The National Forest Inventory: differ Dutch forest owners in their forests and harvest?

Robin Zwijgers

February 2023

WAGENINGEN
UNIVERSITY & RESEARCH

The National Forest Inventory: differ Dutch forest owners in their forests and harvest?

Robin Zwijgers

February 6th, 2023

Wageningen

MSc thesis (FEM-80436)

Student registration number: 1044736

Supervisors:

Dr. ir. M. Schelhaas

Vegetation, Forest, and Landscape Ecology team – Wageningen Environmental Research

Dr. ir. J. den Ouden

Forest Ecology and Forest Management group – Wageningen University

The author declares that this work is his own and that he has adhered to The Netherlands

Code of Conduct for Academic Practice.

ABSTRACT

National forest inventories (NFI) are a common tool to measure the state of a forest and monitor its development over time. Previous Dutch NFIs showed that harvest quantities differed per owner, indicating differences in the way different owners manage their forests. Forest management is thought to be crucial in creating resilient forests that maintain their provisioning of ecosystem services, also under a changing climate. With rising expectations (e.g., recreation, timber production, and biodiversity) from and pressures on the Dutch forests, it is key to understand how forest management is currently performed to know what eventually could be changed to deal with these pressures. Furthermore, harvest in the Dutch forest has always been a hot topic for public debate, with some 'usual suspects' as the wrongdoer. This thesis aims to understand 1) how we can describe and classify Dutch owners and their forests and 2) what the differences in harvest are and how the measured variables explain these.

Variables included in this thesis were related to forest structure, species composition, site conditions, and spatial variables, such as Natura 2000 and SNL-type (subsidy), as derived from the Dutch NFI. I expressed harvest as the proportion of basal area that was removed. The factor analysis of mixed data showed no clear or recognisable clusters of forest owners, but the forests of Large Nature Management organisations and State Forest Service forests were the most distinctive. The differences in the forest were mainly due to site conditions, and the differences between owners were associated with variables related to the forest's development phase. The random forest identified SNL type as the most important variable to predict owner. The zero-inflated beta-distribution regression model showed that both the production-oriented SNL-type and increment were associated with an increase in harvest intensity. Dominant tree species ash and poplar had a significant associated increase in harvest intensity. The owner group did not significantly influence harvest intensity but affected the presence of harvest. Only one owner group (Other Publicly Owned) was associated with increased harvest presence, while all other groups were similar. The volume of dead wood was associated with a lower presence of harvest.

This thesis is the first to apply a multivariate analysis, a machine-learning approach, and a beta-distribution to the Dutch NFI dataset. Overall, it showed that the site conditions determine, for a large part, the type of forest present and that ownership differences are small and probably related to a forest's age and historical context. The SNL-type shows important correlations with the harvesting intensity, which confirms that this scheme is an important policy tool to influence forest management in the Netherlands. Variables related to a forest's age were also associated with affecting the harvest intensity or presence. Concluding, the forest's site conditions shape the 'appearance' of the forest (i.e., forest structure and species composition), while the SNL-type and age shape how the owners manage their forest.

Keywords: The Netherlands, forest inventory, forest owners, harvest, multivariate analysis, random forest, beta-distribution

FOREWORD

The aim of a thesis is often written as the scientific question one wants to answer and how to contribute to the academic society. The educational aim is, however, to learn about conducting research practices and working on your own project. After various courses with group work, a thesis is a different story. I was my project manager, junior researcher, secretary, and statistician. I was lucky enough to be surrounded by the right people to perform well on these tasks, who I want to thank.

First, my supervisors Mart-Jan Schelhaas and Jan den Ouden, were available for questions, feedback, and time. Both of you were understanding at the difficult start of my thesis. Jan, it was great that you made time for me to discuss results or progress when I dropped by your office. I highly valued your expertise when you asked interesting or difficult questions. Mart-Jan, many thanks for your patience and hospitality in letting me work in the offices of VBL. Seeing people grabbing their WUR-card when I entered the room was a pleasure. I appreciated your tireless support and guidance throughout the thesis, especially in moments when you helped me with setting and reaching deadlines. At a certain moment, we got the right recipe, it seemed.

Second, I need to thank Louis König for his support through all my statistical questions while working on his PhD-project. After every chat, I needed to change my methods and approach, but it benefitted this thesis in the end. I also want to thank Sara Filipek for her feedback on the thesis proposal and draft report, and for every time I borrowed your card for coffee.

Lastly, I want to be grateful for my family and friends. In particular, my parents, sister, and girlfriend, Ben, Elly, Lizet, and Fleur. It is tremendously hard to write this summation of people without Peter. I now know that plans can change and that you need the right people to cope with these changes. I am ending this section with sharing the same song Peter shared with me after our grandmother had passed.

Zoveel Dingen - Akwasi

Table of Contents

Abstract ii

Foreword iii

Table of Contents iv

1. Introduction 1
   - 1.1 Context 1
   - 1.2 Objectives and research questions: 2
2. Study area and data preparation 3
   - 2.1 The Dutch forest 3
   - 2.2 Dutch NFI 4
   - 2.3 Data preparation 5
3. Describe the Dutch forest and its owners based on the NFI-7 9
   - 3.1 Methods 9
   - 3.2 Results 9
   - 3.3 Discussion 13
4. Predict owner groups based on the data of the NFI-7 15
   - 4.1 Methods 15
   - 4.2 Results 15
   - 4.3 Discussion 17
5. Investigate harvest intensity based on the repeated inventories 19
   - 5.1 Methods 19
   - 5.2 Results 21
   - 5.3 Discussion 25
6. Discussion 27
   - 6.1 Describe and predict owners 27
   - 6.2 Harvest intensity and owners 28
   - 6.3 Recommendations 29
   - 6.4 Implications for management or society 29
7. Conclusion 30
8. References 31
9. Appendix 35
   - 9.1 Appendix A: Data preparation 35
   - 9.2 Appendix B – additional figures of chapter 3 (FAMD) 38
   - 9.3 Appendix C – additional figures of chapter 4 (RF) 42
   - 9.4 Appendix D – additional figures of chapter 5 (ZIBD) 43
   - 9.5 References 50

1. INTRODUCTION

1.1 CONTEXT

In 2019, the Dutch government imposed a new forest strategy aimed at a 10% forest area increase (37.000 ha) by 2030 (LNV et al., 2020), which presents additional challenges in this small country where every square meter has multiple functions. Dutch forests are used for recreation, nature conservation, and wood production, and these pressures create a balancing act for forest managers while keeping climate change in mind (Boosten, 2016). These pressures are not only a Dutch issue. New forest policies and management styles are suggested, such as Climate Smart Forestry (Nabuurs et al., 2017) or the EU Forest Strategy, which pledges to plant 3 billion trees by 2030 (European Commission, 2021). However, to incorporate new ideas about management effectively, one should know the present status of forests and what drives their management.

For centuries, humans have managed and transformed European forests and landscapes (Kaplan et al., 2009; McGrath et al., 2015). Although the Netherlands was once referred to as "Holtlant"[a], indicating a wooded area, only 11% of the Netherlands' land surface is currently forested (Schelhaas et al., 2022). Most Dutch forests have been cleared, replanted, and managed in diverse ways with various objectives (Staatbosbeheer et al., 2021). Despite the history of our forests and their management, the public, conservationists, and forestry professionals have been arguing about Dutch forest management and the effects and size of harvest (den Ouden et al., 2020; Havermans, 2021; Naaijen et al., 2021). Harvest is a complex topic to communicate since society values forests as vital to nature and sees harvest as disruptive (Arts et al., 2014). Especially State Forest service has been criticised for its harvest methods and forest management (ten Hooven, 2020).

Forest management covers a multitude of silvicultural activities, such as harvest (thinning and final harvest), species selection, replanting, and site preparation (den Ouden et al., 2010). Therefore, stand-level decisions of forest managers are the root of management (Duncker et al., 2012). Others have tried to unravel decision-making in forest management by conducting interviews (Hoogstra et al., 2009) or running exercises in marteloscopes (silvicultural training sites) (Cosyns et al., 2019; Joa et al., 2020). Hoogstra et al. (2009) found that despite the common adage in forestry of "looking ahead for the long run", the surveyed Dutch and German forest managers use a horizon of 15 years in practice. The marteloscope exercises often revealed that decision-making is not only a rational practice but is also based on experience and preferences (Cosyns et al., 2019; Joa et al., 2020). Both make sense since different forest owners have different objectives and incentives for their management (Hengeveld et al., 2017). Nevertheless, mimicking a decision situation or asking managers what they do might alter the process: people might answer or behave to what is desired[b]. That is why a forest inventory is a great opportunity: it is non-experimental observational data where we can see what owners and managers actually do rather than say they do.

National forest inventories are a common tool to measure the state of a forest, monitor forest changes over time, and report to international commitments, such as the land-use, land-use change and forestry (LULUCF) under the Kyoto Protocol (Tomppo et al., 2010; Schelhaas et al., 2014; Arets et al., 2019). The Dutch national forest inventories, hereafter NFI, have been recurring since 1938 at different time intervals and with different aims (Daamen et al., 2010; Oldenburger et al., 2016). The general idea behind the NFI is to gather information about the state of the forest, for instance, standing stock, forest composition, and total forested area. By randomly selecting the sampling points in all types of forests, it creates information about the entire Dutch forest.

The database of the NFI encompasses a multitude of data and provides information about ownership, site conditions, germination year, basal area and tree-specific measurements such as DBH and species. Furthermore, the SNL-type[c] of each plot is added, enabling insight into the type of nature and possible subsidies. Since the NFI database notes when a tree is harvested and removed from a permanent plot, it is possible to

[a] https://etymologiebank.nl/trefwoord/holland accessed 13-05-2022
[b] This type of bias is called the Observer Expectancy Effect. https://thedecisionlab.com/biases/observer-expectancy-effect accessed 01-06-2022
[c] Provincial nature and landscape subsidies are based on the SNL. https://www.bij12.nl/onderwerpen/natuur-en-landschap/subsidiestelsel-natuur-en-landschap/ accessed 06-02-2023

detect this aspect of forest management directly. In contrast, other management activities, such as active planting, are not directly detectable.

In general, most researchers investigating forest management have utilised models to predict a certain type of management based on information which typically stems from one of the following categories: biotic, abiotic, socioeconomic, or political. Hengeveld et al. (2012) present a European-wide map with potential forest management approaches based on variables such as dominant tree species, slope, proximity to cities, and Natura 2000 sites. However, much of the research on NFI data uses variables such as DBH and owner class solely. In an investigation of European-wide NFI data, Schelhaas et al. (2018b) found that harvest probabilities are influenced by, among other things, site productivity, subsidies, and ownership. Levers et al. (2014) showed that European spatial patterns of harvest intensity were related to forest structure and composition, site conditions, and country-specific characteristics. Usually, studies on forest management focus on one side: prediction (modelling potential management) or description (investigating management with NFI data). With this thesis, I aim to contribute to the study area by using NFI data to describe forest management and the owners while combining it with a modelling approach to investigate management.

1.2 OBJECTIVES AND RESEARCH QUESTIONS:

The latest inventory (NFI-7) adds new data to existing knowledge, and to date, few in-depth studies have examined the relationship between harvest and the measured variables in the Dutch case. Instead of investigating all management activities, I will concentrate solely on harvest (expressed as intensity) as it is one of the main activities in Dutch forest management, clearly recognisable in the database and a central topic in the public debate. Therefore, this thesis aims to understand how Dutch forest management can be characterised based on the most recent observation (NFI-7) and the repeated NFI data (MFV, NFI-6, and NFI-7).

I formulated the following questions to explore and understand the possible differences in management between forest owners and the variables that might explain these differences.

1.2.1 How can we describe and classify owners and their forests in the Netherlands based on the single observations of the 7th NFI?

I expect differences in the forest structure and composition between owners since I assume that different owners manage their forests differently. The current structure and species composition of the forest will reflect the management in the past. Furthermore, I expect that owners are clustered mainly on a production ↔ nature-oriented axis (with associated variables such as commercial species and amount of dead wood) and to a lesser extent on a professional–private axis (Schelhaas et al., 2018a; Arets et al., 2019).

1.2.2 What are the differences and patterns of harvest intensity observed in the Netherlands using the repeated NFI data, and how can the measured variables explain these patterns?

I expect that harvest intensity differs between forest and nature management organisations and private owners. Private forests have a lower harvest than others, probably due to the inactivity of some owners (Schelhaas et al., 2018b). Furthermore, I expect variables related to policies (such as Natura 2000) to influence harvest intensity levels for nature management organisations. For Staatsbosbeheer, I expect harvest levels to be affected by variables that are more related to production (such as commercial species or subsidy schemes).

My thesis is composed of seven chapters. The second chapter is concerned with the methodology for data preparation for all other chapters. Three main chapters are following, which address for each applied analysis per chapter the methods, results, and discussion. Chapters three and four investigate the first and chapter five the second research question. It will then go on to a general discussion and the last chapter concludes this work.

2. STUDY AREA AND DATA PREPARATION

2.1 THE DUTCH FOREST

The Dutch landscape shows great variability from a river delta with clay and peatlands in the west and north to dunes along the coast with the North Sea and high-lying areas with sandy soils in the southern and eastern parts of the country (Figure 1.). These soils differ significantly in fertility. Fertile clay and peaty soils have mostly been converted to agricultural use, while the less productive, dry sandy soils host the forests (Figure 2). Scots pine dominates most of these forests, which were planted to reclaim degraded (heath)lands over 1900-1940 (Barends, 2010, p. 147; den Ouden et al., 2010, p. 516). Some remnants of old forests remained, dominated by oak and beech, but these were converted to Douglas fir and larch in many places. After the land reclamation campaigns, creating the polders in the middle of the country (1940-1970). Forests were planted on these richer soils using a variety of broadleaved species (poplar, ash, oak, beech) (Arnoldussen, 1982).

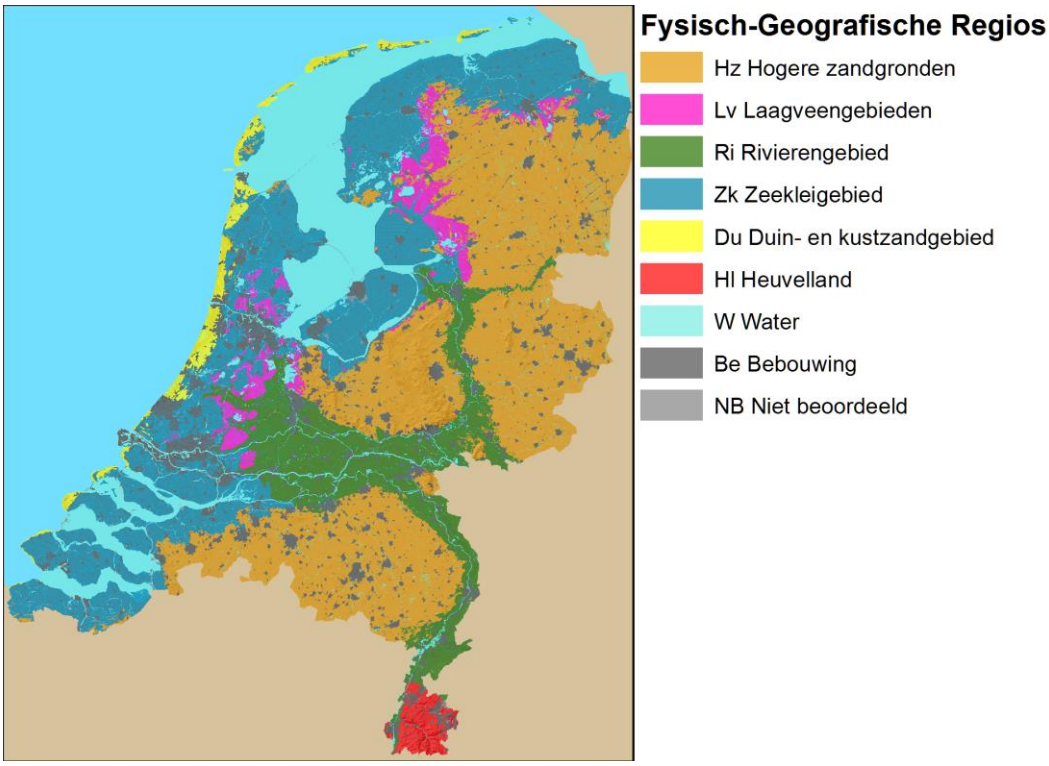

Figure 1. Overview of the physical-geographical regions, copied from https://landschapsleutel.wur.nl/ accessed 20-12-2022. Hz: high-lying sandy soils; Lv: Fen (peat-accumulating wetland); Ri: riverscape/fluvial clay, Zk: tidal marsh/marine clay; Du: dunes; Hi: loess landscape; W: water; Be: built-up; NB: not assessed.

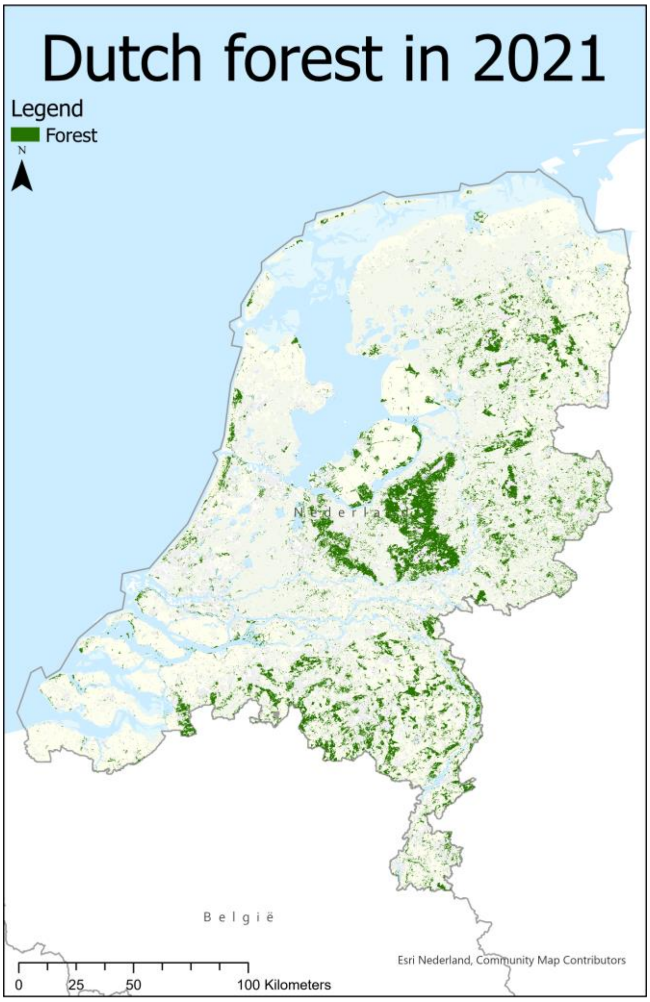

Figure 2. Map of Dutch forests in 2021, based on Land Use, Land-Use Change and Forestry (LULUCF) data from Arets et al. (2022). The forest situates mainly on high-laying sandy soils (called Hz in Figure 1).

2.2 Dutch NFI

For this thesis, I used data from public spatial sources and the publicly available Dutch NFI database [d]. The Dutch National Forest Inventory (NFI) gathers information about the state of the forest in the Netherlands for international reporting and national decision-making (Schelhaas et al., 2022).

The database consists of measurements from the 5th, 6th, and 7th NFI. The 5th NFI (officially referred to as "Meetnet FunctieVervulling"; MFV) was measured between 2001 and 2005 (Schelhaas et al., 2014). The 6th NFI (2012-2013; NFI-6) and the 7th NFI (2017-2021; NFI-7) used, in general, the same methods as the MFV. The NFI uses an unaligned systematic sample design with a density of 1 plot per 100 ha (Schelhaas et al., 2022). Plots were selected that were located in forests, according to the LULUCF maps (Figure 2) and aerial photographs. During MFV and NFI-6, half the plots were marked as permanent, and half were marked as temporary, while in NFI-7, all plots were treated as permanent (Schelhaas et al., 2022). In permanent plots, the same trees are re-measured every time, giving insight into increment, harvest and mortality. The number of plots selected in each inventory is roughly 3600 (Figure 3), while in practice, due to, for example, prohibited access, around 3150 plots could be measured.

Whereas other national forest inventories have a fixed plot radius, often between 5 and 25m (Schelhaas et al., 2018b), the Dutch NFI uses a variable plot radius. The field crew chose a radius that included at least 20 trees (DBH > 5cm) around the centre of the sampling point. Schelhaas et al. (2014), Schelhaas et al. (2022), and Daamen et al. (2019) provide detailed information about plot selection and other methods of the NFI.

---

[d] https://www.probos.nl/publicaties/overige/1094-bosinventarisaties accessed 13-07-2022.

