussed. In the results, there will be described what information the resulting layouts of these metrics give about vegetation structure and in the discussion, there will be discussed why certain metrics are reliable enough to be used in the species distribution models of RQ2 and RQ3. In previous research, such as in the papers of Adhikari et al. (2023), Bakx et al. (2019), Kissling et al. (2023) and Koma et al. (2022), most vertical complexity metrics (*p25*, *p95*, *mean*, *SD*, *CV*, *kurtosis* and *skewness*) have already been created or described. Also the created horizontal heterogeneity metrics (*VR_total* and *VR_low*) and the vector-based metric *CCP* have been earlier implemented (Table 1). The other metrics, which have not been found to be calculated in previous research (*distance_inside*, *distance_outside*, *distance_ST*) were mainly created because their aspects of vegetation structure could give more insight into the habitat preferences of woodpeckers if they would be used in the next steps of this research (RQ2 and RQ3). In general, for the calculation of the metrics, several functions of the R package ‘lidR’ have been used.

Table 1: Overview of the 14 LiDAR metrics. The metrics are divided into three categories. For each metric, a description is given, as well as references that served as an inspiration for the creation of this metric.

Metric	**Description**	**References**
Vertical complexity metrics		
p25	25th percentile of vegetation height	Adhikari et al., 2023; Bakx et al., 2019; De Vries et al., 2021; Hill & Broughton, 2009; Ioki et al., 2014; Kissling et al., 2023; Koma et al., 2022; Li et al., 2022; Martinuzzi et al., 2009
p95	95th percentile of vegetation height	Adhikari et al., 2023; Bakx et al., 2019; De Vries et al., 2021; Hill & Broughton, 2009; Ioki et al., 2014; Kissling et al., 2023; Koma et al., 2022; Li et al., 2022; Martinuzzi et al., 2009
UC	Understorey/crown ratio	Adhikari et al., 2023; Bakx et al., 2019; De Vries et al., 2021; Hill & Broughton, 2009; Ioki et al., 2014; Kissling et al., 2023; Koma et al., 2022; Li et al., 2022; Martinuzzi et al., 2009
Mean	Mean of vegetation height	Adhikari et al., 2023; Bakx et al., 2019; Ioki et al., 2014; Kissling et al., 2023; Koma et al., 2022; Martinuzzi et al., 2009;
SD	Standard deviation of vegetation height	Adhikari et al., 2023; Bakx et al., 2019; Kissling et al., 2023; Koma et al., 2022; Martinuzzi et al., 2009
CV	Coefficient of variation of vegetation height	Adhikari et al., 2023; Bakx et al., 2019; Kissling et al., 2023
Kurtosis	Kurtosis of vegetation height	Adhikari et al., 2023; Bakx et al., 2019; Kissling et al., 2023; Koma et al., 2022; Martinuzzi et al., 2009
Skewness	Skewness of vegetation height	Adhikari et al., 2023; Bakx et al., 2019; Kissling et al., 2023; Martinuzzi et al., 2009
Horizontal heterogeneity metrics		
VR_low	Low vegetation roughness	De Vries et al., 2021; Koma et al., 2022

VR_total	Total vegetation roughness	De Vries et al., 2021; Koma et al., 2022
Vector-based metrics		
distance_inside	Distance to inside forest areas	Kosinski & Winiecki, 2004
distance_outside	Distance to outside forest areas	Kosinski & Winiecki, 2004
distance_ST	Distance to standalone trees	Kosinski & Winiecki, 2004; Puverel et al., 2019
CCP	Canopy cover percentage	Bakx et al., 2019; Kosinski & Winiecki, 2004; Davison et al., 2023

2.3.1 Vertical complexity metrics

The following metrics all provide information about the vertical complexity of vegetation. Vertical complexity represents the variability of vegetation in the direction of the vegetation height (vertical direction) (De Vries et al., 2021; Koma et al., 2022). These metrics are all calculated with the use of the 'pixel_metrics' function in R. For each metric, a different formula is applied to a point cloud, resulting in an output raster. The function uses the vertical column above a 2D raster cell. For example, if the cell size used in the function is 1 meter, the function takes alle the points in areas of 1 by 1 meter and applies the formula to these points. These per-cell operations are called local operations (ArcGIS Desktop, n.d.b). The entire calculation of all vertical complexity metrics can be found in appendix Aa01.

<u>Percentiles of vegetation height</u>

In this research, two metrics were calculated by taking the quantiles out of the LiDAR point cloud and assigning to raster cells the value of the data point that is located at that specific quantile. In case of the 25th percentile, the 0.25th point was selected per raster cell and its height value was assigned. The same was done for the 95th percentile. These two metrics were created to gain more insight into different vegetation layers. Namely, also in similar research, percentile values were used to indicate vegetation layers (Koma et al., 2022). In this thesis research, the 25th percentile of height was determined (*p25*) as the percentile that represents understorey vegetation. Understorey vegetation is vegetation that does not have direct access to sunlight due to being located under the canopies of taller vegetation (Hill & Broughton, 2009). Examples are shrubs, bushes and young small trees. Vegetation that does have direct access to sunlight, mostly trees, is called overstorey vegetation. The 25th percentile was chosen, as it already has been used earlier as a representation of lower vegetation (Adhikari et al., 2023). Moreover, the 95th percentile was calculated to represent overstorey vegetation (*p95*). The 95th percentile was used instead of all data points (100%) in order to remove big outliers from the dataset (Hill & Broughton, 2009; Koma et al., 2022). There was distinguished between understorey vegetation and the tree crown layer (overstorey layer), as understorey vegetation is important for many birds species regarding aspects such as foraging, whereas overstorey vegetation is often important when birds search for a tree to create a nest in (Davison et al., 2009).

<u>Understorey/Crown ratio</u>

Based on *p25* and *p95*, the understorey/crown ratio (*UC*) was calculated. The goal of this metric is to show the ratio between understorey vegetation height and tree crown height. The metric was calculated by dividing the *p25* raster by the *p95* raster and could give an indication of where vegetation can be found under trees. An *UC* of close to 1 would indicate that there is either no understorey layer and only trees or that there are no trees and only an understorey layer.

Mean and standard deviation of vegetation height
Variables that are often calculated in previous research is the mean and standard deviation of vegetation (Koma et al., 2022; Martinuzzi et al., 2009). These variables, similarly to *p25* and *p95*, give information about the distribution of LiDAR points in the vertical direction. For their calculation, the 'mean' and 'sd' functions were used. The standard deviation is a measure of the amount of variation around the mean value. A low standard deviation indicates that the elevation values of the LiDAR points are not very spread out and lie closer to the mean, meaning that there is less variation. Thus, if raster cells have a low standard deviation, this means that the LiDAR points representing biomass are located closer together and that vegetation is more densely distributed in the vertical direction than vegetation in cells with a high standard deviation.

Coefficient of variation of vegetation height
The coefficient of variation (*CV*) of vegetation height provides similar information as the standard deviation. The *CV* is the variability of the standard deviation from the mean and is calculated by dividing the standard deviation by the mean. If the mean is low and the standard deviation is high, this will result in a high *CV*. When *CV* values are between 0 and 1, the standard deviation is smaller than the mean.

Kurtosis and skewness of vegetation height
kurtosis and *skewness* are indicators of the shape of the vertical distribution of vegetation height values. Positive *kurtosis* values indicate that the distribution of height values is more tailed than a normal distribution, meaning that there is a higher quantity of outliers (Scribbr, 2024). This is often in combination with a higher peak. Lower *kurtosis* values indicate that there are less outliers in the distribution and commonly there is a flat peak. Commonly, a *kurtosis* value of 3 indicates that the distribution does not have more outliers than a normal distribution. This metric is relevant for this research, because it gives information about how close together the vegetation points lie. It could be assumed that trees with dense canopies have a higher *kurtosis* and a lower standard deviation, meaning that there are more outliers, than in areas where LiDAR pulses were able to penetrate the vegetation more and are more spread out. Moreover, *skewness* is metric that shows the asymmetry of a dataset (Scribbr, 2022). High *skewness* values (*skewness* > 0) indicate that the distribution of elevation values is skewed to the right (a peak to the left) and low values (*skewness* < 0) indicate that the distribution is skewed to the left (a peak to the right). Thus, *skewness* can give information about the location of the values in a distribution.

2.3.2 Horizontal heterogeneity metrics
Whereas vertical complexity metrics describe the distribution of vegetation points in the height/vertical direction, horizontal heterogeneity describes the distribution of vegetation points in the horizontal direction. The two types of metrics are calculated differently. Vertical complexity metrics are calculated by directly taking a point cloud as input and doing calculations based on the elevation values of the points. This has resulted in several raster layers. However, horizontal heterogeneity metrics are calculated by taking an already existing raster as input and calculating new values based on its raster values. In previous research, it was shown that horizontal heterogeneity has an influence on the distribution of other bird species such as the Savi's warbler (Koma et al., 2022). Moreover, besides in research on bird species, horizontal heterogeneity can also be applied in research on other species. In research on butterflies in the Netherlands was determined that horizontal heterogeneity, as well as vertical complexity, are important factors influencing the distribution of butterflies (De Vries et al., 2021).

Vegetation roughness
Both of the following metrics were based on research of De Vries et al. (2021). At first, *VR_low* was

created to show the differences in vegetation height for the understorey layer. Moreover, *VR_total* was created to show the differences in vegetation height in the tree crown layer. *VR_low* was created with the *p25* raster layer as input. A window of 3x3 moved over the whole raster, doing the same calculation for each cell. This predefined window is called a kernel. A calculation where each cell receives the value of a calculation done for all cells in a certain neighbourhood (kernel), is called a focal operation (ArcGIS Desktop, n.d.a). In this case, for each cell, the difference of the central cell with its 8 neighbouring cells was calculated. This resulted in 8 difference values. From these 8 values, the highest value, which is the maximum difference, was selected. Then, this maximum difference value was assigned to the middle cell and subsequently the moving window moved onto the next cell. The same calculation was done for *VR_total*, but then the *p95* raster was used as input. After the focal operations were done, the raster layers were smoothed with a window of 9x9. This was done to give more highlight to the areas with high heterogeneity. The entire calculation of both metrics can be found in appendix Aa01.

2.3.3 Vector-based metrics

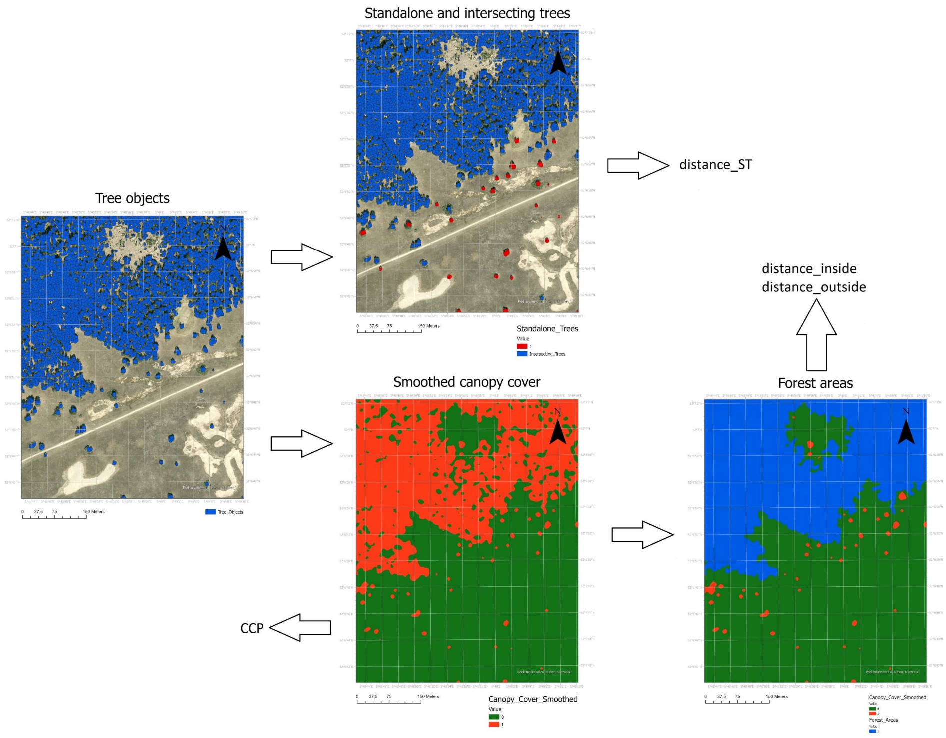


Figure 3: General workflow of the creation of the following metrics: distance_ST, distance_inside, distance_outside and CCP.

The following metrics were all calculated by taking polygons representing tree canopies as input (Figure 3). Trees were detected using several functions in the 'lidR' package. At first, a canopy height model was created. A canopy height model is a raster which values represent the height of the canopy of the vegetation. The raster is created by subtracting the digital terrain model (DTM) from the digital surface model (DSM) (Earth Lab, 2020). The CHM was created with the 'pitfree' algorithm. During the creation of LiDAR points, some pulses can penetrate through the canopy without giving a first return. This results in that first return points can be located at a too low elevation, resulting in pits in the canopy. The 'pitfree' algorithm fills many of these pits by creating multiple CHMs for different canopy layers, resulting in a CHM with fewer pits (Khosravipour et al., 2014). The created CHM was used as input into the 'locate_trees' function. With a window size of 5 meters, the function searches for the highest points in the area and selects the tree tops. Then, the Dalponte2016 algorithm was used to create an algorithm representing the relationship between the presence of the tree tops and the CHM (Dalponte & Coomes, 2016). This algorithm was chosen, because it can be derived from a CHM and can be used to group points clouds into tree units. Namely, the algorithm was applied to points clouds and the points were segmented into groups with the same ID, each group belonging to a separate tree. Using the segmented point cloud, the 'crown_metrics' function was used to create polygons representing trees. Then, this vector dataset was used as the input for the calculation of the metrics described below.

Distance to standalone trees

Literature has suggested that the black woodpecker and middle spotted woodpecker prefer more isolated trees (Kosinski & Winiecki, 2004; Puverel et al., 2019). That is why a metric was created that shows the distance to the nearest standalone/isolated tree (*distance_ST*). This metric was calculated by doing several steps. At first, it was taken into account that trees can have multiple tree tops. There are cases where multiple tree canopies are detected by the 'locate_trees' function that all belong to one tree. Therefore, tree crown polygons that are located within 1 meter of each other were merged in order that they would count as one tree. This was done by giving all trees a buffer of 0.5 meter and creating a union of these buffered trees. Then, the surface area of all polygons was calculated and after having visually checked the general surface area of multiple trees with more than one crown, a threshold of 210 m^2 was determined. All merged polygons with a higher surface area than 210 m^2 were filtered out. After this, the 'st_is_within_distance' function was used to find the trees that were located within 10 meters of each other. This distance threshold of 10 meters was determined by visually analysing the more isolated trees and their minimum distances from other trees in the study area. Trees within 10 meters were defined as intersecting trees and trees with a distance greater than 10 meters were defined as standalone trees. Eventually, after all standalone tree polygons were rasterized, the 'distance' function was used to calculate for each raster cell the nearest distance to a standalone tree. This calculation of distance is called a global operation (ArcGIS Desktop, n.d.b). The entire calculation of standalone tree polygons can be found in appendix Aa02. The merging of the standalone tree rasters and the use of the distance function can be found in appendix Aa03.

Canopy cover

Before the next three metrics could be calculated, buffers were created around the tree objects. The size of the buffer was based on the height of the tree crowns. Taller trees received a bigger buffer around their crown. In this way, the effect of tree height in combination with its crown was taken into account. Namely, taller trees with the same crown area as smaller trees have a bigger shadow than smaller trees. In the next step, the crowns of the trees were rasterized and this raster was smoothed. The resulting raster represented the canopy cover of forest areas. This raster was used as the initial input of the calculation of the following three metrics. The entire calculation can be found in appendix Aa01.

Distance to inside and outside forest areas

Besides the presence of isolated trees, forest edges could also be an important factor for woodpeckers (Kosinski & Winiecki, 2004). For example, there is indicated that the green woodpecker lives and forages on the forest edge (BirdLife International, 2024a). That is why it is interesting to research this statement for the green woodpecker and also for the other woodpecker species. The *distance_inside* metric was calculated because it indicates how big distances birds are willing to fly to reach forest areas and it indicates if they prefer trees outside forest areas more than trees inside forest areas. The metric was calculated as follows. With the canopy cover layer as input, the 'patches' function was used to detect areas with connected tree canopies. Then, small open areas of less than 3500 m^2 were filled up and therefore counted as part of a bigger forest areas, whereas areas with a surface area higher than 3500 m^2 were defined as open areas and were not filled up. Moreover, areas of trees smaller than 3500 m^2 were deleted, as these tree areas were not big enough to count as a forest area. In this way, the study area was simplified into areas with forest and areas with no forest. The thresholds were kept the same for defining forest areas as well as open areas, because then the simplification of the study area was kept consistent. After the simplification, the 'distance' function was used to calculate the nearest distance to forest areas for each raster cell (global operation). The same calculations were done for the *distance_outside* metric, except for that distances to open areas were calculated instead of distances to forest areas. This metric was created in order to get more insight into how deep inside the forest woodpeckers have their habitat locations. A low distance would indicate that a woodpecker prefers a habitat inside the forest that is close to the forest edge. The calculation of both metrics can be found in appendix Aa03.

<u>Canopy cover percentage</u>
The canopy cover percentage metric (*CCP*) is relevant due to that woodpeckers might have a preference for forests with a more open or more closed canopy. A reason for woodpeckers to prefers a habitat with less canopy cover could be that the trees in that habitat are more accessible when trees are standing less close together (Kosinski & Winiecki, 2004). Moreover, areas that are more open (more sunlight can shine through the canopy) could contain different plant species and food resources, which could be beneficial for some birds (Davison et al., 2023). Canopy cover percentage is calculated by assigning each raster cell the percentage of neighbouring cells that contain canopy cover for a certain area (focal operation). In this case, the moving window used is 41 by 41 meters. For the metrics entire calculation, see appendix Aa01.

2.3.4 Upscaling of LiDAR metrics

Now that the calculation of all 14 metrics has been described, the workflow of upscaling these metrics to the whole study area will be explained (Appendix B). The squares represent process steps and the cylinders represent data. As mentioned earlier, 60 LiDAR tiles were downloaded in order to cover the whole study area. The grey parts of the workflow were repeated for each LiDAR tile. The white parts represent processes on the scale of the 60 LiDAR tiles all together. The workflow starts with one LiDAR tile as input. Then, the LiDAR tiles were split into 9 sub-tiles. This was done, as the calculation of point clouds for the whole LiDAR tile would be computationally too heavy for the laptop used in this thesis. For each sub-tile, a buffer of 10 meters was created. Namely, some metric calculations contained focal operations, which operations sometimes required the environment around the extent of the original sub-tiles. These buffers were created in order to be deleted after the calculations were done. The coloured parts of the workflow represent for-loops that were run for each sub-tile. Each colour represents a different for-loop. All in all, in the upscaling process, the same for-loop that includes the parts up to and including the grey parts in part 3 was run 60 times.
Part 1 displays the process of creating all vertical complexity metrics except UC. Besides, tree crown polygons were created. At first, each sub-tile was normalized in a for-loop. Normalization was done with the 'normalize_height' function. The function's output was a LASCatalog collection containing normalized all return points (nlas_AR). Thereafter, first return points were filtered out and a new collection, nlas_FR, was created. The point clouds used to calculate all 14 metrics were normalized. Besides that normalization is a common practice in previous research such as Kissling et al. (2023), this choice was made due to that insights are preferred to be obtained about the height of the different vegetation layers instead of heights that give information about the height of vegetation and the ground together. Moreover, metric layers would not be comparable with each other if some layers were normalized and others not. Additionally, some metrics are dependent on each other, for example *UC* is dependent on *p25* and *p95*. Nevertheless, metrics were either calculated with all return data (nlas_AR) or first return data (nlas_FR). The input of *p25* and *p95* is nlas_AR, as all return points needed to be used in order to represent different vegetation layers. The *mean*, *SD*, *CV*, *kurtosis* and *skewness* only use first return points as these metrics were created to show the vertical complexity of the top of the vegetation. In part 2, the seven already created metrics (indicated as 'First metrics (7)') were cropped per sub-tile, in order that the earlier created buffer was deleted. The raster layers of the sub-tiles were merged and smoothed with a moving window of 3 by 3 meters in order that the differences between areas with different values were highlighted more. For

smoothing, a median filter was used as this filter preserves the detail in the raster image, whereas when a mean filter is used, outliers can give a less realistic view of the neighbourhood cells. Besides these steps, *UC*, *VR_low* and *VR_total* were created with the *p25* and *p95* raster layers. After this, all the so far created metrics were cropped again by 10 meters. This was done because all downloaded LiDAR tiles initially had a buffer of 10 meters. Furthermore, the earlier created tree object polygons were used to create canopy cover rasters and the metric *CCP*. These two types of rasters were subsequently cropped by 20 meters. In part 3, the same tree objects were used to calculate *distance_ST*. Moreover, canopy cover was used to calculate *distance_inside* and *distance_outside*. These three metrics were directly calculated for the extent of the whole study area. Moreover, the remaining 11 metrics were merged to the extent of the study area. Thereafter, all 14 metrics were clipped to the exact borders of the study area and this resulted in the raster layers of the final metrics. All metrics except for *distance_ST*, *distance_outside* and *distance_inside* were calculated with a resolution of 1 meter. The three distance metrics got a lower resolution of 2 meters, because computationally the calculation would have been to heavy (for the laptop used in this thesis) when a resolution of 1 meter would be used. Eventually, also the other metrics were resampled to a resolution of 2 meters. All in all, the duration of running the scripts of all 60 tiles was in total approximately 60 hours (1 hour per tile). The scripts of the whole upscaling process can be found in appendix Aa01, appendix Aa02 and appendix Aa03.

