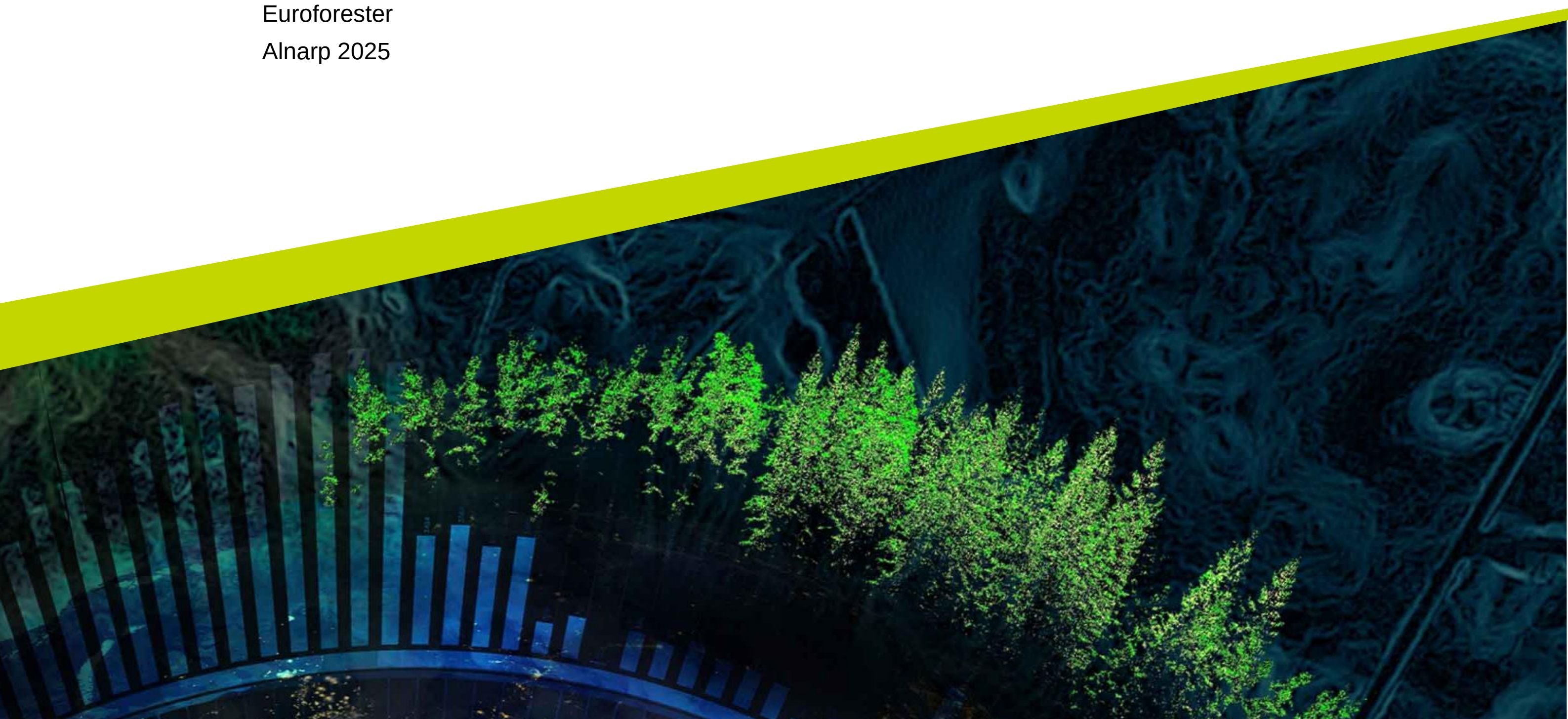
Accuracy comparison of models for cervid forage cover prediction

Modelling cover of oak (Quercus robur), pine (Pinus sylvestris), and birch (Betula pendula and B. pubescens) using remote sensing data in Sweden

Marcin Grądzki

Master's thesis • 30 credits
Swedish University of Agricultural Sciences, SLU
Southern Swedish Forest Research Centre
Euroforester
Alnarp 2025



Accuracy comparison of models for cervid forage cover prediction

Marcin Grądzki

Supervisor:	**Lukas Graf, the Swedish University of Agricultural Sciences, Southern Swedish Forest Research Centre**
Assistant supervisor:	Annika M. Felton, Swedish University of Agricultural Sciences, Southern Swedish Forest Research Centre
Examiner:	Maksym Matsala, the Swedish University of Agricultural Sciences, Southern Swedish Forest Research Centre

Credits:	30
Level:	A2E
Course title:	Master’s thesis in Forest Science
Course code:	EX0984
Programme/education:	Euroforester – Master’s Programme
Course coordinating dept:	Southern Swedish Forest Research Centre
Place of publication:	Alnarp
Year of publication:	2025

Keywords:	remote sensing data, random forest, regression models, cervid forage cover

Swedish University of Agricultural Sciences
Faculty of Forest Sciences
Southern Swedish Forest Research Centre

Abstract

Estimation of cervid forage cover is one of the tasks of forest resource management. It improves the browsing damage prognosis and allows for planning more precise mitigation strategies. In this thesis, I compared the accuracy of four different types of models in predicting the percentage cover of cervid forage. I used data on the Swedish Laser Scanning survey (SLSS), climatic variables (annual temperature and precipitation) from the wordclim database and tree species volume proportions from SLU species maps to train models. The Swedish National Forest Inventory (NFI) was a source of the data about forage cover. Canopy height, canopy cover and elevation were either taken from or calculated based on data from SLSS. I fitted two generalized linear mixed effect models with a beta distribution, one generalized linear additive mixed effect model and one random forests model with forage cover of Scots pine (*Pinus sylvestris*), oak (*Quercus robur*) and birch (*Betula pendula and B. pubescens*) as the response variable. Results varied both among species and methods, but Random Forest was the most accurate model for all species while GAMM performed the worst. The pine models achieved the best r^2 values, but r^2 values were relatively low in all cases. This suggests that in addition to the height of the canopy, canopy cover, species composition, mean annual precipitation, mean annual temperature and elevation, other predictor variables may be needed. Future studies creating predictive models for the percentage cover of these forage plants should utilize additional predictor variables.

Keywords: remote sensing data, random forest, regression models, cervid forage cover

Table of contents

List of tables

List of figures

Abbreviations

Abbreviation	Description
NFI	National Forest Inventory
ALS	Airborne Laser Scanning
LiDAR	Light Detection And Ranging
SLSS	Swedish Laser Scanning Survey
GLMM	Generalized Linear Mixed Model
GAMM	Generalized Additive Mixed Model
RMSE	Root Mean Square Error
RMSE	relative Root Mean Square Error

1. Introduction

Maintaining numbers of cervid populations while minimising the damage they cause to crops and young forests is a challenging balance to strike. A diet of the Cervidae (for example *Dama dama* — fallow deer, *Capreolus capreolus* — roe deer, *Alces alces* — moose, *Cervus elaphus* — red deer) includes green plant material such as tree leaves and needles (Felton et al. 2017). Browsing on trees caused by cervids causes economic damage to forest owners (Felton et al. 2022). It was proven that even a 2% browsing damage level for a stand can decrease profitability (Nilsson et al. 2016). These concerns are exacerbated by the high density of cervids in Europe (Hardalau et al. 2024).

The availability of forage is one of the key factors in reducing browsing damage on economically important timber species (Felton et al. 2022). Higher pine density means higher forage availability, and it decreases browsing on this species. It was proven that the proportion of browsed trees was lower in areas with larger forage cover (Pfeffer et al. 2021). This proves that there is a link between forage availability and browsing damage. There are other factors, such as the distance from a forest stand to the edge of the forest or soil type and fertility, that are relevant (Jalkanen 2001).

In Sweden, the amount of cervid forage is monitored through the Swedish National Forest Inventory (NFI). The NFI provides a long-term programme of monitoring various forest characteristics through annual inventories. Numerous features are measured – for instance, details about forest and land use, carbon sequestration or timber and non-timber forest resources (Fridman et al. 2014). The NFI has over 100 years of tradition and it is an important data source supporting the decision-making process related to subjects such as forestry and environmental policies in Sweden. For example, it was shown that combining data extracted from the NFI database with Airborne Laser Scanning (ALS) data may be applied for predicting site index on grid cells of size 12.5m x 12.5m (Appiah Mensah et al. 2023). Similarly, combining NFI data with other data sources was utilized in Sweden to perform predictive modelling for, among others, stem volume, basal area, and weighted mean tree height (Nilsson et al. 2017) and ecosystem services like bilberry and cowberry yields (Bohlin et al. 2021).

In my thesis, I combined LiDAR data with NFI data describing the cover of oak (*Quercus robur*), birch (*Betula pubescence and B. pendula*) and pine (*Pinus sylvestris*) in selected plots for modelling forage cover. Pine is a staple winter food for moose in many regions and is also consumed by other cervids, while birch is eaten frequently throughout the whole year (Månsson et al. 2007). Oak is a forage highly preferred by cervids (Felton et al. 2022)

In my thesis, I aimed to evaluate the accuracy of four different models for each species: pine, birch, and oak. The main research objective of my thesis is to evaluate which model is best suited for each species.

2. Materials and methods

2.1 NFI data

A National Forest Inventory is carried out as a sample inventory. Sweden is covered by a network of inventory tracks that consist of sample plots. Those are laid out in rectangular or square patterns, and numerous parameters are recorded for them (RIS 2023). I used NFI data collected in the years 2020–2023. I chose to narrow my research to three species — birch, pine and oak — mainly due to their relevance as forage for moose and other cervids, but also due to their cultural and economic value factors (Felton et al. 2022). I chose to model the percentage cover of the cervid forage. I used NFI data about game forage in the form of twigs, branches and small trees available from 0.3m to 2.5m above ground level, collected on non-permanent (7m radius) and permanent (10m radius) plots. Collected features included, for instance, the species of game forage and the cover of the game feed (RIS 2023). The NFI inventory team estimated the horizontal cover of each species per plot using the "diffuse cover" technique. This means that vegetation patches were counted as covered within a periphery of forage that the NFI inventory team estimated as sufficiently dense based on their experience. Such an approach expedites area estimation. Forage cover was measured in square metres. Therefore, in order to normalize this quantity, I calculated the percentage share of forage cover for each species with respect to the total area of permanent and non-permanent plots. For every plot, I divided its area covered with forage by its whole area and then multiplied by 100 to obtain percentages (Table 1). That allowed me to merge data collected for permanent and non-permanent plots for purposes of statistical modelling, since comparing non-processed values may lead to misconceptions. I illustrate it by example — since the total area of permanent plot equals approximately 314 m^2 and the total area of non-permanent plot equals approximately 154 m^2, direct comparison of non-processed values may not take into account those differences — for example, an area of 154 m^2 of forage would cover 100% of the non-permanent plot and around 49% of the permanent plot. Such issues need to be resolved.

