

Predicting natural forest regeneration: a statistical model based on inventory data

Horst Kolo¹  · Donna Ankerst² · Thomas Knoke¹

Received: 14 March 2017 / Revised: 10 July 2017 / Accepted: 25 September 2017
© Springer-Verlag GmbH Germany 2017

Abstract Regenerating forest stands through natural seedlings is a commonly accepted silvicultural strategy in Germany. To plan for and increase natural regeneration within a given stand, foresters must be able to estimate whether the stand in its current state can produce sufficient saplings to replace the overstory. In this paper, we present two approaches to building a model that can estimate the probability of natural regeneration occurring, based on variables that are typically readily available from forest inventories. To estimate model parameters we used the large database of the third National Forest Inventory, which covers forest stands and sites across the whole of Germany, as well as weather and soil data. We examined how these variables impact the emergence of natural regeneration, ultimately fitting a model that can predict the occurrence of natural regeneration in 72% of cases. The influence of the variables on the predicted occurrence of natural regeneration was mixed,

with most stand variables contributing only minor impact and most likely influencing natural regeneration via complex interactions. The exception was vertical structure (number of stand layers), which accounted for a large proportion of the goodness-of-fit of the model. An important finding was that forest ownership structure is a key variable for the prediction of the presence of regeneration. Data from this study support the assumption that some forest owners manage their stands in a way that fosters natural regeneration.

Keywords Natural regeneration · Federal Forest Inventory · Logistic regression · Ownership structure · Number of stand layers

Introduction

In Germany, natural regeneration of forest stands has become a well-accepted alternative to planting. The advantages of natural regeneration are avoiding high costs of planting large numbers of saplings and having an often better adaptation to specific microsites (Burschel and Huss 2003; Nyland 2007). Common planting methods in Germany are estimated to require between 1 and 4 € per seedling (Bauer et al. 2009), with recommendations for planting density ranging between 500 and 8000 seedlings per hectare (e.g., Bavarian Ministry for Nutrition, Agriculture and Forestry 2010). In comparison, natural regeneration occurs in much higher densities. For instance, Jonášová et al. (2010) reported densities up to 67,000 plants per hectare for natural regeneration. However, natural regeneration cannot be planned as accurately as plantations. Additionally, its silvicultural initialization and maintenance may be difficult. This includes, but is not limited to, treatment of parent trees, the opening of the canopy

Communicated by Lluís Coll.

Electronic supplementary material The online version of this article (doi:10.1007/s10342-017-1080-1) contains supplementary material, which is available to authorized users.

✉ Horst Kolo
horst.kolo@tum.de

Donna Ankerst
ankerst@tum.de

Thomas Knoke
knoke@tum.de

¹ Institute of Forest Management, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany

² Chair of Mathematical Modeling of Biological Systems, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany

just enough to create light conditions to foster regeneration or preparation of the seedbed.

Given this background, many models have been built to predict natural regeneration, with different model types developed for different applications. The earliest models include regeneration in an existing growth model (Ferguson et al. 1986). These models did not predict the probability of natural regeneration, but instead the probability that the regeneration exceeds specific heights or diameters and is recruited in the future stand. A tree was regarded as recruited when it is quite certain that it will become part of the future overstory. Such recruitment models have frequently been described. For example, models exist for boreal regions (e.g., Fortin and DeBlois 2007; Lexerød 2005; Lexerød and Eid 2005; Tegelmark 1998), for temperate forests (Klopčič and Boncina 2012; Klopčič et al. 2012) and for tropical forests (Vanclay 1992). Another modeling approach examines the probability that a specific density of seedlings emerges in a stand under given conditions. This approach is favorable for supporting the planning of regeneration because silvicultural management requires a minimum number of seedlings to emerge to ensure a healthy future stand promising reasonable timber quality. Examples of such models are found in Lynch et al. (2003), who modeled regeneration of shortleaf pine in Oklahoma (USA), and Nilsson et al. (2002), who described natural regeneration of Norway spruce, Scots pine and birch in southern Sweden.

This study aims to develop an empirical model to predict natural regeneration of a specific density in Germany. This is of interest for ecological and forest science, as well as for forestry practice, because a sound statistical analysis can reveal important variables that help to predict the emergence of natural regeneration. This is an interesting scientific question as such, but the investigation may also help to improve the efficiency of inventories to record regeneration. In addition, such a model may support practitioners in estimating if a stand is ready and suitable for natural regeneration. To the best of our knowledge, there is no such model for Germany, and internationally no approaches exist that quantify the potential effect of forest ownership structure, comprehensive site variables and forest structure on natural regeneration. The German federal forest inventory (FFI) (Thünen-Institut 2012) provided data for this study. It covers variables describing almost every aspect of forest stands at over 46,500 individual plots. With inventory plots equally distributed across the country, the fitted model covers a wide diversity of sites including different ecological parameters and growing conditions.

In this paper we will show that (1) it is possible to predict natural regeneration with a limited number of variables describing the forest stand and the site conditions, and that (2) there are some key variables that are very important for the occurrence of natural regeneration. In the following

section, we provide a description of the data, variables and applied methods, in “Results” section, the results, and in “Discussion and conclusion” section, a discussion.