2.4 RQ2 and RQ3: Maxent model

After all metrics were calculated, the created LiDAR metrics were used to derive information about the habitat preferences of the woodpecker family in general and the separate woodpecker species. At first, the methodology that RQ2 and RQ3 have in common will be explained. Thereafter, the differences between the two sub-questions will be described. In order to research the habitat preferences of woodpeckers, the Maxent model was used to get insight into the variable importance of the different metrics for the different woodpecker species. Maxent (Maximum Entropy) is a species distribution model that predicts the probability of a species being present at certain locations in a study area (Esri, 2024a). The model uses presence-only data and environmental data (in this case the metric rasters) and gives as output a probability distribution of a species being present for different environmental conditions (Phillips et al., 2006). This model was chosen over other possible models such as a GLM, because this model was specifically developed for situations where there is only presence data of a species available (which is the case in this research).

2.4.1 Background points

Besides presence data, the model also takes randomly generated background points as input. Background points, also called pseudo-absences, are points that show at which environmental conditions a species can either be present or absent (Barbet-Massin et al., 2012). This means that background points can either be located at the locations of woodpecker observations or at locations where the woodpecker was not observed. The following amount of observation and background points were used in the Maxent model. The background points were determined by multiplying the amount of observation points by 2, as then was made sure that a sufficient amount of background points would be potentially at the locations of the presence points and areas with no presences (absence areas).

Table 2: The amount of observation points and background points used in the Maxent model of the black woodpecker, great spotted woodpecker, middle spotted woodpecker, lesser spotted woodpecker, green woodpecker and all woodpecker species.

Species	Observation points	Background points
Black woodpecker	44	88
Great spotted woodpecker	537	1074

Middle spotted woodpecker	20	40
Little spotted woodpecker	73	146
Green woodpecker	38	76
All woodpecker species	712	1424

2.4.2 VIF

Besides that background points were created, the values of the 14 different metrics were extracted for both the presence and the background points. A diameter of 50 meters was used in order to make sure that the surroundings of the points were taken into account but at the same time that the details of the metrics were not lost. Before the Maxent models could be run, variables with a high collinearity with other variables were excluded. To check if there is multicollinearity, the variance inflation factor (VIF) was calculated (Naimi et al., 2014). The VIF is an indicator of the inflation of the standard errors due to multicollinearity (Pradhan, 2016). When a variable has a VIF of 1, it means that there is no correlation between that variable and other variables. A VIF of 1 to 5 indicates a moderate correlation. A VIF of higher than 5 indicates a stronger correlation and a VIF higher than 10 indicates a high correlation. A VIF of 3 or below is considered low enough for these variables to be used in the model and variables with a VIF higher than 3 were excluded. Metrics that were already deemed as unreliable beforehand (RQ1) were not included in the subset of variables to be tested on its VIF value. Why they were unreliable will be discussed in the discussion. On the other hand, priority was given to *p95* and *CCP*, which meant that these two variables could not be excluded after the VIF test. This was done because these two variables were of interest to acquire more knowledge about. For example, *p95* was chosen over *mean* as *p95* filtered out big outliers and *mean* did not. Moreover, canopy cover percentage could show the influence of rasters that are derived from detected tree polygons.

2.4.3 Model output

After the metrics were prepared, the models were run. The Maxent model randomly changes the values of the presence points (Elith et al., 2011). By measuring the differences in accuracy of the model when metric values were changed, the relative variable importance, also called permutation importance, could be derived per metric. This specific accuracy measure, AUC, will be explained later in the methodology. All in all, before predicting new suitable areas, the model already looks at how much the variables contribute to the presence of the already known observation points. In this thesis, only this descriptive part of the Maxent model will be run, as a goal of the thesis is to explain the already known occurrences of the woodpeckers. Besides variable importance values, the model also produces response curves, which show the relationship between metric values and the probability of a species being present (Phillips, 2021). The x-axis shows the range of values to which the metric was randomly changed. The variable on the y-axis of the plot is called the 'cloglog output', which is an estimation of the probability of the woodpecker to be present between 0 and 1 (Phillips, 2021). If the probability values stay approximately the same over the whole range of metric values (no preference for specific values), this indicates that this metric does not have a high importance. Response curves should be interpreted by looking at the steepness of the curve and the differences in probability values. When certain metric values have the highest probability values, while there is a big difference from the lowest probability values and there is a clear trend (steep curve), this indicates that this metric has a high importance and that there is a clear preference for specific metric values. Moreover, the Maxent model gives two different types of response curves as output (Phillips, 2021). The first one shows how the probability of the presence of a species changes when the values of a variable are randomly changed, while the values of the other variables stay at their mean values. Here, the influence of the variable is shown in combination with the influence of the other variables. This means that if other variables have a higher influence on the probability, this could influence the probability values of that certain variable. This type of response curve shows the metric values that are preferred and the most suitable for the species. In this thesis, this type of response curve will be called a total response curve and its values will be called suitable preferred values. In the second type of response curve, the values are shown that a species prefers when it only takes one specific metric into account in its habitat selection. This type shows just like the total response curve probability values, but then only one variable is used in the Maxent model and dependencies between that variable and other variables are not included. This type will be called an individual response curve (as only one individual metric is taken into account) and its values will be called individual preferred values. In this thesis, the total response curves will be given the most importance, as these curves show the influence of the different variables when other variables are also taken into account and thus show what values a species will eventually choose for its habitat. All in all, the total response curves show a trade-off between the values of the different variables/metrics used in the Maxent model. The species might tolerate the less preferred individual values of the less important variables in order to still have the individual preferred values of the more important variables. Thus, the Maxent models give as output the metric value combinations that are overall the most suitable for the species. These values can differ from the values in the individual response curves. Big differences between total response curves and individual response curves could be an indication of that the variable has a low importance in the model compared to the other variables. That is why comparing total and individual response curve could give an additional view of the importance of certain variables for the habitat selection. Additionally, when translating the models to ecological terms, the total response curves show the eventual habitat preferences for different metric values, the individual response curves show the habitat preferences of the metrics while the preferences of other metrics are already taken into account (which means that there is no influence of other variables) and the habitat selection of a species could be represented by the combination of the most suitable preferred values of all metrics.

2.4.4 Accuracy

Before more information can be given on permutation importance, certain accuracy measurements of the model have to be explained. At first, the model automatically measures the amount of omissions and commissions. Omissions are presence points that are misclassified as non-presence points. Commissions are rightly classified presence points. The omission rate is the proportion of presence points that were misclassified (Esri, 2024a). Furthermore, the model creates a graph in which the receiver operation characteristic (ROC) is plotted (Wei et al., 2018). This graph describes the relationship between the sensitivity and 1 minus the specificity. The sensitivity (1 minus the omission rate) is the proportion of presence points that were correctly classified. This should be maximized for a good model accuracy. 1 minus the specificity is the background or absence points that were falsely classified as potential presence points. This value should be minimized. The area under the ROC curve (AUC) is a measure of accuracy of the Maxent model. The AUC value is an indicator of how good the model can discriminate between presence points and absence or background points (Elith et al., 2011). An AUC value of 0.5 is comparable to a random prediction. Higher values indicate that the presence of the points can be explained by the variables. According to Swets (1988), values between 0.5-0.6 are failing, 0.6-0.7 are poor, 0.7-0.8 are fair, 0.8-0.9 are good and 0.9-1 are excellent. Nevertheless, the AUC values calculated by the model could be more reliable when absence points are used instead of background points. Namely, for the calculation of 1-specificity, background points that are located at the locations of presence points can be indicated as falsely classified presence points while they should be indicated as correctly classified presence points (Tesfamariam et al., 2022). When absence points are used, there is a certainty that the points can not be found at the location of the presence points and the indication of falsely classified presence point should then be true. Thus, when background points are used in Maxent, the AUC values can be lower than actually should be the case. That is why poor AUC values (0.6-0.7) could also still be acceptable.

2.4.5 Violin plots

Besides variable importance plots and response curves, violin plots were created. Violin plots show the distribution of values with the use of density curves (Atlassian, 2024). A bigger density (wider curve) indicates that the frequency of points with values in that range is higher. Violin plots show just like individual response curves which metric values are more preferred individually (no other variables are taken into account). They give more insight into if the values found at the observation points differ from the values found at random background points in the study area. For the representation of the study area, the background points of all woodpecker species were used (1424 points). Violin plots can not be used to infer information about the variable importance of the different metrics. They can only be used to get insight into the values species prefer for each variable individually and to compare the different woodpecker species with the general study area and with each other. If a specific metric has a high variable importance, the individual preferred values of this metric match more with its suitable preferred values, meaning that in this case its violin plot can also give more insight into the suitable preferred values of this metric. However, just like the individual response curves, if the values of the violin plot of a specific metric differ much from the suitable preferred values in the total response curves, it can indicate that that metric is not very important in the total habitat selection. To get more insight into the general differences between observation and background points (only in RQ2), it was tested for the collection of all woodpecker species if there are significant differences between the two groups. This was done using either a t-test or a Wilcoxon rank-sum test. The t-test was used if both groups have a normal distribution. The Wilcoxon rank-sum test was done if both groups do not have a normal distribution, for example a skewed distribution with the tail to the left or right. If the chosen test results in a p-value of smaller than 0.05, there is a significant difference between the two groups.

2.4.6 Differences between RQ2 and RQ3

In RQ2, there will be focused much on the importance of the different aspects of vegetation structure represented by LiDAR metrics. In the results, the variable importance values of each metric will be discussed. Thereafter, the total response curve of each metric will be described, resulting in information about the suitable preferred values of each metric. Moreover, the total response curve of each metric will be compared to its individual response curve and violin plot, in order to get more insight into the importance of the metrics. Then, in the discussion, the importance of the metrics will be discussed. Moreover, the suitable preferred values of each metric, which represent the habitat preferences of the woodpecker family, will be discussed. In RQ3, there will be given more importance to the differences in habitat preferences between the five woodpecker species. In the results, the derivation of the importance of the different variables will be analysed less elaborately. There will be focused more on describing the different suitable preferred metric values of the total response curves and the total response curves will only occasionally be compared with the individual response curves and violin plots if needed. Then, in the discussion, the suitable preferred values of each woodpecker species, which represent the habitat preferences of that species, will be discussed and in this way the habitat preferences of the different woodpecker species will be compared. Thus, RQ2 and RQ3 are similar, but the biggest differences between the two sub-questions is that in the results of RQ2 a more elaborate explanation of the importance of the metrics will be given and that in the discussion of RQ3 there will be focused more on the ecological explanations for the found habitat preferences of the different woodpecker species.

3. Results

In this part, the results of RQ1, RQ2 and RQ3 will be described. At first, the layouts of the created metrics will be described (RQ1). Then, the results of the Maxent model of all woodpecker species will be displayed (RQ2). Finally, the results of the Maxent model of the separate woodpecker species will be described (RQ3). For both RQ2 and RQ3, occasionally the additional information of violin plots will be described. Finally, the accuracy results of the models used in RQ2 and RQ3 will be shown.

3.1 RQ1: Layouts of LiDAR metrics

In this section, an overview will be given of the resulting layouts of the 14 calculated metrics. In figure 4, layouts are displayed of the seven most important variables. These layouts are chosen, because they have the highest variable importance according to the results of the Maxent model of all woodpecker species combined. These model results will be explained in RQ2 and RQ3 of the results. For now, the values displayed on the layouts will be described and the metric values will be compared to the reference study area (Figure 4). The layouts of the seven remaining metrics will also be described in this part of the results and can be found in appendix C. The study area shown in figure 4 is a small part of the entire study area. This smaller area is chosen, as it displays a diverse area with dense forest, less dense forest and open areas. For example, in the south, a forest with dense canopies can be seen. Moreover, in the northeast, a forest containing trees with a less dense canopy (and thus more openings) is observed. Furthermore, in the west, there is a big open field and in the north, two smaller open fields with one containing water can be seen.

3.1.1 Vertical complexity metrics

In figure 4, there are two vertical complexity metrics. At first, *p95* shows low values in the open fields and high values in the forest areas. Individual trees can be observed in the open fields and in the east, taller trees are located than in the west. Moreover, *CV* shows that low values can be found in the middle of the tree canopies and that at the edges of the trees and other vegetation units, higher *CV* values are found. For example, in the northeast, the forest seems to be more yellow than the area in the south. This indicates that higher standard deviations are found compared to the mean vegetation height at the edges of vegetation units such as trees and bushes. This also indicates that the points are more spread out at these locations. Additionally, NA values can be found at locations with no vegetation, such as the open fields, openings between canopies and areas with water, as dividing the standard deviation by 0 leads to NA values. In appendix C, the *SD* metric shows a similar pattern as *CV*, with higher values in more open vegetation areas and lower values in the middle of vegetation units. Nevertheless, unlike *CV*, *SD* shows values of 0 meter in open fields and openings instead of NA values. The fact that *SD* does not have NA values at these locations is due to the fact that even though the mean at these locations is 0 meter, the standard deviation can still be higher than 0 meter. Moreover, the *mean* shows similar patterns as *p95,* but in general a bit lower values. Furthermore, *p25* has high values in the denser forests, which values are quite similar to the values of *p95*, in the range of 10 to 20 meters. Moreover, in the less dense vegetation areas, many areas with a height of 0 meter are found. This resulted in that *UC* shows values close to 1 in the areas with dense canopies and values of 0 in the more open areas where *p25* also has values of 0 meter. Moreover, in the layout of the metric *kurtosis* can be seen that high values (higher values than 3) can be found at locations were the point density is high. This is usually at locations with a low standard deviation. The lowest standard deviations can be found in open areas with very low vegetation such as grass. Most raster cells there have a mean height of 0 meter. Also *p95* showed values of 0 meter as the outliers are cut off by the 95th percentile. Nevertheless, as the points are very densely distributed near the ground, the outliers will be bigger and if a maximum height layout of the area was made, values that are a bit higher than 0 meter would be seen due to the presence of these outliers. In the canopies of trees, the points are also densely distributed but still less dense and with higher standard deviations than in grasslands, with the highest values on the edges of the canopies.

The points are still spread out enough in order that less outliers are observed than in a normal distribution, resulting in that lower *kurtosis* values than 3 are most commonly found in the canopies of trees. Overall, the highest *kurtosis* values can be found at locations with very low vegetation such as grass. Close to the edges of vegetation units such as trees and shrubs, some very low vegetation can be found and this results in that higher *kurtosis* values can be observed just outside the edges of vegetation units. Thus, high standard deviations can be found on the edge of vegetation units and high *kurtosis* values can be found just outside the edge of vegetation units and in general in areas with very low vegetation such as grass. In most open areas with no vegetation, many NA values can be found. Furthermore, the *skewness* layout shows that inside denser forests, there are values around 0, meaning that the distribution is not more skewed than a normal distribution. Nevertheless, just like for *kurtosis*, very low vegetation such as grass has a higher *skewness*, which indicates that the distribution has a peak in elevation values to the left and a tail to the right in these areas. This means that most points in these areas have elevation values close to 0 meter. Besides this, stripes with higher *kurtosis* and *skewness* values can be observed. Why this is the case will be explained in the discussion.

3.1.2 Horizontal heterogeneity metrics

The *VR_total* layout shows that higher roughness values are found in the areas containing less vegetation. Especially at the edges of the vegetation units and in between the units, higher values are found. A high roughness means that the height differences are bigger, meaning that at the edges of the trees and shrubs, bigger height differences are found. *VR_low* shows similar results, only that much more areas with a roughness of 0 meter can be found (Appendix C).

3.1.3 Vector-based metrics

In the second row of figure 4, three vector-based metrics can be seen. At first, *distance_ST* displays the distances to the nearest standalone tree. Raster cells with a distance value of 0 meter indicate that they are at the location of a standalone tree. In the layout, it can be seen that several standalone trees are present in the open fields while in the forests higher distances from standalone trees can be found. Furthermore, *distance_outside* shows that the highest distances to outside the forest can be found deeper inside the forest and the lowest values are found closer to the open fields. This also means that in the open field areas, the distance values are 0 meter. The opposite is the case for the *distance_inside* metric, where higher distances away from the forests are found in the middle of the fields and distances of 0 meter can be found in the forest areas. Finally, the last vector-based metric is *CCP*. The layout displays the highest values in the forest, where the most canopy cover is present. Intermediate values can be found at the edges of the forest or areas with more openings in the canopy. Values of 0% can be found in open areas with no trees.

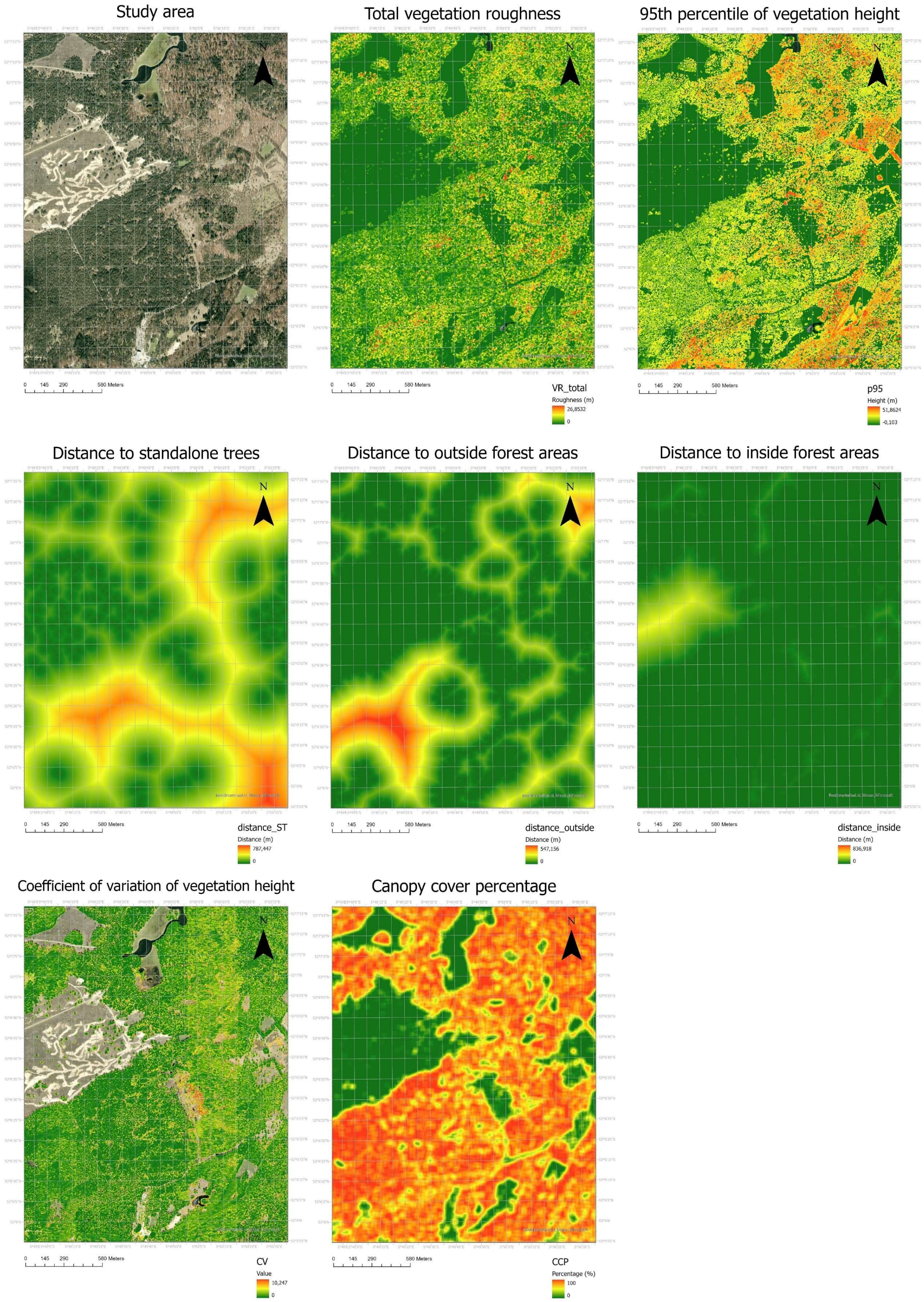


Figure 4: Layouts showing the seven most important LiDAR metrics, as well as the study area as a reference area. The metrics displayed, from left to right and top to bottom, are: VR_total, p95, distance_ST, distance_outside, distance_inside, CV and CCP.