Table 1. Descriptive statistics for the response variables.

Notes: Minimum and maximum, mean, median, and standard deviation for the response variables, before using transformations. The data comes from Swedish NFI. There were 19142 observations.

Variable name	Min	Max	Mean	Median	SD
Forage cover: birch (proportion)	0	0.961	0.032	0.006	0.073
Forage cover: pine (proportion)	0	0.792	0.011	<0.001	0.040
Forage cover: oak (proportion)	0	0.896	0.002	<0.001	0.012

2.2 Predictors of forage

2.2.1 Airborne Laser Scanning Data

Light detection and ranging (LiDAR) works by emitting laser pulses and measuring the time those pulses take to return to the sensor after being reflected. The distance between the sensor and the reflection point is obtained by multiplying that time by the speed of light and then halving the result, since the laser pulses travel both ways. This allows for the precise determination of the position of that point using the known position of the sensor (Figure 1) (Faridhouseini et al. 2011). Airborne laser scanning (ALS) data refers to the LiDAR data collected from airborne platforms, such as planes or drones (Wehr and Lohr 1999).

ALS data used in this thesis was collected from 2018 to 2023 as a part of the Swedish laser scanning survey (SLSS) (Swedish Land Survey 2020). Data was gathered using laser scanners ALS80-HP, Terrain Mapper ALS, City Mapper and City Mapper 2 (Lantmäteriet 2022). All of the scanners provide four or more returns from a single laser pulse. Point density was 1-2 points per square metre, flying altitude ca 3000 metres, scanning angle: maximum ± 20º, side overlap of at least 20 % (Lantmäteriet 2022). Predictor variables, which were used to train models, are canopy height, canopy cover and elevation. The calculation of LiDAR-based metrics was conducted within 7-meter radius buffers around the centres of NFI plots. The LidR package (Roussel et al. 2020) in R (R Core Team

2024) was used for that purpose. Canopy height was calculated as the 95th height quantile from the LiDAR data (Figure 2). Canopy cover was calculated by dividing the number of first returns above 10 metres by the number of all first returns and then multiplying the result by 100 to obtain percentage values (Figure 3). Elevation was extracted from Lantmäteriet's digital elevation model, which has a spatial resolution of 2x2 meters (Figure 4). I calculated the mean elevation inside a 7m buffer around the centre of the NFI plot.

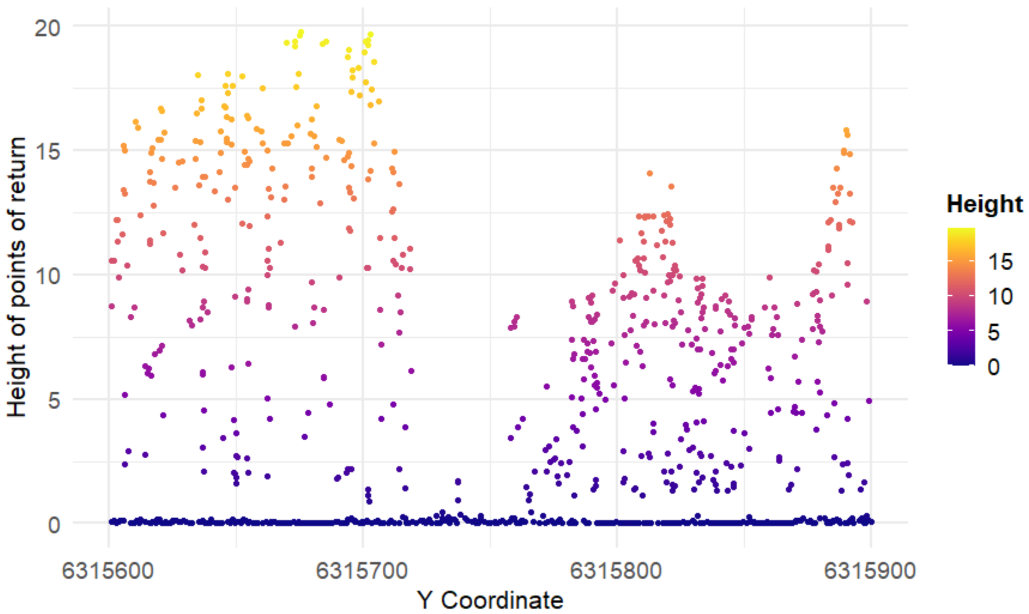


Figure 1. An example cross-section of ALS points of return.

Notes: Each dot represents one point of return. They have known three-dimensional coordinates (X,Y,Z). The X-axis represents changes in the Y-coordinate.

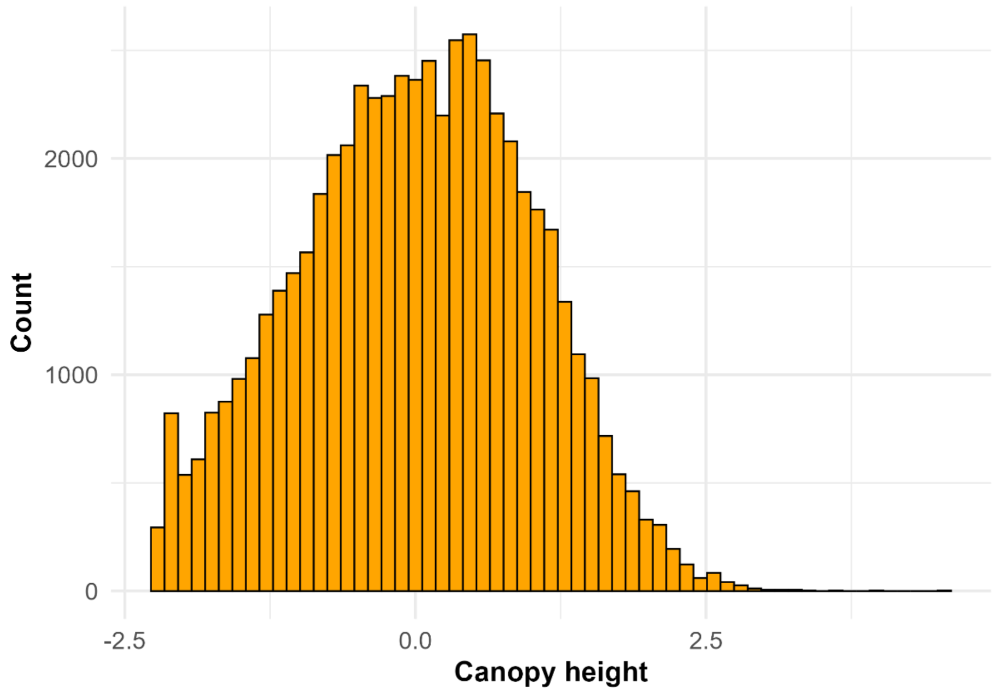


Figure 2. A histogram of canopy height.

Notes: It shows the data that underwent scaling but without division into species-specific datasets. For each predictor variable, I subtracted the mean and divided it by its standard deviation. X-axis represents those scaled values, while Y-axis shows number of observations. The histogram shows data before it was split into training and testing datasets for each species.

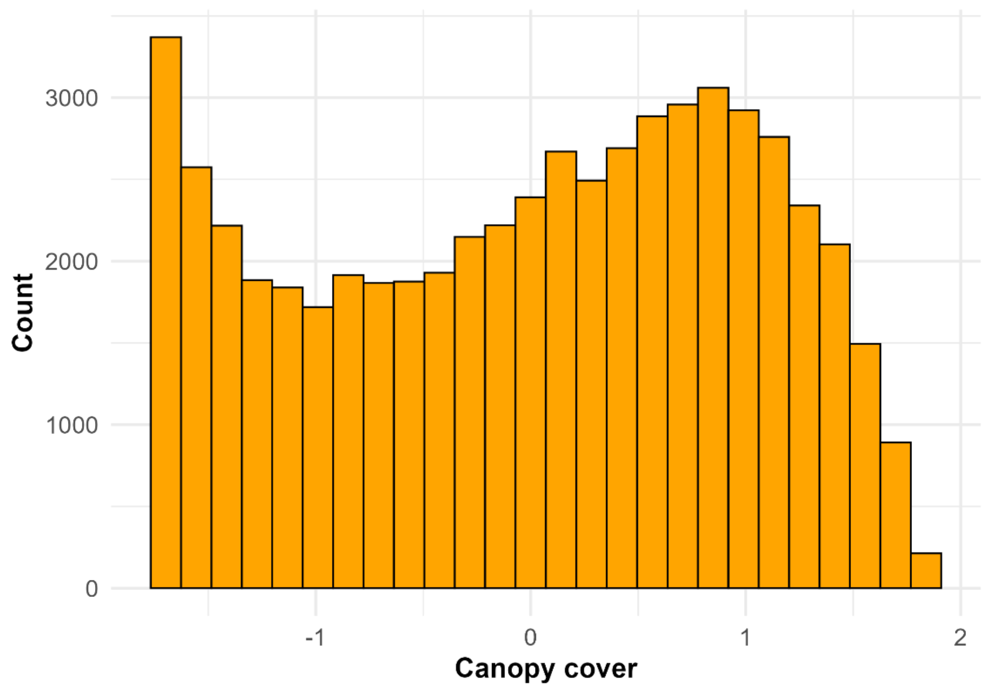


Figure 3. A histogram of canopy cover.

Notes: For more details, see figure 2.

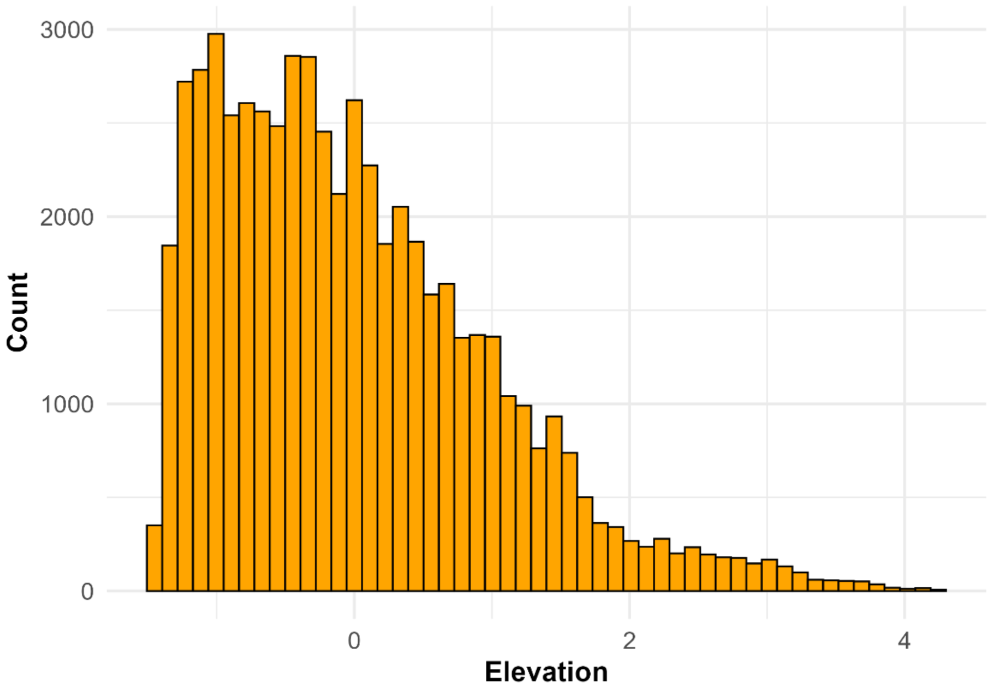


Figure 4. A histogram of elevation.