Materials and methods

Data for the model development

The data set derived from the FFI (Thünen-Institut 2012) is freely available as an Access database. For this inventory, about 46,614 forest plots were surveyed (Polley et al. 2015). They were distributed evenly throughout Germany and were sampled between 2011 and 2012. The inventory methods are described in Polley (2011). We exported variables from this database, using them either directly or transforming them as described below (See Table 1). For each plot, three different sampling schemes for the different variables are used as a standard. Within a radius of 10 m from the center of the plot, we sampled every tree smaller than 4 m height and the ground vegetation. Trees larger than 4 m in height were sampled with a relascope based on Bitterlich’s angle-count method (Bitterlich 1984). The standard counting factor to measure basal area is four. However, depending on the number of trees counted with the counting factor four, a second angle-count measure with area counting factor of one or two for trees with a dbh (diameter at breast height) larger than 7 cm is used in the FFI. We selected the presence or absence of only the natural regeneration as the dependent variable for our study. The FFI reports the number of naturally regenerated saplings with heights between 20 and 130 cm in a sample plot with a radius of 1 or 2 m. This sample plot is usually located 5 m north of the center of the main plot. The larger radius is used either when no plants within the height range between 20 cm and 50 cm are found in the smaller circle or to count trees with a height larger than 50 cm and a dbh of up to 6.9 cm. Natural regeneration was deemed to be present, with the modeling outcome variable set to one, when at least two seedlings or saplings (i.e., trees between 20 and 130 cm in height) were present in the sample plot with radius 2 m, and was deemed to be absent, with modeling outcome set to zero, otherwise. Details of the inventory methods can be found in Polley et al. (2015). With a plot size of $\sim 12.566 \text{ m}^2$ (2-m radius), we obtained an inclusion probability of a plot of ~ 0.00126 when considering 1 hectare. We achieve upscaling to 1 hectare by dividing the number of recorded trees per plot by means of this inclusion probability. Consequently, two young trees recorded in a plot represent 1591 plants per hectare, when using the 2-m radius. Using 1591 stems per hectare as the minimum density for regeneration is justified because some studies have used densities even below 200 stems per hectare (Stewart et al. 2001), while others used densities between

Table 1 Variables considered for modeling

| Site characteristics | |
|--|---|
| Absolute altitude above sea level ^a | In meter |
| Browsing survey ^e | Whether or not ^f there is a browsing survey in the federal state |
| Exposure ^b | North ^f , East, South, West |
| Natural altitudinal belt ^b | Planar, Colline ^f , sub-Montane, Montane, sub-Alpine |
| Ownership structure ^b | Municipal forest, private forest ^f or state forest |
| Potential natural vegetation ^b | (Base level: Luzulo Fagetum) |
| Protected area ^b | Set to 1 if plots lie within any kind of protected area (e.g., a Natura 2000 area or national park) and 0 ^f otherwise. |
| Shrub cover | Extent to which the ground is covered with either blackberry (<i>Rubus</i> spp.), Fallopia japonica or Jewelweed (<i>Impatiens glandulifera</i>). Levels were 0% ^f , < 10%, 10–50%, > 50% |
| Size of property ^b | 1–20 ha ^f , 21–100 ha, 101–1000 ha, > 1000 ha |
| Slope ^b | Inclination in % |
| Terrain ^b | Hillside, plain ^f , plateau and valley |
| Soil and climate | |
| Annual precipitation ^c | In mm |
| Mean annual temperature ^c | in °C |
| Mean temperature in the vegetation period ^c | Vegetation period from May to September (°C) |
| Mean temperature in winter ^c | Winter from December to February (°C) |
| Precipitation in the vegetation period ^c | Vegetation period from May to September (mm) |
| Soil texture ^d | clay–loam, clay–silt ^f , loam–sand, loam–silt, marsh, normal-loam, other, pure sand, sand–loam, silt–clay, silt–sand |
| Soil type ^d | Alisols, chernozem, floodplain/mud flat, marsh, other soils, Pararendzina, Pelosol, Podsol, Pseudogley/Gley, Renzina, secondary Podsol |
| Stand characteristics | |
| Basal area ^b | in m ² per hectare |
| Main tree species ^b | Gives the most common tree species on a given plot. Every species commonly occurring in German forests. The base level was <i>Picea abies</i>) |
| Mean dbh ^b | Mean diameter at breast height (cm) |
| Mean tree age ^b | in years |
| Mean tree height ^b | in meter |
| Operation mode ^b | Timber forest, Coppice forest with short ^f and medium rotation |
| Number of Stand layers ^b | Converted from FFI into three values: one layer ^f , two layers (FFI reports the levels “2 Layers,” “Achiever over current stand” stand (i.e., trees from the last forest generation which were kept during the harvest due to their extraordinary performance) and selective cutting. The layers had to be a minimum dbh of 6.9 cm to count as an additional layer |
| Stand development phase ^b | dbh < 20 cm ^f , 21–35, 36–50, 51–70 and > 70 cm |
| Stand type ^b | Coniferous (CPS) ^f or deciduous pure stands (DPS), mixed stands with leading deciduous trees (DC) or leading coniferous trees (CD), mixed stands with equal share of coniferous or deciduous trees (DC50) |

^aData taken from Jarvis et al. (2006)

^bData taken from the federal forest inventory

^cData taken from the German Weather Service

^dData taken from the German Federal Institute of Geoscience and Resources

^eData newly created for this analysis

^fIndicates the base level

740 and 2500 stems per hectare as lower thresholds (Hallikainen et al. 2007; Larsen et al. 1997; Lynch et al. 2003). We did, however, test our results with higher thresholds in order to assess the impact of this decision.

As potential explanatory variables, we tested several site characteristics that have been associated with regeneration success in the literature. Table 1 shows their definitions and units. Following other authors, we used *absolute*

altitude above sea level (Dodson and Root 2013; Ferguson 1996; Hyppönen et al. 2005; Príncipe et al. 2014), *slope* (Ferguson 1996; Hallikainen et al. 2007; Príncipe et al. 2014), *exposure* (Ferguson 1996; Hallikainen et al. 2007) and *terrain* (Ferguson 1996; Hallikainen et al. 2007; Hyppönen et al. 2005). Ferguson (1996) linked habitat type with regeneration success. In our study, the *potential natural vegetation* (vegetation type which would be present without human interaction) serves as indicator for the habitat type (see for example Schweiger and Sterba 1997).