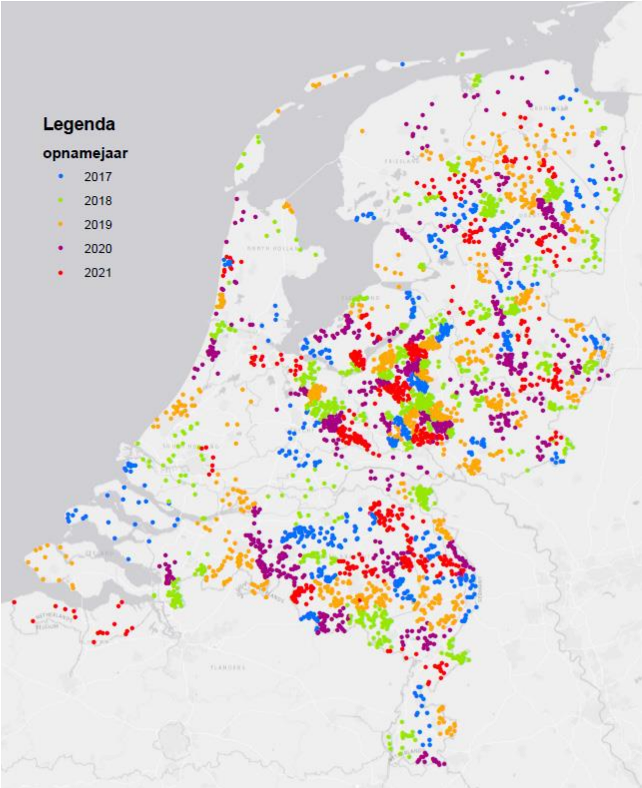

Figure 3. Map with all plots from the NFI-7, coloured by observation year. Copied from Schelhaas et al. (2022). A total of 3.601 plots were selected during the NFI-7 and measured over 2017-2021.

The public NFI database contains data at two levels: plot and tree-level data. Plot-level covers, amongst others, the dominant tree species, anonymised ownership, germination year, and inventory date. It includes, in addition, information about the location of the plot: anonymised grid location, the permanency of the plot, municipality, and province. The table with tree-level data holds measurements of the trees inside each plot, for example, tree number, diameter at breast height (DBH), species, and tree and harvest class.

2.3 Data preparation

For the data preparation and analysis, I used ArcGIS Pro 3.0.2 (Esri inc., 2022) for spatial analysis, Microsoft® Access® (Microsoft, 2022) for extracting and calculating additional plot variables, and RStudio (Posit team, 2022) for data exploration and statistical analysis. I combined owners into groups defined by Schelhaas et al. (2018a) to ease comparison (Table 1). In addition, the database provides the dominant tree species per plot (Table 2), for which I adopted the classification as used by the latest NFI report (Schelhaas et al., 2022). When I refer to a species in this thesis, I used the classification of Table 2 to group species.

Table 1. Classification of owner groups (after Schelhaas et al. (2018a))

Owner group	*Abbreviation*	*Original owner class*
State Forest Service	SBB	Staatsbosbeheer
Large Nature Management organisations	LNM	Natuurmonumenten, Waterleidingduinen, (provincial) landscape organisations
Other Publicly Owned	OPO	Ministry of Defence, Ministry of Finance, municipalities, provinces, water boards etc.
Private Organisational Owned	POO	Estates, foundations, churches, companies, etc.
Private Individuals	PIN	Private owned

Table 2. Classification of species groups (after (Schelhaas et al., 2022)). A detailed version with all species classified per group is in Appendix 9 of Schelhaas et al. (2022).

Species group	*Description*	*Scientific names*
AL	Alnus	*Alnus incana, Alnus glutinosa*
AO	American oak	*Quercus rubra*
AP	Austrian pine	*Pinus nigra var. nigra*
AS	Ash	*Fraxinus excelsior*
BE	Beech	*Fagus sylvatica*
BI	Betula	*Betula pubescens, Betula pendula*
CP	Corsican pine	*Pinus nigra* var. *maritima*
DF	Douglas fir	*Pseudotsuga menziesii*
EB	Exotic broadleaves	Amongst others: *Aesculus hippocastanum, Prunus serotina, Robinia pseudoacacia*
HA	Harvest area (Dutch: "bos in open fase")	
IB	Indigenous broadleaves	Amongst others: *Ulmus, Castanea sativa, Populus tremula*
LA	Larch	*Larix kaempferi, Larix decidua*
MA	Maple	*Acer*
NS	Norway spruce	*Picea abies*
OA	Oak	*Quercus robur, Quercus petraea*
OC	Other conifers	Amongst others: *Abies alba, Abies grandis, Juniperus communis* and *Pinus pinaster*
PO	Poplar	Populus
SH	Shrubs	Amongst others: *Amelanchier lamarckii, Salix caprea, Sorbus aucuparia*
SP	Scots pine	*Pinus sylvestris*
WI	Willow	*Salix*
XX	Not visited	

Table 3 gives an overview of the variables I calculated per plot and per inventory to quantify the forest in terms of forest structure, species composition, and harvest. Appendix A presents more details on how these variables are calculated. These variables are used to describe and classify owners (research question 1), to explain harvest intensity (research question 2), or both. When I indicate that the variable is not used, it was dropped during the data analysis due to, for example, multicollinearity. I selected only living trees for all calculations except the volume of dead wood.

I used the latest inventory (NFI-7) to describe and classify owners and their forests (Chapters 3 and 4). For the harvest intensity analysis, I used the MFV, NFI-6, and NFI-7 (Chapter 5). I used the tidyverse package (Wickham et al., 2019) for data handling, transformation, and visualisation and the ggcorrplot (Kassambara, 2022) for constructing the correlation matrices. The data, GIS-models, and R-scripts are available upon request.

Table 3. Plot-level variables included in this thesis. I grouped the variables based on their relation to the forest and the original data. I specified for which research question I used the variable, but when I stated 'none', I dropped this variable due to, for example, multicollinearity.

Group	*Variable*	*Abbreviation*	*Unit*	*Used in RQ 1 or 2:*	*Notes*
Forest structure	Quadratic Mean Diameter	QMD	mm	Both	
	Stem density	Dens	#trees ha^{-1}	Both	
	Basal area	BA	m^2 ha^{-1}	Both	
	Maximum of DBH	Max_{DBH}	cm	None	
	Standard deviation of DBH	SD_{DBH}	cm	None	
	Gini coefficient of BA	Gini_BA	Unitless	Both	Index for inequality (heterogeneity) of BA within a plot.
	Coefficient of variation of DBH	CV_{DBH}	Unitless	None	Standardised index of dispersion of distribution (i.e., relative variation) of DBH within a plot.
	Regeneration of dominant tree species	DTS_Regeneration	#trees ha^{-1}	1	Without shrub species.
	Regeneration of other tree species	Other_Regeneration	#trees ha^{-1}	1	Without shrub species.
	Volume of dead wood	DW_sum	m^3 ha^{-1}	Both	
Species composition	Shannon Diversity Index	ShannonIndex	Unitless	2	Index for species diversity within a plot.
	Shannon Evenness Index	SEI	Unitless	2	Index for the relative evenness of the species distribution (i.e., relative abundance amongst species).
	Share of BA per species (19)	$BA\__{i}$	Unitless	1	Proportion contribution per species (*i*) to the plot's basal area.
	Dominant tree species	DTS	Category	2	See Table 2.
Abiotic	Albos class	Alb_class	Category	1	Site conditions: poor - medium - rich.
Spatial	SNL-type	SNL_Unrestricted	Binary	Both	The assigned nature type and reclassified as: [1] Multifunctional (SNL 16 and no SNL assigned); [0] Nature oriented (all other SNL-types). Following Schelhaas et al. (2018a).
	Natura 2000	N2000	Binary	Both	Plot is within Natura 2000 area [1] or not [0].

Other	Walking distance	Walking_Distance	Binary	Both	Plot is within an average walking distance around city limits [1] or not [0].
	National parks	National_Parks	Binary	Both	Plot is in a national park [1] or not [0].
	Time between inventories	Time_Between	Years	None	Rounded to the first decimal.
	Age	Age	years	1	
	Increment (volume)	V_incr	m^3 ha^{-1} yr^{-1}	Both	
	Owner type	Owner_Group	Category	2	See Table 1.
	Harvest Intensity	HI_sum	Unitless	2	Proportion removed basal area. Sum of basal area of trees indicated as harvested by the consecutive inventory, divided by the plot's basal area.

3. DESCRIBE THE DUTCH FOREST AND ITS OWNERS BASED ON THE NFI-7

My first research question is: *"How can we describe and classify owners and their forests in the Netherlands based on the single observations of the 7th NFI?"*. I divided this question into a descriptive part (describe the Dutch forest based on the variables, this chapter) and a predictive part (classify the Dutch forest and predict the owner based on the variables, Chapter 4).

3.1 METHODS

To describe the Dutch forests and their owners, I used a multivariate analysis to investigate which variables contribute to explaining the variance in the NFI-7 data. I focussed my analysis on forests where management is likely not restricted by special (historic) features or a predominant other land use, for example, leaving out lanes, forest in build-up areas, holiday resorts and golf courts. The analysis therefore includes only plots, classified as even- or uneven-aged forest or as large-scale regeneration areas (in Dutch: "Kapvlakte"). The analysis covers both permanent and temporary plots.

The explanatory variables include both qualitative variables (Albos class) and quantitative variables (e.g., basal area). I used a factor analysis of mixed data (FAMD) which combines a Principal Component Analysis (PCA) on the quantitative and a Multiple Correspondence Analysis (MCA) on the qualitative variables (Kassambara, 2017).

Since multiple variables were based on the same measurements (e.g., DBH) and to limit the number of involved variables, I used a pair plot (Figure A1) to check all pairwise relationships between the variables to investigate the collinearity. When two variables were highly correlated (i.e., pairwise Pearson correlation of > 0.7 see Dormann et al. (2013)), I selected the variable that was the least correlated with other variables. As a result, I chose one DBH-based index for structural diversity: Gini coefficient of basal area over the standard deviation of DBH (SD_DBH) and the coefficient of variation of DBH (CV_DBH). Next I chose two commonly used DBH-based summaries to describe the plot: basal area and quadratic mean diameter over maximum DBH (Max_DBH). I selected the share of basal area per species group (19 species, see Table 2) over dominant tree species, Shannon diversity index and Shannon evenness index, as a variable regarding species composition. Chapter 2 presents an overview the used variables in this first research question (Table 3).

Like a PCA, FAMD draws new axes to explain the variance in the data. After constructing new dimensions, I coloured each observation (sampling plot of the NFI-7) based on the owner group, which eased visual inspection. I did not use owner group as an input variable; I only used it afterwards to investigate possible clusters. For each owner group, the FAMD function calculates the group's mean and a 95% confidence interval ellipse. I visually compared the means with the orientation of the variables to assess differences in owners and correlating variables. In a second step, I coloured the observations based on the site conditions (Albos class), investigating clusters and correlating variables.

PCA and other factor analyses are used to perform a dimensionality reduction. An arbitrary threshold of 70%-90% of the total explained variance is commonly used when the first dimensions are dominating the explained variance and are the obvious source of variation (Jolliffe, 2002). I chose a threshold of 50% on the performance of the first dimensions of the FAMD. I used the packages FactoMineR (Lê et al., 2008) and factoextra (Kassambara et al., 2020) to perform the factor analysis of mixed data and plot the results.

3.2 RESULTS

For the factor analysis of mixed data, 2.143 plots were available, of which 20 were classified as large-scale regeneration areas, 685 as uneven-aged, and 1438 as even-aged. Observations were not equally distributed amongst all owner groups and Albos classes (Figure 4). Total observations were for Large Nature Management organisations: 420; Other Publicly Owned: 477; Private Individuals: 338; Private Organisational Owned: 305; and State Forest Service: 603.

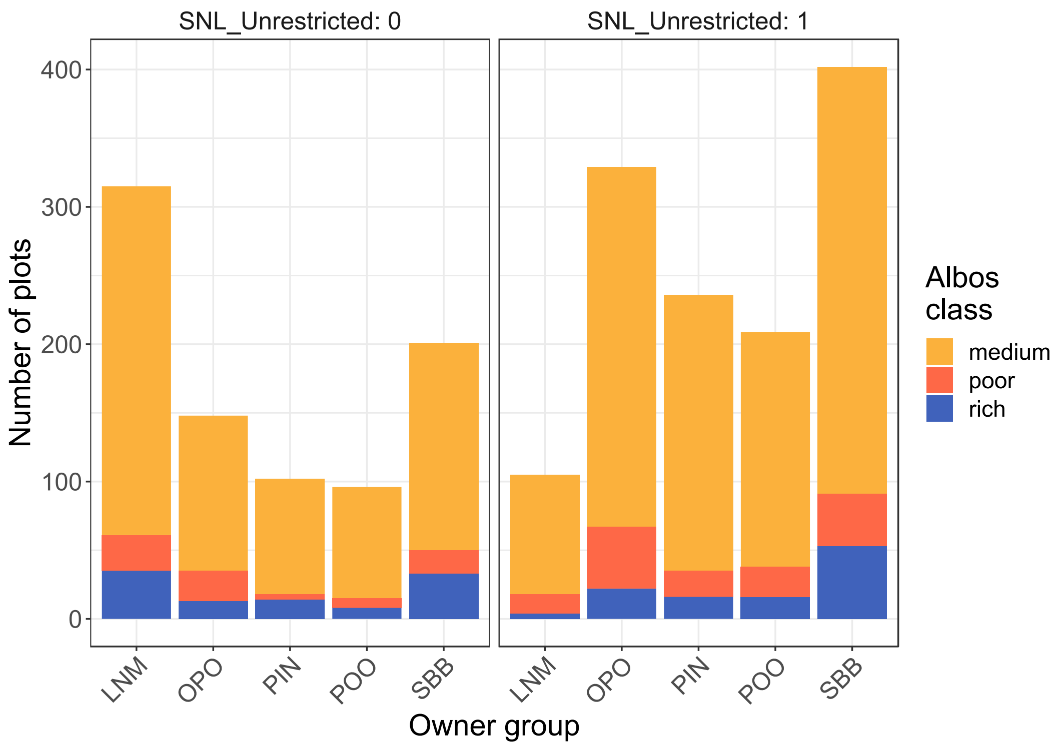

Figure 4. Overview of the observations of NFI-7 as used for the factor analysis of mixed data, grouped by SNL, Albos and owner group: Large Nature Management organisations (LNM), Other Publicly Owned (OPO), Private Individuals (PIN), Private Organisational Owned (POO), and State Forest Service (SBB). SNL_Unrestricted [1] equals forests with SNL 16 (production) or without any SNL-type assigned. SNL_Unrestricted [0] equals forests with all other SNL-types assigned.

The FAMD fitted 33 dimensions, the same number as the used variables. The threshold of 50% explained variance resulted in 12 dimensions (table 4). The table with all dimensions and the (cumulative) explained variance can be found in Appendix B (table A1).

Table 4. Table with eigenvalues for each dimension kept in the factor analysis of mixed data (FAMD). With 12 dimensions, a minimum of 50% of the variance in the NFI-7 data is explained. A table with all dimensions can be found in Appendix B (table A1).

Dimension	*Eigenvalue*	*Variance (%)*	*Cumulative variance (%)*
Dim.1	2.75	8.09	8.09
Dim.2	2.41	7.09	15.18
Dim.3	1.84	5.40	20.58
Dim.4	1.53	4.50	25.08
Dim.5	1.46	4.30	29.38
Dim.6	1.24	3.65	33.03
Dim.7	1.22	3.59	36.62
Dim.8	1.15	3.39	40.01
Dim.9	1.13	3.33	43.34
Dim.10	1.11	3.25	46.59
Dim.11	1.08	3.16	49.76
Dim.12	1.05	3.08	52.84

The first two axes of the FAMD explained 15% of the total variance in the data (table 4). The first dimension correlated most with stem density, and the second dimension correlated most with basal area (Figure 5). The top three contributors to both the first and second dimension are quadratic mean diameter, increment, and stem density (Figure A2). When I include the first 12 dimensions (Figure 6), the contribution to explained variation of species increased (e.g., Scots pine and Douglas fir), compared to the first or second dimension only (Figure 5).

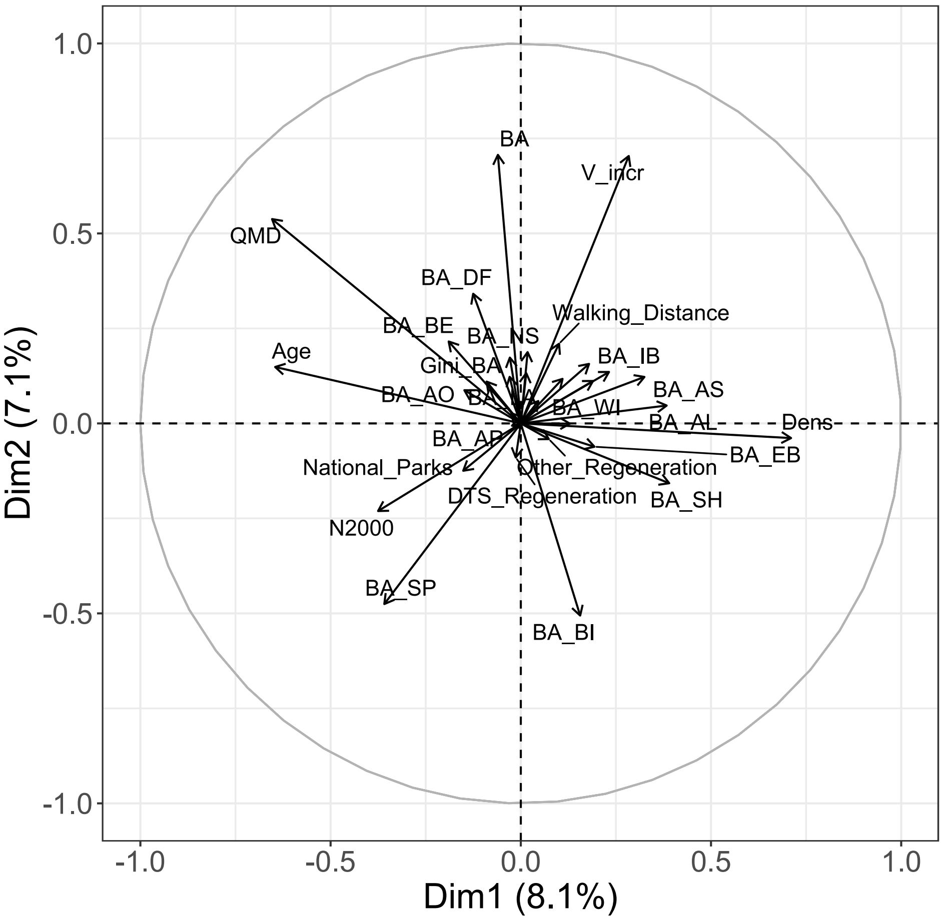

Figure 5. FAMD diagram of the quantitative variables of the factor analysis of mixed data. The first dimension explains 8% of the total variance in the NFI-7 data and the second dimension explains 7% of the total variation. The length and direction of an arrow depict the correlation between a variable and a dimension. Stem density (Dens) contributes the most to the first dimension, and basal area (BA) to the second dimension. For a description of the abbreviated variable names see Table 3.

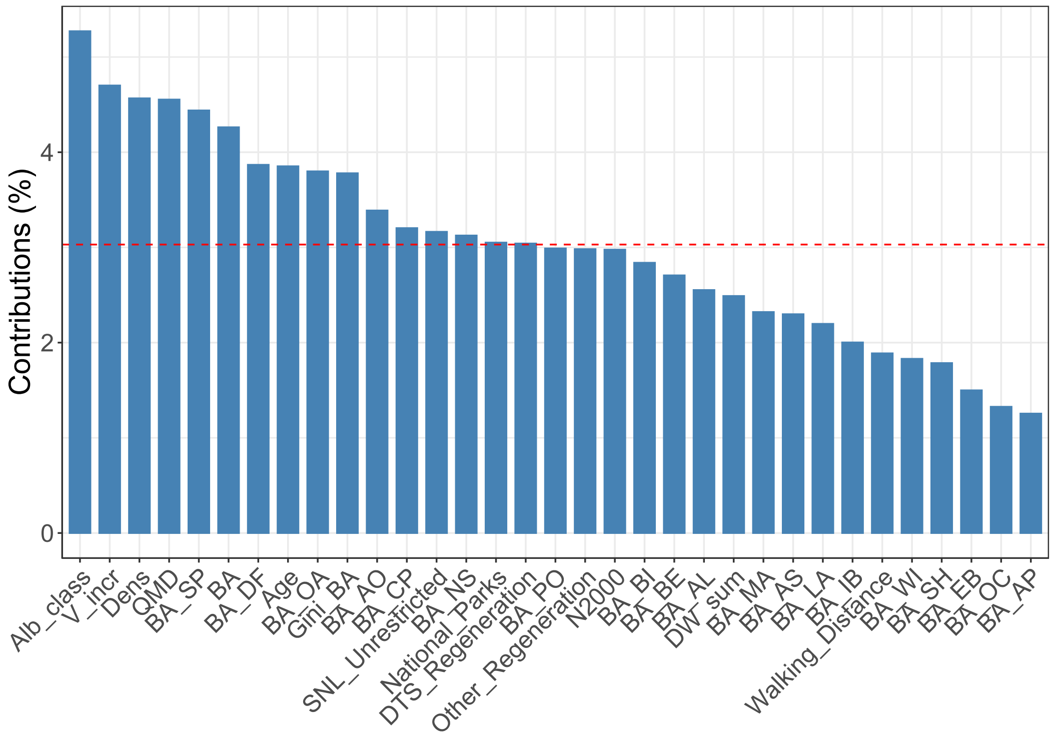

Figure 6. Overview of all 33 variables used in the FAMD and their contribution to the first 12 dimensions (53% explained variance). The red line indicates the expected average if all variables contributed equally.