3.2 RQ2: Model results of woodpecker family

After metric values were extracted for all woodpecker observations, VIF values were calculated in order to derive if there is multicollinearity between the metrics. In general, the following variables were excluded from the subsets used in the Maxent models: *mean*, *SD*, *skewness*, *kurtosis*, *p25*, *VR_low* and *UC*. The *mean* and *SD* were excluded due to high collinearity with other metrics (see correlation plots in appendix D). The other metrics were not included due to that they were deemed not reliable enough to be used in the model. Why this is the case will be explained in the discussion of RQ2. Eventually, 11 variables were excluded and for most species, seven variables remained: *VR_total*, *p95*, *distance_ST*, *distance_outside*, *distance_inside*, *CV* and *CCP*.

In this part of the results, the results of the Maxent model of the collection of all woodpecker species will be described. There will be focused on the differences in importance between the variables used in the model. At first, the permutation/variable importance results of the model will be described. Moreover, the total response curve of each metric will be analysed in order to derive the suitable preferred values (habitat preferences) of the woodpecker family. Furthermore, the total response curve of each metric will be compared with its individual response curve and violin plot. This is done to show how much the individual preferred values (of the individual response curve and the violin plot) of certain metrics are affected by the other variables in the model. If, for a certain metric, the individual preferred values of the individual response curve and violin plot are very similar to the suitable preferred values of the total response curve, this could indicate that this variable is quite important in the model. If there are many differences, other variables could be more dominant in the model.

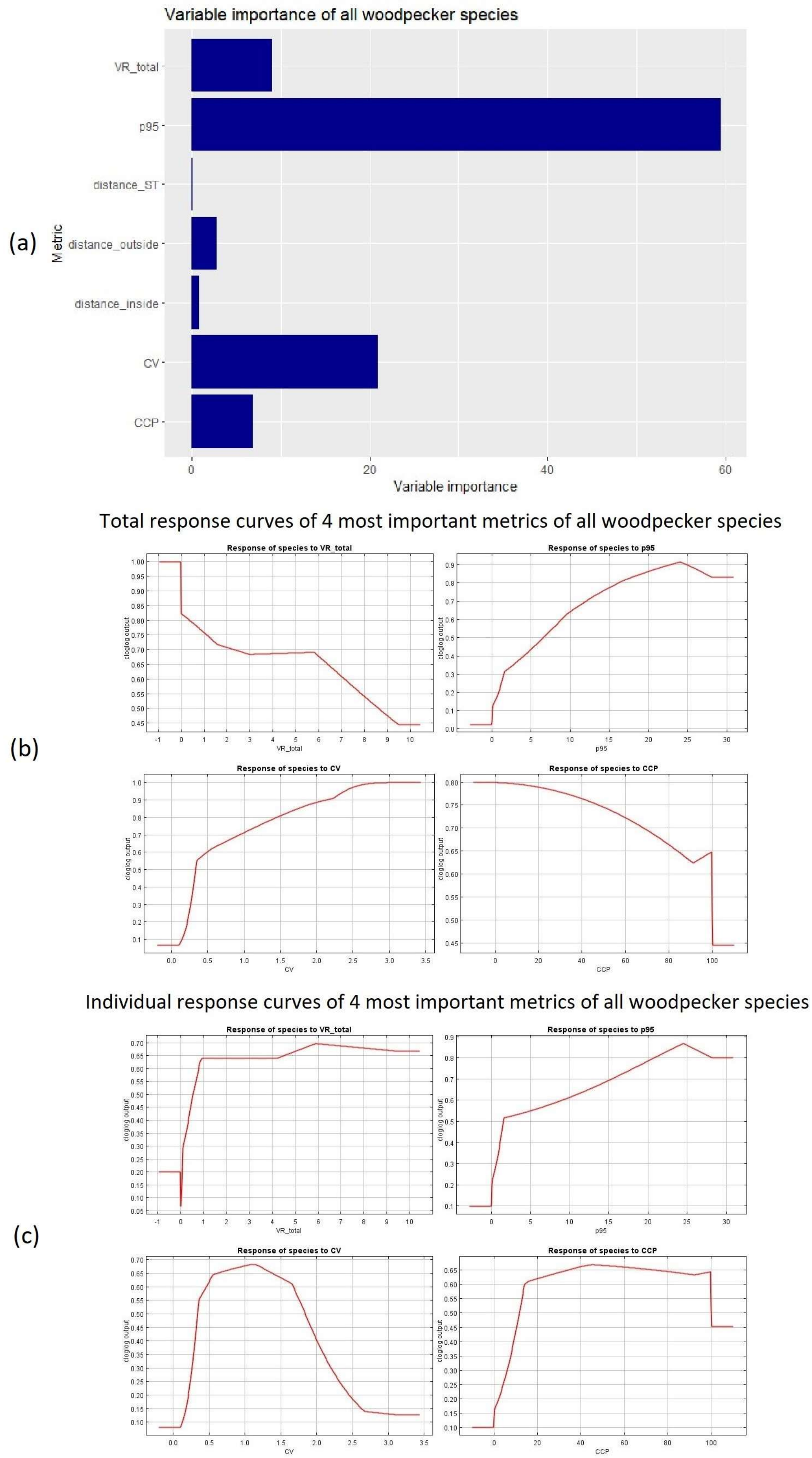


Figure 5: (a) Variable importance of each metric for the collection of all woodpecker species. (b) Total response curves of the four most important metrics for the collection of all woodpecker species. (c) Individual response curves of the four most important metrics for the collection of all woodpecker species.

3.2.1 Variable importance

Figure 5a shows the permutation importance values resulting from the Maxent model run for the presence points of all five woodpecker species combined. The barplot shows that the metric *p95* has the highest variable importance out of all metrics (59.4%). The second most important variable is *CV* (20.9%). Besides these metrics, *VR_total* and *CCP* have a smaller but still considerable importance of 9.1% and 6.9% respectively. The remaining variables, *distance_ST*, *distance_outside* and *distance_inside* have a low significance in this model with importance values ranging from 0.15 to 2.8%. The results of this Maxent model show that the biggest change in AUC value occurs when the values of *p95* are changed randomly, meaning that *p95* contributes the most out of all variables to the presence of the five woodpecker species at their chosen habitat locations. Figure 5b shows the total response curves and figure 5c shows the individual response curves of the four most important metrics. Moreover, in figure 13 in appendix G, violin plots are displayed of the four most important variables used in the Maxent model. Each plot contains a violin plot of the background points, representing the whole study area, and a violin plot of all woodpecker observations. The violin plots of the separate woodpecker species are also added for comparison and will be discussed in the results of RQ3. Furthermore, in appendix H, an overview is given of the statistical tests used to test if there is a significant difference between observation and background points for each metric. Moreover, the median and mean values of the observation and background points, as well as the p-values resulting from the statistical tests, are included.

3.2.2 VR_total

In the total response curve of *VR_total*, the highest probability can be observed for *VR_total* values in a range from 0 to 6 meters. Even though values close to 0 meter are preferred in the total response curve, they are not preferred in the individual response curve. Additionally, the violin plot of this metric shows that *VR_total* values are significantly but only slightly higher for observations compared to background points. The median is 2.92 compared to 2.69 meters. Thus, for *VR_total*, its individual preferred values are slightly higher than the study area and are not close to 0. Moreover, it can be observed that the suitable preferred values are the same, but differ slightly as values close to 0 are also still suitable. This difference between the total response curve and the individual response curve/violin plot shows that *VR_total* does not have a very big variable importance in the model. Overall, from all this information can be inferred that areas with some vegetation roughness (0-6 meters) are preferred by woodpeckers.

3.2.3 p95

For *p95*, the total response curve shows a clear preference for higher values. *p95* has its highest probability around 25 meters. Compared to the other metrics, the biggest differences in probability values (the curve has the steepest line) is observed for this metric, showing the specific preference for certain metric values, in this case higher *p95* values, and thus also the importance of this metric. Other metrics have less steep lines and show a less specific preference. Moreover, in general, the total response curves of metrics with a high variable importance do not differ much from their individual response curves, as their individual preferred values have a higher importance and are therefore also more suitable when other variables are taken into account. This can be observed for *p95*, which has a high variable importance. The individual response curve of *p95* shows not much difference from the total response curve, which means that the individual preferred values match the suitable preferred values. Also in its violin plot, this preference for higher elevation values can be clearly seen. Namely, the density of the observation points is the biggest for higher values compared to the peak density of the background points, meaning that woodpecker observations are generally located at a higher 95^{th} percentile of vegetation height than random points in the area. The density of the violin plot of the background points has two peaks, one at the bottom (close to 0) and one around the median of 9.4 meters, whereas the density of the observation points only has only one peak, around the median of 11.5 meters. Moreover, the observation points have a significantly higher median.

3.2.4 CV

CV has the highest probability values at *CV* values ranging from 0.5 to 3.5, meaning that the suitable preferred values are: 0.5-3.5. In this total response curve, also values higher than approximately 1.7 are deemed as suitable, while the values in the individual response curve have a decreasing probability after 1.7. In the violin plot of *CV*, a significant difference between observation and background points is displayed, with a median of 0.7848 compared to 0.7867 and a mean of 0.86 compared to 1.01. This means that the individual preferred *CV* values of the woodpecker are a bit smaller than the study area. Overall, from these three plots can be derived that the most suitable preferred *CV* values have a broad range. Only values close to 0 are not preferred. Just like *VR_total*, this difference between total and individual response curves shows that *CV* does not have a very big importance compared to the other variables.

3.2.5 CCP

The fourth most important metric, *CCP*, does not show much differences in probability when values are changed. The probability slightly goes down but only near 100%, the probability makes a spike down. In its individual response curve, it can be seen that when *CCP* is independently used in the model, there is a more clear preference for higher values. In its violin plot, woodpecker observations were also found at higher canopy cover percentage values than the study area. Namely, the median (80.20 compared to 78.05%) is significantly higher. From all this can be inferred that all *CCP* values except for values close to 100% can be seen as suitable preferred values. Just like *VR_total* and *CV*, the difference between suitable preferred values and individual preferred values confirms the lower importance of *CCP*. Finally, besides that *p95* has the highest variable importance out of all metrics, *p95* is the only metric that does not display much difference in suitable and individual preferred values, indicating that this metric is the most important/dominant in the model.

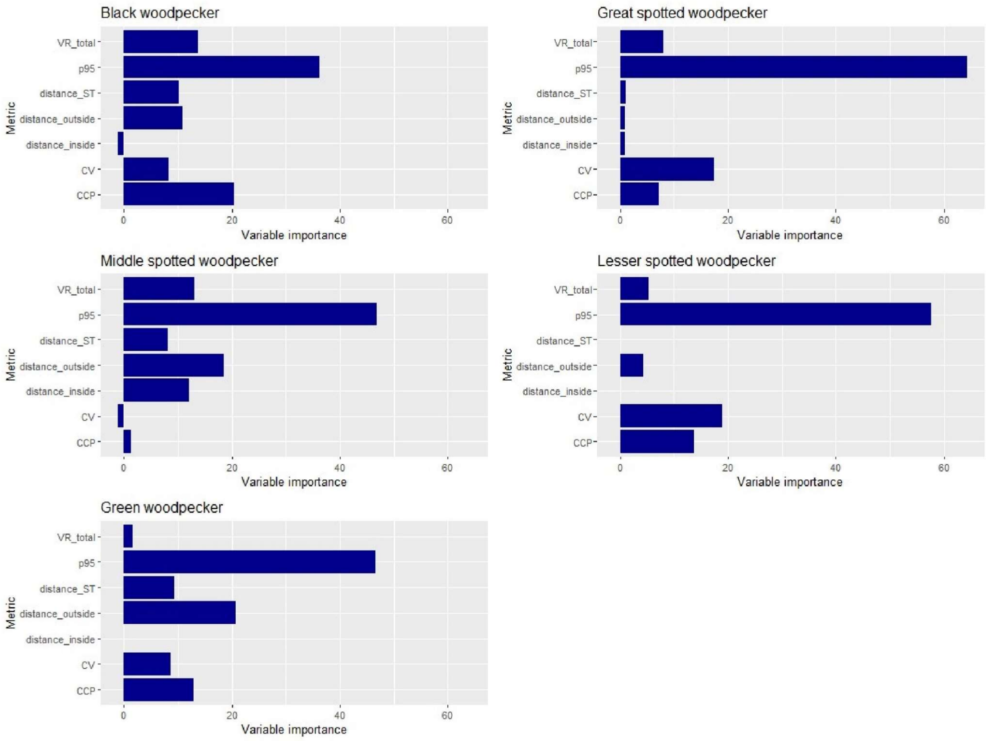
3.2.6 Remaining metrics

In figure 10 in appendix E, the total response curves of the remaining metrics are displayed. Additionally, in figure 14 in appendix G, the remaining violin plots can be found. *distance_ST* has suitable preferred values for all distances from standalone trees except for 0. Moreover, the total response curve of *distance_outside* displays all distances up to 260 meters from forest edges. Also for *distance_inside*, a broad range of distances from the forest edge (0-340 meters) are suitable. Nevertheless, these three metrics have a low variable importance and that is why broad ranges and not specific values are preferred in the total response curves. Furthermore, the violin plots of the remaining variables all show significant differences between observation points and background points. The values of *distance_outside* are higher and of *distance_inside* are lower than the background points. Finally, *distance_ST* values are higher compared to the background points.

3.3 RQ3: Model results of separate woodpecker species

Variable importance of separate woodpecker species

(a)



Total response curves of 4 most important metrics of separate woodpecker species

(b)

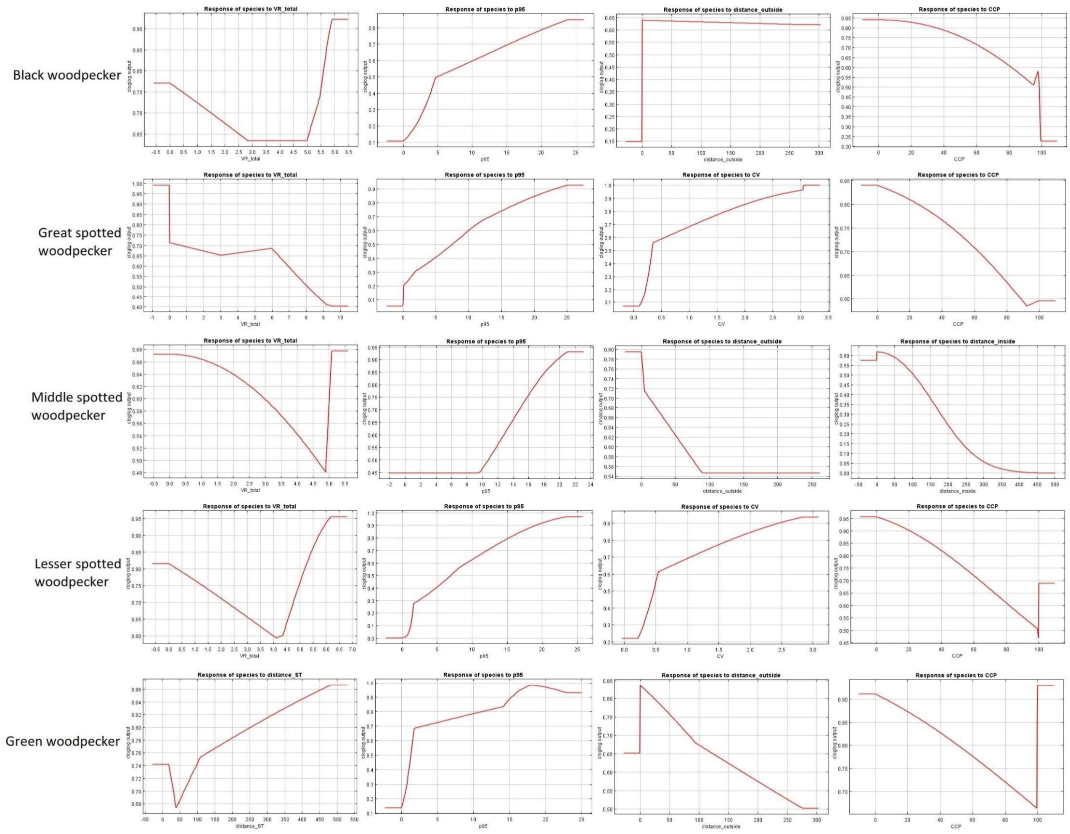


Figure 6: (a) Variable importance of the black woodpecker, great spotted woodpecker, middle spotted woodpecker, lesser spotted woodpecker and green woodpecker. If a variable has a value of -1 , it indicates that this variable was not used in the Maxent model due to high collinearity with other variables. (b) Total response curves of the four most important metrics for the black woodpecker, great spotted woodpecker, middle spotted woodpecker, lesser spotted woodpecker and green woodpecker.

In the results of RQ2, much information was given about the importance of certain metrics by displaying their variable importance values and describing the differences between total response curves and individual response curves/violin plots. In this part of the results (RQ3), the importance of the variables will be described less elaborately and there will be focused more on the differences in suitable preferred values between the different woodpecker species. This means that mostly the variable importance plots and the total response curves will be described and that the individual response curves and violin plots are only occasionally mentioned if needed.

3.3.1 Black woodpecker

The black woodpecker has the highest variable importance value for *p95* (Figure 6a). Its total response curve shows a peak at approximately 25 meters. Also the individual preferred values in its violin plot match these suitable preferred values, showing peak densities at higher *p95* values than the background points (Figure 13 in appendix G). *CCP* also has a high importance. Its total response curve shows a slightly decreasing probability for almost all values except for values around 100%. Moreover, *VR_total* , *distance_outside*, *distance_ST* and *CV* have similar importances that are all lower but still considerable. From the total response curve of *VR_total* can be inferred that a broad range of values are suitable from 0 to 6 meters. The total response curve of *distance_outside* shows that the black woodpecker has no preference for locations either closer or further away from the forest edge. Also the violin plot shows a distribution over all values, which is different than the other species, which have a more specific individual preference of distance values (Figure 14 in appendix G). Furthermore, no specific preference for distances from standalone trees can be observed in the total response curve of *distance_ST* (Figure 11 in appendix E). Additionally, most *CV* values (except for values close to 0) would be suitable for the black woodpecker, but higher *CV* values are the most suitable for the species.