Notes: For more details, see figure 2.

2.2.2 Climatic variables

As plants are susceptible to climatic regimes (Kelly and Goulden 2008), I added two bioclimatic variables to the data. I used the variables annual mean temperature (Figure 5) and mean annual precipitation (Figure 6) based on the values from 1970 to 2000. Both of these have an impact on the pace of vegetation growth and species distribution (Drobyshev et al. 2008; Kellomäki and Kolström 1994), which directly impacts forage cover. Different tree species have distinct preferences in terms of optimal temperature and the amount of rainfall. Places with optimal temperature and rainfall for a certain species are more likely to contain this species. Data was extracted from datasets available on “worldclim.org” (Fick et.al 2017). The spatial resolution is 1km x 1km. I extracted the values using 7m radius buffers.

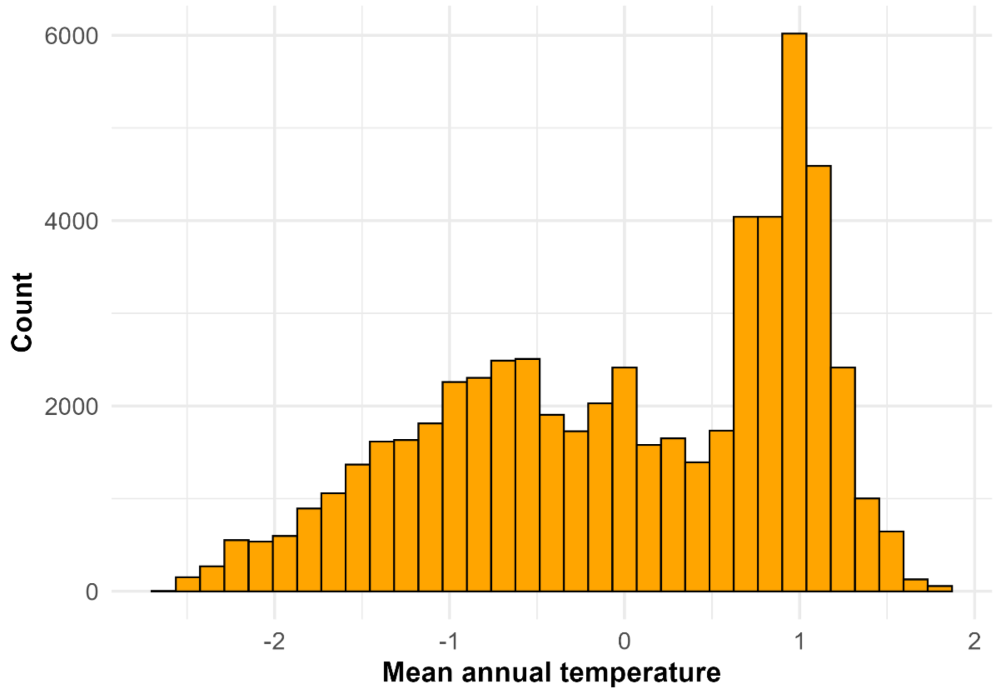


Figure 5. A histogram of mean annual temperature.

Notes: For more details, see figure 2.

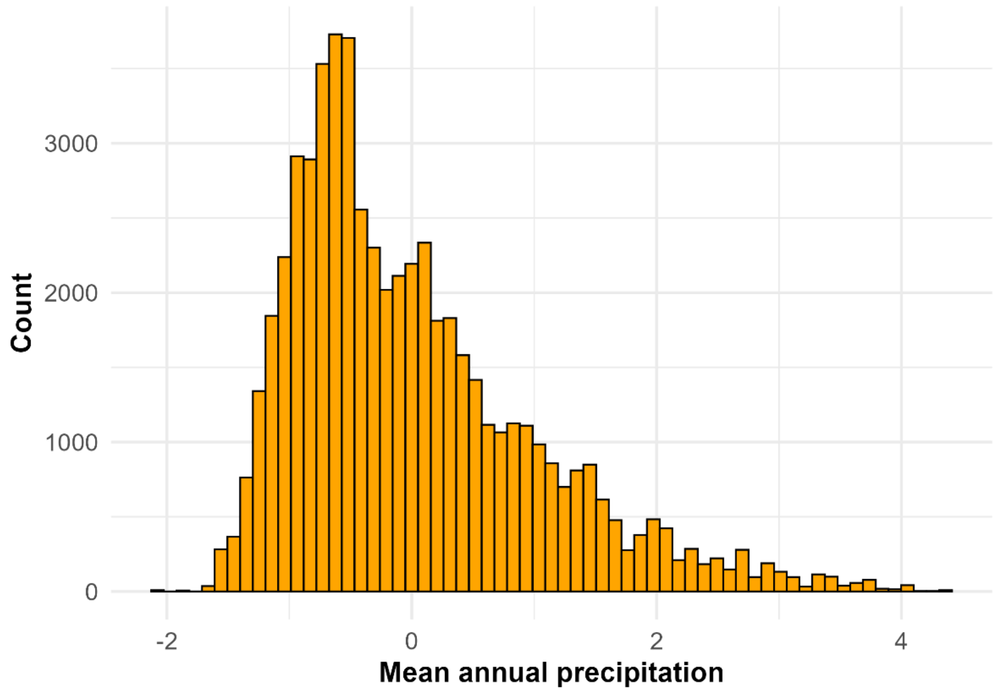


Figure 6. A histogram of mean annual precipitation..

Notes: For more details, see figure 2.

2.2.3 Tree species proportions

Species composition was included as a predictor variable due to its impact on forage cover (Götmark et al. 2005). Models for birch and oak forage cover used the share of pine and spruce (Picea abies) trees volume in the stands' total volume as predictor variable, while models for pine forage cover used only the proportion of spruce volume from the total standing volume (Figure 7, 8). Percentages were calculated using updated species-specific volume maps based on LiDAR data and Sentinel-2 data produced by SLU with a 12.5x12.5m spatial resolution (Nilsson et al. 2017).

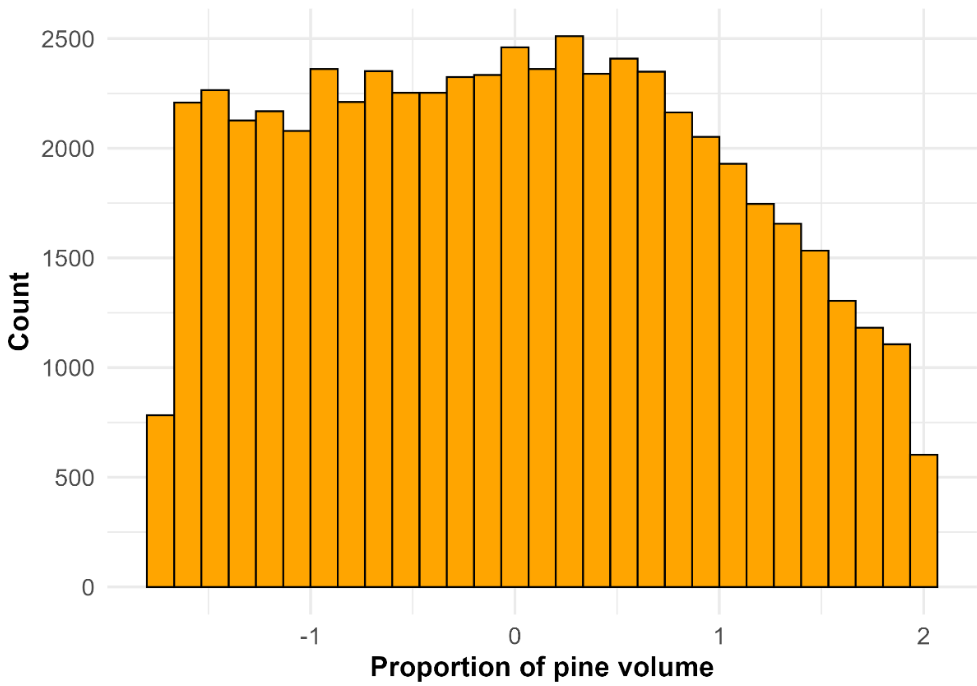


Figure 7. A histogram of pine volume percentage.

Notes: For more details, see figure 2.

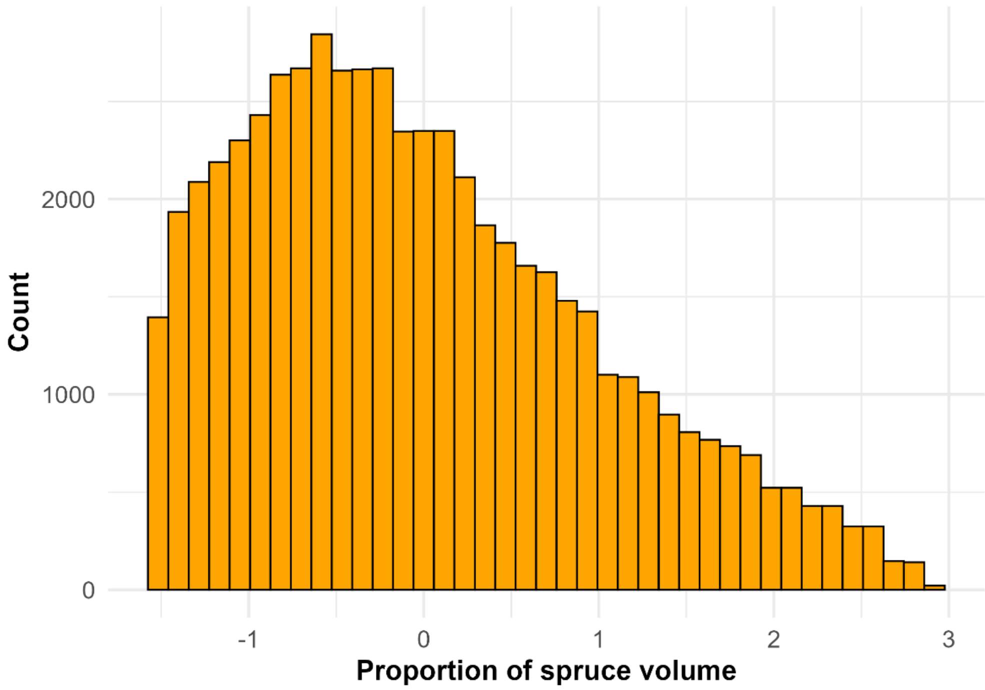


Figure 8. A histogram of spruce volume percentage.