In addition to variables commonly used in existing literature, we tested other independent variables that are readily available in the FFI and could potentially influence regeneration. We selected the ground cover percentage of shrubs that is known to hinder natural regeneration: *Blackberry* (*Rubus* spp.), *Jewelweed* (*Impatiens glandulifera*) and *Japanese Knotweed* (*Fallopia japonica*) and merged them into a variable called *shrub cover*, which indicates what percentage of ground is covered by any of these species. We included the variable *browsing survey* (yes/no) because it is possible that game management is different in forests where the browsing stress is regularly assessed. This variable indicates whether a plot is located in a federal State of Germany where such inventories are conducted because not all federal states conduct browsing surveys. *Ownership structure* (state, municipal or private forest) and *size of the property* in which the inventory plot is situated (i.e., how large is the whole property of the owner of the stand) were also included because it is likely that the forest management differs depending on the size of the property and *ownership structure*. These variables were aggregated to the levels given in Table 1. Finally, as the FFI reports many different forms of protected areas, such as Natura 2000 sites or national parks, the variable *protected area* was also tested. This variable indicates whether the plot falls inside any kind of protected land; all forms of protected areas were aggregated to one variable, i.e., protected area yes/no.

Characteristics of the existing forest stand should be the most influential factors on the occurrence of regeneration. Stand variables frequently used to predict regeneration include the height of trees in the overstory (dominant trees), tree diameter at breast height (dbh) (Hallikainen et al. 2007), basal area as an indicator of overstory density (Klopčič et al. 2012; Lynch et al. 2003; Nilsson et al. 2002) and tree age as an indicator of stand maturity (Lynch et al. 2003). *Mean tree age*, *mean tree height* and *mean dbh* were, thus, tested as independent variables and calculated as arithmetic means of all dominant overstory trees at a plot. The variables *number of stand layers*, *stand type*, *main tree species* and *stand development phase* were considered for the model analysis, because they are frequently used in German forest inventories to describe a forest stand. These four variables are all

referring to the main stand, and regeneration is not included in any of them.

In addition, climate data were included. Using data provided by the German Weather Service (2016), we calculated a 30-year mean for *annual precipitation* and *mean annual temperature* for every plot. We also calculated the 30-year *mean temperature in the vegetation period* and the *precipitation in the vegetation period* from May to September [full spring to beginning of the full autumn; defined by the phenological clock (German Weather Service 2017)] and the 30-year mean *temperature in the winter* (from December to February) for every plot. The additional inclusion of the *mean temperature in winter* is justified by the possible influence of very low temperatures on seedling survival. Some of these variables have also been used in other studies (Klopčič et al. 2012).

Hallikainen et al. (2007) and Hyppönen et al. (2005) considered *soil texture* as an independent variable in their study, but did not include it in their final model. *Soil type* is a similar indicator also used in other studies (Miina and Saksa 2013). Both soil type and texture were considered in our analysis. The soil data were obtained from the German Federal Institute of Geoscience and Resources. We used the general soil map (BÜK1000) (Hartwich et al. 1998) and the surface soil map (Stange 2007) for our model. The *absolute altitude above sea level* was extracted from a digital terrain model (Jarvis et al. Jarvis et al. 2006). Soil maps, a digital height model and weather data were available in a georeferenced format, allowing connection with the FFI data via location of the inventory plots using ArcGIS (Esri Inc. 2014). The resulting data set comprised about 46,614 individual plots. After removing plots with incomplete data, approximately 19,000 remained. The data from 20,700 plots have missing values in the variable exposition and are distributed evenly throughout Germany. The remaining missing values are distributed over all variables, but no variable has more than 10% missing values except *size of property* which had 15% missing values. Therefore, we assumed that there is no bias related to the missing values in the excluded data.

Methods for model estimation and testing

We calculated the mean, standard deviation and range for every continuous variable to detect differences between sites with and without regeneration (Table 2). For the categorical variables (factor variables), we reported the total number and relative frequencies of cases where regeneration is present versus absent per factor level. For each factor variable we set the base level to the factor that had the highest relative share of plots where regeneration is absent to analyze the maximum difference. We calculated the Wilcoxon test for the continuous variables and the χ^2 test for the factor variables to show group differences between plot with and without

Table 2 Descriptive Statistics of the used variables

| Continuous variables | Regeneration present mean (SD) (range) ($N = 5530$) | Regeneration absent mean (SD) (range) ($N = 13,310$) | Wilcox test number of cases (out of 1000 test runs) with $p < 0.05$ |
|-----------------------------------|---|--|--|
| Number of seedlings per hectare | 24310 (45590.18) [1592, 620700] | | |
| Basal area (m ²) | 30.29 (14.22) [4, 100] | 32.72 (15.65) [4, 132] | 979 |
| Mean DBH (cm) | 39.71 (14.35) [1.2, 105.60] | 32.74 (13.60) [1.4, 118.1] | 968 |
| Mean tree age (years) | 91.07 (38.79) [10, 385] | 69.65 (36.32) [8, 346.3] | 983 |
| Slope (% inclination) | 25.33 (17.57) [0, 180] | 25.97 (19.61) [0, 180] | 110 |
| Factor variables | Regeneration present number of cases (relative frequency in %) | Regeneration absent number of cases (relative frequency in %) | X^2 test X^2 (p value) |
| Ownership structure | | | |
| Municipal | 2198 (39.75) | 5283 (39.7) | 1272.20 (< 0.001) |
| Private | 2488 (45.0) | 6612 (49.7) | 1868.90 (< 0.001) |
| State | 844 (15.3) | 1415 (10.6) | 144.33 (0.06) |
| Number of stand layers | | | |
| 1 layer | 290 (5.2) | 6669 (50.1) | 5847.30 (< 0.001) |
| 2 layers | 2859 (51.7) | 5337 (40.1) | 749.20 (< 0.001) |
| Selective cutting | 2381 (43.1) | 1304 (9.8) | 314.77 (< 0.001) |
| Stand type | | | |
| Coniferous–deciduous–mixed–stands | 1580 (28.5) | 4533 (34.0) | 1426.50 (< 0.001) |
| Coniferous pure stands | 938 (17.0) | 3150 (23.7) | 1196.90 (< 0.001) |
| Deciduous–coniferous–mixed–stands | 1481 (26.8) | 2810 (21.1) | 399.43 (< 0.001) |
| Deciduous pure stands | 1531 (27.7) | 2817 (21.2) | 380.36 (< 0.001) |

regeneration. Due to the large sample size and the increased power of the Wilcox test, we repeated the Wilcox test 1000 times with a smaller sample of the whole data set ($n = 100$ for each repetition). We reported the number of repetitions where the H_0 of no difference was rejected. A high number of rejections support a significant difference.