3.3.2 Great spotted woodpecker

In the variable importance plot of the great spotted woodpecker can be seen that *p95* has a very dominant presence in the model. The total response curve shows that, similarly to the black woodpecker, values around 25 meters give the highest probability of presence. Also in its violin plot, it can be seen that most observations were found at higher elevation values compared to the background points, which matches the suitable preferred values (Figure 13 in appendix G). Moreover, *VR_total* and *CCP* have little importance, while *distance_outside* and *distance_inside* have almost redundant influence on the model. *CV* still has a considerable importance and has as similar total response curve for this metric as the other woodpecker species. Furthermore, the total response curves of *VR_total* and *CCP* show that both metrics have high probability values for a broad range of values. The most suitable preferred values of *VR_total* are just like the black woodpecker in the range of 0-6 meters. Its highest values can be found closer to 0 meter. Moreover, *CCP* values are suitable for the whole range of values but slightly decrease in probability for values close to 100%.

3.3.3 Middle spotted woodpecker

Just like the other woodpecker species, *p95* has the highest importance. This species has a similar total response curve with the highest probability values for higher *p95* values. Furthermore, *distance_outside* has the second highest importance. In its total response curve, the highest probability was found from 0 to 100 meters. Also *distance_inside* has a high importance. For this metric, the response curve shows that the highest probabilities are found at lower distances.

Moreover, similarly to the other species, a broad range of *VR_total* values (0-5 meters) are important. The total response curve shows and unexpected vertical peak near 5. This spike looks like a big change in probability. Nevertheless, the range of probability values on the y-axis is relatively small compared to other *VR_total* reponse curves. That is why it can be derived that the total response curve of this species shows relatively low probability values for approximately the whole range of *VR_total* values. Furthermore, *distance_ST* plays a role in the model and for this metric, higher distances from standalone trees are preferred (Figure 11 in appendix E). Besides these metrics, *CCP* has very little influence in this model.

3.3.4 Lesser spotted woodpecker

For the lesser spotted woodpecker, similar to the other species, *p95* has the highest importance, with preferences for high *p95* values. *CCP* and *CV* also have a considerable contribution while *distance_outside* and *VR_total* have small contributions. The total response curve of *CCP* shows that a broad range of values are suitable, but that lower values are the most suitable. Moreover, similar to the other species, higher *CV* values have a higher probability. Meanwhile, similar probability values can be seen for most *VR_total* values, as the change in probability values is not very big for the range of *VR_total* values. Just like the other species, the roughness values range from 0 to 6 meters. Additionally, the total response curve of *distance_outside* shows that probabilities are relatively low over a broad range, but that the highest probabilities can be found closer to 0 meter (Figure 11 in appendix E).

3.3.5 Green woodpecker

The most important metric for the green woodpecker is *p95*. Besides this metric, *distance_outside* also has a high influence on the model. *CCP*, *CV* and *distance_ST* have a similar moderate influence. Moreover, *VR_total* has a low influence. Compared to the other woodpecker species, the total response curve of *p95* shows slightly different results. Namely, in the response curve, high probability values can also be observed at lower vegetation height values, with a range of high probability values from 2 to 25 meters. Additionally, the violin plots of *p95* show that the green woodpecker observations have lower *p95* values compared to the other species, which means that the individual preferred values (which in this case match the suitable preferred values due to high variable importance) also show that green woodpeckers can be found at lower vegetation heights (Figure 13 in appendix G). Furthermore, in the total response curve of *distance_outside*, a clear high probability for low distance values can be observed. Additionally, in its violin plot, the biggest density is located at low distance values (Figure 14 in appendix G). It is also observed that, compared to the other woodpecker species, the green woodpecker has observations with the lowest distance values. Moreover, the total response curve of *distance_ST* displays that higher distance values have a higher probability. Furthermore, the total response curve of *CCP* shows that all *CCP* values are suitable. Just like for the *VR_total* metric of the middle spotted woodpecker, the range of the probability axis is small. That is why the big spike observed at 100% does not represent a big change in probability. From this can be derived that there seems to be a high probability for all *CCP* values. Finally, from the total response curve of *CV* can be inferred that the green woodpecker has similar probability values as the other species (Figure 11 in appendix E).

3.4 Accuracy

Table 3: Accuracy results of the Maxent models run for the black woodpecker, great spotted woodpecker, middle spotted woodpecker, lesser spotted woodpecker, green woodpecker and all woodpecker species combined. This table shows the AUC values, amount of observation points and amount of background points per model.

Maxent model	AUC	Observation points	Background points
Black woodpecker	0.718	44	88
Great spotted woodpecker	0.645	537	1074
Middle spotted woodpecker	0.759	20	40
Lesser spotted woodpecker	0.713	73	146
Green woodpecker	0.750	38	76
All woodpecker species	0.648	712	1424

The Maxent models gave as output AUC values, which are a measure for the accuracy of the model. Table 3 shows these results. The AUC values of all woodpecker species combined and the great spotted woodpecker are in the category of 'poor'. The other individual species are in the category of 'fair'. Moreover, it is observed that there seems to be a trend between the amount of observation points and the AUC values. Namely, models run with less observation points have the highest AUC values. In the model, the AUC values were derived by calculating the area under the ROC curve. The ROC curves of all six Maxent models can be found in appendix I.

4. Discussion

In this part, the results of RQ1, RQ2 and RQ3 will be discussed. At first, the creation and reliability of the LiDAR metrics will be discussed. Then, the results of the Maxent model of the woodpecker family will be explained. Thereafter, the differences in Maxent results of the separate woodpecker species will be discussed. Finally, the accuracy and limitations of the results will be analysed.

4.1 RQ1: LiDAR metrics

In the results, the values displayed in the layouts of the 14 LiDAR metrics were described and they were compared to vegetation structure aspects observed in a reference area. Based on these observations, in this part of the discussion, there will be discussed how to interpret the values of the 14 metrics in order to derive information about vegetation structure. Furthermore, the reliability of the metrics for usage in the Maxent models of RQ2 and RQ3 will be discussed and there will be explained how certain metrics could have been improved.

4.1.1 Vertical complexity metrics

The results have shown that the metric *p95* provides information about the height of vegetation in the study area. Higher values indicate a higher vegetation height. From this metric, areas with trees (high values) can be derived well. Also for the *mean*, different vegetation areas can be distinguished well. Furthermore, it was observed that *SD* values are higher in vegetation areas that are more open. For example, forest areas in which the tree canopies are not very closed displayed higher *SD* values. This could be explained by that most LiDAR points are located in the middle of the trees and less at the edges of the trees, resulting in less densely distributed points at the edges and higher standard deviations. Moreover, LiDAR pulses could more easily hit the ground in between trees than straight through the canopy, which could result in bigger height differences of the points located around the edges of the trees and thus bigger standard deviations. As *CV* represents the proportion of standard deviation compared to the mean, both *CV* and *SD* could be used as indicators for areas in which there are openings in between vegetation units, such as trees and shrubs. The calculations of *p95*, *mean*, *SD* and *CV* are not very complex and are very similar to calculations done in previous research. For example, in Kissling et al., 2023, the exact same calculations were done to create these metrics. This gives these metrics enough reliability to be used in the Maxent models of RQ2 and RQ3. The results of both *kurtosis* and *skewness* have shown that areas with higher values indicate that the vegetation is more open. Even though the two metrics could be reliable due to that their calculations in this thesis research match with calculations done in previous research, the two metrics were not deemed as reliable enough. Namely, vertical stripes with a width of approximately 300 to 400 meters with higher *kurtosis* and *skewness* values were observed in the whole study area. Between each stripe, there is approximately a distance of 850 meters. This occurrence could be explained by the way LiDAR data are created. The plane flies in straight lines (AHN, n.d.b). If the plane flies lower or higher relative to the surface of the study area, this has an influence on the width of the stripes and the amount of LiDAR points created. Due to overlap between flight routes, higher point densities can be found in these areas (see appendix K). *kurtosis* and *skewness* are dependent on the distribution of points in the vertical direction. For example, for *kurtosis*, the amount of outliers in the point distribution are the highest in areas with not much vegetation, as there is a high point density in a small range of values and divergent point more easily become outliers. In those areas with very low vegetation, other metrics such as *p95* give values close to 0 meter and are not affected by this higher point density. Thus, more outlier points are present in the flight overlap areas, resulting in higher *kurtosis* values. *CV* and *SD* are dependent on the values of the points and not specifically on the shape of the distribution (for example the normality or skewedness of a distribution) That is why these two metrics were kept. Another metric that was not included in Maxent models is *p25*. *p25* has similar height values in denser forests and more values of 0 in less dense vegetation areas compared to *p95*. While *p95* seems to give a reasonable representation of overstorey vegetation, *p25* values are either much higher or much lower than expected. There could be multiple explanations. It seems that during the creation of the AHN4 (LiDAR) data, the pulses did not penetrate all the way through tree crowns. This could have resulted in that most of the created points were located in the canopy of the trees and that 25th percentile points were still located almost as high as 95th percentile points. This could explain why the *p25* values in trees with denser crowns have similar elevation values as *p95*. Research of Martinuzzi et al. (2009) could support this statement about AHN4 data. Namely, it was mentioned that understorey vegetation calculated with LiDAR data is commonly less accurate than other methods, due to that the amount of points that reach lower vegetation layers decreases when they are located under thicker canopies. Other research also confirms this statement (Hill & Broughton, 2009). Besides the thickness of the canopy, the percentile threshold chosen to represent understorey vegetation also plays a role, although in this thesis, a higher or lower threshold would still give unreliable results. Namely, if a higher threshold is chosen, most likely higher values would appear at the locations where understorey vegetation was falsely indicated as 0, but then also even higher values would appear at the locations where understorey vegetation already has unrealistically high values. Moreover, the period in which planes have flown to create LiDAR point clouds can influence the ability to map understorey vegetation. Namely, in leaf-on season, canopies are thicker and less penetrable than in leaf-off season. The data used in this thesis were obtained in February of 2022 (ArcGIS, 2024). Even though it was leaf-off season, it was not managed to map a realistic representation of understorey vegetation. This could be explained by that De Veluwe contains much area with coniferous trees (see figure 2). Coniferous trees lose their needles over a longer period of time and still have a dense canopy in the winter, whereas deciduous trees have already lost their leaves in the winter. When observing the *p25* layout, areas with coniferous trees tend to have higher values than areas with deciduous trees, as the thicker canopies of the coniferous trees are less penetrable. This could explain why understorey vegetation still could not be correctly mapped in this area. Additionally, in appendix J, a table with a comparison between the density of first and all return points per tile is given. In all 3 tiles, a dominant proportion of all points consists of first return points. This shows that, even though all return points were used for the calculation of these metrics, most of the pulses did not go through canopies and only gave one return. As the 25th percentile height raster was supposed to be a representation of understorey vegetation and was used to calculate *UC* and *VR_low*, these three metrics were deemed as not very reliable. That is why the metrics were not included in the Maxent models of RQ2 and RQ3.

4.1.2 Horizontal heterogeneity metrics

From the results can be derived that higher *VR_total* values indicate forest areas with openings in its canopy. Namely, the biggest height differences are found between the top of trees and the ground of openings in the forest. When a canopy is more closed, less height differences are present and *VR_total* values are lower. This metric seems suitable for further use in Maxent models, as from its values information is provided about which areas are more or less heterogeneous. The calculation of this metric was based on De Vries et al. (2021). An alternative method for the calculation of this metric could be that the standard deviation of the 95th percentile of vegetation height is calculated per cell for a certain window size (Koma et al., 2022). This is also a focal operation, but instead of looking at the biggest height differences between a cell and its neighbouring cells, the standard deviation of the elevation values of the middle cell and all neighbouring cells together is calculated. Subsequently, this value is assigned to the middle cell. This method might give a similar view of heterogeneity of vegetation, except for that the range of output values is different.

4.1.3 Vector-based metrics

In order to create tree crown polygons, tree tops were detected by using a commonly used window size of 5 meters in the 'locate_trees' function. Nevertheless, the detection of tree tops could have been more accurate if a variable window size was used (Roussel et al., 2023). Namely, it is more suitable to detect taller trees with a bigger window size and to detect smaller trees with a smaller window size. A function could be developed in which the window size is dependent on the elevation of raster cells. In this way, variable window sizes could have been used to detect trees. Nevertheless, the creation of such as function is a topic for further research. Moreover, the layout of the metric *distance_ST* shows clearly the distances to trees that are located minimally 10 meters from other trees. This metric was initially based on the idea that black woodpeckers prefer more isolated trees. Nevertheless, the literature should have been interpreted a bit differently. In the paper of Puverel et al. (2019), isolated trees were defined differently than implemented in this thesis. The distance to the nearest neighbouring trees of each cavity tree was measured from trunk to trunk and not from tree crown edge to tree crown edge. Moreover, much smaller distances between trunks were already used as indicators of isolated trees. Namely, the distance between the cavity tree trunk that was labelled as an isolated tree and the nearest tree was on average 5 meters. This means that the distance of 10 meters between tree crowns chosen in this thesis might have been too high in order to find the isolated trees that black woodpeckers would prefer. In this thesis, standalone trees were defined as trees that were located outside big areas of forest. After these new revelations, it can be said that standalone trees should have been defined as trees inside forest areas that do not have closely neighbouring trees. In further research, distances between canopies in the range of 1 to 3 meters seem more suitable. Moreover, *distance_outside* displays the distances to open areas. Raster cells with higher values are located deeper inside the forest. *distance_inside* displays the opposite, which is distances to forest areas. For this metric, raster cells with higher values are located deeper inside open areas and further away from forest areas. The values of both metric are quite reliable, as they simply show the proximity to certain raster cells. Nevertheless, the detection of open areas and forest areas could be improved. In this thesis, the study area was simplified into bigger areas with forest and bigger open areas. Open areas with a surface area smaller than 3500 m^2 were filled up and became part of the bigger forest areas. Moreover, forest areas smaller than 3500 m^2 were counted as part of bigger open areas. These thresholds were chosen based on trial and error. It was tested for which threshold forest areas were entirely filled up. Additionally, the same was tested for when open areas were entirely open. Nevertheless, these tests were done for a small part of the forest, resulting in that the threshold was based on a specific part of the study area. For other parts, this threshold might have been too small or too large. That is why in future research, more time could be spent on defining a suitable threshold. Furthermore, the metric *CCP* showed the differences in canopy cover in the study area. Canopy cover percentage was calculated using a focal operation. For a moving window of 41 by 41 meters, the percentage of cells containing canopy cover was calculated. Nevertheless, in previous research also other methods were used to calculate canopy cover metrics. In Kissling et al. (2023), a local operation was done instead of a focal operation. Per raster cell, the percentage of points above the mean height was calculated in order to show the density of the upper vegetation layer. Higher percentages above the mean height would indicate that the upper layer is more dense and that there is more canopy cover. Nevertheless, this metric also takes other vegetation than trees into account, such as shrubs. However, the method for *CCP* used in this thesis only takes detected trees into account. Thus, the canopy cover metric of this thesis gives a different type of information than the metric described in Kissling et al. (2023). The window size used for the focal operation in this thesis, could be adjusted more. Namely, now a window size was chosen that was not too big but also not too small, in order to take the closer environment into account but also not loose detail. Even though the window size could be adapted more, the metric still shows a representation of canopy cover. Even if another window size was used and a different range of values was displayed for this metric, the areas indicated as having more or less canopy cover would still be comparable and informative.

4.2 RQ2: Habitat preferences of woodpecker family

In the results section of RQ2, the total and individual response curves of the Maxent model of the collection of all woodpecker species were analysed, as well as violin plots of the different metrics. It was derived which metrics were the most important and which metrics were less important. Moreover, information was acquired about the suitable preferred values of each metric. In this part, it will be discussed why the most important metrics of the Maxent model could be important for the habitat selection of the woodpecker family. Moreover, it will be explained more concretely what information the suitable preferred metric values give about woodpeckers' habitat preferences (certain versions (values) of aspects (metrics) of vegetation structure). Additionally, ecological explanations will be given for why woodpeckers could have these specific habitat preferences.

4.2.1 p95

The 95th percentile of vegetation height, *p95*, has a very dominant variable importance (59.43%), which is more than three times as high as the second most important variable, *CV*. The total response curve showed that woodpeckers have higher probability values to be present at a higher vegetation height, with a peak around 25 meters. Additionally, as the values in its individual response curve and violin plot match with the values in its total response curve, the violin plot of *p95* can be used to infer information about the preferences of the woodpecker family compared to the general study area. In this case, the differences in violin plot densities and in medians between observation and background points indicate that the vegetation heights found at the observation points were most commonly in the range of tree heights and that woodpeckers prefer taller trees than the trees at the locations of the background points. Moreover, less points were found in the range of shrubs/bushes (closer to 0). Thus, these results indicate that the woodpecker family prefers areas containing taller trees (around 25 meters). Davison et al. (2023) did research into the importance of vegetations structure on bird species in Denmark. The results of this research showed that canopy height is the most important vegetation structure metric and that a positive correlation was found between canopy height and forest bird species richness. This preference for higher trees could be explained by that taller trees provide a bigger habitat than smaller trees (Davison et al., 2023). Furthermore, woodpeckers often select trees with a trunk that has a higher diameter than other trees in the area (Basile et al., 2020). The woodpeckers in this research, located in the Black Forest of Germany, preferred a diameter at breast height (DBH) of approximately 15-20 cm bigger than other trees. They also preferred old trees, as the wood of old trees has decayed more and therefore is softer and more suitable for cavity excavation. The fact that the woodpeckers can have a preference for trees with a big DBH and trees that are older could indicate that they could have a preference for taller trees, as taller trees often have a trunk with a bigger diameter and often are older. Furthermore, Menon et al. (2021) did research on woodpeckers in the Himalaya and the results showed that most woodpecker species preferred trees that have a large DBH, as well as that most woodpecker species preferred trees that are taller than the surrounding area. Although the study area of this research is very different than De Veluwe, it does still give an indication about woodpeckers and cavity excavating birds in general. Thus, research supports that the height of vegetation and in this case *p95* has a high importance in the habitat selection of woodpeckers.

4.2.2 CV, VR_total and CCP

Compared to *p95*, *CV*, *VR_total* and *CCP* have a much lower variable importance in the Maxent model. This means that the preferences derived from these metrics that will be described in this section are less important than the preferences derived from *p95*. It is important to take into consideration that *p95* is much more important for the habitat selection of woodpeckers. The second most important metric is *CV*. As earlier explained, *CV* is a metric representing the vertical complexity of vegetation. Literature indicates that vertical complexity is an important factor for the distribution of birds. For example, for certain bird species, the mean of vegetation height in combination with the vertical variability of vegetation are important factors for their spatial distribution (Moudrý et al., 2023). A commonly used metric for vertical variability is the standard deviation. This would mean that the combination of the mean and standard deviation of vegetation height could be important in researching bird species distribution. The metric *CV* was calculated by dividing the standard deviation by the mean of vegetation height. This could explain why *CV* has a some importance in the model. The results of *CV* showed that values between 0.5 and 3.5 are preferred. When *CV* values are higher than 1, the standard deviation is higher than the mean, meaning that there is quite some variability in vegetation height at locations with these values. From this can be inferred that woodpeckers prefer areas that are vertically complex. As explained earlier in the discussion, higher *CV* values and thus more vertically complex vegetation can be found in the edges of vegetation units with open areas. For example, in a forest with small openings in its canopy, the highest *CV* values are found in the edges of these canopies. Thus, the results of *CV* indicate that the woodpecker prefers areas with vegetation, such as trees or bushes, that is not too densely located to each other and has some openings.

Furthermore, as earlier explained, *VR_total* shows the maximum differences in 95th percentile vegetation height between neighbouring raster cells. It is a measure of horizontal heterogeneity. As earlier discussed, in previous research, horizontal heterogeneity was calculated differently but provided similar information. In Koma et al. (2022), the metric representing total vegetation roughness had a small but still considerable importance for the Savi's warbler. This bird preferred homogeneous reedbeds and thus a low horizontal heterogeneity. Even though woodpeckers breed in trees and not in reedbeds, this example shows that horizontal heterogeneity metrics can be important for birds. The results of *VR_total* in this thesis showed that roughness values up to 6 meters are preferred by the woodpecker. Even though smaller roughness values are preferred the most, this broad range of preferred values showed that woodpeckers prefer habitats with many height differences between trees and the ground or shrubs, indicating that these areas are not densely populated with trees and have some openings in between tree canopies. Thus, in contrast to the Savi's warbler, woodpeckers prefer areas that are more horizontally heterogeneous.