Notes: For more details, see figure 2.

2.3 Modelling

Generalized Linear Models are an extension of linear regression that allow analysing the influence of predictor variables (Table 2) on the outcome without assumption of normal distribution (Su, Yan, and Tsai 2012). Generalized Linear Mixed Models (GLMM) extend that approach by taking into account hierarchical structure of the data (Bolker 2015). A hierarchical structure means that data is divided into groups in such manner, that members of each group are more likely to be similar. Some of my mixed-effect models contained “cluster” and “block” variables as random intercepts (Barr et al. 2013) to take into account the influence of the location on variation of forage cover. Each cluster was uniquely determined by tract number of the NFI and year in which the inventory was conducted. The variable “block” refers to the scanning blocks of SLSS (Figure 9). There were 2879 unique clusters and 252 unique blocks.

Table 2. Overview of predictor variables.

Notes: Used predictor variables with definitions, ranges, sources and explanations as to their expected links to cover of pine, birch and oak.

Variables	Definitions	Data ranges	Data sources	Reasoning
Canopy height	the 95th quantile of height of first echoes	0.00-42.28 (m)	SLSS by National Land Survey	Highly correlated to age and light availability
Canopy cover	Proportion of first echoes over 10m to all first echoes	0-100 (%)	SLSS by National Land Survey	Highly correlated to age and light availability
Percent pine	Proportion of Pine volume to total projected volume	0-100 (%)	SLU species map	Influences regeneration
Percent spruce	Proportion of Spruce volume to total projected volume	0-100 (%)	SLU species map	Influences regeneration, can highly limit understory light availability
Elevation	Height above sea level	-0.35-936.30 (m above the sea level)	SLSS by National Land survey	Influences species distribution due to their limitations
Annual mean temperature	Mean annual temperature at the plot	-2.23-8.33 (°C)	Wordclim.org	Influences growth
Annual precipitation	Mean annual precipitation at the plot	420-1198 (mm)	Wordclim.org	Influences growth

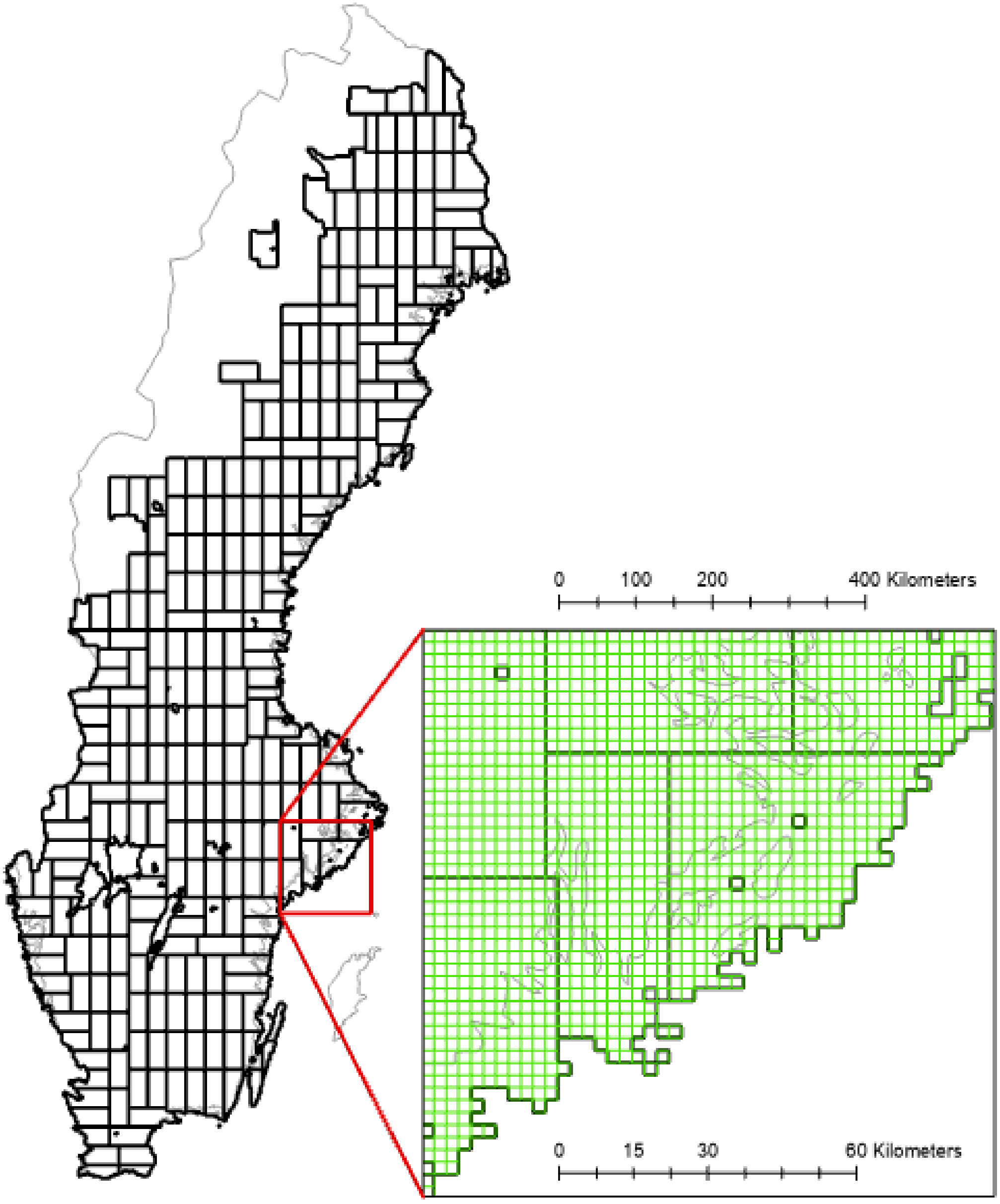


Figure 9. Scanning blocks of the Swedish Laser Scanning Survey.

Notes: Small squares (green) are the closest unit that shares a distinct set of attributes from SLSS. Attributes used in this study included canopy cover and height. NFI plots utilized within this work are distributed across the laser scanning blocks, but due to the data sharing policies of the NFI I was not allowed to have raw coordinates. The sampling intensity of the NFI decreases towards the north areas of Sweden.

For reproducibility during models training process and the sake of fair model testing, the dataset for each species was split using random sampling into training (80% of the data) and testing (20% of the data) sets (Figure 10). I used the caret library (Kuhn 2007) for that purpose. I used the seed 123 to initialize pseudorandom number generation for the split. All training datasets contained 15330 observations, while testing datasets contained 3812 observations.

The response variable of all models was forage cover. I applied arcsine square root transformation (Lin and Xu 2020) in an attempt to normalize the distribution of the response variable. This transformation applied arcsine function to the square root of the forage cover.

The response variable was restricted between 0 and 1, but included both. Since modelling response variable using beta regression model requires continuous data taking values from open interval [0,1] (Olea 2011), I transformed the data using the following formula :

$$y'' = \frac{y'(n-1) + 0.5}{n},$$

where n is the sample size and $y'$is the response variable before transformation. More details concerning this transformation were presented by Smithson and Verkuilen (Smithson and Verkuilen 2006).

In order to improve convergence of the models, my predictor variables were normalized using the “scale” function in R. For each predictor variable, I subtracted the mean and divided by its standard deviation. First, I fitted GLMMs with linear effects for every species, which means that I assumed a linear relationship between the predictor variables and the response variable using the glmmTMB package (Brooks et al. 2017).

Second, for every species I fitted GLMMs with quadratic polynomial terms added for some predictor variables to better account for non-linear trends. Moreover, I included interactions between variables and their squared values were included in the model.

For instance, a similar approach was used in studies of relationships between environmental traits and structures of forest plant communities (Rolhauser, Waller, and Tucker 2021).

Third, I fitted Generalized Additive Mixed Models (GAMM) for every species. I fitted smoothing splines for predictors to take into account the non-linear trends based on visual examination of the data.

Smoothing splines partition the model into intervals that allow local flexibility within the model and allow to capture more complex non-linear relationships. I used the mgcv (Wood 2000) library to include splines. The addition of splines makes such models into GAMM. For instance, they were applied for the identification of environmental variables that affect the abundance of tree species (Antúnez et al. 2017).

Lastly, I fitted random forests using the randomForest package (Breiman et al. 2002). Random forests are machine-learning algorithms used in classification and forecasting (Salman, Kalakech, and Steiti 2024) that creates multiple decision trees. It then takes the average values from the results of those multiple decision trees to make predictions. It is noteworthy that random forest models, as opposed to the rest of the models I fitted, do not use beta regression.

To determine which model best predicted forage cover, I fitted and compared all four different models described above, separately for Oak, Pine and Birch, resulting in 12 fitted models in total.

I used three metrics to compare the accuracy of models. I calculated the Root Mean Square Error (RMSE), relative Root Mean Square Error (rRMSE) and the square of the Pearson correlation coefficient (r^2).

The RMSE measures how far the predicted values are from the observed values in the dataset. Smaller RMSE values indicate a better model fit.

The rRMSE allows a standardized approach through the division of RMSE by the mean of observed values. Results are presented as a unitless proportion (Farooq, Imteaz, and Mekanik 2025).

The coefficient of determination (r^2) shows the proportion of variance of the response variable that can be explained by the predictor variables. It is used to show how properly the data fits the model. It can take values between 0 and 1, and it could be treated as a percentage of model correctness (Chicco, Warrens, and Jurman 2021). It is a measure of linear correlation strength between predictions for each model and actual values from respective test datasets.

All statistical analyses were performed using R version 4.4.1 (R Core Team 2024).