Logistic regression was used to predict the probability with which regeneration occurs, yielding estimated coefficients and their significance levels.

$$\text{logit } P(X) = \alpha + \sum \beta_i X_i$$

Exponentiation of the estimated coefficients of the “logit” form leads to odds ratios, which describe the changed odds of the occurrence of natural regeneration for a unit change in continuous variables or for a change in factor levels compared to the base level (Hosmer et al. 2013). All statistical analyses were performed using the R statistical software package (R Core Team 2016).

There are several alternatives to the logistic regression method used in this paper. The most popular are artificial neural networks and linear discriminant analysis. We adopted logistic regression as a statistical method for several reasons. Firstly, unlike discriminant analysis, which can also discriminate datasets with many variables in two or more groups, logistic regression can deal with variables coded as factors. Discriminant analysis also requires that the two

populations have the same covariance matrix (Backhaus et al. 2016; Rencher 2002). While neural networks often have better discrimination than logistic models, they are difficult to interpret because of their “black box” behavior (Dreiseitl and Ohno-Machado 2002) and the tendency to “overfit” (Tu 1996). The advantage of logistic regression over the other two approaches is that results are easy to interpret in the form of the odds ratio, which quantifies the effect of a change in one variable holding all other variables equal.

To select variables for inclusion in the model, we conducted a modified stepwise forward logistic regression, where two goodness-of-fit tests were implemented at each step. The first test, the Hosmer–Lemeshow test, is based on the distribution of the predicted probabilities (Lemeshow and Hosmer 1982). The second one is the Le Chessie–van Houwelingen test, which has the advantage of considering every observation equally and can detect deviations in all directions (Le Chessie and van Houwelingen 1991). The first test we used is implemented in the R package “Resource selection” (Lele et al. 2015) and the second in the R package “rms” (Harrell 2015). To check for the predictive power of the different model versions, we plotted the receiver-operating-characteristic (ROC) curve and calculated the area under this curve (AUC) with the R package “ROCR” (Sing et al. 2005).

The ROC curve plots the proportion of correctly classified cases for situations where regeneration occurred (true

positive rate) over the proportion of cases where regeneration did not occur that it also classifies correctly (specificity = 1 – false positive rate) across varying cut-points of the predicted probability value (Kleinbaum and Klein 2010). The cut-point is the probability threshold that is used to classify regeneration as occurring. For example, an intuitive cut-point would be 0.50. Given this cut-point, for any site with a probability of regeneration beyond 0.50, the presence of regeneration is therefore predicted. The AUC describes the ability of the model to discriminate between cases where regeneration is present versus the cases where regeneration is absent (Hosmer et al. 2013). Additionally, we checked how often the model predicts the presence and absence of natural regeneration correctly.

To check the stability of the estimated coefficients and their significance, we repeated the fitting process 500 times for every forward selection step. For each repetition, we randomly divided the whole data set in a training set (75% of the data) and a test set (25% of the data), not used for estimating the coefficients of the model. Each repetition was checked for its predictive ability with the test data set using the cross-validated AUC. Therefore, we calculated the average AUC together with its standard deviation.

The final model was selected based on the highest predictive ability using the cross-validated AUC and both goodness-of-fit tests. We included interactions between all variables and quadratic and logarithmic transformations of the continuous variables in the forward selection process to check if this improves the predictive ability of the model and the goodness-of-fit tests. This inclusion of transformations was done by replacing one variable per time with its transformed equivalent. Interactions were also included stepwise to check if the model improves. We used the variance inflation factor (VIF) to detect multicollinearity and excluded any variable with VIF values larger than 5. Variables were excluded if the likelihood ratio test of the nested models was not significant in at least 50% of the repetitions. We also evaluated the number of runs where the coefficients were significant. Continuous variables were included only if their coefficient was significant in at least 50% of the runs. Because factor variables had more than one coefficient and the significance of the individual coefficients varied considerably between the runs, we relied on the likelihood ratio test of nested models.

Results

Descriptive statistics

The results in Table 2 show that there was no significant difference in the means of *slope* on plots with and without natural regeneration. The *mean dbh* and *the mean tree age*

were significantly higher on plots with regeneration. The *basal area* was significantly lower on sites where regeneration was present. Regarding the *number of stand layers*, natural regeneration was more common in stands with two layers or more and less common in stands with only one layer. Natural regeneration was more often absent than present on plots within municipal or private forests. Plots within state-owned forests containing natural regeneration were less frequent, but the difference was not significant. Across all *stand types*, there were more plots where regeneration was absent than those with regeneration present.

Model estimation for all seedling species

The predictive model for regeneration obtained by forward selection with the criteria of predictive ability and the goodness-of-fit tests resulted in a model consisting of the variables *absolute altitude above sea level*, *basal area*, *browsing survey*, *mean tree age*, *mean tree diameter*, *mean tree height*, *number of stand layers*, *ownership structure*, *slope*, *soil texture*, *stand type* and *terrain*.

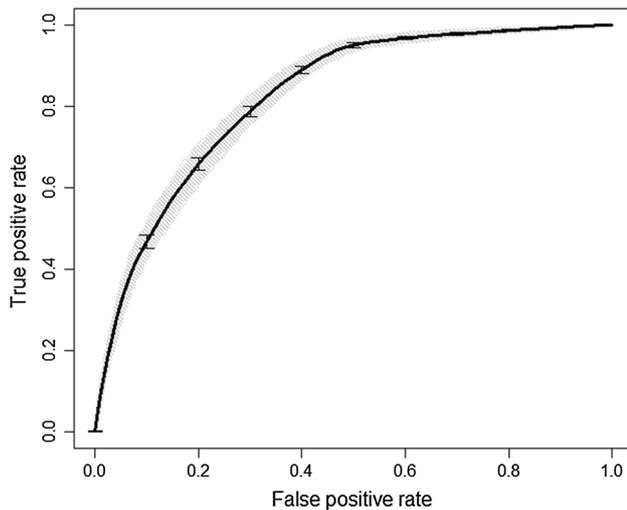
The variables *natural altitudinal belt*, *soil type*, the five climate variables, as well as *stand development phase* were multicollinear (VIF > 5) and were therefore discarded from the analysis.