Structural variables, such as stem density, basal area, and quadratic mean diameter, contributed the most to the first two dimensions (Figure 5). However, Albos class (site conditions) was the most important contributor when assessed over the first 12 dimensions together (Figure 6), followed by increment and stem density (Figure 6). Albos class forms three clearly recognisable clusters (Figure 7).

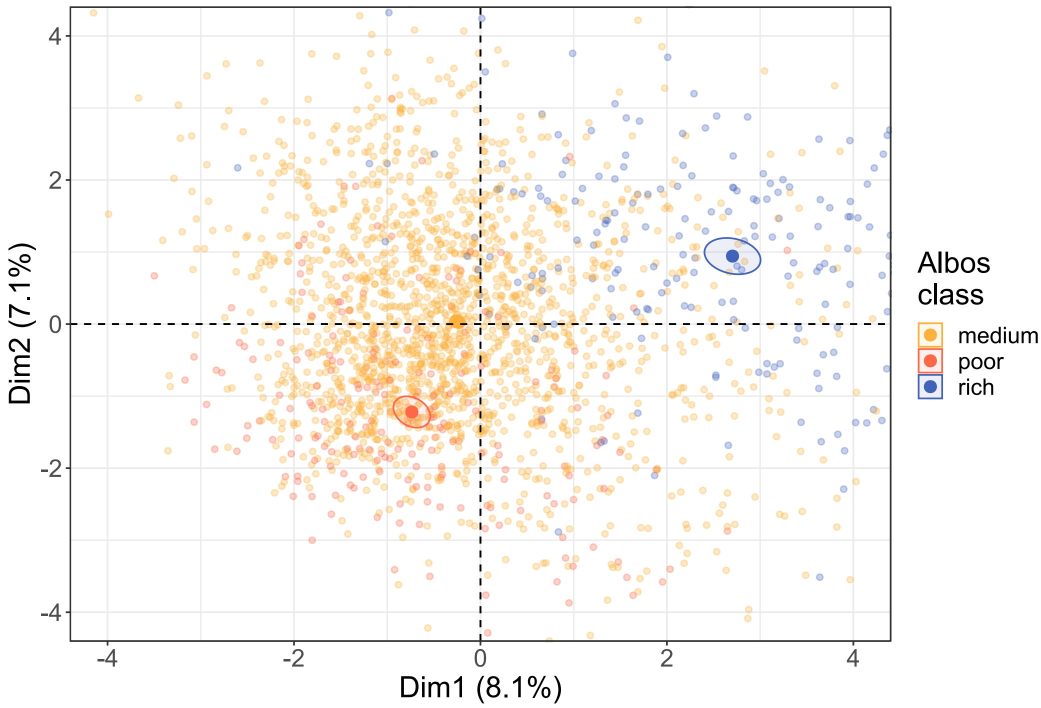

Figure 7. Scatter plot of all individual observations (NFI-7 plots; light dots) in the factor analysis of mixed data. The three darker dots are the mean of each Albos class with their 95% confidence ellipse. For readability this plot is zoomed in; a full version can be found in Appendix B (Figure A3).

On the contrary, no clusters are visible when looking at the five owner groups and their means do not differ greatly (Figure 8). Large Nature Management organisations (LNM) and State Forest Service (SBB) are most dissimilar, while Private Individuals (PIN) and Private Organisational Owned (POO) are located very close to each other (Figure 8). Both private classes have their mean within each other's confidence ellipse, meaning they are highly alike. The owner groups' means are distributed mainly around the first dimension (Figure 8).

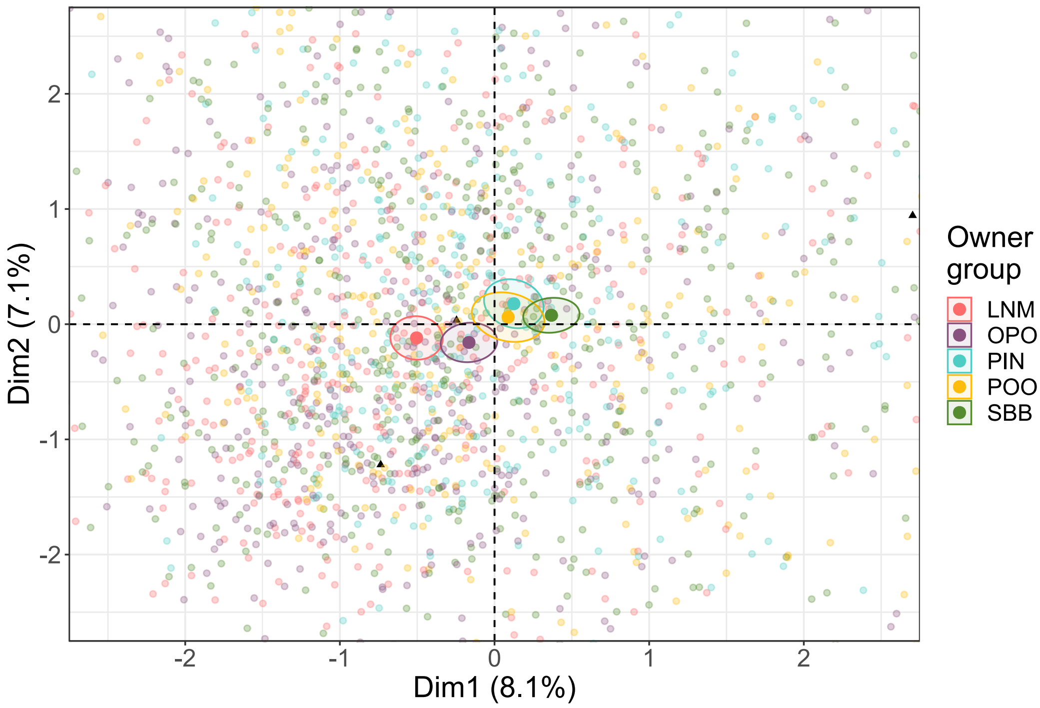

Figure 8. Scatter plot of all individual observations (NFI-7 plots; light dots) in the factor analysis of mixed data. The five darker dots are the mean of each Owner group with their 95% confidence ellipse. For readability this plot is zoomed in; a full version can be found in Appendix B (Figure A4).

3.3 Discussion

The variance in the NFI-7 data seems to be best explained by site conditions (Figures 6 and 7). Albos class does contribute the most to the first 12 dimensions (Figure 6), and the observations are clearly clustered in three groups (Figure 7). In contrast, ownership formed no clear clusters (Figure 8). Owner did not contribute to any dimension, since I used owner group to manually colour the observations. I wanted to see if owners were clustered based on only the structural, compositional, abiotic, and spatial variables, without the effect of owner.

The gradient amongst the mean of owner groups was strongly associated with the first dimension (Figure 8). Stem density on a plot is the major contributor towards the first dimension, followed by quadratic mean diameter and age (Figure 5). However, age and quadratic mean diameter are in the opposite direction compared to stem density; older forests have a lower stem density with thicker trees. Therefore, the first dimension seems to be linked to a forest's development phase, and as a result, SBB seems to own younger forests compared to LNM and OPO.

The gradient amongst owner groups' means was hardly expressed by the second dimension (Figure 8). This dimension's major contributors were basal area, increment, and quadratic mean diameter. Basal area and quadratic mean diameter are related to forest structure and increment depends on the species and fertility (site conditions).

My results suggest a relationship between the occurrence of particular species and the abiotic conditions (Albos class). The share of basal area per species is for specific groups associated with the mean of a particular Albos class: some species occur more at certain site conditions (Figure 5). Most striking is the share of Scots pine on poor sites and the share of alder and ash on rich sites (Figures 5 and 7). Which is in line with Ellenberg (1988): Scots pine is associated with marginal sites, and alder and ash are associated with wet and neutral to alkaline site conditions.

Therefore, site conditions are essential to explain the variance in forest data. Nonetheless, the different owners seem to be characterised by their associated variables. Large Nature Management organisations (LNM) and Other Publicly Owned (OPO) forests are more clustered towards the Natura 2000 areas, national parks, and share of Scots pine. Especially, Scots pine on the poor soils is in historical context understandable. The new forests from the 1850s onwards had a production purpose, and especially conifers were planted (Staatbosbeheer et al., 2021). The state-driven projects from the 1930s were mainly executed with Scots pine to reclaim heathlands (Barends, 2010, p. 147; den Ouden et al., 2010, p. 516). As a result, Scots pine dominates on poor sandy soils (den Ouden et al., 2010, p. 514). The objective for these relatively old forests has been shifting from production towards conservation.

Comparing Figure 5 and 8, indicates that LNM and OPO own the older Scots pine dominated forests. LNM consists of provincial landscape organisations, Natuurmonumenten and Waterleidingduinen. The reason for establishing Natuurmonumenten[e] was to buy and protect already existing nature. Schelhaas et al. (2022) also indicated that nature management organisations (LNM), estates (POO), and other private organisations (POO) manage the oldest forests.

In comparison, based on my results, forests of the State Forest Service are younger, denser, and more often located on richer sites (Figures 5 and 8). Schelhaas et al. (2022) indicate that forests planted after 2000 are relatively often owned by State Forest Service (SBB), Other Publicly Owned (OPO), companies (POO), and estates (POO). The younger forests of SBB are likely to result from the large share of new forests in Flevoland (e.g., Horsterwold). State Forest Service has more than double the number of plots on rich sites (86) compared to other owners (32 on average). A bar graph with observations per owner, province, and site class, is located in Appendix B (Figure A5).

Based on their mean, POO and PIN are remarkably similar, and, especially, POO is close to the plot's centre, meaning their forests are not clustered in a specific direction (Figure 8). Private forest owners are a varied group, and multiple classifications are used in literature, such as economically dependent owners, indifferent owners, or farmers (Ficko et al., 2019). It is, therefore, assumable that this variety reflects in the forests that they own. Less variety is in the forests of LNM and SBB compared to the three other classes.

To conclude, the factor analysis of mixed data indicates that site conditions contribute the most to the variance in the Dutch forest, owners do not own very distinctly different types of forest, and that the differences between owners follow a gradient along the first dimension. Specifically, variables related to the forest's developmental phase contributed the most towards the first dimension.

[e] https://www.natuurmonumenten.nl/ontstaansgeschiedenis, accessed 12-01-2023.

4. PREDICT OWNER GROUPS BASED ON THE DATA OF THE NFI-7

My first research question is: *"How can we describe and classify owners and their forests in the Netherlands based on the single observations of the 7th NFI?"*. I divided the question into a descriptive part (describe the Dutch forest based on the variables, chapter 3) and a predictive part (classify the Dutch forest and predict the owner based on the variables, this chapter).

4.1 METHODS

I used a random forest to predict a plot's owner group since random forests and other machine-learning techniques are frequently used in forest ecology since can handle large datasets, complex interactions, and non-linear data (Liu et al., 2018). To build onto the results of the factor analysis of mixed data, I used the same data and variables from the NFI-7.

I partitioned the data randomly in a training (80%) and a test (20%) dataset. I used the cForest function, which is able to handle both continuous and categorical predictor variables with varying scales, (Hothorn et al., 2006; Strobl et al., 2007; Strobl et al., 2008).

A random forest uses a multitude of randomly generated decision trees, using variables from a random subset. The result of the response variable (i.e. owner group) is the average of the prediction of all trees (Strobl et al., 2007). To balance computational time and predictive power, I set the number of randomly chosen variables to be six and 1000 trees to grow a forest (mtry=6 and ntree=1000, respectively).

After training the random forest, I applied the model to the test dataset, using the predict function from the R Stats Package (R Core Team, 2022). To evaluate the performance of the random forest, I used caret's confusion matrix function (Kuhn, 2022) on the test dataset. Two overall statistics derived from the confusion matrix are the accuracy and the kappa. The accuracy is the percentage of observations correctly classified (i.e., observed accuracy). Kappa corrects for the probability of randomly correct classified observations and enables comparisons between differing datasets:

$$kappa = \frac{(observed\ accuracy - expected\ accuracy)}{(1 - excpected\ accuracy)}$$

Afterwards, I used the vi_model function from the vip Package (Greenwell et al., 2020) to calculate the importance of each variable for the random forest. In general, variable importance for a random forest is calculated by the change in the overall accuracy of the random forest when the variable is dropped from the analysis. High importance means that the variables were essential to predict the true owner group.

4.2 RESULTS

The original NFI-7 dataset consisted of 2.143 plots (Chapter 3). The training dataset had 1.701 and the test dataset 442 randomly drawn plots (Table 5), and the accuracy and kappa were 0.45 and 0.28 after model evaluation (Text box 1). The random forest correctly predicted the owner group of about 45% of the plots. Similar to accuracy, kappa is expressed as a value between 0 and +1.0. A value of 0.28 is often described as poor (Fleiss et al., 1981) or fair (Landis et al., 1977).

The model had the highest sensitivity (true positives) for Large Nature Management organisations (LNM) and State Forest Service (SBB), meaning that it was better at correctly predicting these classes (Text box 1). The sensitivity of POO was 0.00: it did not correctly predict any true POO plots (Text box 1). Sensitivity is the true positive rate, whereas specificity is the true negative rate. POO has a high specificity since almost all reference observations that were not POO were also predicted as not POO; only one reference observation that was SBB was predicted as POO Text box 1).

Table 5. An overview of the number of observations per owner for the train and test datasets. The abbreviations for owner groups are Large Nature Management organisations (LNM), Other Publicly Owned (OPO), Private Individuals (PIN), Private Organisational Owned (POO), and State Forest Service (SBB).

Owner Group	*Training*	*Test*	*Original*
LNM	327	93	420
OPO	384	93	477
PIN	262	76	338
POO	251	54	305
SBB	477	126	603
Total	1.701	442	2.143

Confusion Matrix and Statistics – test data

\ Prediction	Reference LNM	OPO	PIN	POO	SBB	
LNM	**62**	17	19	14	17	129
OPO	13	**41**	15	9	22	100
PIN	1	8	**14**	3	5	31
POO	0	0	0	**0**	1	1
SBB	17	27	28	28	**81**	181
	93	93	76	54	126	

```
Overall Statistics:
             Accuracy : 0.448
               95% CI : (0.401, 0.4957)
                Kappa : 0.2755

Statistics by Class:
                 LNM       OPO        PIN       POO          SBB
Sensitivity     0.6667    0.44086    0.18421   0.000000     0.6429
Specificity     0.8080    0.83095    0.95355   0.997423     0.6835
```

Text box 1. Confusion matrix of the random forest applied on the test dataset. Reference refers to the (true) observations from the test dataset (Table 5) while prediction refers to the prediction of the model. In bold cases in which the random forest correctly predicted the owner group. For abbreviations of the owner group, see to Table 1.

The three most important variables were SNL-type, Natura 2000, and National Parks, which are all spatial variables (Figure 9). The variable importance plot shows multiple groups of importance. SNL-type stands out (importance of 0.038), followed by Natura 2000 (0.013). Importance of National Parks, the basal area of Scots pine, age, and the basal area of exotic broadleaves are all around 0.009. An important exotic broadleaved species is black cherry (*Prunus serotina*), covering 83% of the number of trees in the group of exotic broadleaves . The other variables have little importance, and the last two have negative importance. Negative importance means that the variable is not predictive enough to classify the owner group.

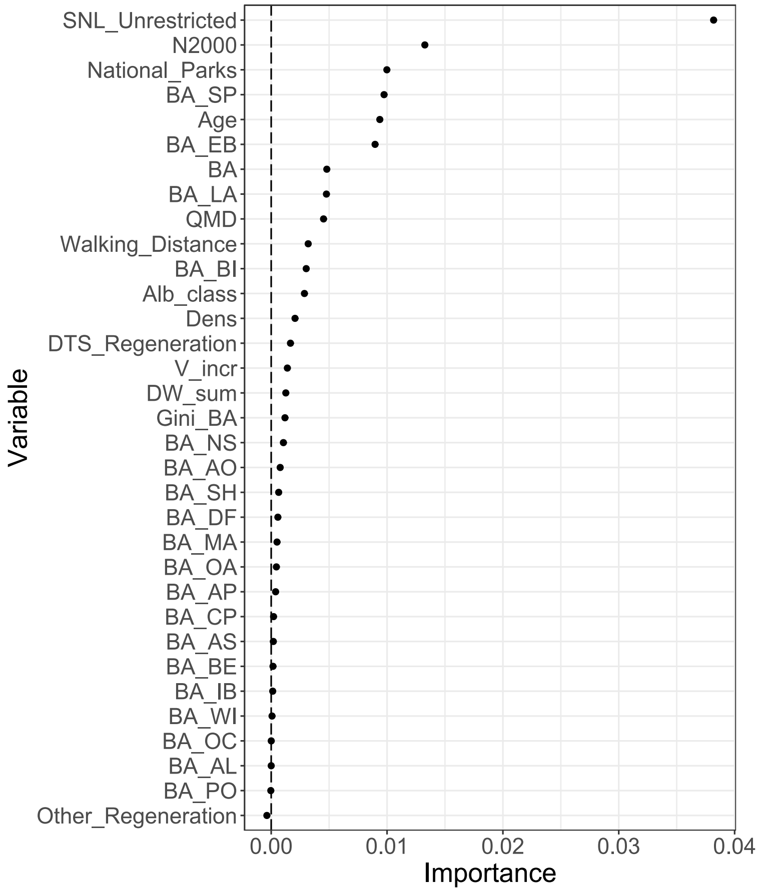

Figure 9. Variable importance plot of the random forest, using the cForest function. For a description of the abbreviated variable names, see Table 3.

4.3 Discussion

The random forest was most sensitive towards forests of LNM and SBB, with OPO as third (Text box 1). Both private classes (PIN and POO) were difficult for the random forest to recognise. Private owners might be too similar to other owner groups, too diverse in owner composition, or might be underrepresented in the training dataset. The latter is always a hurdle when working with imbalanced datasets. Nonetheless, the results indicate certain predictability of LNM, SBB, and OPO owners. Disregarding non-important variables of no interest might improve the random forest.

Accordingly, to the variable importance plot, the top three were spatial variables: SNL-type, Natura 2000, and National Parks (Figure 9). Moreover, National Parks, the share of Scots pine and exotic broadleaves, and age are a cluster of variables of equal importance. This cluster could be a pattern in the data between the age of the plot and the predictability of specific owners. The Dutch reforestation plans of the 20th century used black cherry as admixture species of Scots pine plantations (Nyssen et al., 2016). Therefore, Scots pine dominated forests with black cherry are associated with (an older) age.

Interestingly, the three spatial variables are no physical description of the forest structure or species composition but what humans expect from or aim for these forests. The importance of National Parks is unexpected; however, it could be linked to the characteristics of a decision tree.

The dataset's number of observations per owner group was not equally distributed. Specifically, when I split the original dataset per spatial variable, I see that some variables occur more often in a particular owner group than others (Table 6). For example, only around 10% of the train data was in a National Park. However, since LNM and SBB had more observations in National Parks compared to other groups, the probability of an observation from the test dataset being LNM or SBB increased when it had the attribute of being in a National Park. That might be why these spatial variables are essential for the random forest to predict the owner correctly.

Table 6. An overview of the number of observations per owner for the original NFI-7 dataset (see chapter 3) for the three most important variables. These three variables are interestingly spatial. It is visible that some combinations occur more frequently. The rows do not add up since the data is split up per variable (i.e., n=1701 per variable). A graphical representation is located in Appendix C (Figure A6).

	SNL Unrestricted		*N2000*		*National Parks*	
Owner	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
LNM	315	105	201	219	327	93
OPO	148	329	292	185	466	11
PIN	102	236	288	50	318	20
POO	96	209	229	76	278	27
SBB	201	402	395	208	521	82

Concluding, the forests from Large Nature Management organisations and State Forest Service are the most predictable, indicated by a higher sensitivity of the random forest. Both private owner groups were complex to classify for the random forest, indicating considerable similarities with other owners. Furthermore, three spatial variables were most important (SNL-type, N2000, and National Parks) for the random forest to predict the owner. Not forest structure or composition but what people have as aim for the forest were most important. With the limited accuracy and low power of predicting the private classes, it should be kept in mind that this method had a data mining aim: looking for patterns in the NFI-7 data and not for robust hypothesis testing. Predicting the owner group is possible and the variable importance adds a better understanding of potential relationships between owner groups and variables.

5. INVESTIGATE HARVEST INTENSITY BASED ON THE REPEATED INVENTORIES

My second research question is: "What are the differences and patterns of harvest intensity observed in the Netherlands using the repeated NFI data, and how can the measured variables explain these patterns?".