Moreover, *CCP* displays the openness of the canopy of a forest. There are several ecological explanations for the importance of this metric. At first, openings in canopies provide sunlight to areas under the forest canopy (Moudrý et al., 2023). This results in the presence of certain vegetation, such as grass, that attracts insects (Kosinski et al, 2004). These insects could be prey for woodpeckers that forage on the ground. Moreover, a more open canopy gives birds more space to fly from and to their cavity, making the tree more accessible (Kosinski & Winiecki, 2004; Puverel et al., 2019). The results of the *CCP* metric showed that generally all canopy cover percentages except values very close to 100% are preferred by the woodpecker. From this can be inferred that woodpeckers prefer areas with canopy cover and thus areas with trees, but the canopy cover should not be too closed off.

Overall, from *CV*, *VR_total* and *CCP* can be inferred that the woodpecker family prefers forest areas with a canopy that is not too closed. The canopy of these forest areas should still have some openings. Besides the earlier mentioned explanations for this preference, the fact that woodpeckers generally prefer breeding in old or dead trees could also be an explanation (Basile et al., 2020; Kosiński, 2006; Nijssen et al., 2020; Olsson et al., 1992). The wood of dead trees is softer and therefore easier to excavate (Smith & Charman, 2012). Dead trees have a less dense canopy than younger trees. That is because some dying branches and leaves have fallen off. The preference of woodpeckers for dead trees could have contributed to the found preferences of woodpeckers in this thesis. Namely, dead trees results in a more open forest, with more space between tree canopies. Finally, compared to forest areas in general in the study area, it does not seem to be the case that woodpeckers prefer a forest that is more open than the general forest areas in the study area. Namely, normal forest most of the time do not have an entirely closed canopy. There are always some openings. With this taken into consideration, it can be said that the woodpecker family prefers forest areas with a quite normal canopy cover. The only big difference from normal forest areas is that taller trees are preferred compared to the average trees in the study area.

4.2.3 Remaining metrics

Metrics that had a low influence on the model were *distance_outside*, *distance_inside* and *distance_ST*. The metric *distance_outside* showed that in general the woodpecker family does not have a specific preference for habitat locations deeper or less deep inside the forest. Moreover, from *distance_inside* can be derived that the most preferred habitat locations can be found closer to the forest edge if woodpeckers would be located outside forest areas. From *distance_ST* can be derived that woodpeckers prefer forest areas over standalone trees. The total response curve of *distance_ST* showed that woodpeckers could find their most suitable habitat for a big range of distances, as long as the habitat is further away from standalone trees. Thus, these tree low importance variable show that woodpeckers in general prefer breeding in forest areas and not in trees in open fields.

4.3 RQ3: Habitat preferences of separate woodpecker species

Now that the importance of the different aspects of vegetation structure for the woodpecker family are discussed, as well as the habitat preferences of the woodpecker family, there will be focused on the differences in habitat preferences of the five woodpecker species. Compared to RQ2, there was focused more in the results of RQ2 on the differences in suitable preferred metric values between the five species. Moreover, there was explained less elaborately why certain metrics were more important than others. In the discussion of RQ3, for each species, possible explanations will be given for the found suitable metric values. Namely, literature describing the ecology of certain woodpecker species can give explanations for these species' preference for certain versions of aspects of vegetation structure. For example, factors such as the preferred foraging locations of woodpeckers could influence their habitat selection. One aspect that all separate woodpecker species have in common is that *p95* has the highest variable importance. The total response curves show for all species that higher vegetation heights/taller trees are the most suitable. This means that just like the collection of all species, all separate species prefer locations with taller trees. In a similar research about the black woodpecker in Poland, the species created cavities in taller trees, around 28 meters tall (Zawadzki & Sławski, 2023). These results are similar to the results of the five woodpecker species, which most suitable heights range from 22 to 25 meters. So, the results of this paper confirm the trends observed in this thesis research. Moreover, the explanations given in the discussion of *p95* in RQ2 can also be applied on the separate woodpecker species' preference for taller trees.

4.3.1 Black woodpecker

The earlier described results of the black woodpecker are similar to the results of the collection of all woodpecker species. Just like for all species, a broad range of roughness values (*VR_total*) are preferred, especially higher values. Furthermore, the total response curve of *CV* indicates that there is a preference for vertically complex vegetation. Moreover, from the *CCP* results can be inferred that forests with a too high canopy cover percentage are not preferred. From all this information can be derived that more open forests are preferred by the black woodpecker. Results that stood out were that *distance_outside* and *distance_ST* have quite high importance values. These two metrics indicate that black woodpeckers are found inside forest areas and that inside forests, they have no specific preference for locations closer or further away from the forest edge. Research was done in a study area in Poland on the habitat preferences of black woodpeckers regarding vegetation structure (Zawadzki & Sławski, 2023). The results indicated that the black woodpecker has a preference for more isolated trees, whereas in this thesis research, no indication of this was found. Nevertheless, as earlier mentioned, the definition of standalone trees used in this research does not match the definition of standalone trees that are suitable for black woodpeckers. Moreover, the fact that *VR_total*, *CCP* and *CV* indicate that the black woodpecker likes forests with more openings and thus more space in between vegetation, could support the preference of black woodpeckers for more isolated trees. There are several explanations for this preference. At first, research of Nijssen et al. (2020) indicates that black woodpeckers prefer a more open vegetation structure. The given explanation was that woodpeckers prefer to forage in dead wood that is either still standing or laying on the ground. Dead trees usually have a less dense crown, resulting in more open forests and more isolated trees. Furthermore, the species prefers to make cavities in trees that are less connected with the branches of other trees in order to be able to reach its cavity faster and to reduce predation risk (Puverel et al., 2019). Additionally, also the results of Zawadzki and Sławski (2023) showed that black woodpeckers prefer cavity trees with a lower crown density surrounding these trees, because they want to have space in between trees to fly around.

4.3.2 Great spotted woodpecker

The results of the great spotted woodpecker showed that *p95* is very important for its habitat selection and that compared to the other woodpecker species, *p95* is the most important for the great spotted woodpecker. Research in Poland showed that great spotted woodpeckers prefer large trees to nest and forage in (Piacentini & Chiatante, 2022). This matches with the *p95* results. The other metrics have much lower importance values. When looked at the total response curve of *VR_total*, even though it has a different shape, the probability values are similar to that of the black woodpecker. Also the total response curve of *CV*, as well as *p95*, are very similar to the black woodpecker. Only for *CCP*, also values of 100% are suitable. Moreover, *VR_total* does not have a spike in probability values at the same higher *VR_total* values. All these results indicate that the great spotted woodpecker prefers a little bit denser (less open) forest than the black woodpecker. Moreover, this species does not seem to have a preference for standalone trees or forest edges, as the total response curves have their highest probable values at distances further away from standalone trees and spread out all over the forest compared to the edge. These findings can be explained by that the species prefers mature trees (Woodland Trust, 2021a). They forage by excavating tree bark, for example they eat beetle larvae, caterpillars and spiders. As this species breeds and forages in trees and does not make much use of the ground, it could be possible that the species is most commonly found deeper inside the forest compared to other species. Compared to the black woodpecker, which species occasionally forages in wood on the ground besides in trees, the great spotted woodpecker has to fly around less commonly from the ground to its cavity and thus needs a less open forest in order to have space to fly around. This could explain why the black woodpecker prefers a more open forest with less vegetation that blocks its flight path.

4.3.3 Middle spotted woodpecker

Besides *p95*, the metrics *distance_outside*, *distance_inside* and *distance_ST* have a high importance. The species prefers habitats close to the forest edge when the species is inside the forest and also when the species is outside the forest. Moreover, the species prefers locations far away from standalone trees. Furthermore, the most preferred habitats for middle spotted woodpeckers are found at lower *VR_total* values compared to the other metrics, meaning that these habitats might be less horizontally heterogeneous than the habitats of the other species. These lower values could indicate that they prefer a more closed canopy compared to the other woodpecker species. However, it can also mean that the suitable areas contain areas with no vegetation or lower vegetation. This is more likely due to that the species can be found close to forest edges (low *distance_outside* values) according to the results and due to that the species prefers less closed off vegetation areas according to literature. Research of Kosinski and Winiecki (2004) gives information about the different habitat preferences of middle spotted woodpeckers and great spotted woodpeckers. Middle spotted woodpeckers select their habitat based on if they are close to areas where they can forage, whereas great spotted woodpeckers give more priority in their selection to the availability of suitable nest sites. The research shows that middle spotted woodpecker have a preference to nest at forest edges, as they like to forage in more open vegetation and more open vegetation can be found more at forest edges. Also in the Netherlands, middle spotted woodpeckers prefer open areas in forests (Vogelbescherming Nederland, n.d.c). They search for food in horizontal branches of trees. A less dense forest gives the branches of trees more exposure to sunlight and this can increase the amount of arthropods found in the branches (Kosinski & Winiecki, 2004). Namely, arthropods are important food for middle spotted woodpeckers. Nevertheless, the *CCP* metric does not convincingly explain the species' preference for areas with lower canopy cover percentage, as the metric has a low importance and the total response curve showed the same relatively low probability values for the whole range of percentages, with the lowest values for percentages close to 0. Maybe the fact that no specific suitability can be observed for higher *CCP* values can indicate that the suitable area does not specifically contain dense forest but can contain a variety of more open and denser forest. Moreover, besides this species, the great spotted woodpecker can find its food at a higher diversity of places in the forest and is therefore less dependent on the forest edge (Kosinski & Winiecki, 2004). Which could explain the earlier discussed preference for distances further away from the forest edge.

4.3.4 Lesser spotted woodpecker

The lesser spotted woodpecker has higher probability values for lower *CCP* values. This indicates that this species prefers a forest with a more open canopy cover. Furthermore, just like for the other species, higher *CV* values and a broad range of *VR_total* values are preferred. This all indicates that forests with a not too high density of vegetation units are preferred. Just like for the other species, the canopy of the forest should not be entirely closed. Furthermore, although having a low importance, the most suitable habitat is found to be at a broad range of distances, but the most suitable locations are closer to the outside of the forest edge when the species is inside the forest (lower *distance_outside*). This could be explained by that the lesser spotted woodpecker prefers similarly to the middle spotted woodpecker open forest areas and forest edges (BirdLife International, 2024b). Lesser spotted woodpeckers forage in tree trunks and branches, looking for invertebrates and insects (Vogelbescherming Nederland, n.d.b; Woodland Trust, 2021c). It is uncertain why the results showed that they have a preference for locations closer to the edge. A reason could be that, similarly to the middle spotted woodpecker, more open forests let in more sunlight and results in a higher abundance of invertebrates and insects in the branches of trees (Kosinski & Winiecki, 2004). Especially because the preferred *CCP* values of this species are lower compared to the other species, this could be an explanation for their location at the forest edge.

4.3.5 Green woodpecker

Similar to the middle spotted woodpecker, *distance_outside* and *distance_ST* are important for the green woodpecker. *distance_outside* has the second highest importance, besides *p95*. The total and individual response curves of *distance_outside*, as well as its violin plot, showed that green woodpeckers have a preference for habitats close to the forest edge. Moreover, the results showed that the species prefers locations further away from standalone trees. Additionally, *CCP* and *CV* results indicate that the green woodpecker prefers similarly to the other species a less dense/more open forest. The violin plot of *CCP* displayed that observations were found at lower *CCP* values compared to the other species. This makes sense, as there is less canopy cover at the edge of the forest. Moreover, higher *CV* values are preferred. This can be explained by the fact that the edge of the forest is the transition zone between trees/shrubs and open areas. As mentioned earlier, the edges of forests and the edges of vegetation units such as trees have higher *CV* values due to that there is a lower points density at these locations. Thus, the green woodpecker prefers a forest that is not too densely populated with vegetation, with its habitat close to the edge of the forest. These results can be explained by that literature suggests that the green woodpecker prefers semi-open habitats (BirdLife International, 2024a). Examples of these habitats are parks, orchards but also forest edges. In large forest areas, green woodpeckers often create nests near the edge of the forest (Vogelbescherming Nederland, n.d.a). Namely, the primary food of green woodpeckers is ants that they find on the ground (Woodland Trust, 2021c). They prefer to breed close to the location where they can find food resources, which is in this case open ground areas near forests. Especially the amount of ants in an area is an important factor for the habitat selection of green woodpeckers (BTO, 2020). Many ants can be found in grasslands. Out of all five species, the green woodpecker is the only species that forages in open fields. This could explain why the results showed that also lower *p95* values are preferred by the green woodpecker, whereas for the other species a more clear preference for higher *p95* values was observed.

4.4 Reliability of results

4.4.1 Accuracy

Table 3 showed the AUC values of the six different Maxent models. The AUC values, ranging from 0.645 to 0.759, are not very high for all models but are still sufficient enough, as they all have AUC values of at least 0.645, while an AUC of 0.5 is similar to random sampling. As mentioned in the methods, when background points are used instead of absence points, AUC values tend to be lower than actually should be the case due to that background points can be wrongly classified as 'falsely classified presence points'. If this problem would be taken into account in the accuracy assessment of the Maxent models, the AUC values would be a bit higher. Therefore, even though the AUC values of the six models are not very high, they are still deemed accurate enough. Besides that the use of background points instead of absence points has likely influenced the AUC values, there are also other factors that could have had an influence on the accuracy assessment of the model. Namely, it was observed that models that have the biggest amount of observation points and background points as input, gave the lowest AUC values as output. This could be explained by that for the species with higher sample sizes a higher number of background points were created. If there are more observation/presence points and more background points, there is a higher chance that presence and background points overlap, resulting in that more background points are wrongly labelled as 'falsely classified presence points'. For species with a small sample size, there is more space for the presence and background points in the study area, resulting in less overlap. Additionally, species with a narrower range in metric values, which is more likely to be found at lower sample sizes, can have higher AUC values than species with a bigger range (Phillips, 2021; Tesfamariam et al., 2022). This all could explain why the woodpecker species with lower sample sizes, such as the medium spotted woodpecker (n = 20), have the highest AUC values and why the species with the most observations, such as the great spotted woodpecker (n = 712), have the lowest AUC values. Moreover, even though Maxent models can handle small sample sizes, the differences in sample sizes between the species are high. It can be assumed that the results of the species with the highest sample sizes are the most reliable due to that more observation points result in more information about the suitable preferred metric values of the species.

4.4.2 Great spotted woodpecker

Out of the five woodpecker species, the results of the great spotted woodpecker are the most similar to the results of the woodpecker family in general. This is because this species has by far the biggest sample size (537). This means that out of all 712 observations of the woodpecker family, 537 observations belonged to the great spotted woodpecker. This most likely has led to that the results of the collection of woodpecker species are influenced much by the great spotted woodpecker. If each of the five woodpecker species had the same sample size, the influence of each species would have been represented more in the Maxent model and a better view would have been created of the habitat preferences of the woodpecker family in the Netherlands. Nevertheless, the fact that the great spotted woodpecker is observed so much more than the other species also means that most woodpeckers in this study area most likely behave similar to the great spotted woodpecker. Thus, even though not all species are evenly represented, the overall results of this thesis research do represent the average woodpecker in this study area.

4.4.3 Response curves

The total response curves of the woodpeckers with the lowest sample sizes, which are the middle spotted woodpecker and green woodpecker, displayed some high spikes. For both species, these spikes could probably be observed due to that the specific metrics had some small changes in probability values over the range of the metric values. For example, for the green woodpecker, it seems like there is a high increase in *CCP* probability very close to 100%. This seems like a big change, but when looked at the y-axis, a relatively small change in probability compared to the other species is observed. The fact that these curves are not very smooth might be because these woodpecker species have quite small sample sizes. An increase in sample size would make the curves smoother.

4.4.4 Extraction buffer

An extraction buffer with a 50 meter diameter was used to extract metric values for the observation and background points. This size was chosen in order to take the environment of each observation into account while also making sure that not too much detail is lost. Namely, some metrics, such as vegetation roughness, are more detailed than other metrics, such as distance to forest edge. Nevertheless, it is uncertain which diameter would be the most suitable for extracting metric values. This will be discussed more in the recommendations.

5. Conclusion

5.1 RQ1

Which metrics that provide information about vegetation structure and are potentially relevant for woodpeckers can be derived from LiDAR data?

The first aim of this research was to create metrics in order to retrieve information about aspects of vegetation structure. While selecting and creating the metrics, it was kept in mind that the metrics should be potentially useful for deriving information about woodpeckers in RQ2 and RQ3. Eventually, 14 different metrics were calculated with the use of LiDAR data. These metrics were placed in the following categories: vertical complexity metrics, horizontal heterogeneity metrics and vector-based metrics. The results showed that there are several metrics, such as *p95* and *CCP*, that are a reliable representation of vegetation structure. Nevertheless, *p25*, *UC*, *VR_low*, *kurtosis* and *skewness* were deemed unreliable to be used in the next steps of this research. Namely, thicker canopies prevented LiDAR beams from penetrating deeper through the vegetation, influencing the values shown in *p25*, *UC* and *VR_low*. Moreover, due to overlap in flight routes during the creation of LiDAR points, the metrics *kurtosis* and *skewness* displayed higher values at these overlap areas compared to other areas. Furthermore, other metrics that are more reliable could still be improved. *distance_ST* did not entirely serve the purpose for which it was created and a too big distance threshold of 10 meters was used to detect standalone trees. Thus, the metric still gives reliable information about distances from certain trees, but not the trees that could be most useful for the woodpecker. In the future, a smaller range of distances, for example 1-3 meters, would be more suitable. Additionally, for the metrics *distance_inside* and *distance_outside*, the threshold for forest areas and open areas was based on a smaller part of the study area. In further research, the threshold could be tested in multiple parts of the study area. All in all, even though some metrics are less reliable for providing information about vegetation structure, they still give an insight into what aspects of vegetation structure potentially could be calculated with LiDAR data if improvements were made. Besides the specific metrics, the methodology of upscaling the LiDAR metrics that was showed in this thesis could serve as an example for similar research in which LiDAR metrics are calculated.

5.2 RQ2

Which aspects of vegetation structure represented by LiDAR metrics have the biggest influence on the habitat selection of the woodpecker family?

The goal of the second sub-question was to derive information about the importance of the different aspects of vegetation structure for the habitat selection of the woodpecker family. The metrics that were deemed as reliable in RQ1 for representing vegetation structure and being potentially informative about the habitat preferences of woodpeckers were tested on multicollinearity. This resulted in that metrics that correlated too much with other metrics were excluded. Then, the extracted metric values at the locations of the observation and background points were used in a Maxent model. The model had the following outputs for each metric: a variable importance value, a total response curve and an individual response curve. The total response curves showed the suitable preferred metric values, representing the habitat preferences of woodpeckers, whereas individual response curves, as well as external violin plots, showed the individual preferred metric values. The results showed that higher variable importance values and much similarity between on the one hand total response curves and on the other hand individual response curves and violin plots indicate that the concerning metric has a high importance. Out of all metrics, *p95* had the highest importance in the model. It was derived that woodpeckers prefer trees that are taller than the average tree in De Veluwe. Moreover, metrics that have a less but still considerable importance in the model and thus in the habitat selection of the woodpecker family are *CV*, *VR_total* and *CCP*. *CV* results showed that woodpeckers prefer areas with some vertical complexity. Areas with vegetation units, such as trees and shrubs, which canopy is not too closed off are preferred. *VR_total* showed that horizontally heterogeneous areas are preferred. This metric indicated, just like *CV*, that the canopy of forests should not be too closed off. Finally, *CCP* values showed that woodpeckers prefer areas with canopy cover and thus areas with trees, as long as the forest canopy is not too open or too closed. It should be taken into consideration that the variable importance values of *CV*, *VR_total* and *CCP* are much lower than the values of *p95*, meaning that the 95th percentile of vegetation height is by far the most dominant aspect of vegetation structure in the habitat selection of the woodpecker family in De Veluwe. From all this information was inferred that woodpeckers most likely prefer areas containing vegetation, preferably taller trees, which canopy is not too closed off. Overall, this description of the most preferred habitat of the woodpecker family is quite similar to normal forest areas in De Veluwe. The only big difference is that woodpeckers have a preference for taller trees.