We excluded the variables *absolute altitude above sea level*, *browsing survey*, *mean tree height*, *soil texture* and *terrain* from the final model because these variables were not significant in almost all repetitions of the fitting process. Because of the possible correlation between the climate variables and the variables describing the terrain, we ran additional tests. One relationship is given by the nature of the data. The climate variables were calculated with a regression of a digital terrain model. Therefore, those variables and the *absolute altitude above sea level* could never be used in one model due to their multicollinearity. We also checked whether the variables *terrain* or *slope* were correlated with the *absolute altitude above sea level*, but we do not find evidence to support that. We did check for possible linear relationships but found none. No interactions or transformations were kept in the final model, because they either did not improve the model or were not significant. Additionally, we recalculated the model using different densities (from 796 to 3000 young trees per hectare) as the minimum threshold for determining regeneration presence (i.e., for deciding whether the dependent variable is set to one or zero). We finally used a threshold of 1500 seedlings per hectare, because with higher thresholds the result of regeneration present slowly became a rare incident.

The obtained model is described in Table 3. All coefficients are significant. The significance level of the factor-level deciduous pure stand is very close to our threshold of

Table 3 Coefficients and odds ratios of the predictive model for regeneration

| Variable | Coefficient (SD) | Z value (p value) | Odds ratio (95% CI) |
|-----------------------------------|------------------|-------------------|----------------------------|
| Intercept | - 3.0089 (0.094) | - 31.88 (< 0.001) | 0.0493 (0.0410, 0.0593) |
| Basal area | - 0.0122 (0.001) | - 9.02 (< 0.001) | 0.9878 (0.09852, 0.9904) |
| Mean dbh | 0.0116 (0.002) | 5.74 (< 0.001) | 1.0116 (1.0077, 1.0157) |
| Mean tree age | 0.0050 (0.001) | 6.49 (< 0.001) | 1.0050 (1.0035, 1.0065) |
| Number of stand layers | | | |
| 2 layers | 2.4528 (0.066) | 36.92 (< 0.001) | 11.6211 (10.2197, 13.2621) |
| Selective cutting | 3.5907 (0.072) | 49.55 (< 0.001) | 36.2588 (31.5062, 41.8616) |
| Ownership structure | | | |
| Municipal forest | - 0.1283 (0.049) | - 3.13 (0.002) | 0.8796 (0.8817, 0.9531) |
| State forest | 0.2659 (0.059) | 3.38 (< 0.001) | 1.2205 (1.0872, 1.3700) |
| Slope | -0.0087 (0.001) | - 8.46 (< 0.001) | 0.9913 (0.9893, 0.9933) |
| Stand type | | | |
| Coniferous–deciduous-mixed-stands | - 0.4656 (0.057) | - 8.10 (< 0.001) | 0.6277 (0.5609, 0.7026) |
| Deciduous–coniferous-mixed-stands | - 0.3096 (0.061) | - 5.09 (< 0.001) | 0.7337 (0.6512, 0.8266) |
| Deciduous pure stands | - 0.1208 (0.001) | - 1.97 (0.049) | 0.9878 (0.9852, 0.9904) |

**Fig. 1** ROC of all repetitions with the final variable selection (gray lines). Mean ROC in black together with the standard deviation of every decile

5%. The odds ratios for the continuous variables indicate the changed odds for an increase in one unit of the variable.

The ROCs of the 500 repetitions with the final variable selection are shown in Fig. 1. The average AUC of the 500 repetitions with the final variable selection was 0.82 (SD = 0.005). Therefore, according to Hosmer et al. (2013), the predictive ability of the final model is very good. The final model predicts the presence of regeneration correctly in 70.01% of all cases in-sample and in 73.51% of all cases of an independent data set. We used the Youden index (Youden 1950) to select the optimal cutoff point at which the model best distinguishes between regeneration presence

and absence. The Hosmer–Lemeshow goodness-of-fit test rejects H_0 of good fit in only 0.6% of the 500 repetitions, and the Le Chessie test rejects H_0 of good fit in only 23% if all repetitions.

The most influential variable is the *number of stand layers*. Increasing the number of layers from one to two raises the probability of regeneration by almost 12 times. Having a third layer increases this probability 36 times compared to a stand with one layer. Figure 2a shows this relation. We carried out further investigations into this high influence of the stand structure. Therefore, we recalculated the model without the factor variable *number of stand levels*. The resulting model is shown in supplement Table 1. The significant coefficients changed. The coefficient of the factor-level municipal forest (*ownership structure*) and coniferous–deciduous-mixed-stands (CD; *Stand type*) changed to nonsignificant. The odds ratio, and with it the influence of the variable *stand type*, increased strongly up to a 60% difference between the odds ratio of the original and the model without the variable *stand layers*. In the case of the *stand type*, in the new model the mixed stands had a positive influence on the probability of finding natural regeneration. The odds ratios of *ownership structure* changed moderately by 10% for municipal forests and about 20% for state forests. The other variables increased only moderately up to 1% in magnitude, but the direction of the influence stayed the same (i.e., all odds ratios kept their positive or negative influence like in the original model). The more reliable Le Chessie goodness-of-fit test rejected the H_0 of good fit for the new model. In-sample, the AUC dropped to 0.69 and the correct predictions decreased to 63%. Out-of-sample, the AUC of the new model was 0.69 and the correct discrimination of the model was 65%. In summary, the model worsened when