5.1 METHODS

To limit the model's complexity, instead of using data from MFV and NFI-6, I only used the NFI-6 data and harvest intensity between NFI-6 and NFI-7 (HI_sum). I still used the harvest intensity between MFV and NFI-7, not as the dependent variable but as a covariate in the model (HI_sum.MFV).

5.1.1 Data harmonisation and data selection

Since I used multiple inventories for the second research question, I needed to harmonise, select a subset of, or aggregate the data:

- In 11 cases, the owner group was the same in MFV and NFI-7 but different in NFI-6. I assumed, for these cases, the owner group from the MFV and NFI-7.
- Since exact historical data was not publicly available, I assumed that Walking distance, Albos class, and Natura 2000 were static and adopted the NFI-7 status for the NFI-6.
- I selected only plots that were measured during all three inventories (MFV, NFI-6, and NFI-7) and were indicated as even- or uneven-aged forests or as large-scale regeneration areas (in Dutch: "Kapvlakte").
- I removed the Albos class as a variable during the analysis of harvest intensity because it was very unequally distributed (84 poor, 723 medium, and 80 rich sites).
- To avoid parametrisation errors for dominant tree species (table 2) with only a few observations, I reclassified dominant tree species with ten or fewer observations as 'rare species' (RS). These species were Alnus (Al), Exotic Broadleaves (EB), Harvest Areas (HA), Indigenous Broadleaves (IB), Maple (MA), Other Conifers (OC), and Willow (WI).

5.1.2 Data analysis

5.1.2.1 Variable selection

I used the data exploration protocol of Zuur et al. (2010) to check for collinearity, variable interaction, and homogeneity of variance as the basis for variable selection. As a result, I again used a pair plot to check all pairwise correlations (see Chapter 3).

Furthermore, I transformed right-skewed variables. I applied either a Log(x+1) transformation to deal with zeros or a square root when I needed a weaker transformation, as proposed by Warton (2022, pp. 38-40). First, I applied the log transformation and visually checked if it was likely that the data were normally distributed. If that was not the case, I used a square root transformation. Variables that I log transformed were: QMD, stem density, dead wood, and increment. I applied the square root transformation to the basal area and standing volume.

I did not transform the Shannon Diversity Index and the Gini coefficient of basal area since they were not right-skewed or seemed bimodally distributed. I compared the results of a Shapiro-Wilk test for normality to compare the transformed with untransformed covariates and selected the one with the highest p-value. HI_sum.MFV had the same p-value as the transformed version. Hence I chose the untransformed one.

Next, I used all (transformed and original) continuous variables in a pair plot to check collinearity (Figure A7). As a result, I selected QMD_t over Density_t (pairwise Pearson correlation of 0.79) since QMD_t was less correlated with other variables and followed a normal distribution (Shapiro-Wilk test for normality, W=0.99748, p-value = 0.1922). Since basal area (m^2 ha^{-1}) and standing volume (m^3 ha^{-1}) were highly correlated (Pearson correlation >0.95), I selected measured data (basal area) over derived data (standing volume). Transformations did not change the collinearity. I used harvest intensity after NFI-6 as the dependent variable for the full model. Dominant tree species, owner, SNL-type, Natura 2000, Walking distance, transformed quadratic mean diameter, transformed basal area, transformed volumic increment, Gini coefficient of basal area, Shannon Diversity Index, transformed volume of dead wood, and the harvest intensity between MFV and NFI-6 as independent variables (Text box 2).

To check for collinearity across both categorical and continuous variables, I used the corvif function from the HighstatLibV13 R-script of Zuur et al. (2021). I used the $GVIF^{\frac{1}{2*df}}$ values and by squaring these, I get values that are comparable to that of a variance inflation factor (VIF). VIF > 5 indicated collinearity (Zuur et al., 2009), which was the case.

Subsequently, I visually inspected possible variable interactions with coplots, as suggested by Zuur et al. (2010). I used transformed QMD (QMD_t), as it was normally distributed, on the x-axis and harvest intensity (HI_sum) on the y-axis. I used the owner group and one categorical variable (i.e., N2000, SNL_Unrestricted, Walking_Distance, DTS) to create a scatterplot to see if the relationship changed per owner per category. As a result, I only applied an interaction between the owner group and the SNL-type.

Since plots were not measured twice or had a different kind of nested design, I assumed that the observed plot-level harvest intensities of the NFI-6 were independent of each other (Zuur et al., 2010). Unfortunately, I could not use the time between inventories as an offset term. The offset corrects for the differences in observations and requires a Log link distribution family, whereas the beta distribution is a Logit link.

I discussed other methods with my supervisors to correct for the time between inventories. First, standardizing harvest intensity per year, which changes a harvest event into a continuous variable. Second, the time between inventories as a covariate, which influences the results in unwanted ways (since it is not directly affecting harvest intensity but the probability of finding a harvest). Third, select only observations that have a minimum time between inventories that corresponds to the average Dutch harvest frequency, which has the probability of covering multiple harvest events. Since the time between inventories was slightly correlated with harvest intensity and from NFI-8 onwards, each plot is measured every five years; we decided to assume that time between inventories had a neglectable effect

5.1.2.2 Fitting the model

Since harvest intensity is continuous proportion data and 60% of the observations were zeros, I used a zero-inflated beta distribution regression. A beta distribution is suitable for applying a regression analysis to proportion data and provides less biased estimators when comparing it to transforming the data (Douma et al., 2019). The function to fit the model could not handle 1 in the data; thus, I reclassified the value 1 as 0.9999999 as a workaround (Douma et al., 2019). I used the glmmTMB function from the similarly named package (Brooks et al., 2017) since it handles the beta family as a distribution and a specified zero-inflation formula.

The general approach of my model fit was to create a full model with all selected variables (see variable selection) and perform a feature selection based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Sheather, 2009, p. 236). Compared to the AIC, the BIC penalises a model's complexity more, and models with a lower AIC and BIC are assumed to be better (Sheather, 2009, pp. 230-233). A model is significantly better if the difference in AIC (ΔAIC) ≥ 2 (Burnham et al., 2004). A commonly used approach in model selection is selecting a model which minimises AIC and BIC and maximises R^2 (Sheather, 2009, p. 233).

The glmmTMB function consists of a conditional part, modelling harvest intensity, and a zero-inflated part, modelling the occurrence of zeros in the dependent variable. To check which type of zero-inflated formula performed best, I started with a full variable model (Text box 2) in the conditional part and three different zero-inflation formulas:

1) a model without zero-inflation (ziformula = ~0),
2) a model with zero-inflation (but the probability of zero was everywhere the same (ziformula = ~1), and
3) a model in which zero-inflation had the same formula as the conditional model (i.e., a full model for zero-inflation).

The first model had the lowest AIC and BIC, but residuals and R^2 performed worse. The second and third models had better residuals and R^2. The third model had the lowest AIC and was my choice for the backward feature selection.

```
HI_sum ~ DTS + Owner_Group × SNL_Unrestricted + N2000 + Walking_Distance +
QMD_t + BA_t + Increment_t + Gini_BA + ShannonIndex + DW_t + HI_sum.MFV
```

Text box 2. The full model formula has harvest intensity (HI_sum) as the dependent variable and the other variables as covariates. The multiplication term (×) denotes interaction between variables. Covariates ending with a "_t" are transformed variables. Table 3. explains the abbreviations. HI_sum is the harvest intensity after NFI-6 and HI_sum.MFV is the one after MFV.

I used the stepAIC function from the MASS package (Venables et al., 2002) to perform a feature selection on the conditional part in both forward and backward directions. This approach resulted in the variable HI_sum.MFV to be nonsignificant, which I manually removed to compare the AIC and BIC. The ΔAIC was < 2. Hence, both models performed equally. However, based on the principle of parsimony, I selected a model with fewer variables (Burnham et al., 2002, p. 443). Moreover, this model resulted in a significantly lower BIC (ΔBIC=14.4).

The stepAIC function did not work on the zero-inflation part of the model. Therefore, I manually performed a backward feature selection for the zero-inflation formula. In each step, I removed the variable with the highest p-value and compared the AIC and BIC with the compare_performance function from the performance package (Lüdecke et al., 2021). The ΔAIC was not always larger than two. At each step, I selected the simplest model because of parsimony, and the BIC favoured the model without the variable. I repeated these steps until only significant covariates remained.

The last four models (m.minN2000, m.minWD, m.minShan, m.minQMD/m.final. see Table A2) had an ΔAIC within the range of two, making them not distinguishable as different. Removing QMD_t increased the AIC, but ΔAIC<2 and BIC was in favour. Based on BIC and the mantra of parsimony, I selected the final model without QMD_t. Table A2. provides an overview of the AIC and BIC during the backward selection for the zero-inflated part.

I checked the model assumption of homogeneity of variance by plotting the residuals and fitted values amongst each used and removed covariate from both the conditional and zero-inflated part (Zuur et al., 2010; Zuur et al., 2016). I used the DHARMa package (Hartig, 2022) for residual analysis. After fitting the model, I explored the influence of each variable from the final model by visualising the fitted model, as suggested by Zuur et al. (2016). I used ggeffects (Lüdecke, 2018) and sjPlot (Lüdecke, 2022) to visualise the fitted model.

5.2 RESULTS

For the zero-inflated beta distribution model (ZIBD), 887 plots were available, of which five were classified as large-scale regeneration areas, 663 as even-aged, and 219 as uneven-aged during NFI-6. The Large Nature Management organisations (LNM) had 192 observations, Other Publicly Owned (OPO) 192, Private Individuals (PIN) 124, Private Organisational Owned (POO) 107, and State Forest Service (SBB) had 272. The observations differed per species and owner group, but Scots pine was the predominant species (Figure 10). Model validation indicated no crucial problems (see Figures A8 - A10 for residual plots).

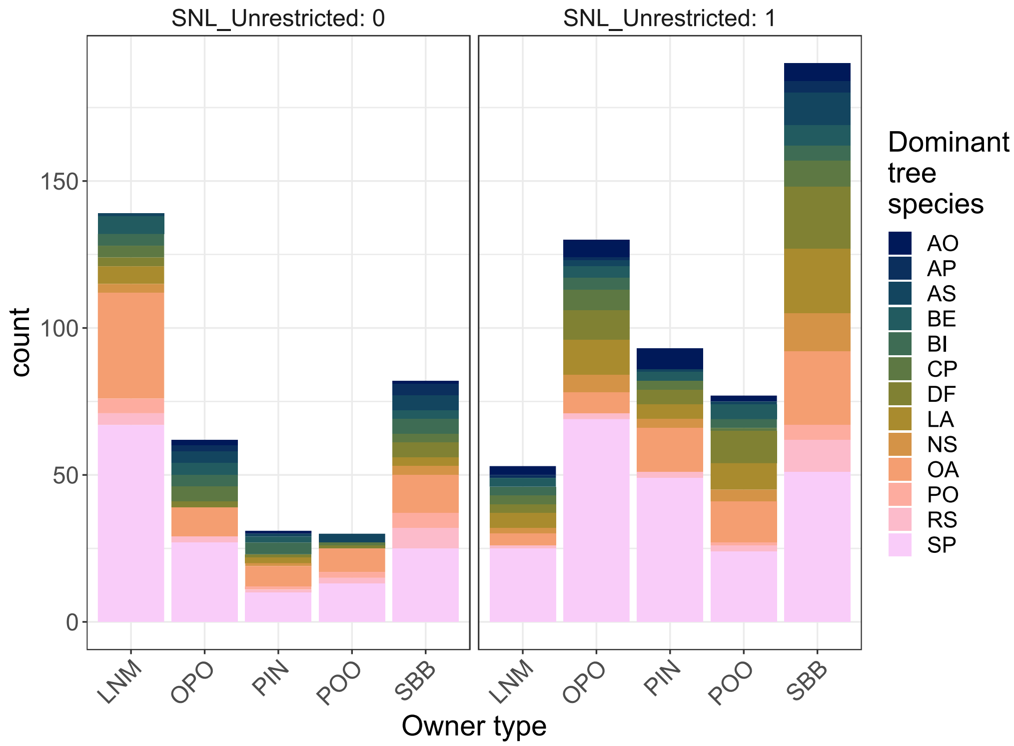

Figure 10. Overview of the observations of NFI-6 used for the model. The observations are coloured by dominant tree species and grouped per owner group and SNL-type. The abbreviations for species are American oak (AO), Austrian pine (AP), ash (AS), beech (BE), birch (BI), Corsican pine (CP), Douglas fir (DF), larch (LA), Norway spruce (NS), indigenous oak (OA), poplar (PO), rare species grouped (RS), and Scots pine (SP). SNL_Unrestricted [1] refers to forests with SNL 16 (production) or without any SNL-type assigned. SNL_Unrestricted [0] refers to forests with all other SNL-types assigned.

The final model's conditional part had dominant tree species, SNL-type, and increment as significant covariates (Table 6). The zero-inflation part had owner group, SNL-type, increment, and volume of dead wood as significant covariates (table 6). Both formulas had no interaction effect.

5.2.1 Effect of variables

StepAIC dropped the covariate owner group from the conditional part of the model (table 6). The intercept for the conditional part (DTS [AO], SNL_Unrestricted [0], other covariates as 0) was significant (z = -3.981, p<0.001). Dominant tree species ash significantly positively affected harvest intensity (estimate = 1.763, z = 3.832, p<0.001), which is the same for dominant tree species poplar (estimate = 2.894, z=4.984, p<0.001). SNL N16 and no SNL-type assigned (SNL_Unrestricted [1]) had a significant increase in harvest intensity (estimate = 0.428, z=2.832, p<0.01). The transformed increment also increased harvest intensity significantly (estimate =0.438, z=2.134, p<0.05).

The intercept for the zero-inflated part (Owner_Group [LNM], SNL_Unrestricted [0], other covariates as 0) was significant (estimate=2.27, z=5.613, p<0.001). The owner group Other Publicly Owned (OPO) was significant in lowering the occurrence of zero (estimate -0.579, z=-2.498, p<0.05). SNL-type N16 and no SNL assigned (SNL_Unrestricted [1]) decreased the zero-occurrence (estimate = -0.811, z=-5.016, p<0.001). Transformed increment is also associated with a significant decrease of the zero-occurrence (estimate -0.612, z=-3.445, p<0.001). Contrarily, the volume of dead wood is associated with an increase of zero-occurrence (estimate = 0.116, z=2.048, p<0.05).

Table 6.Summary table of the fitted zero-inflated beta distribution (ZIBD) model. The model consists of a conditional part, modelling the harvest intensity (HI_sum) and a zero-inflation part, modelling the presence of harvest. The conditional part uses an intercept of (DTS [AO] and SNL_Unrestricted [0]. The zero-inflation part an intercept of Owner_Group [LNM] and SNL_Unrestricted [0].

Family:	beta (logit)
Formula:	HI_sum ~ DTS + SNL_Unrestricted + Increment_t
Zero inflation:	~ Owner_Group + SNL_Unrestricted + Increment_t + DW_t

Predictor	*Estimate*	*Std.*	*Error*	*z value*	*Pr(>|z|)*
Conditional model:					
(Intercept)	-2.286387	0.574322	-3.981	**6.86E-05**	***
DTS [AP]	-0.15667	0.673179	-0.233	0.81597	
DTS [AS]	1.762626	0.459925	3.832	**0.000127**	***
DTS [BE]	0.243321	0.478002	0.509	0.610726	
DTS [BI]	0.761292	0.629557	1.209	0.226567	
DTS [CP]	0.004702	0.445517	0.011	0.991579	
DTS [DF]	0.512214	0.416477	1.23	0.218745	
DTS [LA]	0.677466	0.405463	1.671	0.094753	.
DTS [NS]	0.735872	0.432078	1.703	0.088549	.
DTS [OA]	0.184733	0.386182	0.478	0.632395	
DTS [PO]	2.893908	0.580646	4.984	**6.23E-07**	***
DTS [RS]	0.30424	0.49431	0.615	0.538235	
DTS [SP]	0.63587	0.355132	1.791	0.073371	.
SNL_Unrestricted [1]	0.427681	0.151003	2.832	**0.004622**	**
Increment_t	0.438275	0.205393	2.134	**0.032856**	*
Zero-inflation model:					
(Intercept)	2.27097	0.40461	5.613	**1.99E-08**	***
Owner_Group [OPO]	-0.57917	0.23184	-2.498	**0.012484**	*
Owner_Group [PIN]	-0.24462	0.2624	-0.932	0.351214	
Owner_Group [POO]	-0.10933	0.27382	-0.399	0.689678	
Owner_Group [SBB]	-0.2914	0.22192	-1.313	0.189158	
SNL_Unrestricted [1]	-0.81087	0.16165	-5.016	**5.27E-07**	***
Increment_t	-0.61168	0.17754	-3.445	**0.000571**	***
DW_t	0.11628	0.05677	2.048	**0.040529**	*
Signif. codes:	0.001 < '***'	0.01 < '**'	0.05 < '*'	0.1 < '.'	

AIC	BIC	logLik	deviance	df.resid
950.5	1065.4	-451.3	902.5	878

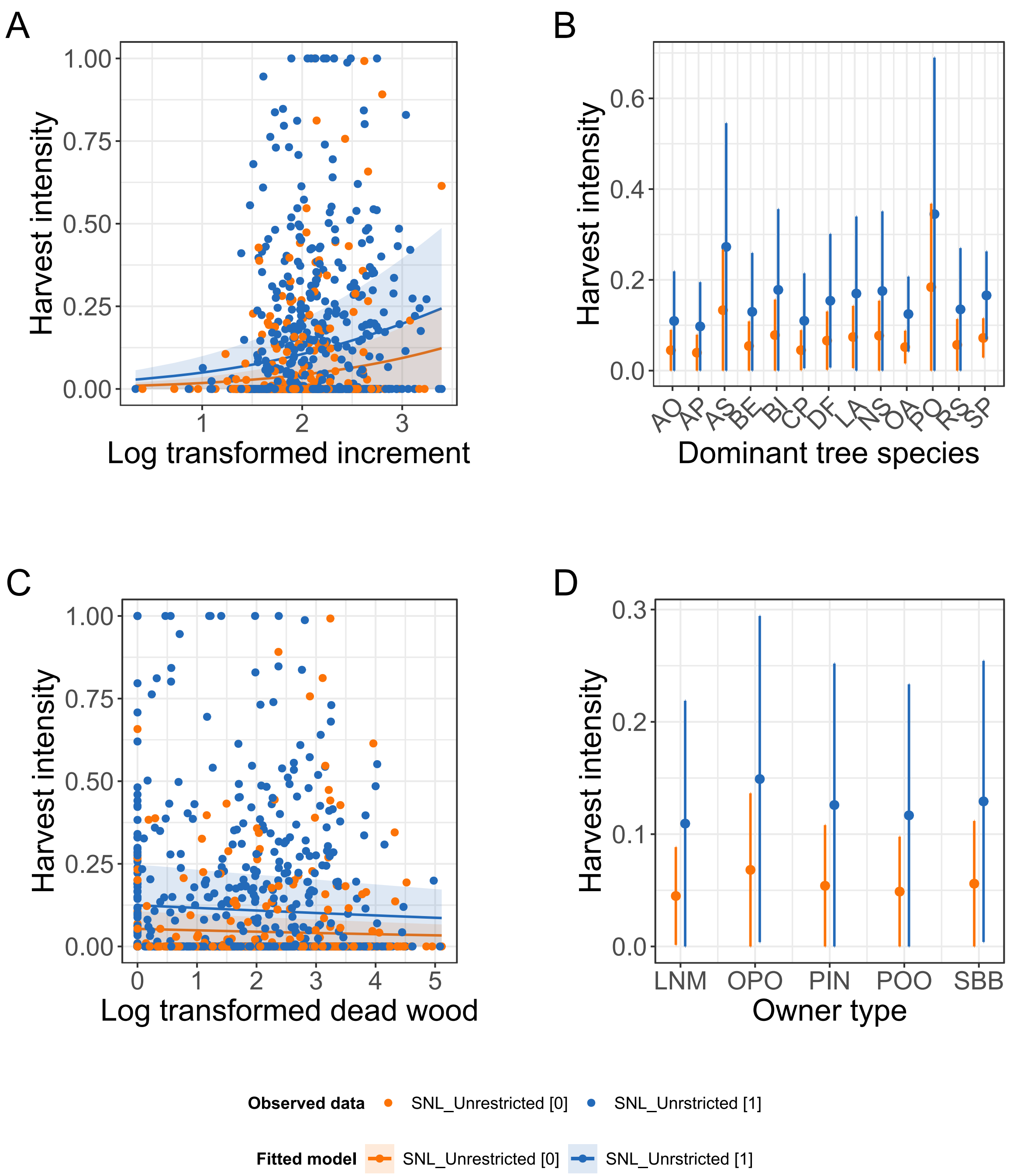

Figure 11. Visual representation of the fitted zero-inflation beta distribution model for harvest intensity between NFI-6 and NFI-7. I visualised model predictions of harvest intensity by changing one variable while keeping the others on the same level. Orange represents the SNL_Unrestricted [0], and blue is the SNL_Unrestricted [1]. A) The fitted model (line) is plotted with a 95% confidence interval, and the dots are the 887 observations. B) Modelled harvest intensity per dominant tree species with their error bars (95% CI). Species' names and the number of observations are American oak (AO; n=28), Austrian pine (AP; n=12), ash (AS; n=29), beech (BE; n=37), birch (BI; n=32), Corsican pine (CP; n=36), Douglas fir (DF; n=62), larch (LA; n=64), Norway spruce (NS; n=35), indigenous oak (OA; n=139), poplar (PO; n=19), rare species grouped (RS; n=34), Scots pine (SP; n=360). C) The fitted model (line) is plotted with a 95% confidence interval, and the dots are the observed data. D) Modelled harvest intensity per owner group with their error bars (95% CI). Owner groups and the number of observations are Large Nature Management organisations (LNM; n=192), Other Publicly Owned (OPO; n=192), Private Individuals (PIN; n=124), Private Organisational Owned (POO; n=107), and State Forest Service (SBB; n=272).