5.3 RQ3

How do the habitat preferences of the five separate woodpecker species differ?

In RQ3, the aim was to acquire information about the habitat preferences of the five separate woodpecker species in De Veluwe. Unlike in RQ2, there was focused less on the importance of the different metrics and more on the specific habitat preferences of the woodpeckers. This means that the variable importance was explained less elaborately and more research was done into explaining the suitable preferred values of each metric. The results showed that, just like the woodpecker family, all species have in common that *p95* has a big importance in their habitat selection. All species preferred areas containing vegetation, preferably taller trees. Nevertheless, the species prefer different locations regarding the edge of the forest. The green woodpecker prefers to breed near forest edges due to the ability to forage in the open area close to its nesting tree. Moreover, the lesser spotted woodpecker and middle spotted woodpecker also prefer forest edges. This is possibly due to that the canopy is more open at these edges, which creates more suitable conditions for prey to live, such as insects and invertebrates. Furthermore, the results of the black woodpecker and great spotted woodpecker showed that they can be found at most locations in the forest. No specific preferences for locations deeper inside the forest or closer to the edge were found. The great spotted woodpecker might only differ from the black woodpecker due to its preference for slightly denser forest areas. Namely, due to that the black woodpecker also occasionally forages on the ground besides in trees, this species might need more space to fly from the ground to their cavity than the great spotted woodpecker.

5.4 Main question

How can the use of airborne LiDAR data improve the understanding of the habitat preferences of five woodpecker species in De Hoge Veluwe, the Netherlands?

Now that the three sub-questions have been discussed, the main research question can be answered. The goal of this research was to acquire knowledge about the additional value of LiDAR data in ecological research into woodpeckers. This research has shown that LiDAR data indeed have an additional value in woodpecker research. Namely, the calculated LiDAR derived metrics gave information about vegetation structure, which eventually led to the acquirement of information about the habitat preferences of woodpeckers. Especially the use of tree segmentation in order to derive metrics from vectors is a method that has not been used much yet in ecological research. Namely, besides that more general metrics such as *p95* have given information about the habitat preferences of woodpeckers, also less common metrics derived from tree polygons, such as *distance_outside* have given valuable insights. Even though improvements could be made in the calculation of certain metrics in order that more reliable inputs were used in the Maxent models, this research has shown what can be done with LiDAR data in research on bird species distributions and can be used as inspiration for further research. Moreover, although the accuracy values of the Maxent models were not very high, the results are still accurate enough in order to derive information about the habitat preferences of woodpeckers. Overall, the methodology of this research, which includes the calculation and upscaling of the LiDAR metrics, as well as the use of the Maxent models, could be used for research in other study areas. This thesis has shown that even though several aspects of the overall workflow could be improved, much information about the habitat preferences of woodpeckers and thus also of birds in general can be derived from LiDAR data.

6. Recommendations

As not all calculated LiDAR metrics were reliable and the modelling of the relationship of these metrics with the presence of woodpecker species could be improved, several recommendations can be done.

6.1 Vegetation density

Four metrics representing vegetation density were first created and eventually not used in this research. Each raster layer would represent the vegetation density in a different height interval. In examples in literature, the different height intervals were based on values in meters. For example, research of Koma et al. (2022) used height intervals of 0-1, 1-2 and 2-3 meters to calculate vegetation density for a reed landscape. In this research, it was estimated how high the reed vegetation would be before defining these intervals. Eventually, the vegetation density metrics would give information about the vertical variability of the vegetation. For this thesis research in De Veluwe, it was unclear what the absolute intervals should be, as it beforehand was unclear what the maximum vegetation height values of the vegetation were. This resulted in that a division into percentiles was chosen to represent different layers. The calculations were as follows. All the points in a certain percentile interval were selected and divided by the total amount of points. Later in the thesis process, it was discovered that the calculation of these four metrics would not give the information that was desired beforehand. Namely, an error was made by using percentiles for the layers instead of values in meters. Namely, if intervals of 25% are used, 25% of the points are divided by the total amount of points and this results in density values around 0.25. This error has led to that these metrics were not suitable to be used for inferring information about the vegetation density in different layers. The calculations of these metrics can be found in appendix Aa01. In future research, it is still very relevant to calculate vegetation density metrics, as these metrics are often used to give information about the vertical complexity of vegetation (De Vries et al., 2022; Koma et al., 2022). Instead of using percentile intervals, absolute intervals should be decided in order to get reliable values. Nevertheless, as earlier discussed, less points can be found closer to the ground under canopies of trees. This means that intervals representing understorey vegetation could still be not very reliable. That is why it is recommended to focus more on the upper layers of vegetation when calculating vegetation density.

6.2 Other metrics

Besides the metrics discussed in this thesis, there are also other metrics that can have influence on the habitat selection of woodpeckers but are difficult to calculate with LiDAR data. In papers such as Ćiković et al. (2014), much is described about the trunk of the nesting tree, such as the elevation of the cavity (its location on the tree), the diameter of the tree at breast height, wood hardness and the presence of fungi. At first, the diameter of tree trunks at breast height has been indicated in previous research as a factor that could influence the habitat selection of woodpeckers (Basile et al., 2020; Menon et al., 2021). It was found that when the diameter of a tree is bigger compared to the surrounding trees, that tree is more suitable for woodpeckers to create cavities in. Furthermore, literature has indicated that woodpecker prefer older trees, because the wood of these trees is softer and easier to excavate (Basile et al., 2020; Kosiński, 2006; Smith & Charman, 2012). Nevertheless, the age of trees has not been mapped much with the use of LiDAR data, mostly due to that other metrics such as tree height already give an indication of the age of a tree. Still, creating a method to calculate the age of trees with LiDAR data would be an interesting topic for future research.

6.3 Extraction buffer

As earlier mentioned, buffer circles with a 50 meter diameter were used to extract metric values. This diameter was chosen, because it is big enough to take the environment into account, but small enough to keep some detail. Nevertheless, there is still uncertainty about which buffer size is the most reliable for certain metrics. In research of McNeil et al. (2023), buffers with a diameter of 100, 200, 500 and 1000 meters were used. Different diameters were suitable for different calculated metrics, as for some metrics, detail is more important than other metrics. The research showed that different buffer sizes were suitable for different metrics. For this thesis research, this same method could also have been applied, as the metrics differ in detail. That is why in future research, in order to test the effect of the size of the buffer, different buffer sizes could be used to extract the values of LiDAR metrics and to eventually run a species distribution model. It could be tested which buffer sizes are the most suitable for a certain set of LiDAR metrics.

6.4 Accuracy and models

In the methodology and discussion was discussed how the use of background points can lower the AUC values of Maxent models. Therefore, Maxent models could give more reliable results if absence points are created. SDMs that make use of presence-absence points are more likely to have a better performance according to previous comparisons between different SDMs (Barbet-Massin et al., 2012). However, it is difficult to create absence points that are just as reliable as presence points. Absence points can be created in different ways. That is why it can be explored in future research what the best method is for the creation of absence points. Furthermore, in this research, a Maxent model was used, because this model seemed to be the most suitable when only presence data is available. Nevertheless, when also absence points are created, other statistical models such as a Generalized Linear Model (GLM) could also be suitable models for research into the distribution of bird species. Finally, in the discussion was described that the amount of observation points and background points could have influenced the AUC values of the Maxent models. The results of this research indicate that there could be a positive relationship between sample size and AUC values. In future research, the influence of sample size on the accuracy of a Maxent model, or another statistical model, could be explored more.

7. References

Adhikari, A., Montes, C. R., & Peduzzi, A. (2023). A Comparison of Modeling Methods for Predicting Forest Attributes Using Lidar Metrics. *Remote Sensing*, *15*(5), 1284. https://doi.org/10.3390/rs15051284

AHN. (n.d.a). *Kwaliteitsbeschrijving.* https://www.ahn.nl/kwaliteitsbeschrijving

AHN. (n.d.b). *AHN: The making of.* https://www.ahn.nl/ahn-the-making-of

ArcGIS. (2024). *AHN Vlieglijnen per jaar.* https://www.arcgis.com/apps/mapviewer/index.html?url=https://services6.arcgis.com/g6LeeIps2PV3ZiQn/ArcGIS/rest/services/AHN_Vlieglijnen_per_jaar/FeatureServer&source=sd

ArcGIS Desktop.(n.d.a). *How Focal Statistics works.* https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-focal-statistics-works.htm

ArcGIS Desktop. (n.d.b). *The types of operations in Spatial Analyst.* https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/performing-analysis/the-types-of-operations-in-spatial-analyst.htm#GUID-183D5280-C8B6-4ABF-A359-23F96B49C282

Atlassian. (2024). *A complete guide to violin plots.* https://www.atlassian.com/data/charts/violin-plot-complete-guide#:~:text=A%20violin%20plot%20depicts%20distributions,plot%2C%20to%20provide%20additional%20information.

Bakx, T. R., 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

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many?. *Methods in ecology and evolution*, *3*(2), 327-338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Basile, M., Asbeck, T., Pacioni, C., Mikusiński, G., & Storch, I. (2020). Woodpecker cavity establishment in managed forests: relative rather than absolute tree size matters. *Wildlife Biology*, *2020*(1), 1-9. https://doi.org/10.2981/wlb.00564

BirdLife International. (2024a). *Eurasian Green Woodpecker Picus viridis.* https://datazone.birdlife.org/species/factsheet/eurasian-green-woodpecker-picus-viridis/text

BirdLife International. (2024b). *Lesser Spotted Woodpecker Dryobates minor.* https://datazone.birdlife.org/species/factsheet/lesser-spotted-woodpecker-dryobates-minor/text

BTO. (2020). *Green woodpecker.* https://www.bto.org/understanding-birds/birdfacts/green-woodpecker

Burns, P., Clark, M., Salas, L., Hancock, S., Leland, D., Jantz, P., ... & 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

Ćiković, D., Barišić, S., Tutiš, V., & Kralj, J. (2014). Nest site and nest-hole characteristics used by Great Spotted Woodpecker Dendrocopos major L. in Croatia. *Polish Journal of Ecology*, *62*(2), 349-360. https://doi.org/10.3161/104.062.0213

Dalponte, M., & Coomes, D. A. (2016). Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in ecology and evolution*, *7*(10), 1236-1245. https://doi.org/10.1111/2041-210X.12575

Davison, C. W., Assmann, J. J., Normand, S., Rahbek, C., & Morueta-Holme, N. (2023). Vegetation structure from LiDAR explains the local richness of birds across Denmark. *Journal of Animal Ecology*. https://doi.org/10.1111/1365-2656.13945

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

Earth Lab. (2020, September 11). *Lesson 4. Canopy Height Models, Digital Surface Models & Digital Elevation Models – Work With LiDAR Data in Python.* https://www.earthdatascience.org/courses/use-data-open-source-python/data-stories/what-is-lidar-data/lidar-chm-dem-dsm/

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*, *17*(1), 43-57. https://doi.org/10.1016/j.gecco.2018.e00477

Endenburg, S., Mitchell, G. W., Kirby, P., Fahrig, L., Pasher, J., & Wilson, S. (2019). The homogenizing influence of agriculture on forest bird communities at landscape scales. *Landscape Ecology*, *34*, 2385-2399. https://doi.org/10.1007/s10980-019-00895-8

Esri. (2024a, April 9). *How Presence-only Prediction (Maxent) works*. https://pro.arcgis.com/en/pro-app/3.1/tool-reference/spatial-statistics/how-presence-only-prediction-works.htm

Esri. (2024b, March 22). *What is Lidar data?* https://pro.arcgis.com/en/pro-app/latest/help/data/las-dataset/what-is-lidar-.htm

Farwell, L. S., Gudex-Cross, D., Anise, I. E., Bosch, M. J., Olah, A. M., Radeloff, V. C., ... & Pidgeon, A. M. (2021). Satellite image texture captures vegetation heterogeneity and explains patterns of bird richness. *Remote sensing of Environment*, *253*, 112175. https://doi.org/10.1016/j.rse.2020.112175

Geotiles. (2024). *GeoTiles: readymade geodata with a focus on the Netherlands.* https://geotiles.citg.tudelft.nl/

Hill, R. A., & Broughton, R. K. (2009). Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne LiDAR data: A case study in lowland Britain. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(2), 223-233. https://doi.org/10.1016/j.isprsjprs.2008.12.004

Ioki, K., Tsuyuki, S., Hirata, Y., Phua, M. H., Wong, W. V. C., Ling, Z. Y., ... & Takao, G. (2014). Estimating above-ground biomass of tropical rainforest of different degradation levels in Northern Borneo using airborne LiDAR. *Forest ecology and management*, *328*, 335-341. https://doi.org/10.1016/j.foreco.2014.06.003

Jackson, J. A., & Jackson, B. J. (2004). Ecological relationships between fungi and woodpecker cavity sites. *The Condor*, *106*(1), 37-49. https://doi.org/10.1093/condor/106.1.37

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., Shi, Y., Koma, Z., Meijer, C., Ku, O., Nattino, F., ... & Grootes, M. W. (2023). Country-wide data of ecosystem structure from the third Dutch airborne laser scanning survey. *Data in Brief*, *46*, 108798. https://doi.org/10.1016/j.dib.2022.108798

Koma, Z., Seijmonsbergen, A. C., Grootes, M. W., Nattino, F., Groot, J., Sierdsema, H., ... & Kissling, W. D. (2022). Better together? Assessing different remote sensing products for predicting habitat suitability of wetland birds. *Diversity and distributions*, *28*(4), 685-699. https://doi.org/10.1111/ddi.13468

Kosiński, Z. (2006). Factors affecting the occurrence of middle spotted and great spotted woodpeckers in deciduous forests—a case study from Poland. In *Annales Zoologici Fennici* (pp. 198-210). Finnish Zoological and Botanical Publishing Board.

Kosinski, Z., & Winiecki, A. (2004). Nest-site selection and niche partitioning among the Great Spotted Woodpecker Dendrocopos major and Middle Spotted Woodpecker Dendrocopos medius in riverine forest of Central Europe. *Ornis Fennica*, *81*(4), 145-156.

Lees, A. C., Haskell, L., Allinson, T., Bezeng, S. B., Burfield, I. J., Renjifo, L. M., ... & Butchart, S. H. (2022). State of the world's birds. *Annual Review of Environment and Resources*, *47*, 231-260. https://doi.org/10.1146/annurev-environ-112420-014642

Li, C., Chen, Z., Zhou, X., Zhou, M., & Li, Z. (2022). Development of Generalized Estimation Models of Forest Inventory Attributes Using an Exhaustive Combination of Airborne Lidar-Derived Metrics. *Available at SSRN 4104346*. https://dx.doi.org/10.2139/ssrn.4104346

Martinuzzi, S., Vierling, L. A., Gould, W. A., Falkowski, M. J., Evans, J. S., Hudak, A. T., & Vierling, K. T. (2009). Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment*, *113*(12), 2533-2546. https://doi.org/10.1016/j.rse.2009.07.002

McNeil, D. J., Fisher, G., Fiss, C. J., Elmore, A. J., Fitzpatrick, M. C., Atkins, J. W., ... & Larkin, J. L. (2023). Using aerial LiDAR to assess regional availability of potential habitat for a conservation dependent forest bird. *Forest ecology and management*, *540*, 121002. https://doi.org/10.1016/j.foreco.2023.121002

Menon, T., & Shahabuddin, G. (2021). Assessing woodpeckers as indicators of bird diversity and habitat structure in managed forests. *Biodiversity and conservation*, *30*(6), 1689-1704.

Moudrý, V., Cord, A. F., Gábor, L., Laurin, G. V., Barták, V., Gdulová, K., ... & Wild, J. (2023). Vegetation structure derived from airborne laser scanning to assess species distribution and habitat suitability: The way forward. *Diversity and Distributions*, *29*(1), 39-50. https://doi.org/10.1111/ddi.13644

Naimi, B., Hamm, N. A., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling?. *Ecography*, *37*(2), 191-203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Nijssen, M., Van den Brug, A., Van Kleunen, A. & Van Manen, W. (2020). *Terreingebruik en voedsel van de Zwarte Specht in Noord-Brabant en Drenthe.* https://pub.sovon.nl/pub/publicatie/18065

Olsson, O., Nilsson, I. N., Nilsson, S. G., Pettersson, B., Stagen, A., & Wiktander, U. (1992). Habitat preferences of the lesser spotted woodpecker Dendrocopos minor. *Ornis Fennica*, *69*, 119-119.

Phillips, S. J., (2021). *A Brief Tutorial on Maxent.* https://biodiversityinformatics.amnh.org/open_source/maxent/Maxent_tutorial_2021.pdf

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*, *190*(3-4), 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Piacentini, E., & Chiatante, G. (2022). Habitat selection, density and breeding of Great Spotted Woodpecker Dendrocopos major in a protected natural area in northern Italy. *Avocetta*, *46*(2).

Pradhan, P. (2016). Strengthening MaxEnt modelling through screening of redundant explanatory bioclimatic variables with variance inflation factor analysis. *Researcher*, *8*(5), 29-34.

Puverel, C., Abourachid, A., Böhmer, C., Leban, J. M., Svoboda, M., & Paillet, Y. (2019). This is my spot: What are the characteristics of the trees excavated by the Black Woodpecker? A case study in two managed French forests. *Forest Ecology and Management*, *453*, 117621. https://doi.org/10.1016/j.foreco.2019.117621

Roussel, J., Goodbody, T.R.H., Tompalski, P. (2023). *The lidR package.* https://r-lidar.github.io/lidRbook/index.html

Scribbr. (2024). *What Is Kurtosis? | Definition, Examples & Formula.* https://www.scribbr.com/statistics/kurtosis/#:~:text=Kurtosis%20is%20a%20measure%20of,(medium%20tails)%20are%20mesokurtic.

Scribbr. (2022). *What Is Kurtosis | Definition, Examples & Formula.* https://www.scribbr.com/statistics/skewness/

Sidemo-Holm, W., Ekroos, J., Reina García, S., Söderström, B., & Hedblom, M. (2022). Urbanization causes biotic homogenization of woodland bird communities at multiple spatial scales. *Global Change Biology*, *28*(21), 6152-6164. https://doi.org/10.1111/gcb.16350

Smith, K. W., & Charman, E. C. (2012). The ecology and conservation of the Lesser Spotted Woodpecker. *British Birds*, *105*(6), 294.

Stitt, J. M., Hudak, A. T., Silva, C. A., Vierling, L. A., & Vierling, K. T. (2022). Characterizing individual tree-level snags using airborne lidar-derived forest canopy gaps within closed-canopy conifer forests. *Methods in Ecology and Evolution*, *13*(2), 473-484. https://doi.org/10.1111/2041-210X.13752

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285-1293. https://doi.org/10.1126/science.3287615

Tesfamariam, B. G., Gessesse, B., & Melgani, F. (2022). MaxEnt-based modeling of suitable habitat for rehabilitation of Podocarpus forest at landscape-scale. *Environmental Systems Research*, *11*(1), 4. https://doi.org/10.1186/s40068-022-00248-6

Torresani, M., Rocchini, D., Sonnenschein, R., Zebisch, M., Hauffe, H. C., Heym, M., ... & Tonon, G. (2020). Height variation hypothesis: A new approach for estimating forest species diversity with CHM LiDAR data. *Ecological Indicators*, *117*, 106520. https://doi.org/10.1016/j.ecolind.2020.106520

4TU. (2022). *Landelijk Grondgebruik Nederland (LGN2022).* https://doi.org/10.4121/688363cc-8c79-439f-bb0e-fe5d0deb3161.v1

Villanúa, D., Lizarraga, A., Campión, D., Senosiain, A., Gorospe, G., & Cabodevilla, X. (2023). Nest-site selection of the Iberian Green Woodpecker (Picus sharpei) in three biogeographical regions. *Journal of Ornithology*, *164*(1), 115-123. https://doi.org/10.1007/s10336-022-02024-w

Vogelbescherming Nederland. (n.d.a). *Groene specht* https://www.vogelbescherming.nl/ontdek-vogels/kennis-over-vogels/vogelgids/vogel/groene-specht#:~:text=Groene%20spechten%20zijn%20standvogels%20van,meestal%20op%20de%20grond%20verzameld

Vogelbescherming Nederland. (n.d.b). *Kleine bonte specht.* https://www.vogelbescherming.nl/ontdek-vogels/kennis-over-vogels/vogelgids/vogel/kleine-bonte-specht

Vogelbescherming Nederland. (n.d.c). *Middelste bonte specht.* https://www.vogelbescherming.nl/ontdek-vogels/kennis-over-vogels/vogelgids/vogel/middelste-bonte-specht

Wei, B., Wang, R., Hou, K., Wang, X., & Wu, W. (2018). Predicting the current and future cultivation regions of Carthamus tinctorius L. using MaxEnt model under climate change in China. *Global Ecology and Conservation*, *16*, e00477. https://doi.org/10.1016/j.gecco.2018.e00477

Woodland Trust. (2021a). *Great spotted woodpecker.* https://www.woodlandtrust.org.uk/trees-woods-and-wildlife/animals/birds/great-spotted-woodpecker/

Woodland Trust. (2021b). *Green woodpecker.* https://www.woodlandtrust.org.uk/trees-woods-and-wildlife/animals/birds/green-woodpecker/

Woodland Trust. (2021c). *Lesser spotted woodpecker.* https://www.woodlandtrust.org.uk/trees-woods-and-wildlife/animals/birds/lesser-spotted-woodpecker/#:~:text=Where%20do%20lesser%20spotted%20woodpeckers,sufficient%20number%20of%20old%20trees.