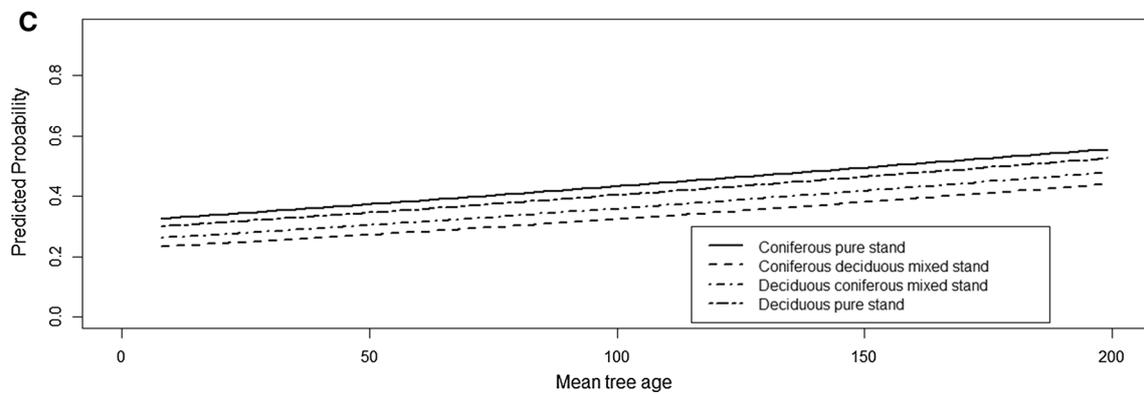
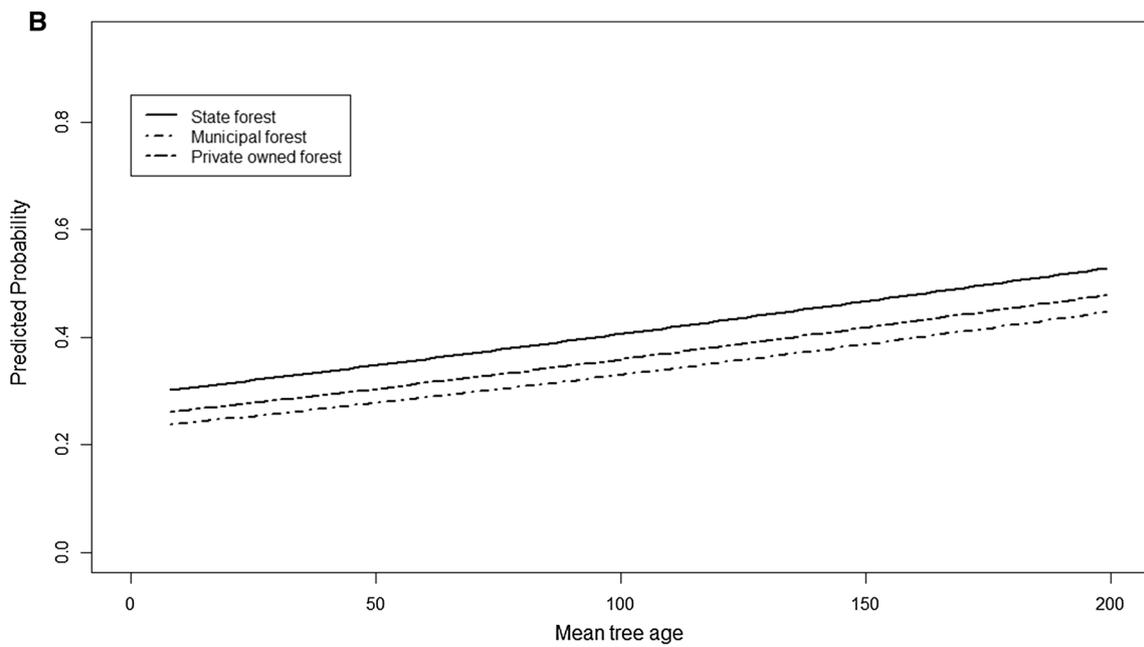
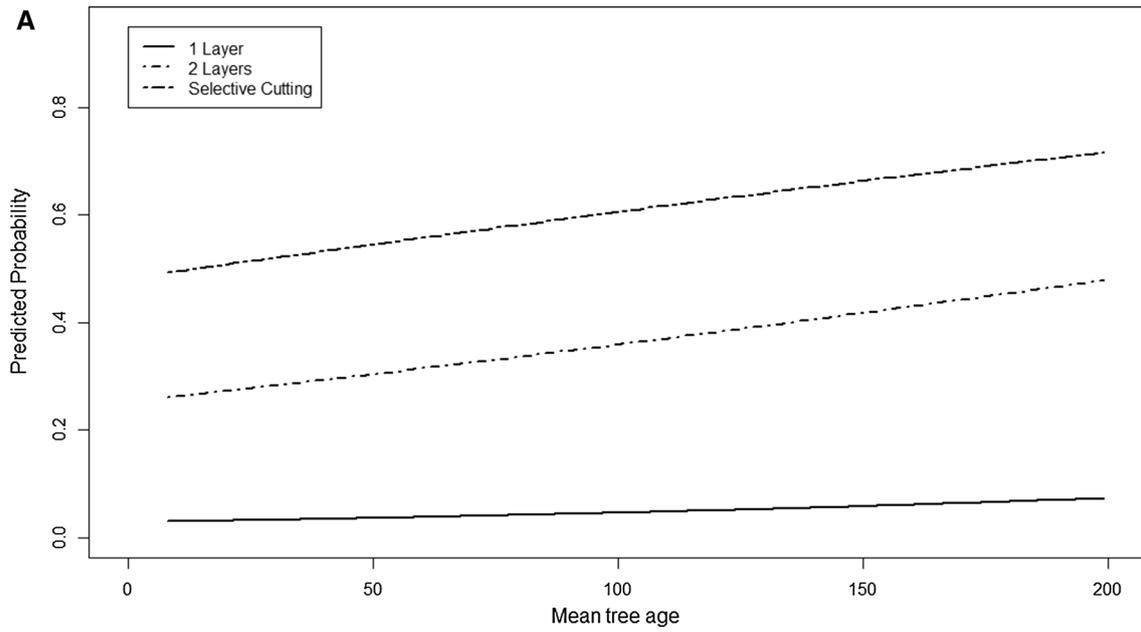


Fig. 2 Predicted probability of natural Regeneration (all continuous variables set to their mean; all factors set to their most frequent factor level; predictions are plotted up to an age of 200 years, because there are very few cases with higher age). **a** Variable *ownership structure* varies; **b** variable *number of stand layers* varies; **c** variable *stand type* varies

excluding the stand vertical structure and the goodness-of-fit test revealed that the reduced model adaptation to our data was worse than the original one.