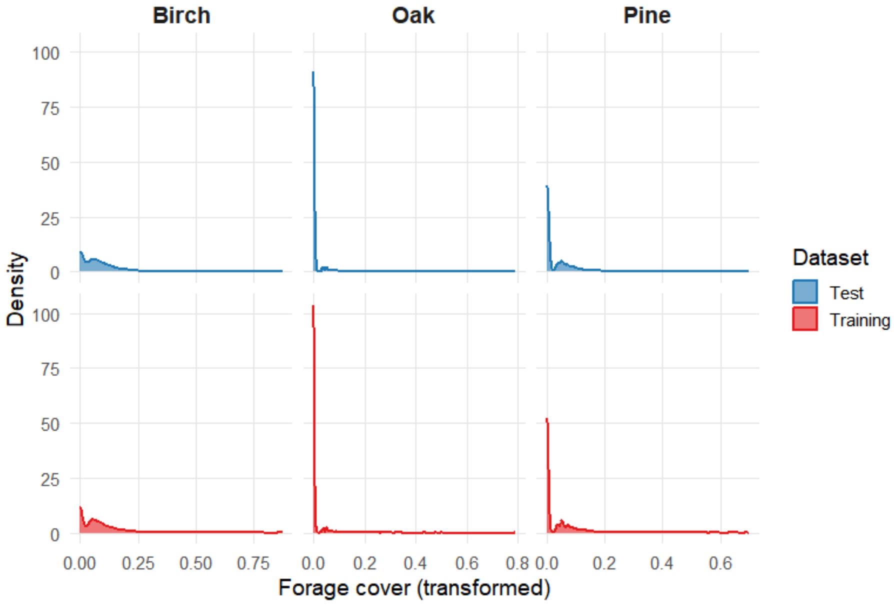


Figure 10. A density histogram of birch, oak, and pine forage cover.

Notes: Values are after transformation, shown separately for training and testing datasets. I chose a density histogram over a frequency histogram due to the different sizes of the compared datasets. The training datasets contain 15330 observations each, while the test datasets contain 3812 observations each.

3. Results

A comparison between results for all of models revealed that in terms of RMSE, rRMSE and r^2, the Random Forest models performed the best (Table 3). They are followed by GLMM with quadratic polynomial terms and then GLMM, which achieved similar results, worse than Random Forest. GAMMs turned out to have the worst accuracy.

Models for oak had the lowest RMSE and highest rRMSE values. Models for birch had highest RMSE values but lowest rRMSE values (with an exception of GAMM).

Pine models had the highest values of r^2 amongst the species. The highest achieved value overall was 0.312 (random forest model for pine) while the lowest was 0.004 (GAMM for birch).

Figures 11, 12 and 13 provide a graphic representation of all model results. Tables 4, 5 and 6 provide estimates, standard errors and p-values for predictor variables that I used to fit GLMM, GLMM with quadratic polynomial terms and GAMM models. Figure 14 provides %IncMSE of the variables used to fit random forest models. More details, including metric descriptions, are provided in respective table and figure legends.

According to the table 4, annual mean temperature and elevation were consistently significant in birch models. These predictors were associated with negative coefficient estimates, with an exception of variable s(Annual mean temperature). According to figure 14, canopy cover, percent pine, canopy height and annual mean temperature were the most important for birch random forest model (~ 47 - 63% IncMSE).

According to table 5, canopy cover and canopy height were consistently significant in pine models. Those predictors were associated with negative coefficient estimates, with an exception of their squared terms. According to figure 14, canopy cover and annual mean temperature were most important for pine random forest model (~ 47.5% IncMSE).

It is noteworthy that I was not able to obtain estimates, standard errors and p-values for predictor variables of oak GAMM model, which is also the model with highest rRMSE among the models. According to table 6, only annual mean temperature was consistently significant in oak models. This predictor was associated with positive coefficient estimates. According to figure 14, annual mean temperature and annual mean precipitation were the most important for oak random forest model (~ 28% IncMSE).

Table 3. Root Mean Square Error (RMSE), relative Root Mean Square Error (rRMSE) and the square of the Pearson correlation coefficient (r^2) for all models.

Species	Model	RMSE	rRMSE (proportion)	r^2
Birch	Linear	0.090	1.147	0.058
	Polynomial	0.090	1.141	0.066
	GAMM	0.172	2.189	0.004
	Random Forest	0.083	1.058	0.189
Pine	Linear	0.057	1.741	0.229
	Polynomial	0.056	1.701	0.274
	GAMM	0.062	1.895	0.004
	Random Forest	0.052	1.568	0.312
Oak	Linear	0.022	3.823	0.128
	Polynomial	0.022	3.806	0.162
	GAMM	0.023	4.029	0.051
	Random Forest	0.019	3.315	0.267

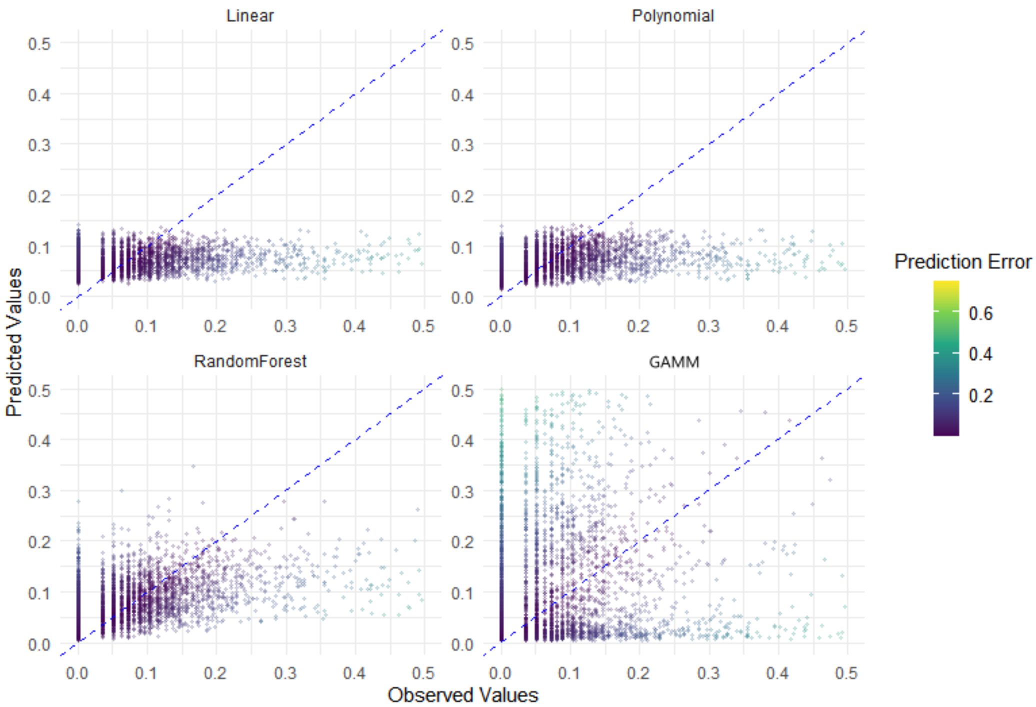


Figure 11. Comparison between predicted and actual values for each model made to predict birch forage cover.

Notes: It contains 4 scatterplots; each of them shows the performance of one model. The x-axis represents observed values while the y-axis represents predicted values. I set the axis limits in order to enhance figure interpretability. The blue dashed line represents an ideal scenario – if all points fell exactly on the line, the model would have 100% accuracy. Models with points more clustered around that line have higher accuracy compared to models with less such clustering. Colours indicate absolute error.

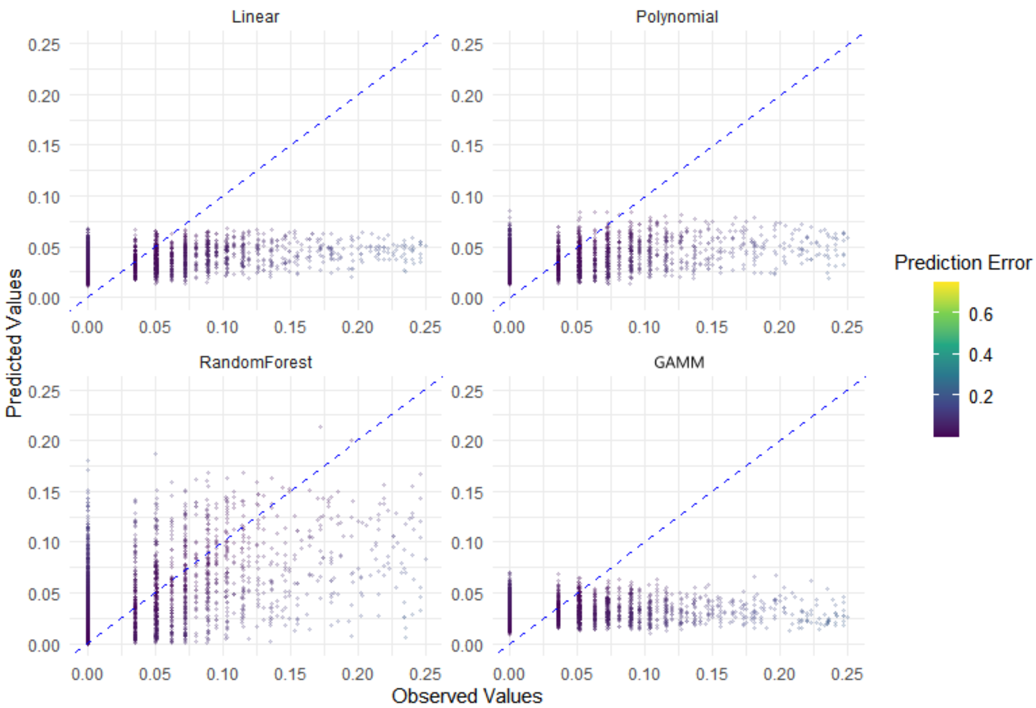


Figure 12. Comparison between predicted and actual values for each model made to predict pine forage cover.

Notes: It contains 4 scatterplots; each of them shows the performance of one model. The x-axis represents observed values while the y-axis represents predicted values. I set the axis limits in order to enhance figure interpretability. The blue dashed line represents an ideal scenario – if all points fell exactly on the line, the model would have 100% accuracy. Models with points more clustered around that line have higher accuracy compared to models with less such clustering. Colours indicate absolute error.