The influence of the different variables is visualised by making model predictions of harvest intensity by varying the variable on the x-axis while keeping the other variables at the same value (Figure 11). Dominant tree species, SNL-type, and transformed increment influenced the conditional part (Figures 11A and 11B). Owner group, SNL-type, transformed increment, and transformed volume of dead wood influenced the zero-inflation part (Figures 11A, 11C, and 11D).

Concluding, harvest intensity was higher in plots with SNL_Unrestricted [1] (either forest with SNL production or no SNL assigned) compared to SNL_Unrestricted [0]. Plots with dominant tree species ash or poplar seemed to have a higher harvest intensity. An increase in increment is associated with an increase in harvest intensity. The effect of the owner was class-dependent. On the contrary, an increase in the volume of dead wood is associated with a decreased harvest intensity.

5.3 Discussion

The main question was which variables are linked to harvest intensity in the Dutch NFI data, and it was of special interest what the differences between owners were.

The main result is that the owner group does not seem to affect harvest intensity significantly, but it does affect the presence of harvest. The differences between owner groups seem to be limited, and only Other Publicly Owned (OPO) forests appeared to be harvested more frequently than the intercept (table 6). Since the residuals of OPO were not uniformly distributed (Figure A10), the OPO group should be analysed with caution. The other three owner groups (PIN, POO, and SBB) do not differ significantly from the intercept (LNM) (table 6), meaning their harvest presences are comparable. Their residuals indicated no problems. Figure 11D shows the modelled harvest intensity, which shows the same patterns; OPO stands out, while the others are in the same range. The Dutch owners, therefore, seem to be harvesting their forests with the same intensity.

In general, plots with SNL-type production or no SNL (SNL_Unrestricted [1]) are modelled to be more intensively harvested. The SNL-type is based on the information from the IMNa[f] (Informatiemodel Natuur) and describes different nature types and subsidy schemes. The Dutch nature subsidies, policies, and management revolve around these management types. We do not know if actual subsidies are paid out, but owners seem to follow the assigned SNL-type, which is in line with Schelhaas et al. (2018a).

Only two dominant tree species significantly increased harvest intensity: ash and poplar. The ash dieback had its effects over the last decade, and harvesting affected trees is one of the recommendations (Siebel et al., 2018). Therefore, this disease could probably explain the relatively high harvest intensity of ash-dominated forests after NFI-6 (2012-2013). The relatively high harvest intensity amongst poplar forests could be due to the harvest of temporary forests. The subsidy scheme for fast-growing temporary forests of poplar and willow in the period 1985-1994 required a final felling after 25 years (Jansen, 2004).

The volume of dead wood is associated with a significant decrease in the presence of harvest (table 6). Forest managers have been changing their management; as a result, more dead wood remains in the forests (den Ouden et al., 2010, pp. 425-435). Also, Schelhaas et al. (2022) report an increase in dead wood over the different inventories, from 10 $m^3\ ha^{-1}$ at the MFV to 19.2 $m^3\ ha^{-1}$ at the NFI-7. The difficulty is that it is hard to disentangle cause and effect. A forest with a high volume of dead wood might be the reason for a forest manager to harvest less intensively, or since a forest is harvested less intensively, more trees could die of competition or age and increasing the volume of dead wood. The volume of dead wood is associated with the age and stage of a forest (Wijdeven, 2006); the volume of dead wood could be a proxy for age or plots with closer-to-nature forest management. The slope of the line is small (Figure 11C), meaning that the association with harvest intensity is small.

Moreover, the increment is associated with a significant increase in harvest intensity, but just as for the volume of dead wood, a correlation does not mean a causal relationship. The increment is based on species and site conditions. The site conditions determine which species are planted and, presumably, if the forest's aim is nature (poor sites) or production (fertile sites). However, the older forests seem to have a lower increment than younger forests (Figure A11), which could be due to the forests' maturing or the history of reforestation programmes. Nabuurs et al. (2013) also suggested that Europe's increment decrease is attributed to, for example, maturing forests.

Other research, such as Hengeveld et al. (2012), often uses Natura 2000 to predict management. My analysis indicates that Natura 2000 might not be as useful, and nature type (SNL-type) seemed to be better linked

---

[f] https://www.bij12.nl/onderwerpen/natuur-en-landschap/natuurgegevens-uniform-uitwisselen-imna/informatiemodel-natuur-imna/ accessed 11-01-2023

to harvest intensity. Nonetheless, further research should use Natura 2000 but with its specific habitat types within the Natura 2000 protection[g]. My analysis used Natura 2000 as a binary variable, probably averaging the effect of habitat type.

Lastly, to make comparisons amongst the different plots and owners possible, I needed to standardise harvest within a plot. I chose for proportion removed basal area, which is highly comparable to the proportion removed standing volume (Figure A13). However, harvest intensity could be the same for two different amounts of (volumic) harvest since the standing stock was also different. This standardisation could imply that I lost possible important differences or nuances in the data.

In conclusion, the owner group was affecting the presence of harvest. Especially, other publicly organisations harvested more frequently, yet between the other owner groups were little differences. SNL-type N16 (production) and plots without SNL-type assigned are associated with a higher harvest intensity. Ash and poplar showed significantly higher harvest intensity compared to the other species. The increment and volume of dead wood are associated with an increase and a decrease in harvest intensity, respectively. However, causality between these variables and harvest intensity is difficult to prove and could have an underlying connection with the forest's age.

[g] https://www.natura2000.nl/profielen/habitattypen accessed 02-02-2023

6. DISCUSSION

This thesis' aim was to investigate the differences between owners and their forest management, specifically harvest. I used data from the NFI-7 to describe (Chapter 3) and to predict the owners (Chapter 4). I used data from the NFI-6 to analyse the effects of different variables on harvest intensity, with a special focus on the owner group. This chapter will focus on the similarities and discrepancies between the three analyses and will end with recommendations and implications for research, management and society.

6.1 DESCRIBE AND PREDICT OWNERS

With my first question, I tried to explain the variation in the Dutch forest and see if different owners also have different forests. I expected that past forest management would result in differences in forest structure and species composition. I hypothesised that owners were clustered mainly on a production ↔ nature-oriented axis (with associated variables such as commercial species and amount of dead wood) and, to a lesser extent, on a professional–private axis.

I expected these axes, based on research about the NFI-6, which showed a difference in management between Private Individuals on the one hand and nature management organisations and State Forest Service on the other (Schelhaas et al., 2018a). Furthermore, Schelhaas et al. (2018a) indicated that harvest is higher in production-oriented forests (SNL_Unrestricted [1]) compared to the nature-oriented forests (SNL_Unrestricted [0]); nonetheless, this difference was negligible for Private Individuals. I did not find a clear professional ↔ private axis or a production ↔ nature-oriented axis between owner groups.

The main result of the factor analysis of mixed data (Chapter 3) was that site conditions (Albos class) explains the variance in the NFI-7 data the best (Figure 6 and 7). Furthermore, the differences between owner groups are mainly correlated to the first dimension, which seems mostly associated with the forests' development phase (Figures 5 and 7). Both private owner groups were highly similar, and State Forest Service and the Large Nature Management organisations were the most distinctive, which is in line with the main results of the random forest (Text box 1). In addition, spatial variables were the most important for the random forest, followed by variables associated with the forest's development phase (Figure 9).

Both analyses reveal that LNM and SBB have forests that are distinctive from the others, being at the terminal of the gradient (Figure 8) or having the highest sensitivity (Text box 1). Other Publicly Owned forests are in both methods also somewhat distinctive (i.e., third place sensitivity (Text box 1) and a mean without other overlapping ellipses (Figure 8)). The private groups (PIN and POO) are, in both methods, hard to describe or predict. POO is closest to the middle of the plot, meaning that these owners occur everywhere in the ordination (Figure 8), which could explain why the random forest is poor at predicting them.

Both methods show that the NFI data and variables are unsuitable to describe, classify, or explain the private owners. A possible reason for the high variability and low predictability of the private owner group is that it usually encompasses a very diverse range of owners (Ficko et al., 2019). Therefore, it should be worthwhile to gather more information that goes beyond the (private) forests' characteristics and look into the demographical and cultural contexts of the owners to investigate their forest management (Weiss et al., 2019).

Another agreement between chapter 3 and 4 is that variables that are associated with forests' age are contributing to either explaining the first dimension (Figure 5) or important for predicting owner (Figure 9). The results of the factor analysis of mixed data indicated that SBB forests are, in general, younger than those of LNM, which aligns with the variable importance; age could therefore be an important part of explaining the differences between the forests and their owners.

Albos class is the most important contributor to the explained variance (Figure 6), however, it was of little importance to predict owners (Figure 9). This difference means that the site conditions are relevant to indicate the variation in the forest but not for explaining the differences between owners. Owners seem to be associated with site conditions (Figure 8) which is likely to be of historical origin. The older forests are often located at the poor soils and owned by Natuurmonumenten or the provincial nature organisations, whereas the younger forests are located on the richer soils and owned by State Forest Service (Chapter 3). This association seems to be of a lower importance for the random forest.

In conclusion, there does not seem to be a production ↔ nature-oriented axis amongst owner groups. However, the differences between owners and their forests are very likely to be associated with site class, the forest's development phase and aim.

6.2 Harvest intensity and owners

My second research question was: "What are the differences and patterns of harvest intensity observed in the Netherlands using the repeated NFI data, and how can the measured variables explain these patterns?". I expected that harvest intensity differed between forest and nature management organisations and private owners. Furthermore, I expected variables related to policies (such as Natura 2000) to influence harvest intensity levels for nature management organisations. For State Forest Service, I expect harvest levels to be affected by variables that are more related to production (such as commercial species or subsidy scheme).

I did not find a significant difference between private owners and nature management organisations. Only Other Publicly Owned (OPO), e.g., municipalities or ministries, had a somewhat higher harvest intensity. Furthermore, owner did not influence the harvest intensity itself, but only the presence of harvest. Natura 2000 did not influence harvest intensity, but SNL _Unrestricted [1] increased harvest intensity. Lastly, two variables that are associated with either an increase (transformed increment) or decrease (transformed volume of dead wood) of harvest intensity, are also associated with the plot's age.

My research focussed on differences between owners. However, owner group had a limited effect on describing the forest structure and composition (Chapter 3) and influencing harvest intensity (Chapter 5). Instead, the factor analysis of mixed data shows that the forests are better clustered based on their site conditions and that the gradient among owners is likely associated with the forest's development phase. Results of the harvest predictions point partly in the same direction, with increment being associated with better sites and both increment, and deadwood associated with the forest's phase. Therefore, owner might not be of high importance in how Dutch forests are managed.

Harvest intensities of both private owner groups (PIN and POO) are around the same as LNM and SBB (Chapter 5). On the contrary, Schelhaas et al. (2018a) showed that harvest of Dutch private forest owners was less compared to SBB or LNM. Generally, private owners are thought to have lower harvest due to their property size or lack of knowledge (Schelhaas et al., 2018b). My results could be influenced by the small number of observations of both groups. However, both private groups seem to be diverse (Chapters 3 and 4), which is in line with literature (see Ficko et al., 2019).

The random forest (Chapter 4) and the zero-inflated model (Chapter 5) both indicate that SNL-type is of major importance. Higher harvest intensities are associated with SNL_Unrestricted [1] and SNL-type improved the accuracy of predicting owners. For example, LNM has more SNL_Unrestricted [0] forests, increasing the owner predictability. The observed (or expected) overall differences between owners, therefore, might be due to the amount of SNL_Unrestricted [1] forests an owner has.

Further research would be interesting with site conditions as a random effect in the model to correct for correlation between site conditions and for example increment or species composition. In addition, including age as a covariate would properly underwrite the idea of age being the most explanatory factor in Dutch forest management instead of owner. Extending the database with extra (historical) inventories, such as the NFI-4 or NFI-8, would make the investigation about harvest intensity more complex but also includes a temporal effect, reducing the probability of one-time influences.

In conclusion, all three analyses are linked to either variables that are associated with the forest's phase or with the assigned nature management type (SNL-type) for subsidies. The effect of owner on forests' structure and composition or harvest is limited. The differences between owners are mostly linked to age and when they harvest, all owners seem to harvest with the same intensity.

6.3 Recommendations

During this thesis, I encountered some (methodological) limitations and, therefore, I suggest some recommendations for further research.

- The use of continuous variables to express abiotic conditions[h] rather than the current Albos class would enable the use of more common (multivariate) methods and would benefit the regression-type of analysis.
- It would be interesting to include disturbances (e.g., wind and drought) to see if it affects harvest.
- Aggregating the data into larger groups to reduce the number of variables and data imbalance. For example, species composition ('BA_') could be further aggregated, perhaps based on ecological principles, such as pioneer vs. climax species.
- Reconsider both private owner groups into typologies that have a more similar background and aim. Especially estates differ in their background and aim from the other private organisations.
- Further research could investigate if there are actual structural and compositional differences between forests with a specific management or protection aim (e.g., Natura 2000 or National Parks) or without those aims.
- Investigate the possibilities of improved/other variable selection and model validation to improve the accuracy.
- I used a zero-inflated beta distribution model, which was complex to fully understand. Predicting presence/absence of harvest and harvest intensity of harvested plots separately would make the analysis more straight-forward. It would require a (more common) binomial distribution and probably removes the zero-inflation.
- I used a linear model, but my data indicates that harvest intensity might not be linear with all variables, which is in line with Levers et al. (2014). Investigate the use of non-linear models to assess the harvest intensity.

6.4 Implications for management or society

Although I investigated harvest and being it a management tool, this research does not directly contribute to applied management implications. Nonetheless, I pose some remarks for Dutch forest management, managers, and the public.

First, the heated debate about harvest in Dutch forests could be nuanced with my results. This thesis does not reveal major differences of harvest between the Dutch forest owners. The observed and assumed differences between owners might be locally but are fading on an aggregated scale.

Second, the average Dutch forest is aging, contradictory to the European average (Vilén et al., 2012). The average plot's age was 62 years at the NFI-6 and 65 years at the NFI-7 (Schelhaas et al., 2022). Silvis et al. (2022) state that due to the end of a replanting-subsidy and lower wood prices, harvest in the private forests declined. Furthermore, final felling disappeared without compensating it with more intense thinnings, resulting in an increase of standing volume and aging forests (Silvis et al., 2022). This thesis indicates that harvest intensity is not following a linear relationship, for example, only 49/887 plots had a harvest intensity of 0.5 or higher. Harvest intensity seems associated with age; old and young forests are less likely to be harvested intensively. If we want to use more Dutch wood, we could mobilise more resources from the older forests. However, harvest in older forests probably has large trade-offs with other aspects, such as protection of biodiversity or (rare) nature, and the public's opinion.

[h] For example, clay or silt content and soil pH from SoilGrids https://www.isric.org/explore/soilgrids/faq-soilgrids-2017

7. CONCLUSION

In this study, my aim was to describe and understand the differences between owners in terms of their forests and their management (harvest). To my knowledge, this thesis is one of the first studies to analyse the Dutch NFI data using data-exploration methods and to establish insight into forest structure and composition, ownership, and harvest. Yet, my approach of applying a multivariate, a machine-learning, and a regression-type of analysis to investigate patterns in NFI data is not limited to the Dutch case.

I have found that generally, site conditions contribute the most to the variance in the Dutch forest, while owners do not own very distinctly different types of forests. The differences between owners follow a gradient along variables that are related to the forest's developmental phase. The forests of the Large Nature Management organisations and State Forest Service were the most distinctive and especially age and site conditions contribute to these differences. The private owners, either individuals or organisations, were complex to describe and predict, indicating that this is a very heterogenous group.

The second major finding was that harvest intensity was not dependent on the owner group but more related to the SNL-type, increment, and dominant tree species. SNL-type was also the most important variable to correctly predict the owner group. The province assigns the SNL-type, and it is used for nature management plans and subsidies. Since harvest was lower in nature-oriented SNL-types, it seems that the owners follow the assigned SNL-type.

The presence of harvest was, nevertheless, affected by owner group. The Other Publicly Owned forests have a somewhat higher harvest presence than the other owner groups. Variables that are associated with age (increment and volume of dead wood) were also affecting the harvest presence, aside from SNL-type. How the forest is managed presumably depends on the subsidy type and forest's age.

Concluding, the forest's site conditions shape the 'appearance' of the forest (i.e., forest structure and species composition) while the SNL-type and age shape how the owners manage their forest.

8. References

Arets, E. J. M. M., Baren, S. A. v., Kramer, H., Lesschen, J. P., & Schelhaas, M.-J. (2022). *Greenhouse gas reporting for the LULUCF sector in the Netherlands : methodological background, update 2022*. Wageningen: Wettelijke Onderzoekstaken Natuur & Milieu.

Arets, E. J. M. M., & Schelhaas, M.-J. (2019). *National Forestry Accounting Plan: Submission of the Forest Reference Level 2021-2025 for the Netherlands.* Ministerie LNV. Retrieved from https://edepot.wur.nl/513199

Arnoldussen, A. H. (1982). Enkele impressies omtrent het functioneren van de populier in de IJsselmeerpolders. *Populier : tijdschrift voor teelt en afzet van populier en wilg*(19), 10 - 13 Retrieved from https://edepot.wur.nl/447823 Retrieved 05-02-2023.

Arts, B., Buijs, A., Elands, B., Groot, M. d., Hoogstra, M. A., Kamphorst, D., & Turnhout, E. (2014). *"Interactief kappen" : Staatsbosbeheer en de maatschappelijke acceptie van houtoogst*. Wageningen: Wageningen Universiteit en Research Centre, Bos- en Natuurbeleid, Alterra.

Barends, S. (2010). *Het Nederlandse landschap : een historisch-geografische benadering* (10e herz. dr. ed.). Utrecht: Matrijs.

Boosten, M. (2016). Nederlands bos(beheer): al 250 jaar in ontwikkeling. *Vakblad natuur bos landschap / Stichting Vakblad Natuur Bos Landschap 13 (127): 12 - 14* Retrieved from https://edepot.wur.nl/393230 Retrieved

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal, 9*(2), 378-400. doi:10.32614/RJ-2017-066

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference : a practical information-theoretic approach*. In. Retrieved from https://link.springer.com/book/10.1007/b97636 doi:10.1007/b97636

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research, 33*(2), 261-304. doi:10.1177/0049124104268644

Cosyns, H., Kraus, D., Krumm, F., Schulz, T., & Pyttel, P. (2019). Reconciling the Tradeoff between Economic and Ecological Objectives in Habitat-Tree Selection: A Comparison between Students, Foresters, and Forestry Trainers. *Forest Science, 65*(2), 223-234. doi:10.1093/forsci/fxy042

Daamen, W. P., Clerkx, A. P. P. M., & Schelhaas, M.-J. (2019). *Veldinstructie Zevende Nederlandse Bosinventarisatie (2017-2021).* Wettelijke Onderzoekstaken Natuur & Milieu. Retrieved from Wageningen: https://edepot.wur.nl/477076

Daamen, W. P., & Dirkse, G. M. (2010). The Netherlands. In E. Tomppo, T. Gschwantner, M. Lawrence, & R. E. McRoberts (Eds.), *National Forest Inventories: Pathways for Common Reporting* (pp. 383 - 389): Springer Netherlands.

den Ouden, J., & Mohren, G. M. J. (2020). *De ecologische aspecten van vlaktekap in het Nederlandse bos.* Rapport voor het ministerie van LNV in het kader van de Bossenstrategie Wageningen University & Research. Retrieved from Wageningen: https://edepot.wur.nl/534859

den Ouden, J., Muys, B., Mohren, G. M. J., & Verheyen, K. (2010). *Bosecologie en Bosbeheer*. Leuven: ACCO.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., . . . Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27-46. doi:https://doi.org/10.1111/j.1600-0587.2012.07348.x

Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution, 10*(9), 1412-1430. doi:https://doi.org/10.1111/2041-210X.13234

Duncker, P. S., Barreiro, S. M., Hengeveld, G. M., Lind, T., Mason, W. L., Ambrozy, S., & Spiecker, H. (2012). Classification of Forest Management Approaches: A New Conceptual Framework and Its Applicability to European Forestry. *Ecology and Society, 17*(4). doi:10.5751/ES-05262-170451

Ellenberg, H. H. (1988). *Vegetation ecology of central Europe*: Cambridge University Press.