The *ownership structure* explained regeneration probability to a certain extent. Stands in state forests had a 22% higher probability of the occurrence of regeneration compared to private forests, whereas municipal forests have a lower probability than private forests (see Fig. 2b). Different management by the different owners could possibly explain different stand structures expressed by the *number of stand layers*. To explore a possible correlation between *ownership structure* and *number of stand layers*, we used a χ^2 test to check the independence of those two variables. Given the large size of our data and the sensitivity of the χ^2 test to large sample sizes, we drew a smaller sample of 500 cases out of our whole data. We calculated the χ^2 test with these 500 cases. This process was repeated 1000 times. The mean of the p values of those repetitions was 0.21, which led to the conclusion that there is no relationship between those two variables. To check the strength of the correlation, we calculated the Cramer's (1999). It returned a result of 0.07 and can be interpreted as almost no correlation.

Mixed stands have a lower probability of the presence of natural regeneration compared to pure coniferous stands, whereas pure deciduous stands have almost equal probabilities (see Fig. 2c and Table 3). An increasing slope decreases the odds of finding natural regeneration. The variables that describe the maturity of the stand, *dbh* and *mean tree age*, improve the probability of occurrence of natural regeneration. A stand which is 10 years older than another has a 5% higher probability of natural regeneration.

The variable *basal area* describes the density of a stand. Our model shows that stands with higher *basal area* had an odds ratios below one, which indicates that the probability of natural regeneration occurring is lower in denser stands. Figure 3 shows the distribution of the age classes of our data.

Model estimation for single tree species

To explore the possibility of predicting natural regeneration of a specific species, we fitted individual models for each species which had sufficient complete data sets for a regression. We used the same forward selection process described before. As in the original model, we tried to include interactions between those variables, but none of them were significant or improved the predictions. The

detailed results are given in the supplement Table 2–7. The model for Norway Spruce (*Picea abies*; Supplement Table 2) consisted of 4 variables, namely *annual precipitation*, *mean tree height*, *number of stand layers* and *stand type*. All coefficients except that of the factor-level deciduous pure stand were significant. In this model, the influence of the number of stand layers is lower than in the overall model. For spruce, mixed forests lead to lower odds of finding natural regeneration. The mean tree height in this model serves as proxy for the site quality, and therefore, a higher mean tree height leads to higher probabilities of finding natural regeneration. Interestingly, in this model, a climate variable, the *annual precipitation*, is included and higher precipitation raises the odds for emerging natural regeneration.

The model for Scots pine (*Pinus sylvestris*, supplement Table 3) consisted of *mean temperature during the vegetation period*, *number of stand layers* and *stand type*. As with the spruce model, all coefficients except one of the factor-level deciduous pure stands were significant. Here, the number of stand layers had a similar, large influence on the odds of finding regeneration compared to the overall model. As with spruce, mixed forests lead to lower odds of finding natural regeneration. Like with spruce, this model contains a climate variable, the mean temperature in the vegetation period, which raises the odds for emerging natural regeneration.

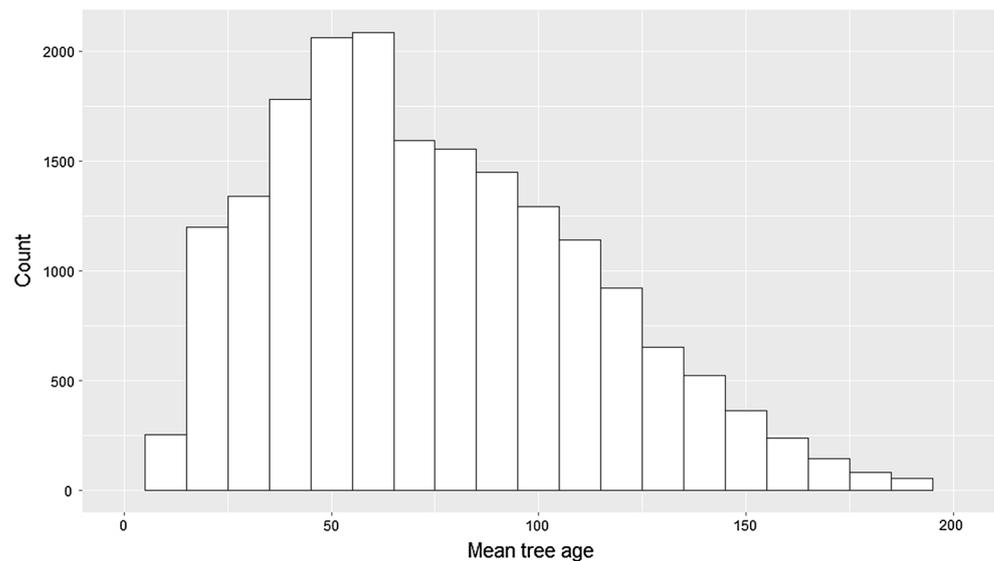
The model for silver fir (*Abies alba*, supplement Table 4) is the smallest single-species model. Its only independent variables were *mean tree age* and *number of stand layers*.

Adding additional variables did not improve the model. Both variables had a positive influence on the probability of finding natural regeneration. All coefficients were significant to a very high level.