Zawadzki, G., & Sławski, M. (2023). Green tree retention as a conservation tool for the black woodpecker in managed forests. *Forest Ecology and Management*, *548*, 121398. https://doi.org/10.1016/j.foreco.2023.121398

8. Appendix

Appendix A: Overview of digital appendix

A zip file with the following files:

Overview document (PDF): Explanation of appendix A

Appendix_Aa (Folder):

- Appendix_Aa01 (R script): Upscaling of LiDAR metrics
- Appendix_Aa02 (R script): Calculation of standalone tree polygons
- Appendix_Aa03 (R script): Script 3: Merging of metric rasters of all tiles. Calculation of *distance_ST*, *distance_inside* and *distance_outside*
- Appendix_Aa04 (R script): Script 4: Extraction of metric values and running of Maxent models

Appendix_Ab (Folder): Figures of report

Appendix_Ac (Folder): Layouts of LiDAR metrics

Appendix_Ad (Folder): Raster files of LiDAR metrics

Appendix_Ae (Folder): Woodpecker observation data and other used data

Appendix_Af (Folder): PDF files of the scientific papers that were used in the report

Appendix_Ag (Folder): Thesis proposal and Powerpoint of midterm presentation

Appendix_Ah (Folder): Final report and Powerpoint of final presentation (the Powerpoint will be added after the final presentation)

Appendix B: Upscaling workflow

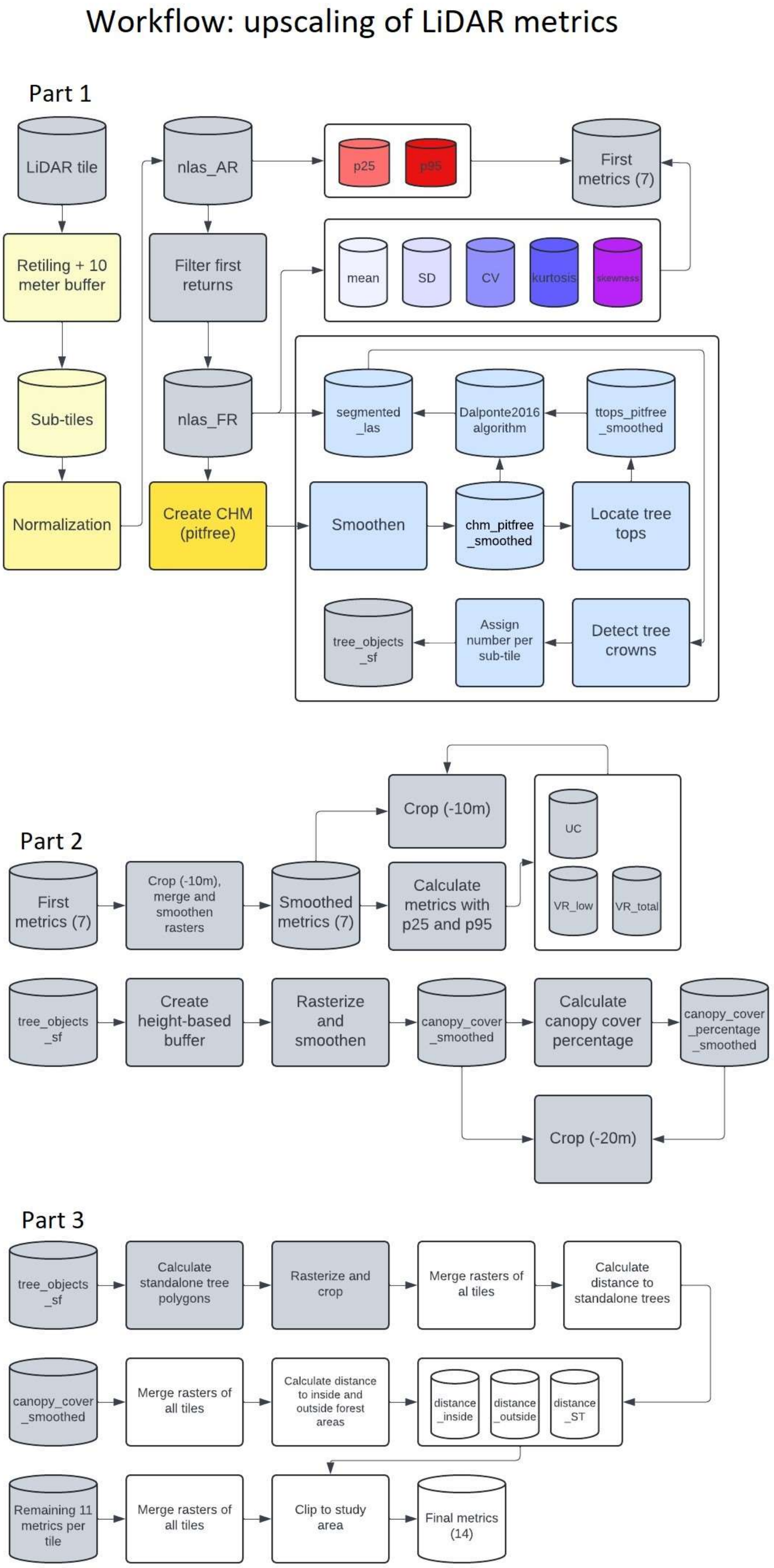


Figure 7: Workflow of the upscaling the LiDAR metrics. Grey parts of the workflow indicate processes that happened at the scale of one whole LiDAR tile. Coloured parts indicate processes that happened at the scale of one sub-tile. White parts indicate processes that happened at the scale of all 60 tiles together (containing whole study area).

Appendix C: Layouts of remaining metrics

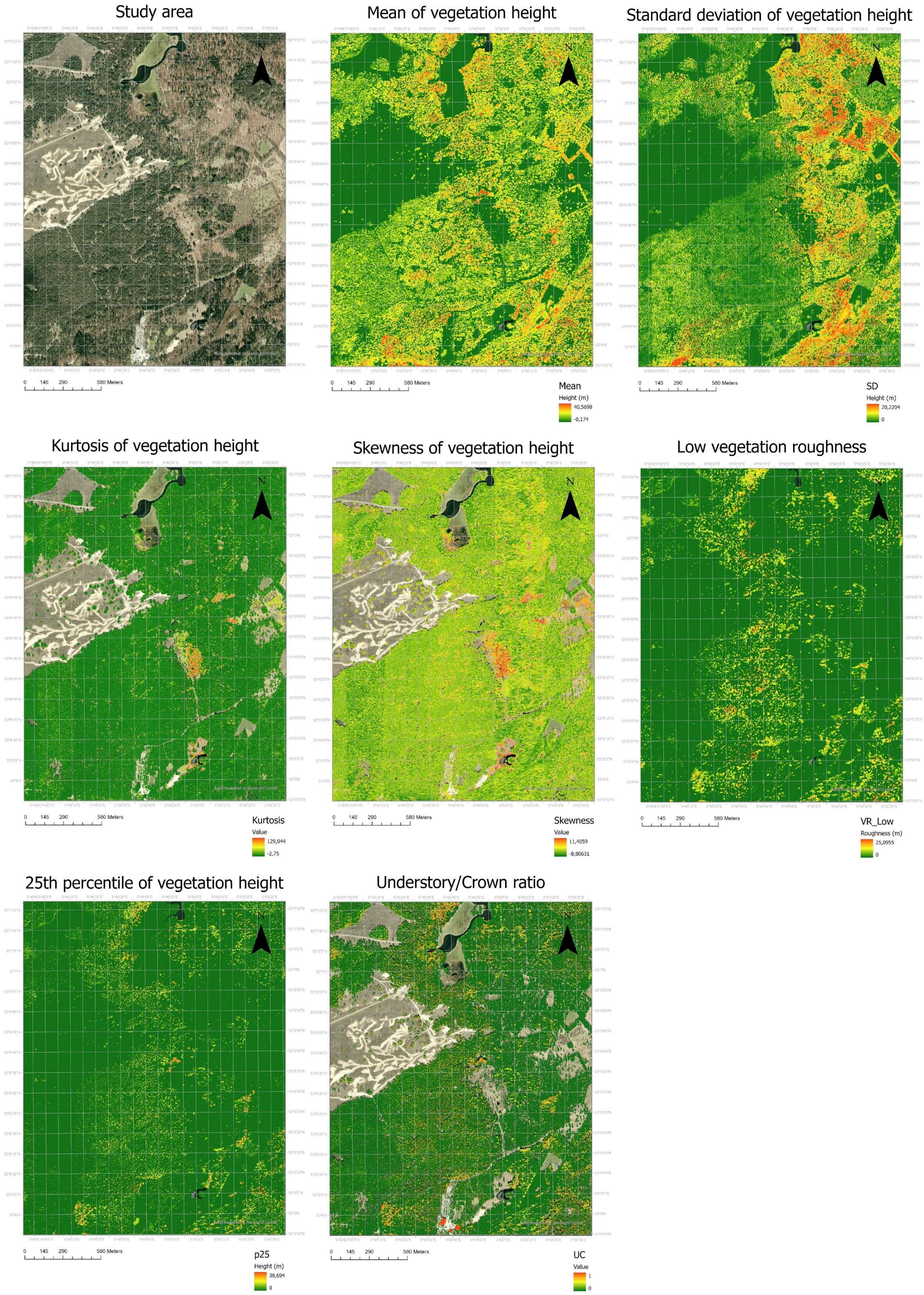


Figure 8: Layouts of the remaining metrics. This figure contains from top-left to bottom-right: mean, SD, kurtosis, skewness, VR_low, p25 and UC.

Appendix D: Correlation plots

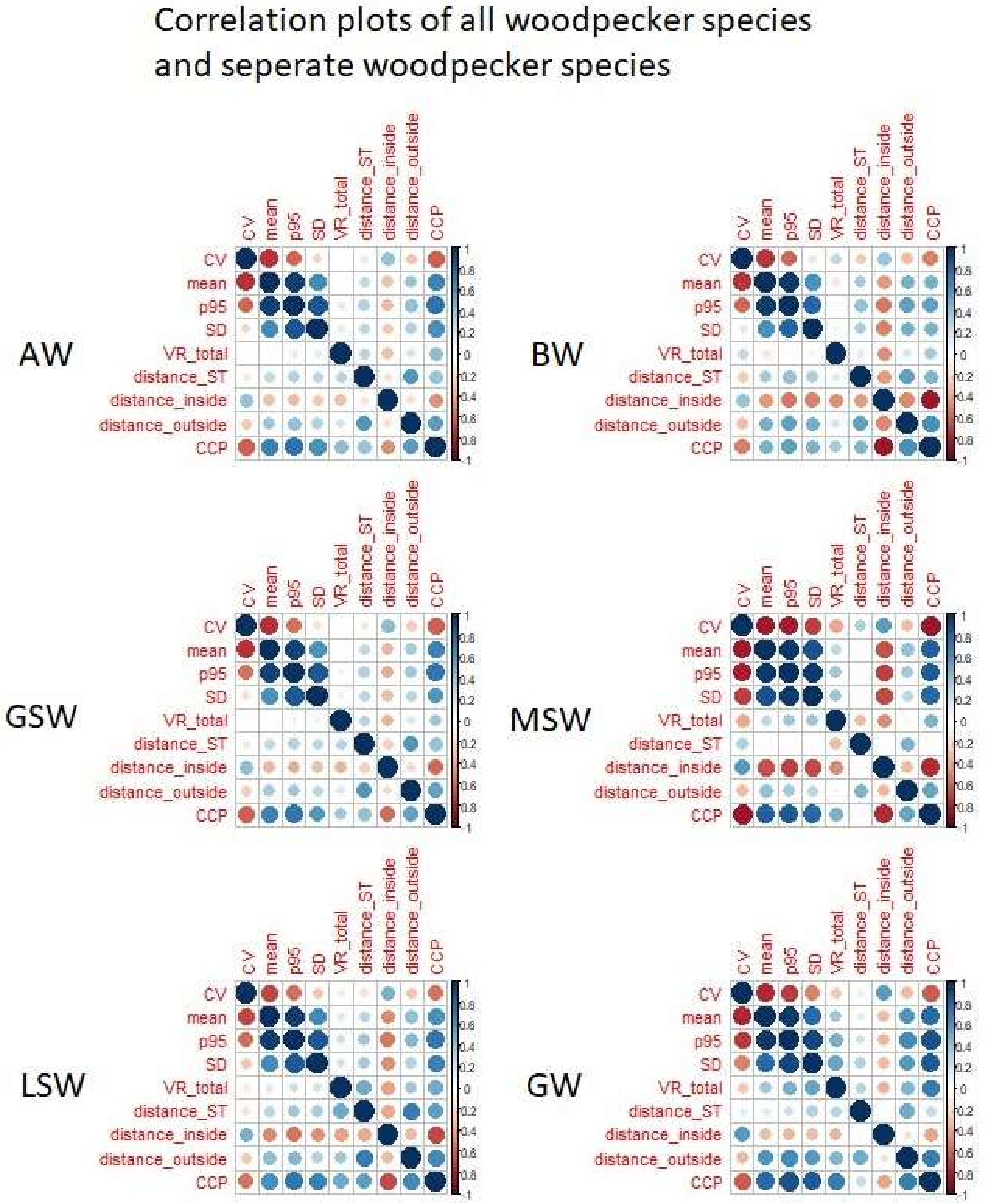


Figure 9: Correlation plots of all woodpecker species and separate woodpecker species. AW: all woodpecker species, BW: black woodpecker, GSW: great spotted woodpecker, MSW: middle spotted woodpecker, LSW: lesser spotted woodpecker and GW: green woodpecker.

Appendix E: Total response curves

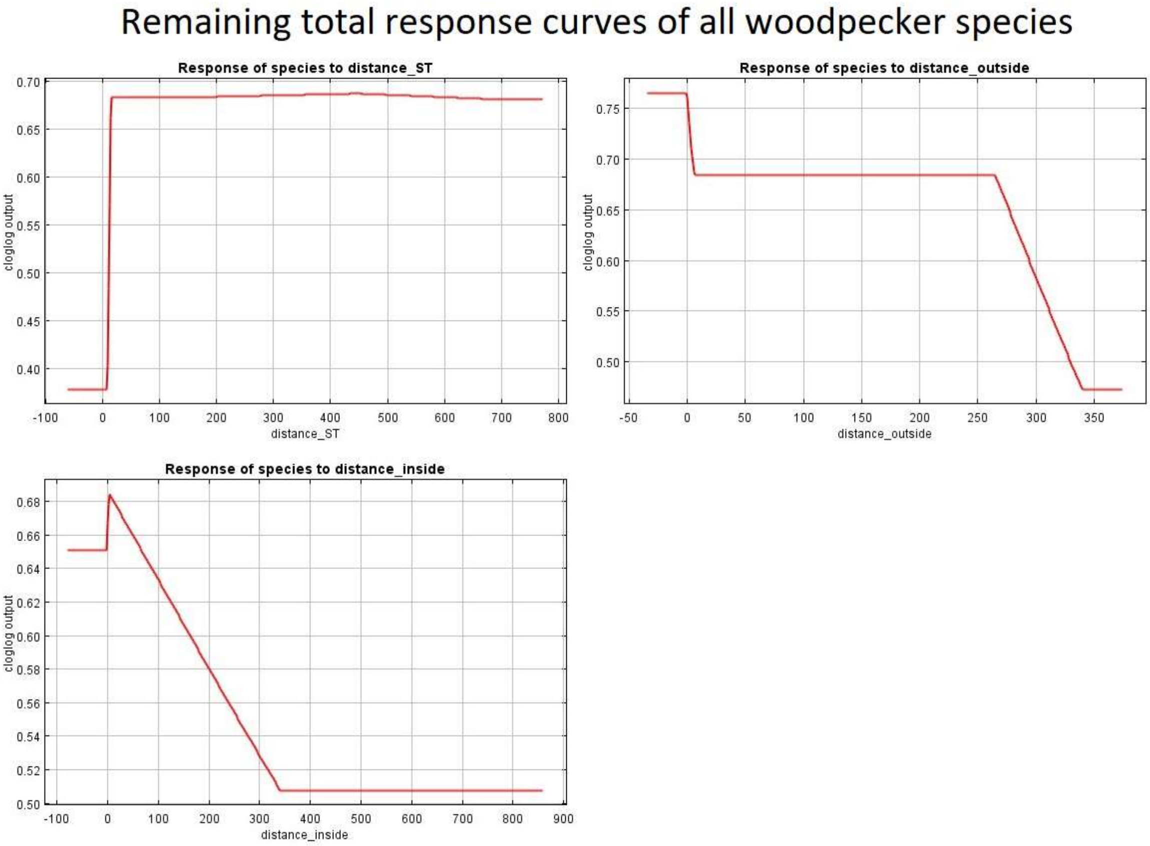


Figure 10: Remaining total response curves of all woodpecker species. This figure contains the metrics distance_ST, distance_outside and distance_inside.

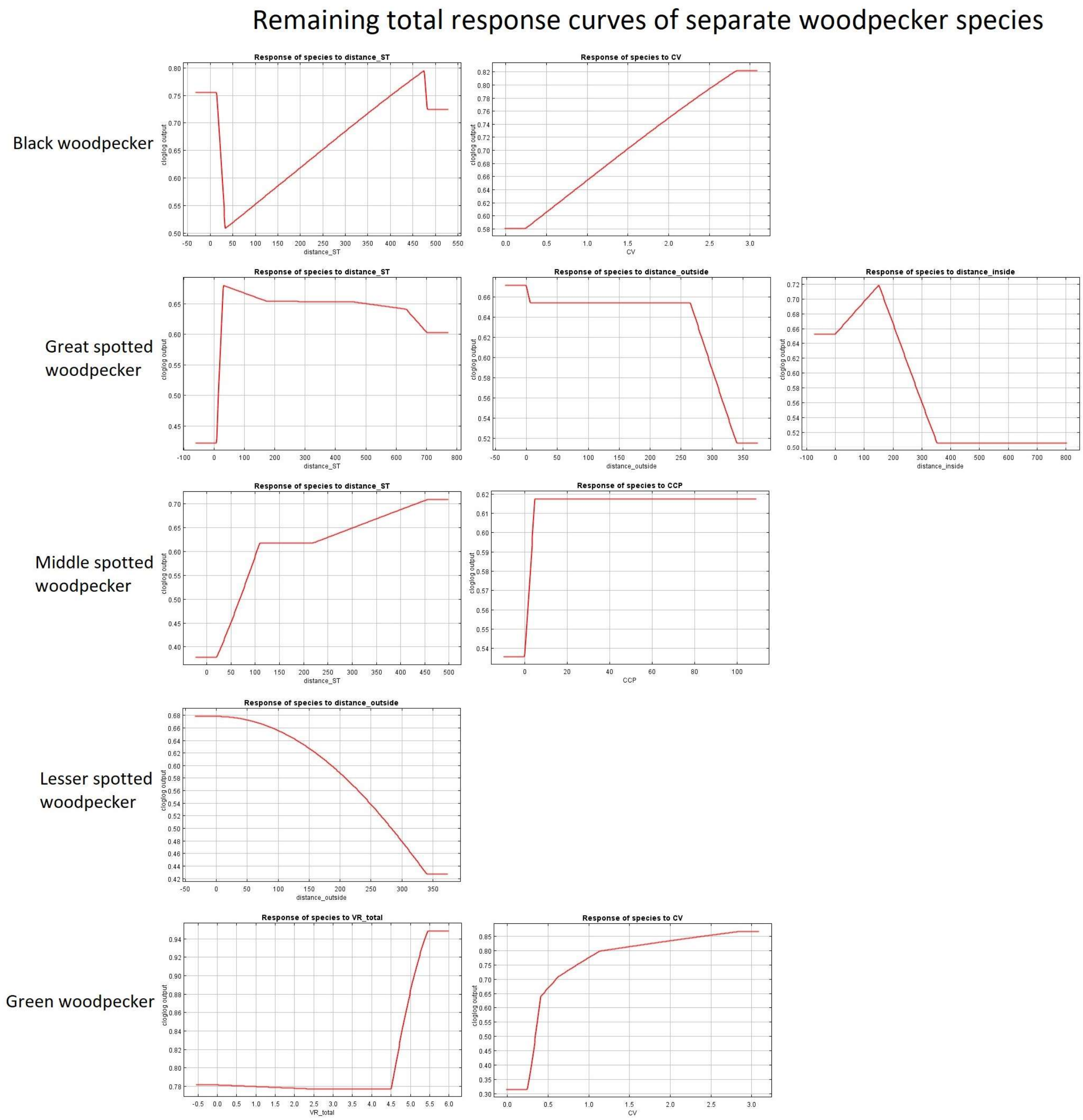


Figure 11: Remaining total response curves of separate woodpecker species. For each woodpecker species, the total response curves that are less important for the habitat selection of the species are displayed.