Esri inc. (2022). ArcGIS Pro (Version 3.0.2). Retrieved from https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

European Commission. (2021). *COMMISSION STAFF WORKING DOCUMENT: The 3 Billion Tree Planting Pledge For 2030.* (52021SC0651). Brussels Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021SC0651.

Ficko, A., Lidestav, G., Ní Dhubháin, Á., Karppinen, H., Zivojinovic, I., & Westin, K. (2019). European private forest owner typologies: A review of methods and use. *Forest Policy and Economics, 99*, 21-31. doi:https://doi.org/10.1016/j.forpol.2017.09.010

Fleiss, J. L., Levin, B., & Paik, M. C. (1981). *Statistical methods for rates and proportions* (Vol. 1981218): New York: Wiley.

Greenwell, B. M., & Boehmke, B. C. (2020). Variable Importance Plots—An Introduction to the vip Package. *The R Journal, 12*(1), 343-366. doi:10.32614/RJ-2020-013

Hartig, F. (2022). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models (Version 0.4.6). Retrieved from https://CRAN.R-project.org/package=DHARMa

Havermans, O. (2021). Ja, bomen kappen is soms beter voor het bos, zeggen wetenschappers. 'We doen het goed in Nederland', Reportage. *Trouw*. Retrieved from https://www.trouw.nl/duurzaamheid-natuur/ja-bomen-kappen-is-soms-beter-voor-het-bos-zeggen-wetenschappers-we-doen-het-goed-in-nederland~b6d4eb2f/

Hengeveld, G. M., Nabuurs, G.-J., Didion, M., van den Wyngaert, I., Clerkx, A. P. P. M., & Schelhaas, M.-J. (2012). A Forest Management Map of European Forests. *Ecology and Society, 17*(4). doi:10.5751/ES-05149-170453

Hengeveld, G. M., Schüll, E., Trubins, R., & Sallnäs, O. (2017). Forest Landscape Development Scenarios (FoLDS)–A framework for integrating forest models, owners' behaviour and socio-economic developments. *Forest Policy and Economics, 85*, 245-255. doi:https://doi.org/10.1016/j.forpol.2017.03.007

Hoogstra, M. A., & Schanz, H. (2009). Future orientation and planning in forestry: a comparison of forest managers' planning horizons in Germany and the Netherlands. *European Journal of Forest Research, 128*(1), 1-11. doi:10.1007/s10342-008-0234-6

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. J. (2006). Survival ensembles. *Biostatistics, 7*(3), 355-373. doi:10.1093/biostatistics/kxj011

Jansen, P. A. G. (2004). *Van tijdelijk bos naar permanent bos.* Stichting Bos en Hout. Retrieved from Wageningen: http://edepot.wur.nl/89804

Joa, B., Paulus, A., Mikoleit, R., & Winkel, G. (2020). Decision making in tree selection – contemplating conflicting goals via marteloscope exercises. *Rural Landscapes, 7*(1), 1-14. doi:10.16993/rl.60

Jolliffe, I. T. (2002). Choosing a Subset of Principal Components or Variables. In I. T. Jolliffe (Ed.), *Principal Component Analysis* (pp. 111-149). New York, NY: Springer New York.

Kaplan, J. O., Krumhardt, K. M., & Zimmermann, N. (2009). The prehistoric and preindustrial deforestation of Europe. *Quaternary Science Reviews, 28*(27), 3016-3034. doi:https://doi.org/10.1016/j.quascirev.2009.09.028

Kassambara, A. (2017). FAMD - Factor Analysis of Mixed Data in R: Essentials. In *Principal Component Methods in R: Practical Guide*: STHDA.

Kassambara, A. (2022). ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2' (Version 0.1.4). Retrieved from https://CRAN.R-project.org/package=ggcorrplot

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses (Version 1.0.7). Retrieved from https://CRAN.R-project.org/package=factoextra

Kuhn, M. (2022). caret: Classification and Regression Training (Version 6.0-93). Retrieved from https://CRAN.R-project.org/package=caret

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software, 25*(1), 1 - 18. doi:10.18637/jss.v025.i01

Levers, C., Verkerk, P. J., Müller, D., Verburg, P. H., Butsic, V., Leitão, P. J., . . . Kuemmerle, T. (2014). Drivers of forest harvesting intensity patterns in Europe. *Forest Ecology and Management, 315*, 160-172. doi:https://doi.org/10.1016/j.foreco.2013.12.030

Liu, Z., Peng, C., Work, T., Candau, J.-N., DesRochers, A., & Kneeshaw, D. (2018). Application of machine-learning methods in forest ecology: recent progress and future challenges. *Environmental Reviews, 26*(4), 339-350. doi:10.1139/er-2018-0034

LNV, & IPO. (2020). *Bos voor de toekomst: Uitwerking ambities en doelen landelijke Bossenstrategie en beleidsagenda 2030.* (1120-001), Ministerie van Landbouw, Natuur en Voedselkwaliteit & Interprovinciaal Overleg. Retrieved from Den Haag: https://open.overheid.nl/repository/ronl-d6ac7db2-0d36-45b0-9507-

f76638a48c0d/1/pdf/Bos%20voor%20de%20toekomst_Uitwerking%20ambities%20en%20doelen%20landelijke%20Bossenstrategie%20en%20beleidsagenda%202030.pdf
Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software, 3*(26), 772. doi:10.21105/joss.00772
Lüdecke, D. (2022). sjPlot: Data Visualization for Statistics in Social Science (Version 2.8.12). Retrieved from https://CRAN.R-project.org/package=sjPlot
Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software, 6*(60), 3139. doi:10.21105/joss.03139
McGrath, M. J., Luyssaert, S., Meyfroidt, P., Kaplan, J. O., Bürgi, M., Chen, Y., . . . Valade, A. (2015). Reconstructing European forest management from 1600 to 2010. *Biogeosciences, 12*(14), 4291-4316. doi:10.5194/bg-12-4291-2015
Microsoft. (2022). Access (Version 2211).
Naaijen, C., & Schmidt, P. (2021). Houtoogst past prima in het Nederlandse bos! *Vakblad natuur bos landschap*(172), 42-45 Retrieved from https://edepot.wur.nl/542399 Retrieved 24/05/2022.
Nabuurs, G.-J., Delacote, P., Ellison, D., Hanewinkel, M., Hetemäki, L., & Lindner, M. (2017). By 2050 the Mitigation Effects of EU Forests Could Nearly Double through Climate Smart Forestry. *Forests, 8*(12), 484. Retrieved from https://www.mdpi.com/1999-4907/8/12/484
Nabuurs, G.-J., Lindner, M., Verkerk, P. J., Gunia, K., Deda, P., Michalak, R., & Grassi, G. (2013). First signs of carbon sink saturation in European forest biomass. *Nature Climate Change, 3*(9), 792-796. doi:10.1038/nclimate1853
Nyssen, B., den Ouden, J., Verheyen, K., & Vanhellemont, M. (2016). Integrating black cherry in forest management in the Netherlands and Belgium. In F. Krumm & L. Vítková (Eds.), *Introduced tree species in European Forests: Opportunities and challenges* (pp. 362-372). Freiburg: European Forest Institute.
Oldenburger, J., & Schoonderwoerd, H. (2016). The Netherlands. In C. Vidal, I. A. Alberdi, L. Hernández Mateo, & J. J. Redmond (Eds.), *National Forest Inventories: Assessment of Wood Availability and Use* (pp. 581-599). Cham: Springer International Publishing.
Posit team. (2022). RStudio: Integrated Development Environment for R (Version 2022.12.0.353). Boston, MA: Posit Software, PBC. Retrieved from http://www.posit.co/
R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/
Schelhaas, M.-J., Clerkx, A. P. P. M., Daamen, W. P. O., J. F. Velema, G. , & Schnitger, P. S., H. Kramer, H. (2014). *Zesde Nederlandse bosinventarisatie: methoden en basisresultaten.* (Alterra-rapport 2545), Alterra, Wageningen-UR. Retrieved from Wageningen:
Schelhaas, M.-J., Clerkx, S., Schoonderwoerd, H., Daamen, W. P., & Oldenburger, J. (2018a). Meer hout uit het Nederlandse bos. *Vakblad natuur bos landschap*(144), 14-17 Retrieved from https://edepot.wur.nl/450212 Retrieved 28/04/2022.
Schelhaas, M.-J., Fridman, J., Hengeveld, G. M., Henttonen, H. M., Lehtonen, A., Kies, U., . . . Nabuurs, G.-J. (2018b). Actual European forest management by region, tree species and owner based on 714,000 re-measured trees in national forest inventories. *PLoS ONE, 13*(11), e0207151. doi:10.1371/journal.pone.0207151
Schelhaas, M.-J., Teeuwen, S., Oldenburger, J., Beerkens, G., Velema, G., Kremers, J., . . . Clerkx, A. P. P. M. (2022). *Zevende Nederlandse Bosinventarisatie: methoden en resultaten.* WOt-rapport (BAPS-projectnummer WOT-04-009-046), Wettelijke Onderzoekstaken Natuur & Milieu. Retrieved from Wageningen: https://doi.org/10.18174/571720
Sheather, S. J. (2009). *A modern approach to regression with R*. In Springer texts in statistics. Retrieved from https://link.springer.com/book/10.1007/978-0-387-09608-7 doi:10.1007/978-0-387-09608-7
Siebel, H., Reichgelt, A., & Blankema, A. (2018). *Praktijkadvies Essentaksterfte.* Vereniging van bos- en natuurterreineigenaren. Retrieved from https://www.vbne.nl/productdetails/praktijkadvies-essentaksterfte
Silvis, H. J., Voskuilen, M. J., & Woltjer, P. J. (2022). *Bedrijfsuitkomsten in de Nederlandse particuliere bosbouw over 2020.* Projectnummer WOT-06-001-007 (978-94-6447-307-0), Wageningen Economic Research. Retrieved from Wageningen: http://dx.doi.org/10.18174/573454
Staatbosbeheer, & Rijksdienst voor het Cultureel Erfgoed. (2021). *Historische schets van bossen en bosaanplant in Nederland.* Retrieved from https://www.staatsbosbeheer.nl/over-staatsbosbeheer/blijf-op-de-hoogte/nieuws/2021/11/historische-schets-van-bossen-en-bosaanplant-in-nederland

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*(1), 307. doi:10.1186/1471-2105-9-307

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(1), 25. doi:10.1186/1471-2105-8-25

ten Hooven, M. (2020). 'Dit is landbouw met bomen'; Het bos in de uitverkoop. *De Groene Amsterdammer*. Retrieved from https://www.groene.nl/artikel/dit-is-landbouw-met-bomen

Tomppo, E., Gschwantner, T., Lawrence, M., & McRoberts, R. E. (2010). *National forest inventories: Pathways for common reporting*: Springer Netherlands.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer.

Vilén, T., Gunia, K., Verkerk, P. J., Seidl, R., Schelhaas, M. J., Lindner, M., & Bellassen, V. (2012). Reconstructed forest age structure in Europe 1950–2010. *Forest Ecology and Management, 286*, 203-218. doi:https://doi.org/10.1016/j.foreco.2012.08.048

Warton, D. I. (2022). *Eco-stats : data analysis in ecology : from t-tests to multivariate abundances*. In Methods in statistical ecology, (pp. 433). Retrieved from https://link.springer.com/content/pdf/10.1007/978-3-030-88443-7.pdf doi:10.1007/978-3-030-88443-7

Weiss, G., Lawrence, A., Lidestav, G., Feliciano, D., Hujala, T., Sarvašová, Z., . . . Živojinović, I. (2019). Research trends: Forest ownership in multiple perspectives. *Forest Policy and Economics, 99*, 1-8. doi:https://doi.org/10.1016/j.forpol.2018.10.006

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. a., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. doi:10.21105/joss.01686

Wijdeven, S. M. J. (2006). *Factsheets dood hout in het bosbeheer.* (1430), Alterra. Retrieved from Wageningen: http://edepot.wur.nl/35486

Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution, 7*(6), 636-645. doi:https://doi.org/10.1111/2041-210X.12577

Zuur, A. F., & Ieno, E. N. (2021). *The World of Zero-Inflated Models*. Newburgh, United Kingdom: Highland Statistics Ltd.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution, 1*(1), 3-14. doi:https://doi.org/10.1111/j.2041-210X.2009.00001.x

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer.

9. APPENDIX

9.1 APPENDIX A: DATA PREPARATION

In this Appendix I explain which variables I extracted from the database and how I calculated the other variables. The variables are grouped as seen inTable 3.

9.1.1 Forest structure

I calculated the following variables: quadratic mean diameter (QMD: mm), basal area (BA: m^2 ha^{-1}), and stem density (Dens: trees ha^{-1}). I chose the quadratic mean diameter over the arithmetic mean since it is directly related to the basal area and gives greater weight to larger trees (Curtis et al., 2000). Furthermore, QMD is a more common measure for the average size of a tree than the arithmetic mean.

$$QMD = \sqrt{\frac{\sum_{i=1}^{n} DBH^2}{n}}$$

Where the DBH (in mm) of the *i*th tree is squared, and a plot has *n* trees. I calculated the basal area per tree (in m^2) as follows:

$$BA_{tree} = \frac{\pi * \left(\frac{DBH}{2}\right)^2}{1000000}$$

and the basal area per plot (in m^2 ha^{-1}) was the sum of *n* alive trees in a plot divided by plot surface (in dm^2).

$$BA_{plot} = \frac{\frac{\sum_{i=1}^{n} BA_{tree}}{plot\ surface}}{1000000}$$

I used the largest DBH (Max_{DBH}) to indicate the thickest tree, the standard deviation of DBH (SD_{DBH}) to capture the variance in a plot, and the Gini coefficient of basal area ($Gini_{BA}$) to investigate the heterogeneity of basal area in a plot (Hakkenberg et al., 2016). The Gini coefficient of basal area is the mean of the difference in basal area between every possible pair of trees in a plot, divided by the mean basal area of a plot (Damgaard et al., 2000). The Gini coefficient is calculated as follows:

$$Gini_{BA} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2\mu}$$

with *n* as the number of trees per plot, x_i and x_j being the basal area of trees *i* and *j*, and μ *as* the mean basal area per tree (Damgaard et al., 2000; Hakkenberg et al., 2016). I calculated the Gini coefficient of basal area with the Gini function of the Reldist package (Handcock, 2022). In addition, I calculated the coefficient of variance for DBH (CV_{DBH}), a standardised index for the dispersion of variation (i.e., relative variation), as standard deviation over the mean (Hakkenberg et al., 2016).

$$CV_{DBH} = \frac{\sigma}{\mu}$$

The $Gini_{BA}$ and CV_{DBH} are both standardised indices explaining heterogeneity or variation.

Moreover, I extracted the volume of dead wood (m^3 ha^{-1}) and regeneration (trees ha^{-1}). I summed the measured lying and standing volumes of dead wood. The NFI crew counted on an 8m radius the number of regenerations (DBH < 5cm and height > 50cm) per species (Schelhaas et al., 2022). I recalculated it as regeneration for the dominant tree species (DTS_Regeneration) or other species per plot (Other_Regeneration). Both without shrub species and in the number of trees per ha.

9.1.2 Species composition

Furthermore, I calculated the Shannon Diversity Index (ShannonIndex) as follows:

$$ShannonIndex = -\sum_{i=1}^{S} p_i \ln p_i$$

where S is the number of species in a plot and p_i is the proportion of individuals belonging to the ith species of the plot (Eichhorn, 2016, p. 49; Hakkenberg et al., 2016). The Shannon diversity index is a way to measure the taxonomic diversity within a community: the higher the index, the higher the diversity. A Shannon diversity index of 0 indicates a single species in a plot. I calculated the Shannon Evenness Index (SEI) as follows:

$$SEI = \frac{ShannonIndex}{ln(S)}$$

where S is the number of species in a plot, and I used an if statement (If ShannonIndex = 0, then SEI = 1) in the Access query to deal with plots of single species. The SEI is an index to measure how evenly the species are present in the community, ranging from zero to one, with one indicating complete evenness. Furthermore, I calculated the share of basal area per species group (i) in a plot ($BA\__{i}$) and extracted the dominant tree species (DTS) from the database.

9.1.3 Abiotic

The database expresses site conditions as the Albos class. Albos classifies the site conditions based on a combination of moisture availability and richness of the soil (de Vries et al., 1992). For the NFI, Schelhaas et al. (2022) have reclassified the site conditions into three groups: poor, medium, and rich. See Schelhaas et al. (2022), Figure 2.5, for an overview of the distinct soil types per Albos class.

9.1.4 Spatial

The exact plot locations are classified but available on request for research purposes. I used these locations to determine: 1) if the plot was in a Natura 2000 area, 2) if the plot was in a National Park, and 3) if the plot was in a zone around cities based on the average walking distance. I downloaded from nationaalgeoregister.nl the Natura 2000 areas (*Natura 2000-gebieden*, 2020) and the National Parks (*Nationale Parken*, 2017). For the city limits, I used the "bebouwde kom" from the BRT TOP10NL (Esri, 2022). To estimate the walking distance, I assumed an average distance, which is half of the average of the categories "toeren, wandelen" and "uitgaan, sport, hobby" over the period 2018-2021 (CBS, 2022). Since data was available per province, I calculated averages per province. These three variables acted as a binary operator ([1] yes, [0] no)

The fourth spatial variable is the assigned SNL-type. SNL-type is a steering instrument for nature management based on the abiotic and spatial conditions, nature values and vegetation[i]. Provinces establish nature management subsidies on SNL-type. I extracted the SNL-type per plot from the database. I reclassified these SNL-types as: [1] no SNL or SNL 16 assigned (multifunctional-oriented) or [0] all other SNL-types assigned (nature-oriented), following Schelhaas et al. (2018).

9.1.5 Other

Other variables that did not seem to be part of forest structure, species composition, abiotic, spatial variables were: age, the time between inventories, increment and harvest intensity. Regarding time, I calculated the time between inventories (rounded to 1 decimal place) and the plot's age as follows:

[i] https://www.bij12.nl/onderwerpen/natuur-en-landschap/index-natuur-en-landschap/natuurtypen/ accessed 24-01-2023.

$$Time\_Between = Round(\frac{Days\ between\ inventories}{365.25}; 1)$$
$$Age = Inventory\ year - Germination\ year$$

Since each tree in a permanent plot received a status regarding being alive, dead, or harvested, I calculated the harvest per plot. I quantified harvest in this thesis as intensity: the proportion of the removed basal area (HI_sum: 0 - 1). I selected only trees that received the harvest status “Harvested, stem removed” in the subsequent inventory (NFI-7) and had a tree status of “dominating tree cohort”, “understory tree cohort”, or “overstory” in the inventory of interest (NFI-6). The inventory of interest was the data source for calculating harvest (and all other variables), while the subsequent inventory acted as a label for the harvested trees. The harvest intensity of NFI-6 (HI_sum) was the dependent variable for explaining the harvest intensity, and, aside from other variables, the harvest intensity of MFV was a covariate (HI_sum.MFV).

$$HI\_sum = \frac{\sum removed\ BA_{tree}}{BA_{plot}}$$

I extracted the yearly cubic increment ($m^3\ ha^{-1}\ yr^{-1}$) per plot from the database.

I used the latest inventory (NFI-7) to describe and classify owners and their forests. To answer what influences harvest intensity, I used the 5th, 6th and 7th inventories (MFV, NFI-6, NFI-7). The data, GIS-models, and R-scripts are available upon request.