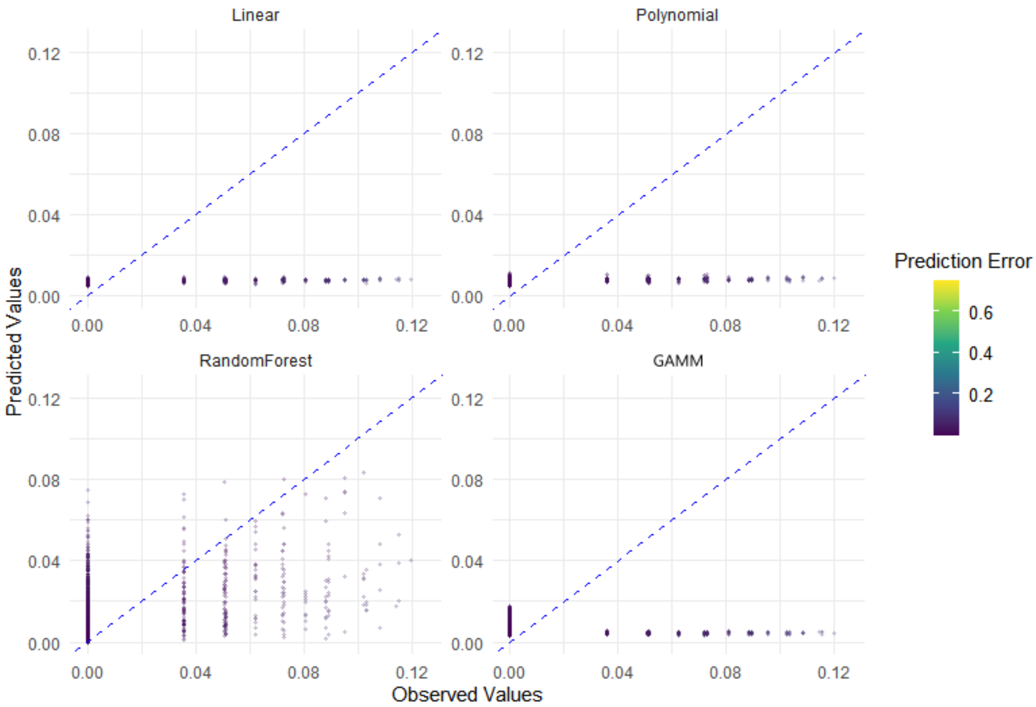


Figure 13. Comparison between predicted and actual values for each model made to predict oak forage cover.

Notes: It contains 4 scatterplots; each of them shows the performance of one model. The x-axis represents observed values while the y-axis represents predicted values. I set the axis limits in order to enhance figure interpretability. The blue dashed line represents an ideal scenario – if all points fell exactly on the line, the model would have 100% accuracy. Models with points more clustered around that line have higher accuracy compared to models with less such clustering. Colours indicate absolute error.

Table 4. A table showing the estimates, standard errors and p-values for predictor variables used to fit all of the birch models (except random forest models).

Notes: "Estimate" shows the estimated coefficients (they represent the effect that the raising of the predictor variable has on the outcome), while "Standard error" shows the standard errors of the coefficients. p-value gives the significance of the predictor, with values below $p < 0.05$ being considered significant. Polynomial models included linear and second-order polynomial terms of predictor variables.

Models	Variables	Estimates	Standard error	p-value
Birch linear	Canopy height	-0.194	0.014	<0.001
	Canopy cover	-0.045	0.014	0.001
	Percent pine	-0.049	0.013	<0.001
	Percent spruce	-0.096	0.012	<0.001
	Annual mean temperature	-0.247	0.026	<0.001
	Annual mean precipitation	0.112	0.020	<0.001
	Elevation	-0.087	0.022	<0.001
Birch polynomial	Canopy height	-0.376	0.020	<0.001
	Canopy height^2	-0.061	0.012	<0.001
	Canopy cover	-0.017	0.025	0.482
	Canopy cover^2	-0.161	0.014	<0.001
	Percent pine	-0.060	0.014	<0.001
	Percent pine^2	-0.127	0.013	<0.001
	Percent spruce	-0.038	0.016	0.016
	Percent spruce^2	-0.030	0.011	0.006
	Annual mean temperature	-0.232	0.028	<0.001
	Annual mean temperature^2	-0.062	0.018	0.001
	Annual mean precipitation	0.055	0.026	0.035
	Annual mean precipitation^2	0.024	0.011	0.033
	Elevation	-0.075	0.024	0.001

	Canopy height x Canopy height^2	0.043	0.005	<0.001
	Canopy cover x Canopy cover^2	-0.012	0.014	0.381
Birch GAMM	Canopy cover	-0.091	0.015	<0.001
	s(Canopy height)	-0.290	0.177	0.102
	s(Percent pine)	0.138	0.111	0.215
	s(Percent spruce)	-0.135	0.106	0.204
	s(Annual mean temperature)	1.176	0.250	<0.001
	s(Annual mean precipitation)	0.216	0.248	0.383
	s(Elevation)	-0.400	0.141	0.005

Table 5. A table showing the estimates, standard errors and p-values for predictor variables used to fit all of the pine models (except random forest models).

Notes: For explanation of the column headings, see table 4.

Models	Variables	Estimates	Standard error	p-value
Pine linear	Canopy height	-0.102	0.012	<0.001
	Canopy cover	-0.280	0.012	<0.001
	Percent spruce	-0.049	0.009	<0.001
	Annual mean temperature	-0.110	0.019	<0.001
	Annual mean precipitation	0.023	0.015	0.112
	Elevation	-0.052	0.017	0.002
Pine polynomial	Canopy height	-0.146	0.017	<0.001
	Canopy height2	0.023	0.009	0.007
	Canopy cover	-0.285	0.013	<0.001
	Canopy cover 2	0.041	0.012	0.001
	Percent spruce	-0.104	0.010	<0.001

	Percent spruce 2	0.056	0.007	<0.001
	Annual mean temperature	-0.176	0.026	<0.001
	Annual mean temperature2	-0.098	0.017	<0.001
	Annual mean precipitation	0.002	0.014	0.880
	Elevation	0.001	0.018	0.969
	Elevation2	-0.040	0.008	<0.001
	Canopy height x Canopy height2	0.021	0.004	<0.001
	Annual mean temperature x Annual mean temperature2	0.019	0.011	0.086
Pine GAMM	Canopy cover	-0.241	0.013	<0.001
	Annual mean precipitation	-0.005	0.014	0.727
	s(Canopy height)	-0.342	0.157	0.030
	s(Percent spruce)	-0.151	0.077	0.052
	s(Annual mean temperature)	-0.273	0.225	0.226
	s(Elevation)	0.045	0.093	0.626

Table 6. A table showing the estimates, standard errors and p-values for predictor variables used to fit all of the oak models (except random forest models).

Notes: For explanation of the column headings, see table 4.

Models	**Variables**	**Estimates**	**Standard error**	**p-value**
Oak linear	Canopy height	0.022	0.012	0.072
	Canopy cover	-0.029	0.012	0.013
	Percent pine	-0.023	0.011	0.039
	Percent spruce	-0.038	0.011	0.001
	Annual mean temperature	0.155	0.016	<0.001
	Annual mean precipitation	-0.063	0.012	<0.001

	Elevation	0.016	0.014	0.241
Oak polynomial	Canopy height	0.031	0.018	0.083
	Canopy height^2	-0.001	0.011	0.951
	Canopy cover	-0.014	0.022	0.537
	Canopy cover^2	-0.012	0.013	0.348
	Percent pine	-0.012	0.012	0.288
	Percent pine ^{2}	-0.007	0.011	0.557
	Percent spruce	-0.027	0.014	0.062
	Percent spruce^2	-0.003	0.010	0.786
	Annual mean temperature	0.189	0.016	<0.001
	Annual mean temperature^2	0.074	0.011	<0.001
	Annual mean precipitation	-0.061	0.016	<0.001
	Annual mean precipitation^2	0.002	0.014	0.892
	Elevation	0.021	0.017	0.211
	Elevation^2	0.019	0.012	0.129
	Canopy height x Canopy height^2	-0.007	0.005	0.205
	Canopy cover x Canopy cover^2	-0.003	0.012	0.826
	Annual mean precipitation x Annual mean precipitation^2	0.001	0.004	0.803
	Elevation x Elevation^2	-0.004	0.004	0.381

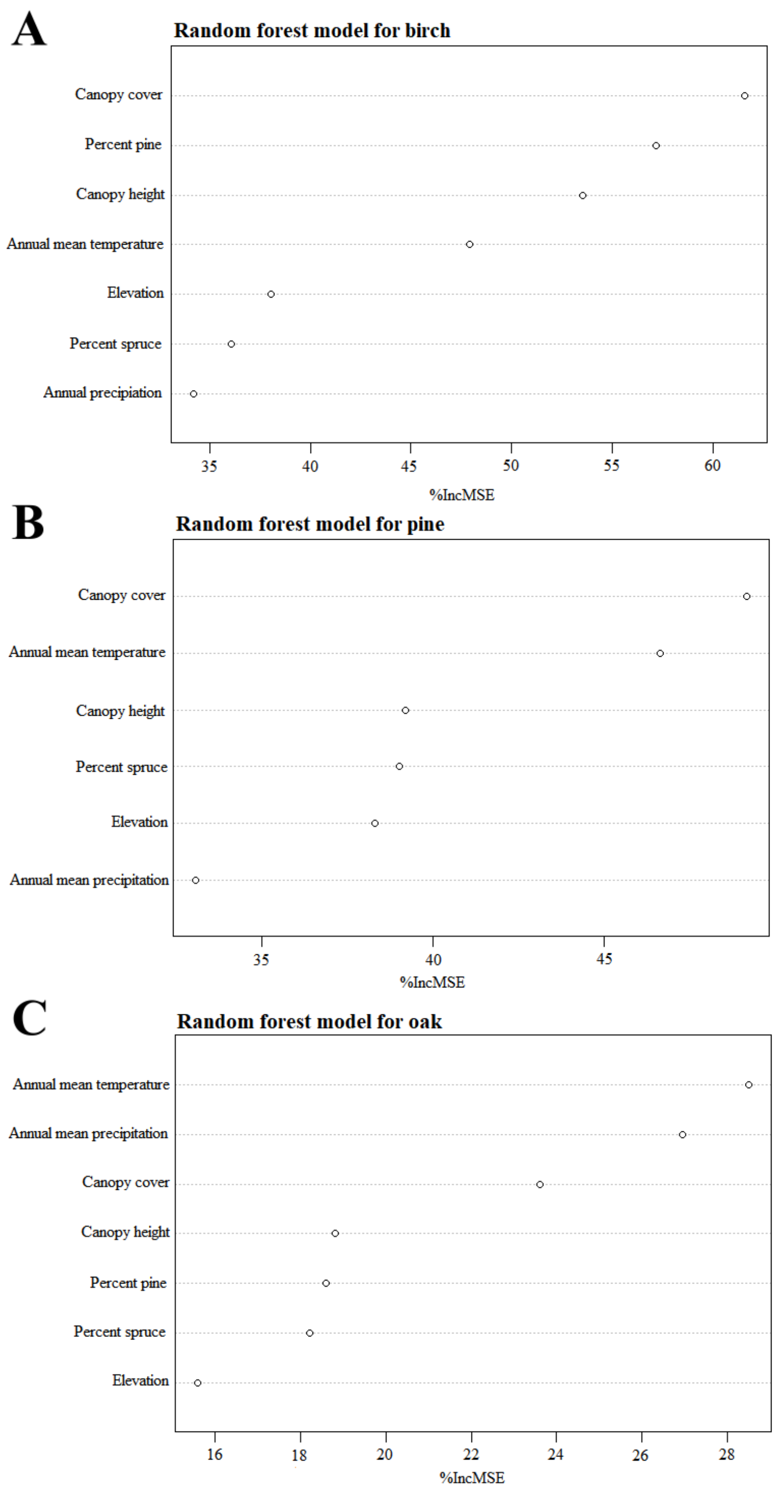


Figure 14. The figure presents the importance of the variables that were used to fit the birch (A), pine (B), and oak (C) random forest models.