Appendix F: Individual response curves

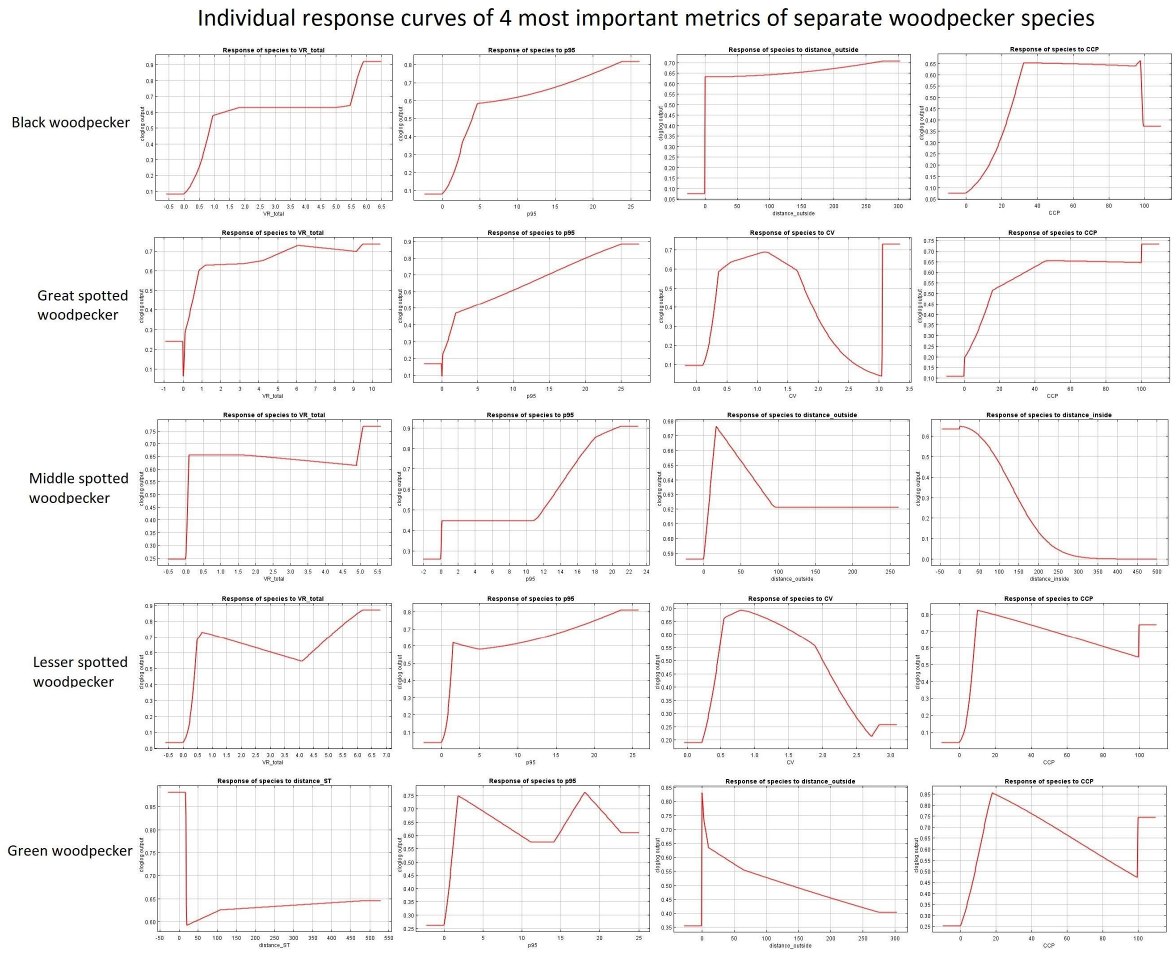


Figure 12: Individual response curves of four most important metrics of separate woodpecker species. For each woodpecker species, the individual response curves of the four most important metrics in its habitat selection are displayed.

Appendix G: Violin plots

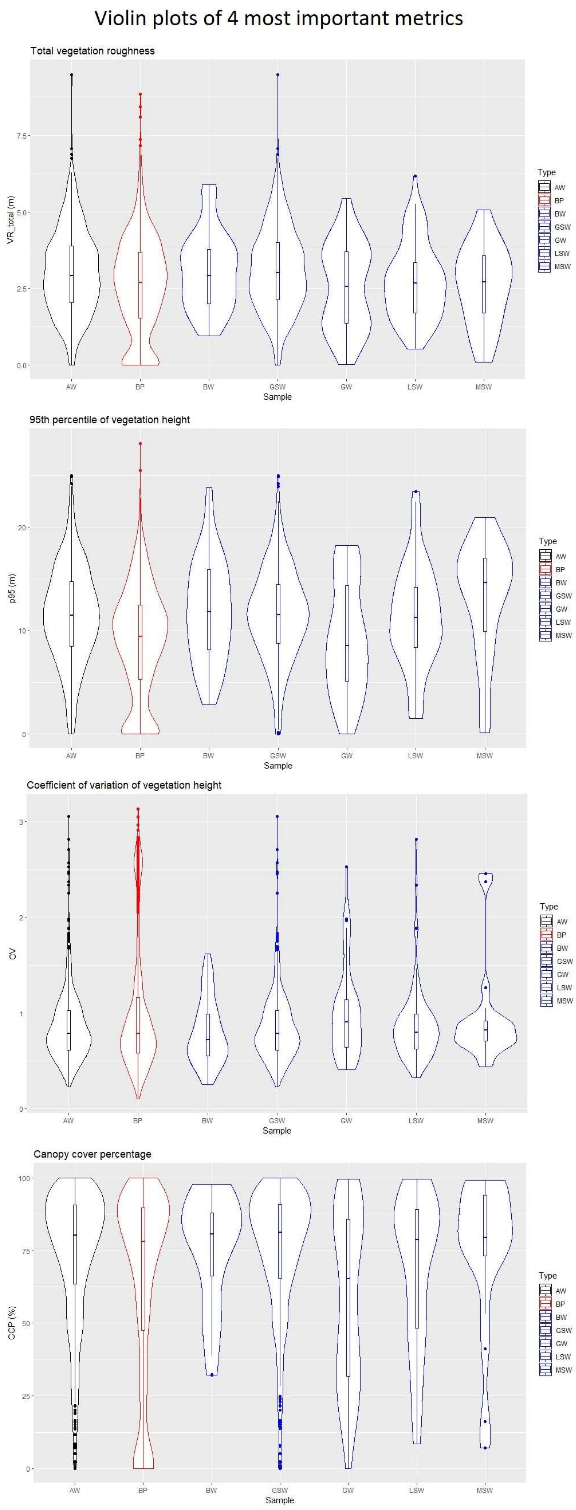


Figure 13: Violin plots of the four most important metrics of all woodpecker species. For each metric, a violin plot showing the distribution of the values of all woodpecker species, a violin plot showing the background points and a separate violin plot for each woodpecker species are displayed. AW: all woodpecker species, BW: black woodpecker, GSW: great spotted woodpecker, MSW: middle spotted woodpecker, LSW: lesser spotted woodpecker and GW: green woodpecker.

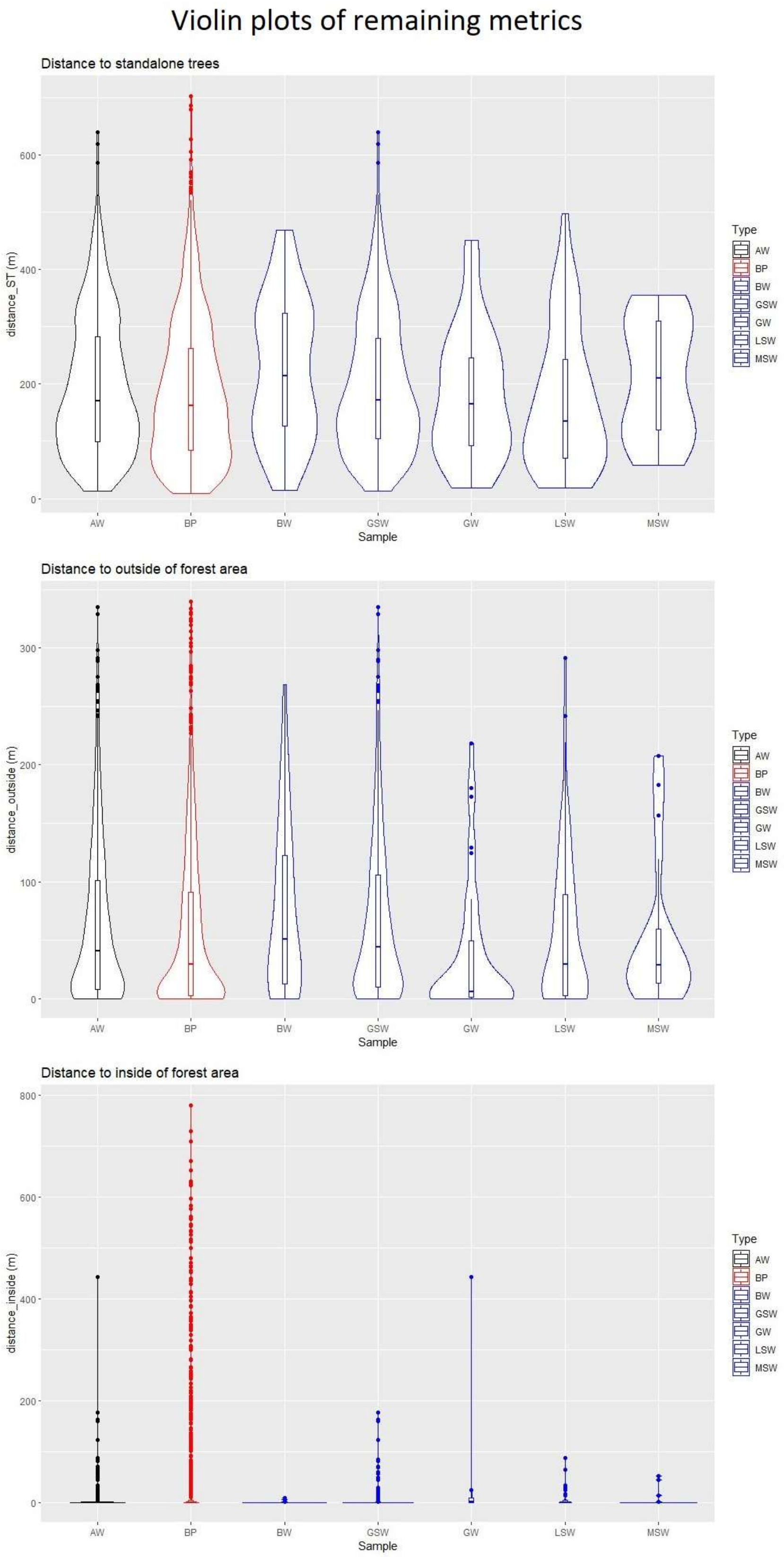


Figure 14: Violin plots of the remaining metrics of all woodpecker species. For each metric, a violin plot showing the distribution of the values of all woodpecker species, a violin plot showing the background points and a separate violin plot for each woodpecker species are displayed. AW: all woodpecker species, BW: black woodpecker, GSW: great spotted woodpecker, MSW: middle spotted woodpecker, LSW: lesser spotted woodpecker and GW: green woodpecker.

Appendix H: Results of statistical tests

Table 4: Comparison of metric values of observation points with metric values of background points. This table contains the following calculated statistics and information for each metric: the median of observation points, the median of background points, the mean of observation points, the mean of background points, the used statistical test, the resulting p-value of the statistical test.

Metric	Median: observation	Median: background	Mean: observation	Mean: background	Statistical test	p-value
VR_total	2.92	2.70	3.02	2.59	t-test	$6.692*10^{-11}$
p95	11.49	9.43	11.47	8.81	t-test	$2.2*10^{-16}$
CV	0.7848	0.7867	0.86	1.01	t-test	$5.517*10^{-11}$
CCP	80.21	78.05	73.75	64.51	Wilcoxon rank-sum	$2.979*10^{-5}$
distance_ST	169.80	162.64	195.40	183.73	t-test	0.03494
distance_outside	41.11	29.63	63.77	57.33	Wilcoxon rank-sum	$4.482*10^{-5}$
distance_inside	0	0	4.42	33.20	Wilcoxon rank-sum	$3.853*10^{-6}$

Appendix I: ROC curves

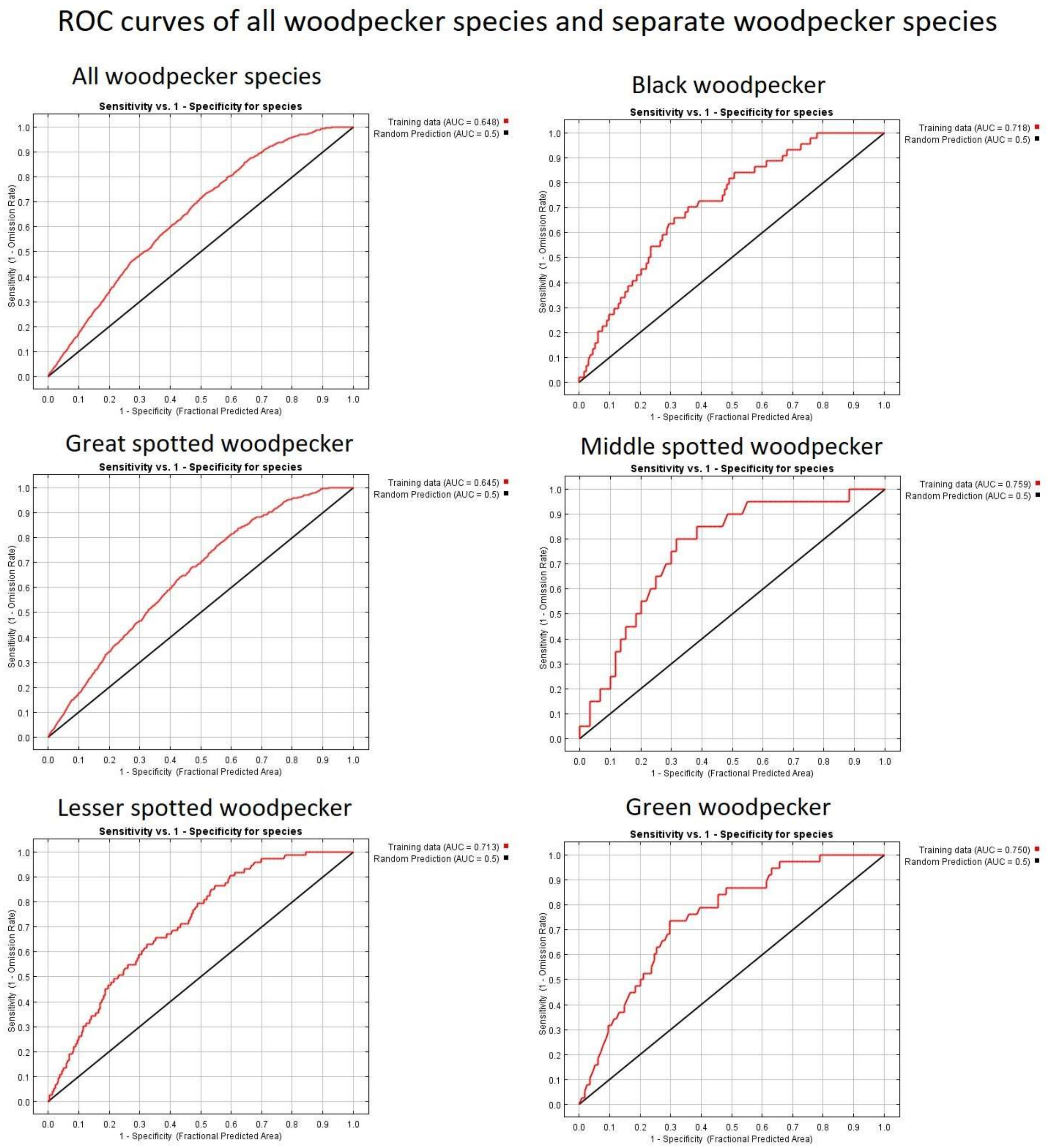


Figure 15: ROC curves of the Maxent models run for the collection of all woodpecker species and the separate woodpecker species. The red line is the ROC curve and the straight black line indicates the line of a random prediction. The area under the red line (AUC) was calculated in order to retrieve the accuracy of the model.

Appendix J: Comparison of first and all return data

Table 5: Comparison between first return and all return point density (points/m^2) for three different LiDAR tiles.

LiDAR tile	First return (points/m^2)	All returns (points/m^2)
33CN1_25.LAZ	19.26	33.98
33CZ2_03.LAZ	15.02	15.27
33CZ2_01.LAZ	20.7	26.53

Appendix K: Flight overlap in kurtosis layout

The influence of flight route overlap on kurtosis values

Figure 16: Layout showing higher kurtosis values at the locations where flight routes overlap. The orange stripes indicate these overlap areas.

Appendix L: LiDAR tiles
The following 60 LiDAR tiles were downloaded from Geotiles in order to cover the whole study area in De Hoge Veluwe (Geotiles, 2024):

<u>33CN1</u>
33CN1_13
33CN1_14
33CN1_15
33CN1_18
33CN1_19
33CN1_20
33CN1_23
33CN1_24
33CN1_25

<u>33CN2</u>
33CN2_11
33CN2_12
33CN2_13
33CN2_14
33CN2_16
33CN2_17
33CN2_18
33CN2_19
33CN2_21
33CN2_22
33CN2_23
33CN2_24

<u>33CZ1</u>
33CZ1_02
33CZ1_03
33CZ1_04
33CZ1_05
33CZ1_07
33CZ1_08
33CZ1_09
33CZ1_10
33CZ1_13
33CZ1_14
33CZ1_15
33CZ1_19
33CZ1_20
33CZ1_24
33CZ1_25

<u>33CZ2</u>
33CZ2_01
33CZ2_02
33CZ2_03
33CZ2_04

33CZ2_06
33CZ2_07
33CZ2_08
33CZ2_09
33CZ2_11
33CZ2_12
33CZ2_13
33CZ2_14
33CZ2_16
33CZ2_17
33CZ2_18
33CZ2_21
33CZ2_22
33CZ2_23
33CZ2_24

<u>40AN1</u>
40AN1_05

<u>40AN2</u>
40AN2_01
40AN2_02
40AN2_03
40AN2_04