9.2 Appendix B – additional figures of chapter 3 (FAMD)

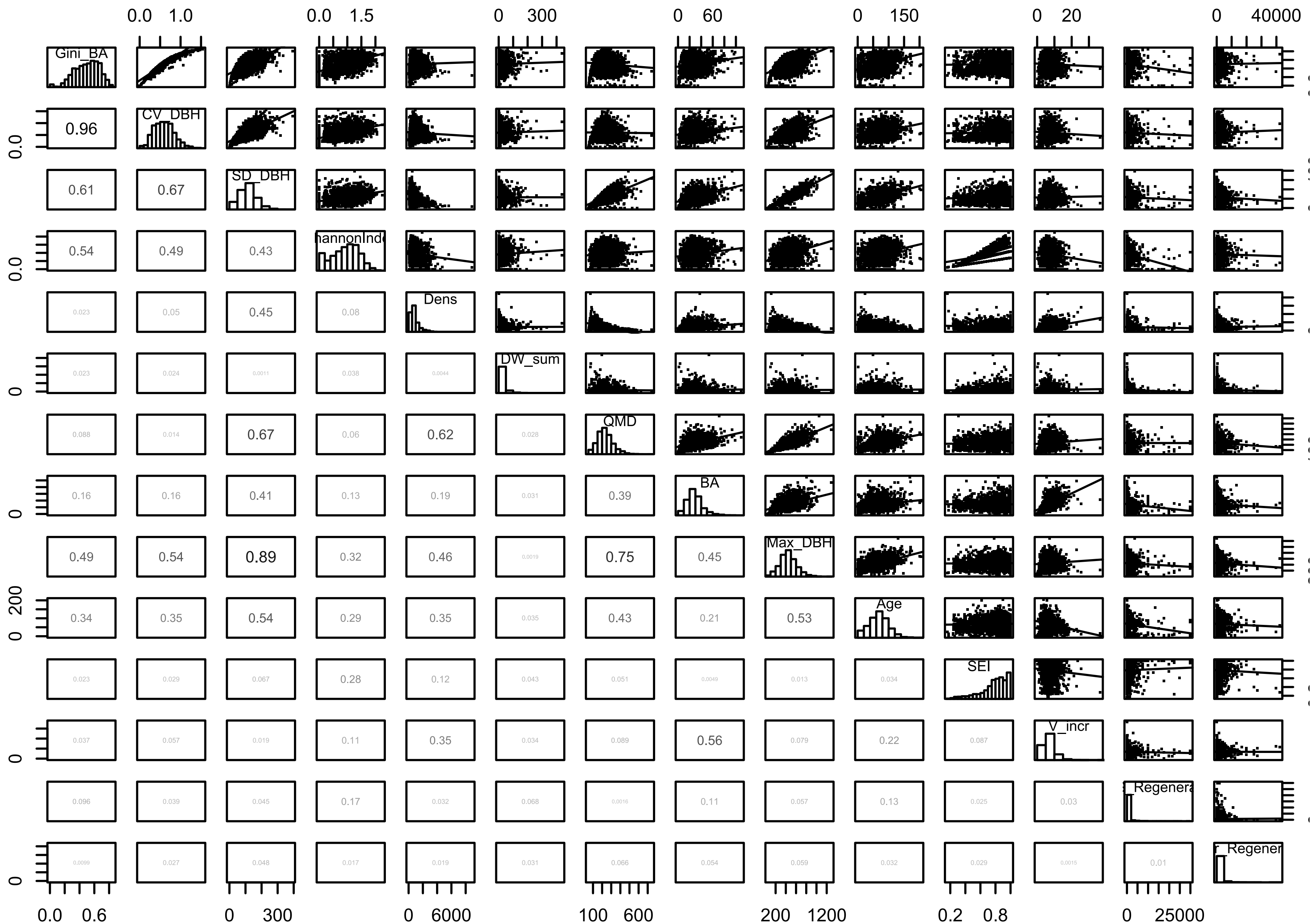

Figure A1. Pair plot to investigate pairwise relationships between all continuous variables from the NFI-7 data in order to assess multicollinearity. See chapter 3 for additional information.

Table A1. Table with eigenvalues for each dimension of the factor analysis of mixed data (FAMD).

Dimension	*Eigenvalue*	*Variance (%)*	*Cumulative variance (%)*
Dim.1	2.75	8.09	8.09
Dim.2	2.41	7.09	15.18
Dim.3	1.84	5.40	20.58
Dim.4	1.53	4.50	25.08
Dim.5	1.46	4.30	29.38
Dim.6	1.24	3.65	33.03
Dim.7	1.22	3.59	36.62
Dim.8	1.15	3.39	40.01
Dim.9	1.13	3.33	43.34
Dim.10	1.11	3.25	46.59
Dim.11	1.08	3.16	49.76
Dim.12	1.05	3.08	52.84
Dim.13	1.04	3.05	55.89
Dim.14	1.03	3.04	58.93
Dim.15	1.00	2.95	61.88
Dim.16	0.99	2.90	64.78
Dim.17	0.98	2.88	67.66
Dim.18	0.97	2.85	70.51
Dim.19	0.94	2.76	73.27
Dim.20	0.91	2.69	75.96
Dim.21	0.90	2.63	78.59
Dim.22	0.88	2.58	81.18
Dim.23	0.86	2.53	83.71
Dim.24	0.83	2.43	86.14
Dim.25	0.79	2.32	88.47
Dim.26	0.76	2.24	90.71
Dim.27	0.73	2.15	92.86
Dim.28	0.66	1.95	94.81
Dim.29	0.58	1.72	96.53
Dim.30	0.47	1.37	97.90
Dim.31	0.37	1.09	98.99
Dim.32	0.21	0.61	99.60
Dim.33	0.14	0.40	100.00

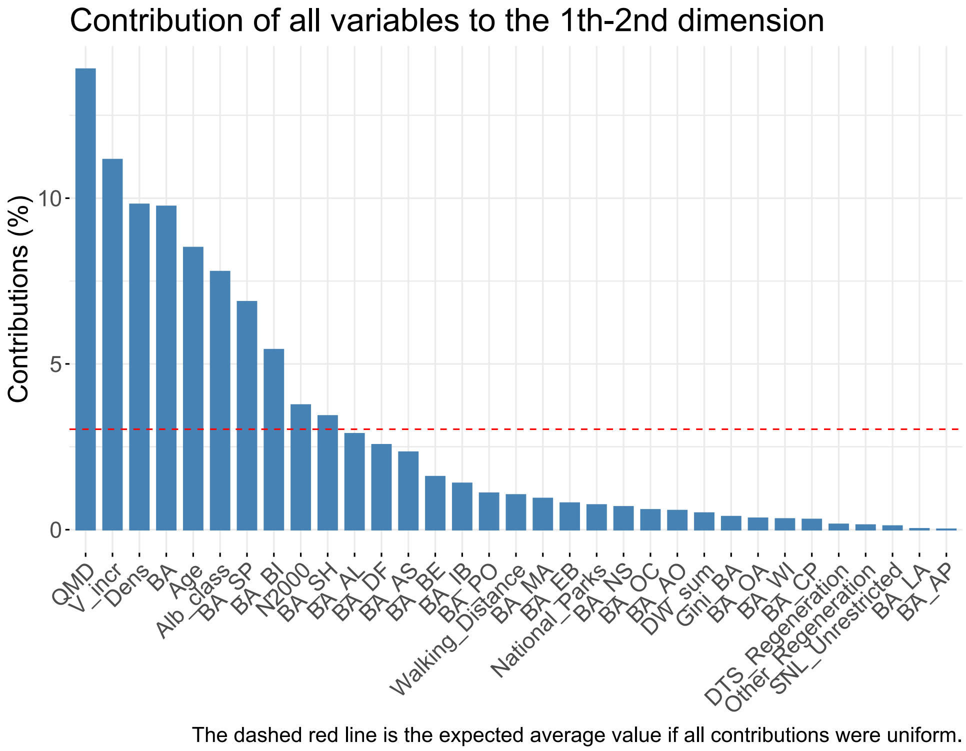

Figure A2. Overview of all 33 variables used in the FAMD and their contribution to the first two dimensions (15% explained variance). The red line indicates the expected average if all variables contributed equally.

Figure A3. Scatter plot of all individual observations (NFI-7 plots; light dots) in the factor analysis of mixed data. The three darker dots are the mean of each Albos class with their 95% confidence ellipse. This is the original plot from Figure 7.

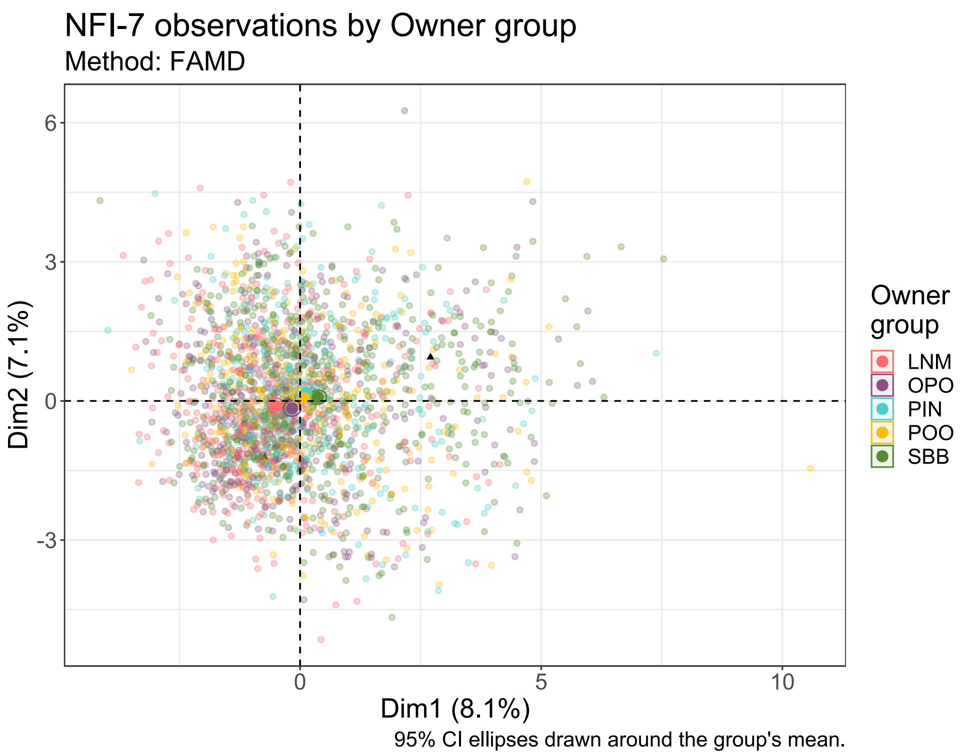

Figure A4. Scatter plot of all individual observations (NFI-7 plots; light dots) in the factor analysis of mixed data. The five darker dots are the mean of each Owner group with their 95% confidence ellipse. This is the original plot from Figure 8.

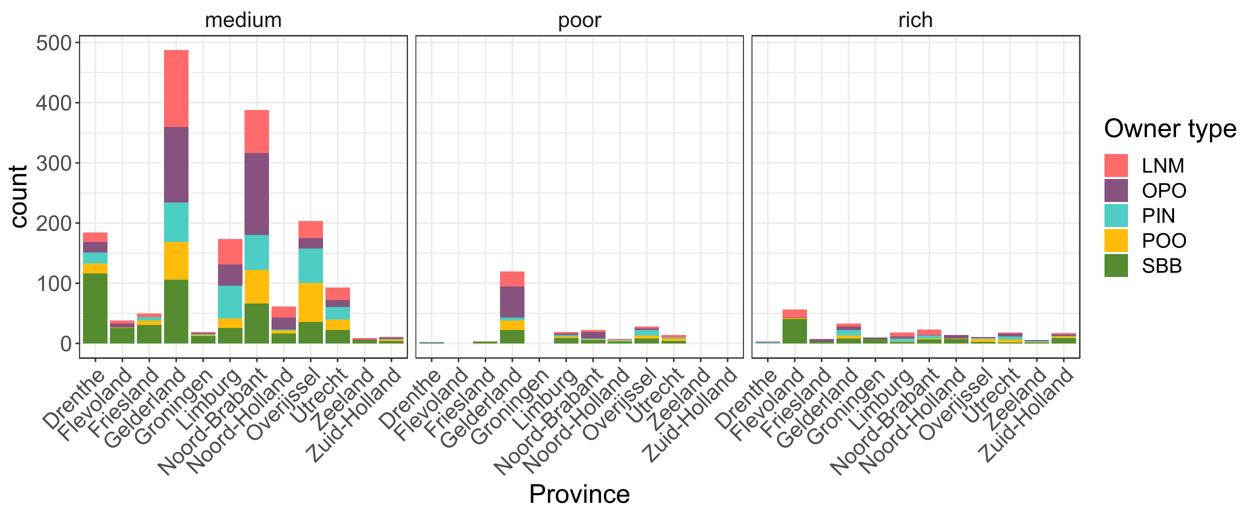

Figure A5. Bar graph with the observations of the selected NFI-7 data for the factor analysis of mixed data, per owner, province and Albos class (site conditions).

9.3 APPENDIX C – ADDITIONAL FIGURES OF CHAPTER 4 (RF)

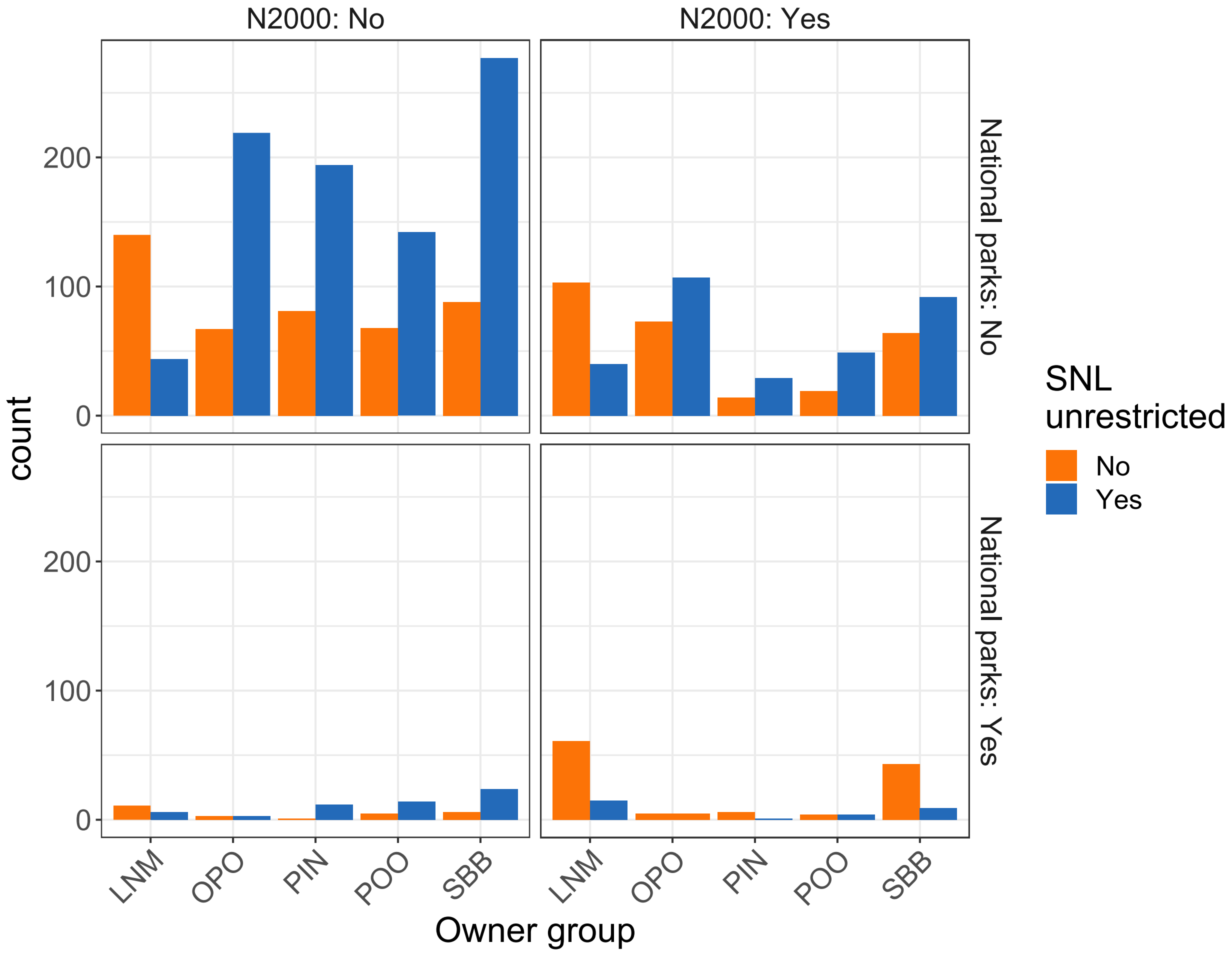

Figure A6. Number of selected NFI-7 observations per owner group, for the three most important spatial variables of the random forest.

9.4 Appendix D – additional figures of chapter 5 (ZIBD)

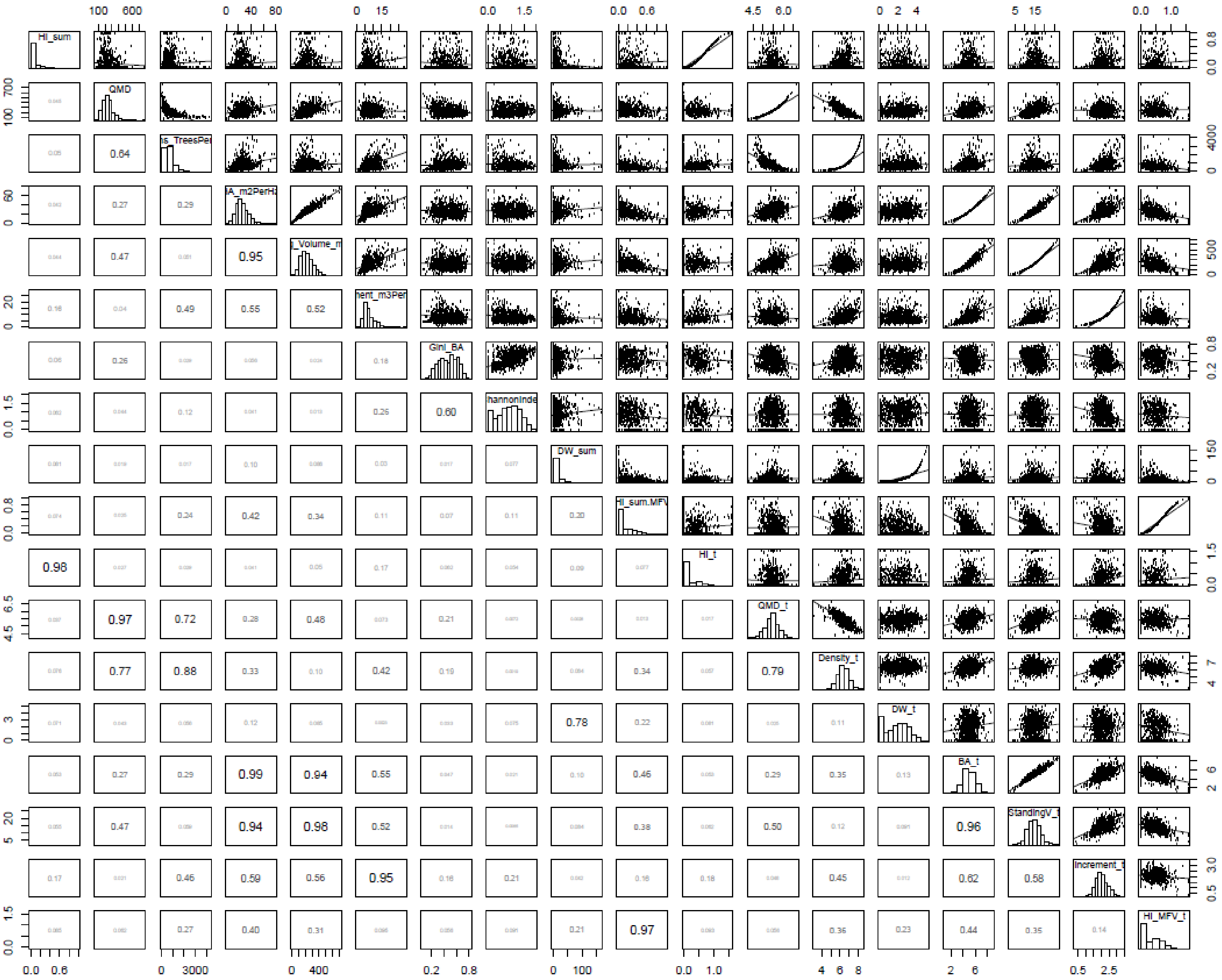

Figure A7. Pair plot to investigate pairwise relationships between all continuous variables from the NFI-6 data in order to assess multicollinearity. See chapter 5 for additional information.

Table A2. Overview of all models during the feature selection. Cond. Refers to the formula of the conditional part of the model and ZI to the zero-inflated part. From model number 5 onwards, I performed a manual backward feature selection on the zero-inflated formula.