Notes: %IncMSE is an indicator that tells how much Mean Squared Error increases after performing random permutation of predictor variables (Tompalski et al. 2015).

4. Discussion

In my thesis, I trained 12 models. Differences in accuracy between types of models were mostly expected at the beginning of my work, but differences between species were not. Due to the consistent NFI methodology, I expected the possible negative impact of under-sampling or incomprehensive sampling on model accuracy (Fei and Yu 2016) to be low. However, it is important to note the major differences in prevalence of the three tree species in Sweden, which could influence the outcome of the modelling. *Pinus sylvestris* covers around 39% of productive forest land in Sweden in comparison to approximately 12% in the case of *Betula spp.* and just around 1% in the case of *Quercus robur* (Samuelsson et. al. 2020). The possible impact of this is further discussed below. On the other hand, such differences between the species' distribution cannot explain why the highest r^2 values were just 0.267 for oak, 0.189 for birch, and 0.312 for pine, especially due to oak achieving higher r^2 than birch. Moreover, descriptive statistics for the response variable achieved highest values for birch (Table 1), a species that is neither most prevalent in Sweden nor achieved highest r^2 values. Those values suggest that chosen predictor variables were not enough, and that future attempts at creating predictive models should be made using adjusted sets of predictor variables.

The height of the canopy, canopy cover, species composition, climatic variables, and elevation were my predictor variables of choice. Global environmental factors, such as climate change and nitrogen deposition, can interact with local factors to affect plants (Hedwall 2021). A deeper understanding of such interactions would help to choose meaningful predictor variables in the future. For example, one study indicates that boreal biome forest structure response to moose presence is higher in areas with higher temperatures (Petersen et al. 2023). The impact of cervidae presence on forest structure and availability of forage cover has not been tested or included as a part of this thesis. Other factors that could be included in future studies are LiDAR metrics describing the canopy in detail, such as the surface roughness, as they have been proven to be useful for creating forest attribute maps (Bohlin et al. 2017).

The theoretical strength of the GAMM lies in its properties that should account for the local data variation better. It seems logical that, for example, the non-parametric and highly flexible smoothing splines could capture variation in a way that could allow for better prediction than using linear terms. However, smoothing splines, by being very sensitive to fluctuations in data, are prone to overfitting (Yandell 1993), especially with a large number of predictor variables. It is possible that with more data and fewer predictor variables, results for GAMMs would have been better than for the GLMMs. However, my results show that they performed worst out of all models tested (Table 3). One indicator of possible reason can be found in table 4. For example, annual mean temperature predictor variable for birch GAMM model had the absolute value of standard error of the coefficient estimate larger than the absolute value of the coefficient estimate itself for annual mean temperature predictor variable of other models. The difference between the accuracy achieved by the Random Forest models and the alternatives is clear. The results of my thesis, along with other works related to environmental niche modeling (Bonsoms and Ninyerola 2024), suggest that Machine Learning (ML) methods should be utilized instead of GAMM and GLMM models.

I briefly considered the usage of the Stacking Ensemble Machine Learning method (Nguyen Van and Lee 2025) that could potentially greatly boost final accuracy in comparison to Random Forest. This method utilizes multiple predictive learning models, such as Random Forest and Linear Regression. Those models, called base-learners, are trained individually on the same datasets. After that, one model, referred to as meta-learner, is trained on the other models' outputs. If done correctly, this should result in an integrated model that outperforms its parts. A comparison between rRMSE results for Random Forest (Nguyen Van and Lee 2025), and less complex methods made me wonder if a similar scale of difference would be seen between the Random Forest and the integrated model. Then again, some more complex models trained on huge datasets require high computational power to fit, and such fitting process may several days. In the end, the choice of the best modelling method depends not only on the desired accuracy but also on available resources and the size and other individual characteristics of available datasets. Utilized methods cannot fully overcome influence of the training data on the metrics of model accuracy, though. A density histogram of response variable for the models (Figure 10) shows that the density of values near 0 in case of oak and pine is larger than in the case of birch. This might partially explain why oak models achieved lowest RMSE, highest rRMSE and higher values of r^2 than birch models.

The future holds many challenges. One such example is related to the climatic data. Annual mean temperature was a consistently significant predictor for birch and oak models other than random forest. % IncMSE of random forest model predictors also indicated at annual mean temperature relative importance amongst other predictors. Taking into consideration the range of climatic conditions across whole Sweden and the fact that my datasets contained NFI information from all over the country, it is not so surprising. It is known that climate change increased the frequency of drought and heat events in the region (Knutzen et al. 2025). European forests are susceptible to damage from such disturbances. Reduced vitality of stands harms their economic and ecological relevance. This puts drive behind planned and tested countermeasures like assisted migration of the seed provenances (Chakraborty et al. 2024). Exact knowledge about the impact of climate on the vegetation and on the variables used to predict vegetation-related measurements can be useful in making informed choices of predictor variables for models. For example, annual mean temperature was associated with negative coefficient estimates for most birch models and with positive coefficient estimates for oak models. A better choice of variables means more accurate models. More accurate models can be referred to with higher confidence when using them as a reference for management decisions in conditions of climate change.

Models with best RMSE, rRMSE and r2, that is pine models, had canopy cover and canopy height as consistently significant predictors. Moreover, values of estimates and %IncMSE further indicate their importance. The results of this thesis suggest that scientists preparing models for forest management should put more focus on testing the impact of variables and the relations between them. Differences between estimates and p-values of predictor variables between models for different species indicate that variables should be chosen separately for each species.

If a similar study were to be conducted, models should be fitted using the same method but with including different set of variables and different variables combinations.

5. Conclusions

Random forest models represented the best accuracy amongst all the models tested and all used performance metrics for prediction of oak, birch, and pine forage cover. Canopy cover, canopy height and annual mean temperature were the most versatile amongst predictor variables I used to fit the models. However, used predictor variables were not enough to achieve high accuracy. Moreover, the choice of predictor variables should have been made separately for each species. Future works in this field should focus on the choice of predictor variables.

References

Antúnez, Pablo, José Hernández-Díaz, Christian Wehenkel, and Ricardo Clark-Tapia. 2017. ‘Generalized Models: An Application to Identify Environmental Variables That Significantly Affect the Abundance of Three Tree Species’. *Forests* 8 (3): 59. https://doi.org/10.3390/f8030059.

Appiah Mensah, Alex, Jonas Jonzén, Kenneth Nyström, Jörgen Wallerman, and Mats Nilsson. 2023. ‘Mapping Site Index in Coniferous Forests Using Bi-Temporal Airborne Laser Scanning Data and Field Data from the Swedish National Forest Inventory’. *Forest Ecology and Management* 547 (November):121395. https://doi.org/10.1016/j.foreco.2023.121395.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. ‘Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal’. *Journal of Memory and Language* 68 (3): 255–78. https://doi.org/10.1016/j.jml.2012.11.001.

Bohlin, Inka, Matti Maltamo, Henrik Hedenås, Tomas Lämås, Jonas Dahlgren, and Lauri Mehtätalo. 2021. ‘Predicting Bilberry and Cowberry Yields Using Airborne Laser Scanning and Other Auxiliary Data Combined with National Forest Inventory Field Plot Data’. *Forest Ecology and Management* 502 (December):119737. https://doi.org/10.1016/j.foreco.2021.119737.

Bohlin, Jonas, Inka Bohlin, Jonas Jonzén, and Mats Nilsson. 2017. ‘Mapping Forest Attributes Using Data from Stereophotogrammetry of Aerial Images and Field Data from the National Forest Inventory’. *Silva Fennica* 51 (2). https://doi.org/10.14214/sf.2021.

Bolker, Benjamin M. 2015. ‘Linear and Generalized Linear Mixed Models’. In *Ecological Statistics*, edited by Gordon A. Fox, Simoneta Negrete-Yankelevich, and Vinicio J. Sosa, 1st ed., 309–33. Oxford University PressOxford. https://doi.org/10.1093/acprof:oso/9780199672547.003.0014.

Bonsoms, Josep, and Miquel Ninyerola. 2024. ‘Comparison of Linear, Generalized Additive Models and Machine Learning Algorithms for Spatial Climate Interpolation’. *Theoretical and Applied Climatology* 155 (3): 1777–92. https://doi.org/10.1007/s00704-023-04725-5.

Breiman, Leo, Adele Cutler, Andy Liaw, and Matthew Wiener. 2002. ‘randomForest: Breiman and Cutlers Random Forests for Classification and Regression’. https://doi.org/10.32614/CRAN.package.randomForest.

Brooks, Mollie, E., Kasper Kristensen, Koen Benthem J.,van, Arni Magnusson, Casper Berg W., Anders Nielsen, Hans Skaug J., Martin Mächler, and Benjamin Bolker M. 2017. ‘glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling’. *The R Journal* 9 (2): 378. https://doi.org/10.32614/RJ-2017-066.

Chakraborty, Debojyoti, Albert Ciceu, Dalibor Ballian, Marta Benito Garzón, Andreas Bolte, Gregor Bozic, Rafael Buchacher, et al. 2024. ‘Assisted Tree Migration Can Preserve the European Forest Carbon Sink under

Climate Change'. *Nature Climate Change* 14 (8): 845–52. https://doi.org/10.1038/s41558-024-02080-5.
Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. 2021. 'The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation'. *PeerJ Computer Science* 7 (July):e623. https://doi.org/10.7717/peerj-cs.623.
Drobyshev, Igor, Mats Niklasson, Olafur Eggertsson, Hans Linderson, and Kerstin Sonesson. 2008. 'Influence of Annual Weather on Growth of Pedunculate Oak in Southern Sweden'. *Annals of Forest Science* 65 (5): 512–512. https://doi.org/10.1051/forest:2008033.
Faridhouseini, Alireza, Ameneh Mianabadi, Mohammad Bannayan, and Amin Alizadeh. 2011. 'Lidar Remote Sensing for Forestry and Terrestrial Applications'.
Farooq, Rashid, Monzur Alam Imteaz, and Fatemah Mekanik. 2025. 'A Pioneering Approach to Deterministic Rainfall Forecasting for Wet Period in the Northern Territory of Australia Using Machine Learning'. *Earth Science Informatics* 18 (2): 224. https://doi.org/10.1007/s12145-025-01724-0.
Fei, Songlin, and Feng Yu. 2016. 'Quality of Presence Data Determines Species Distribution Model Performance: A Novel Index to Evaluate Data Quality'. *Landscape Ecology* 31 (1): 31–42. https://doi.org/10.1007/s10980-015-0272-7.
Felton, Annika M., Adam Felton, Joris P. G. M. Cromsigt, Lars Edenius, Jonas Malmsten, and Hilde Karine Wam. 2017. 'Interactions between Ungulates, Forests, and Supplementary Feeding: The Role of Nutritional Balancing in Determining Outcomes'. *Mammal Research* 62 (1): 1–7. https://doi.org/10.1007/s13364-016-0301-1.
Felton, Annika M., Per-Ola Hedwall, Adam Felton, Fredrik Widemo, Märtha Wallgren, Emma Holmström, Erik Löfmarck, Jonas Malmsten, and Hilde Karine Wam. 2022. 'Forage Availability, Supplementary Feed and Ungulate Density: Associations with Ungulate Damage in Pine Production Forests'. *Forest Ecology and Management* 513 (June):120187. https://doi.org/10.1016/j.foreco.2022.120187.
Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37 (12): 4302-4315
Fridman, Jonas, Sören Holm, Mats Nilsson, Per Nilsson, Anna Ringvall, and Göran Ståhl. 2014. 'Adapting National Forest Inventories to Changing Requirements – the Case of the Swedish National Forest Inventory at the Turn of the 20th Century'. *Silva Fennica* 48 (3). https://doi.org/10.14214/sf.1095.
Götmark, Frank, Jonas Fridman, Göran Kempe, and Björn Norden. 2005. 'Broadleaved Tree Species in Conifer-Dominated Forestry: Regeneration and Limitation of Saplings in Southern Sweden'. *Forest Ecology and Management* 214 (1–3): 142–57. https://doi.org/10.1016/j.foreco.2005.04.001.

Hardalau, Darius, Codrin Codrean, Daniel Iordache, Mihai Fedorca, and Ovidiu Ionescu. 2024. 'The Expanding Thread of Ungulate Browsing—A Review of Forest Ecosystem Effects and Management Approaches in Europe'. *Forests* 15 (8): 1311. https://doi.org/10.3390/f15081311.

Hedwall, Per-Ola & Uria-Diez, Jaime & Brunet, Jörg & Gustafsson, Lena & Axelsson, Anna-Lena & Strengbom, Joachim. (2021). Interactions between local and global drivers determine long-term trends in boreal forest understorey vegetation. Global Ecology and Biogeography. 30. 10.1111/geb.13324.

Jalkanen, Anneli. 2001. 'The Probability of Moose Damage at the Stand Level in Southern Finland'. *Silva Fennica* 35 (2). https://doi.org/10.14214/sf.593.

J. Månsson, C. Kalén, P. Kjellander, H. Andrén, H. Smith 2007 'Quantitative estimates of tree species selectivity by moose (Alces alces) in a forest landscape' Scand. J. For. Res., 22 (5) (2007), pp. 407-414, 10.1080/02827580701515023

Kellomäki, Seppo, and Marja Kolström. 1994. 'The Influence of Climate Change on the Productivity of Scots Pine, Norway Spruce, Pendula Birch and Pubescent Birch in Southern and Northern Finland'. *Forest Ecology and Management* 65 (2–3): 201–17. https://doi.org/10.1016/0378-1127(94)90171-6.

Kelly, Anne E., and Michael L. Goulden. 2008. 'Rapid Shifts in Plant Distribution with Recent Climate Change'. *Proceedings of the National Academy of Sciences* 105 (33): 11823–26. https://doi.org/10.1073/pnas.0802891105.

Knutzen, Florian, Paul Averbeck, Caterina Barrasso, Laurens M. Bouwer, Barry Gardiner, José M. Grünzweig, Sabine Hänel, et al. 2025. 'Impacts on and Damage to European Forests from the 2018–2022 Heat and Drought Events'. *Natural Hazards and Earth System Sciences* 25 (1): 77–117. https://doi.org/10.5194/nhess-25-77-2025.

Kuhn, Max. 2007. 'Caret: Classification and Regression Training'. https://doi.org/10.32614/CRAN.package.caret.

Lantmäteriet. 2022. Laser Data Download—Forest. Last accessed February 28, 2025. https://www.lantmateriet.se/globalassets/geodata/geodataprodukter/hojddata/e_pb_laserdata_nedladdning_skog.pdf

Lantmäteriet. 2022. Quality Description of Lidar Data. Last accessed February 28, 2025. https://www.lantmateriet.se/globalassets/geodata/geodataprodukter/hojddata/quality_description_lidar.pdf

Lin, Lifeng, and Chang Xu. 2020. 'Arcsine-based Transformations for Meta-analysis of Proportions: Pros, Cons, and Alternatives'. *Health Science Reports* 3 (3): e178. https://doi.org/10.1002/hsr2.178.

Nguyen Van, Linh, and Giha Lee. 2025. 'Optimizing Stacked Ensemble Machine Learning Models for Accurate Wildfire Severity Mapping'. *Remote Sensing* 17 (5): 854. https://doi.org/10.3390/rs17050854.

Nilsson, Mats, Karin Nordkvist, Jonas Jonzén, Nils Lindgren, Peder Axensten, Jörgen Wallerman, Mikael Egberth, et al. 2017. 'A Nationwide Forest Attribute Map of Sweden Predicted Using Airborne Laser Scanning Data and Field Data from the National Forest Inventory'. *Remote Sensing of*

Environment 194 (June):447–54. https://doi.org/10.1016/j.rse.2016.10.022.
Nilsson, Urban, Mattias Berglund, Jonas Bergquist, Hampus Holmström, and Märtha Wallgren. 2016. 'Simulated Effects of Browsing on the Production and Economic Values of Scots Pine (*Pinus Sylvestris*) Stands'. *Scandinavian Journal of Forest Research* 31 (3): 279–85. https://doi.org/10.1080/02827581.2015.1099728.
Olea, Ricardo A. 2011. 'On the Use of the Beta Distribution in Probabilistic Resource Assessments'. *Natural Resources Research* 20 (4): 377–88. https://doi.org/10.1007/s11053-011-9153-1.
Petersen, Tanja K., Anders L. Kolstad, Jari Kouki, Shawn J. Leroux, Lynette R. Potvin, Jean-Pierre Tremblay, Märtha Wallgren, et al. 2023. 'Airborne Laser Scanning Reveals Uniform Responses of Forest Structure to Moose (*Alces Alces*) across the Boreal Forest Biome'. *Journal of Ecology* 111 (7): 1396–1410. https://doi.org/10.1111/1365-2745.14093.
Pfeffer, Sabine E., Navinder J. Singh, Joris P.G.M. Cromsigt, Christer Kalén, and Fredrik Widemo. 2021. 'Predictors of Browsing Damage on Commercial Forests – A Study Linking Nationwide Management Data'. *Forest Ecology and Management* 479 (January):118597. https://doi.org/10.1016/j.foreco.2020.118597.
R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
RIS 2023 Riksinventering av skog. Fältinstruktion 2023. [National forest inventory. Field instructions] ; in SLU: Institutionen för skoglig resurshushållning, Institutionen för mark och miljö, https://www.slu.se/globalassets/ew/org/centrb/rt/dokument/faltinst/ris_fin_2 023.pdf
Rolhauser, Andrés G., Donald M. Waller, and Caroline M. Tucker. 2021. 'Complex Trait–environment Relationships Underlie the Structure of Forest Plant Communities'. *Journal of Ecology* 109 (11): 3794–3806. https://doi.org/10.1111/1365-2745.13757.
Roussel, Jean-Romain, David Auty, Nicholas C. Coops, Piotr Tompalski, Tristan R.H. Goodbody, Andrew Sánchez Meador, Jean-François Bourdon, Florian De Boissieu, and Alexis Achim. 2020. 'lidR: An R Package for Analysis of Airborne Laser Scanning (ALS) Data'. *Remote Sensing of Environment* 251 (December):112061. https://doi.org/10.1016/j.rse.2020.112061.
Salman, Hasan Ahmed, Ali Kalakech, and Amani Steiti. 2024. 'Random Forest Algorithm Overview'. *Babylonian Journal of Machine Learning* 2024 (June):69–79. https://doi.org/10.58496/BJML/2024/007.
Sanna Black-Samuelsson Andreas Eriksson Jonas Bergqvist 2020 'The second report on The state of the world´s forest genetic resources Sweden ' https://www.skogsstyrelsen.se/globalassets/om-oss/rapporter/rapporter20222021202020192018/rapport-2020-3-forest-genetic-resources-in-sweden--- 2nd-report.pdf
Smithson, Michael, and Jay Verkuilen. 2006. 'A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent

Variables.' *Psychological Methods* 11 (1): 54–71. https://doi.org/10.1037/1082-989X.11.1.54.
Su, Xiaogang, Xin Yan, and Chih-Ling Tsai. 2012. 'Linear Regression'. *WIREs Computational Statistics* 4 (3): 275–94. https://doi.org/10.1002/wics.1198.
Swedish Land Survey (2020). Digital elevation model http://www.lantmateriet.se
Tompalski, Piotr, Nicholas C. Coops, Joanne C. White, and Michael A. Wulder. 2015. 'Augmenting Site Index Estimation with Airborne Laser Scanning Data'. *Forest Science* 61 (5): 861–73. https://doi.org/10.5849/forsci.14-175.
Wehr, Aloysius, and Uwe Lohr. 1999. 'Airborne Laser Scanning—an Introduction and Overview'. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3): 68–82. https://doi.org/10.1016/S0924-2716(99)00011-8.
Wood, Simon. 2000. 'Mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation'. https://doi.org/10.32614/CRAN.package.mgcv.
Yandell, Brian S. 1993. 'Smoothing Splines-A Tutorial'. *The Statistician* 42 (3): 317. https://doi.org/10.2307/2348812.

Popular science summary

Trees are food for such species as, for example, Moose or Deer. Those animals can cause economic losses by simply feeding themselves. Information about availability of food for them can enhance decision-making aimed at mitigating economic impact of their eating. At the same time availability of this information is limited to sample plots. By combining information from those sample plots with other data, such as data that comes from drone laser scannings, we can create statistical models that allow us to predict the amount of available forage on large areas. After all, it would be impossible to manually inventory entire Sweden. In my work I fitted four types of predictive models for three different species — oak, pine and birch. My goal was to compare accuracy between the resulting models.

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file. If you are more than one author, the checked box will be applied to all authors. You will find a link to SLU's publishing agreement here:

- https://libanswers.slu.se/en/faq/228318.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.