Nr.	Model	Part	Formula	AIC	BIC	Notes
1	m.ZI0	Cond.	HI_sum ~ DTS + Owner_Group * SNL_Unrestricted + N2000 + Walking_Distance + QMD_t + BA_t + Increment_t + Gini_BA + ShannonIndex + DW_t + HI_sum.MFV	-12686.3	-12528.4	Has the lowest AIC and BIC, however also the lowest R2 and the residuals showed deviations and clear patterns.
		ZI.	ziformula = ~0			
2	m.ZI1	Cond.	as nr. 1	1034.8	1197.6	
		ZI.	ziformula = ~1			
3	m.Zifull	Cond.	as nr. 1	994.1	1305.3	Has a lower AIC and BIC compared to the ZI1 model and the same R2 but better looking DHARMa residuals
		ZI.	ziformula = ~DTS + Owner_Group * SNL_Unrestricted + N2000 + Walking_Distance + QMD_t + BA_t + Increment_t + Gini_BA + ShannonIndex + DW_t + HI_sum.MFV			
4	m.reduced	Cond.	HI_sum ~ DTS + SNL_Unrestricted + Increment_t + HI_sum.MFV	965.4	1204.8	reduced conditional model with full ZI formula
		ZI.	ziformula = ~DTS + Owner_Group * SNL_Unrestricted + N2000 + Walking_Distance + QMD_t + BA_t + Increment_t + Gini_BA + ShannonIndex + DW_t + HI_sum.MFV			
5	m.reducedMinMFV	Cond.	as nr. 4 -HI_sum.MFV	965.4	1190.4	AIC and AICc are comparable with m.reduced (with HI_sum.MFV), meaning the models are perfoming equally good/bad. However, BIC is better for the model without HI MFV and due to model parsimony I choose the simpler model.
		ZI.	as nr. 4			
6	m.minInter	Cond.	as nr. 5	961.7	1167.6	

		ZI.	as nr. 5 without interactions			Model without interaction effect between SNL and Owner. Gini_BA has the highest P-value: 0.708734
7	m.minGini	Cond. ZI.	as nr. 5 as nr. 6 - Gini_BA	959.9	1161	DTS has the highest p-values (0.969809 for Austrian Pine and 0.965104 for Larch) Only SP is almost significant (0.065308)
8	m.minDTS	Cond. ZI.	as nr. 5 as nr. 7 - DTS	954.2	1097.8	BA_t has the highest p-value (0.563430)
9	m.minBA	Cond. ZI.	as nr. 5 as nr. 8 - BA	952.5	1091.3	HI_sum.MFV has the highest p-value (0.333586)
10	m.minHI	Cond. ZI.	as nr. 5 as nr. 9 - HI_sum.MFV	951.4	1085.5	N20001 has the highest P-value (0.275481)
11	m.minN2000	Cond. ZI.	as nr. 5 as nr. 10 - N2000	950.6	1079.9	Walking_Distance has the highest P-value (0.253060)
12	m.minWD	Cond. ZI.	as nr. 5 as nr. 11 - Walking_Distance	949.9	1074.4	ShannonIndex has the highest P-value (0.23549)
13	m.minShan	Cond. ZI.	as nr. 5 as nr. 12 - ShannonIndex	949.3	1069	QMD_t has the highest P-value (0.074715)
14	m.minQMD	Cond. ZI.	as nr. 5 as nr. 13 - QMD	950.5	1065.4	All variables are significant, but AIC(c) increases. Since the delta AIC<2 I stick to the simplest model. Which is without QMD
	Final model	Cond. ZI.	HI_sum ~ DTS + SNL_Unrestricted + Increment_t ziformula = ~Owner_Group + SNL_Unrestricted + Increment_t + DW_t			

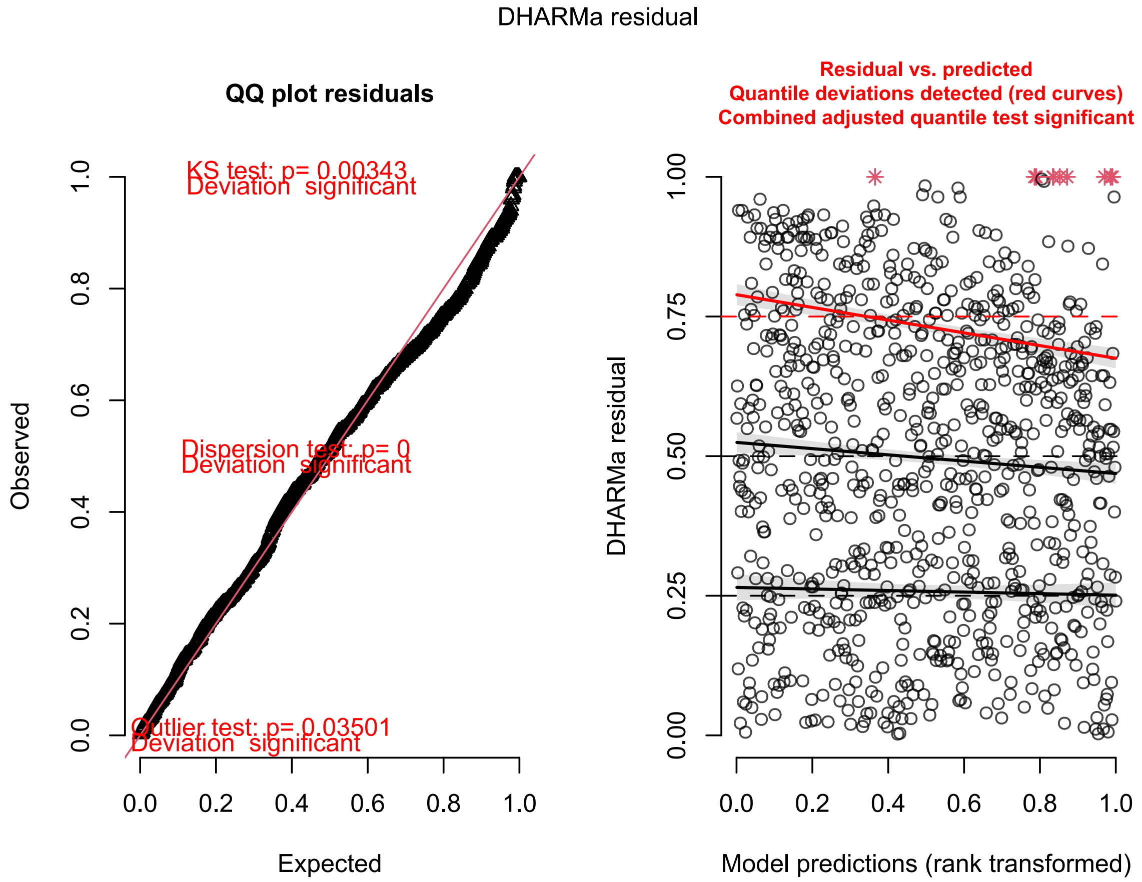

Figure A8. The left panel is a qq-plot to detect overall (residual) deviations from the expected distribution. The right panel is a plot of the residuals against the predicted value. Red stars in the right panel indicate outliers.

Due to time constraints, I was unable to improve to model to deal with the deviations in the residuals. My expectation is that the model could be improved by removing some of the outliers (which I did not do during my thesis) and more observations with a higher harvest intensity (i.e., HI_sum>0.5) are needed.

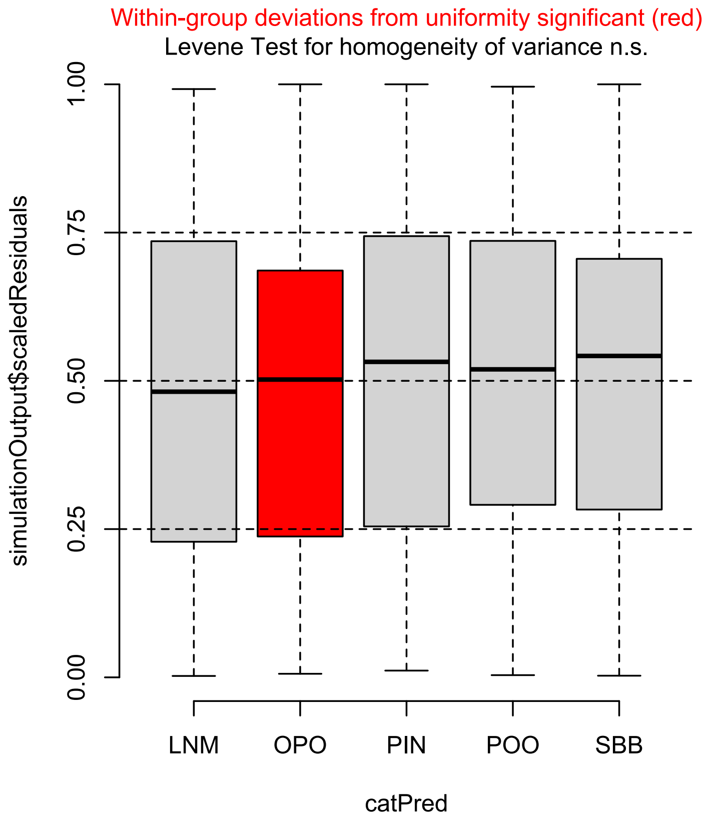

Figure A9. Residual test for the categorical variable owner group. See chapter 2 for abbreviations.

The test alerts that the residuals within the OPO group are not uniformly distributed and deviate from the model's assumptions. However, since only one group was deviating, I assumed that it was not crucial for model's functioning.

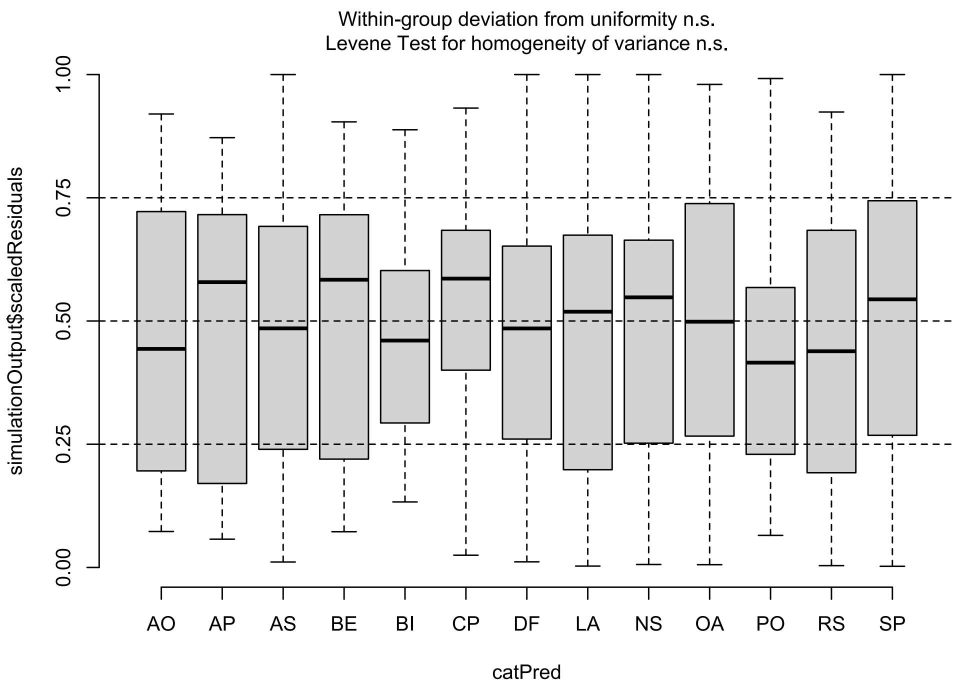

Figure A10. Residual test for the categorical variable dominant tree species. See chapter 2 for abbreviations.

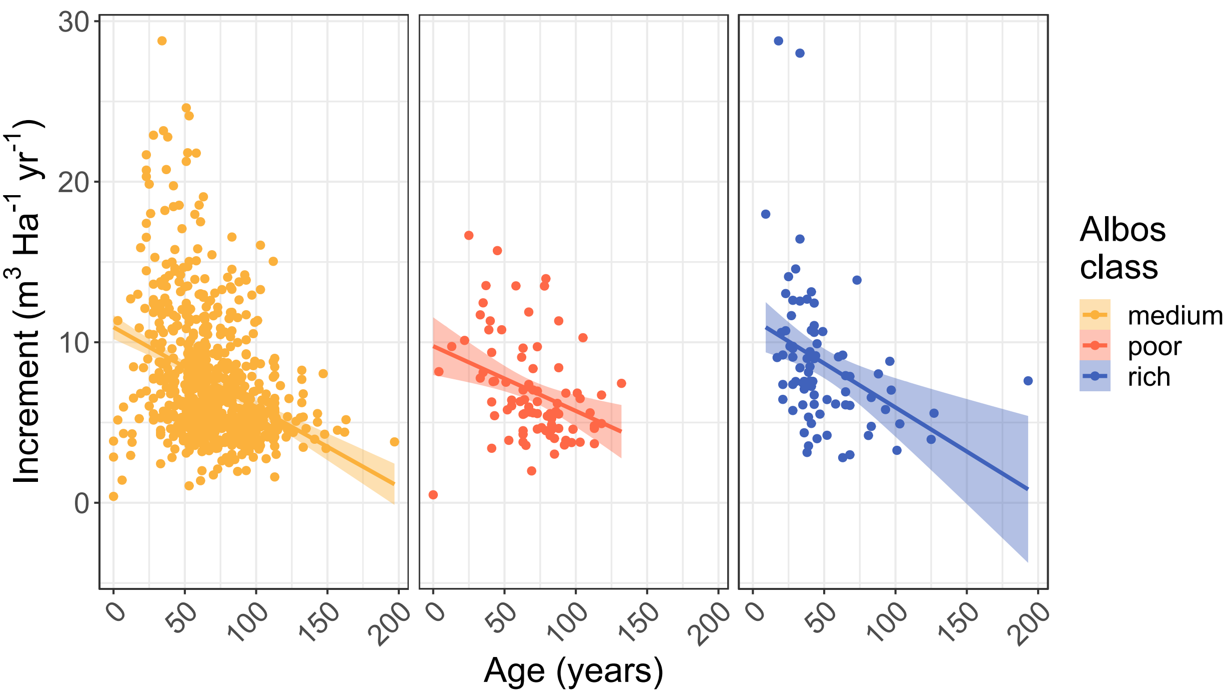

Figure A11. Scatter plot of the used NFI-6 observations, plotted over age and increment, per Albos class. I used a linear smoother to investigate the relationship between transformed annual increment and plot's age.

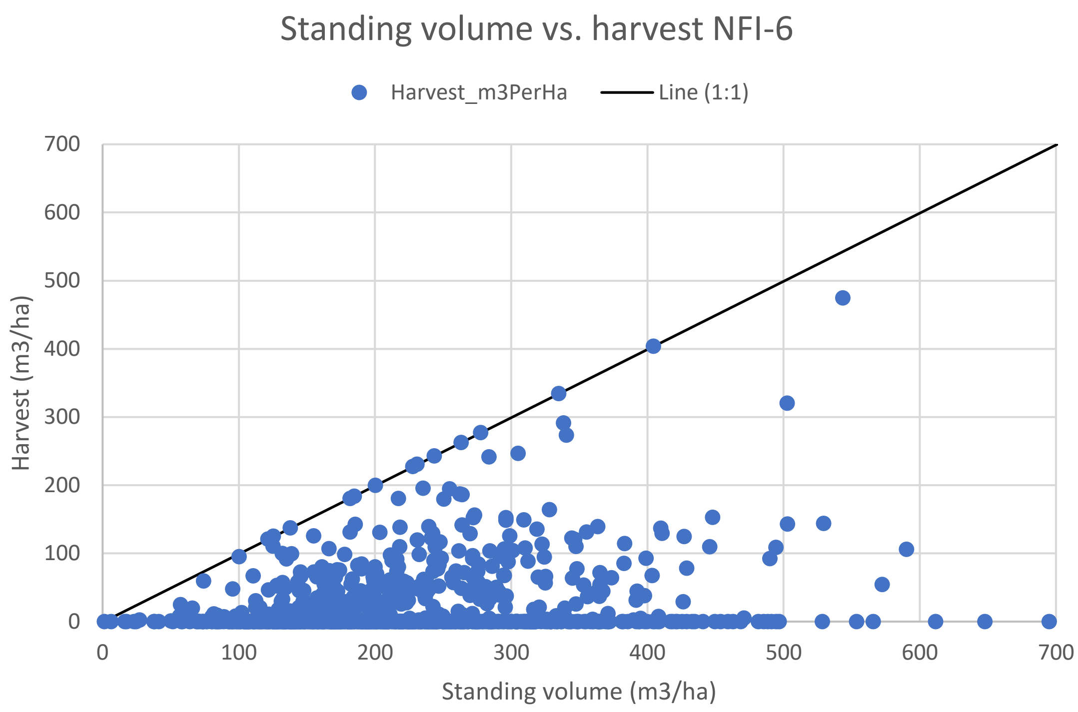

Figure A12. The volumic harvest over the standing volume. The black line indicates when harvest intensity is 1.

The data shows that Dutch forests are harvested but that there is a non-linear relationship with standing volume. It seems like harvest flattens and forests with a high standing stock are less harvested.

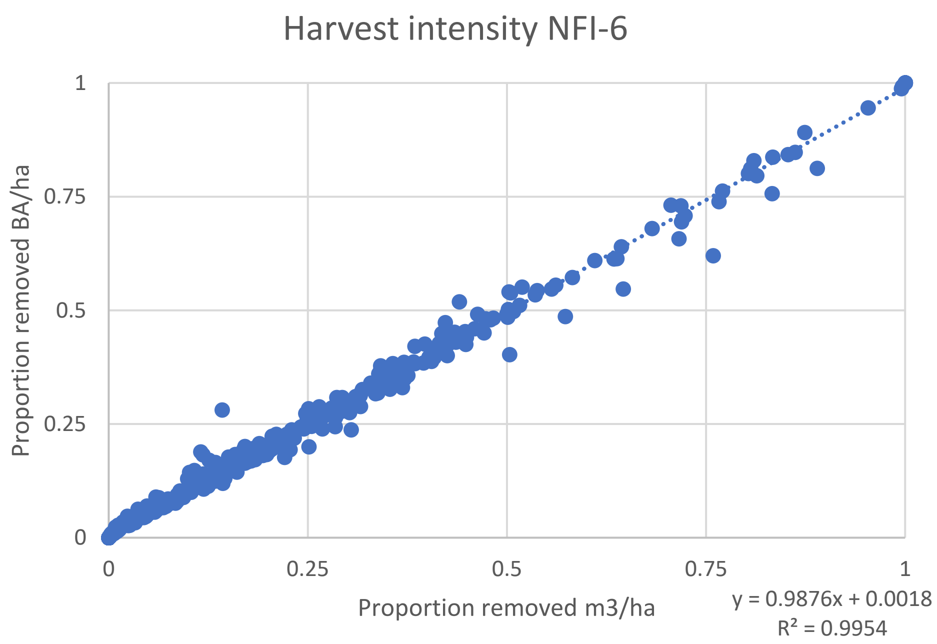

Figure A13. Harvest intensity as the proportion removed volume (x-axis) over the proportion removed basal area (y-axis).

Harvest intensity as proportion removed basal area is highly correlated with harvest intensity as proportion removed standing volume. Little observations are present after both harvest intensities of 0.5, presumably resulting in less accurate model predictions for high harvest intensity.

9.5 References

CBS. (2022). *Mobiliteit; per verplaatsing, vervoerwijzen, motieven, regio's*. Retrieved from: https://opendata.cbs.nl/#/CBS/nl/dataset/84711NED/table?ts=1658316955435 Retrieved

Curtis, R. O., & Marshall, D. D. (2000). Technical Note: Why Quadratic Mean Diameter? *Western Journal of Applied Forestry, 15*(3), 137-139. doi:10.1093/wjaf/15.3.137

Damgaard, C., & Weiner, J. (2000). DESCRIBING INEQUALITY IN PLANT SIZE OR FECUNDITY. *Ecology, 81*(4), 1139-1142. doi:https://doi.org/10.1890/0012-9658(2000)081[1139:DIIPSO]2.0.CO;2

de Vries, F., & Al, E. J. (1992). *De groeiplaatsgeschiktheid voor bosdoeltypen in beeld met ALBOS.* Rapport / DLO-Staring Centrum, 0927-4499 ; 234 (0927-4499 ;), DLO-Staring Centrum. Retrieved from Wageningen:

Eichhorn, M. P. (2016). *Natural systems : the organisation of life*. Chichester, UK ;: John Wiley & Sons.

Esri. (2022). *Basisregistratie Topografie TOP10NL*. Retrieved from: https://basisregistraties.arcgisonline.nl/arcgis/rest/services/BRT/BRT_TOP10NL/FeatureServer/135 Retrieved 20/07/2022

Hakkenberg, C. R., Song, C., Peet, R. K., & White, P. S. (2016). Forest structure as a predictor of tree species diversity in the North Carolina Piedmont. *Journal of Vegetation Science, 27*(6), 1151-1163. doi:https://doi.org/10.1111/jvs.12451

Handcock, M. S. (2022). Relative Distribution Methods (Version 1.7-1). Los Angeles, CA. Retrieved from https://CRAN.R-project.org/package=reldist

Nationale Parken. (2017). Retrieved from: https://nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/4961d305-fbb5-426a-9ba3-53e1ca5f3b18 Retrieved 20/07/2022

Natura 2000-gebieden. (2020). Retrieved from: https://nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/8829e5dd-c861-4639-a6c8-fdbb6e3440d2? Retrieved 20/07/2022

Schelhaas, M.-J., Clerkx, S., Schoonderwoerd, H., Daamen, W. P., & Oldenburger, J. (2018). Meer hout uit het Nederlandse bos. *Vakblad natuur bos landschap*(144), 14-17 Retrieved from https://edepot.wur.nl/450212 Retrieved 28/04/2022.

Schelhaas, M.-J., Teeuwen, S., Oldenburger, J., Beerkens, G., Velema, G., Kremers, J., . . . Clerkx, A. P. P. M. (2022). *Zevende Nederlandse Bosinventarisatie: methoden en resultaten.* WOt-rapport (BAPS-projectnummer WOT-04-009-046), Wettelijke Onderzoekstaken Natuur & Milieu. Retrieved from Wageningen: https://doi.org/10.18174/571720