Our prediction model for European beech (*Fagus sylvatica*; supplement Table 5) is the largest single-species model. It included *mean tree age*, *mean tree height*, *precipitation in the vegetation period*, *number of stand layers* and *stand type*. *Mean tree age* and *mean tree height* both had odds ratios above one, indicating that age as an indicator for maturity and height as a proxy for the site quality, both positively influenced the emergence of natural regeneration. The variable *precipitation in the vegetation period* had a positive influence too. All coefficients were significant.

The model for oaks is somewhat different, because it is fitted to data from three oak species (*Q. robur*, *Q. petraea* and *Q. rubra*). This was done because there were not a sufficient number of complete cases for any of the three individual species to complete an analysis. Given the approximately similar ecological niche of the three species, we decided to combine them in one data set. The resulting model is shown in supplement Table 6. This model consisted of three variables, the *mean temperature during the vegetation period*,

Fig. 3 Age distribution of the training data set. Values over 200 years ($n = 39$) were omitted for a better presentation



the number of stand layers and the stand development phase. All coefficients had a positive influence.

The last single-species model was fitted for sycamore maple (*Acer pseudoplatanus*; supplement Table 7). It included the variables the number of stand layers, stand development phase and stand type. All coefficients except the stand development phase dbh > 70 cm and the stand type coniferous–deciduous–mixed-stands were significant to a high level.

The predictive abilities, the number of complete cases and the results of the goodness-of-fit tests for all single-species models are summarized in Table 4.

Discussion and conclusion

Plausibility of our findings for the overall model

A key finding from the factor analysis is the difference in natural regeneration between different *ownership structures*. After adjustment for other factors, the odds of finding regeneration increased by 22% in state-owned forests compared to private forests. There are two possible explanations for this increased probability of regeneration. First, state-owned forests may be situated in areas with climate and site conditions that are more conducive for natural regeneration. Of the variables included in the initial model, *potential natural vegetation*, *soil type* and *soil texture* describe the growing conditions regarding the available water and nutrients and overall site characteristics. Although not all of these variables are part of the final model, they can serve as indicators to compare the stands of the different *ownership structures*. Comparing the relative frequencies of these factors across

Table 4 Summary of the performance of the single-species model in- and out-of-sample together with the number of complete cases and the results of the Hosmer–Lemeshow test (HL) and the Le Chessie test (ICh)

| Species | In-sample performance | | Out-of-sample performance | | Number of complete cases | NNumber of repetitions where the HL rejects H0: good fit (%) | NNumber of repetitions where the ICh rejects H0: good fit (%) |
|----------------|-----------------------|----------------------------|---------------------------|----------------------------|--------------------------|--|---|
| | Mean AUC | Mean correct specification | Mean AUC | Mean correct specification | | | |
| Norway spruce | 0.71 | 0.67 | 0.72 | 0.67 | 2837 | 94 | 4 |
| Scots pine | 0.76 | 0.69 | 0.76 | 0.69 | 836 | 8 | 26 |
| Silver fir | 0.68 | 0.65 | 0.66 | 0.63 | 1168 | 3 | 81 |
| European beech | 0.77 | 0.71 | 0.77 | 0.71 | 5351 | 13 | 10 |
| Oaks | 0.68 | 0.63 | 0.68 | 0.62 | 3744 | 29 | 6 |
| Sycamore maple | 0.71 | 0.69 | 0.7 | 0.64 | 2649 | 2 | 1 |

ownership structures revealed that all *ownership structures* were located on similar sites. Relative frequencies differed by less than 6% for all site factors. The only exception was the *soil texture* clay–silt, where the proportion is much lower in state-owned forests than in municipal or private forests (25 and 18%, respectively). The variables *slope* and *absolute altitude above sea level* did not differ significantly, which also indicates that there is no difference regarding the sites.

Given the lack of difference in site characteristics described by the variables available from the FFI, a second explanation for the higher probability of the occurrence of natural regeneration in state-owned forests could be that these forests are managed in a way that promotes natural regeneration. It is worth mentioning that the odds of finding natural regeneration are lower in municipal forests than in private forests, which suggests that the management in municipal forests is not as focused on promoting natural regeneration. Here, the relative frequencies of the site characteristics also differed by less than 6%, again suggesting that management objectives, rather than site characteristics, are responsible for the difference in regeneration levels. This is unexpected because, like state-owned forests, municipal forests are usually managed by trained foresters who would be expected to manage stands in a way that fosters natural regeneration.

The second key finding of this study is the effect of *the number of stand layers* on natural regeneration. While the relationship is itself not surprising, the magnitude of the effect is unexpected. A forest stand with two layers was 12 times more likely to contain natural regeneration than a stand with one layer, while stands with selective cutting were 36 times more likely to contain regeneration than a single layer stand. According to our model, the vertical stand structure is the most important factor determining the ability of a stand to produce natural regeneration. Although it is a crucial variable for the model, the model still performs reasonably well without the variable.

The coefficients and the corresponding odds ratios of the remaining variables are mostly as expected and similar to those reported in other studies. The negative impact of the slope on natural regeneration found in this study is in line with findings of Ferguson (1996) and Schweiger and Sterba (1997). Steeper slopes increase the speed of water runoff and, in higher altitudes, the snow offset, which hinders the successful establishment of seedlings.

The *basal area* of a stand describes the density of trees. It is clear that a higher stand density prevents regeneration through shading and competition for water and soil nutrients. This finding is consistent with the studies of Larsen et al. (1997) and Lynch et al. (2003). Although this is the usual relationship between basal area and seedling survival and growth in middle European ecosystems, this relationship can be reversed under certain conditions. Lucas-Borja et al.

(2016) found evidence that a medium tree density could foster seedling emergence at least in drier years. This shows that a high basal area could help in retaining good conditions for saplings and seedlings under very dry and hot conditions. The odds ratio for *mean dbh* is above one.

Therefore, holding all other variables equal, a higher dbh increases the predicted probability of natural regeneration. Klopčič et al. (2012) reported the opposite relationship and reasoned that a high stem diameter lowers light availability on the ground, therefore, lowering the chances of regeneration emerging. On the other hand, the dbh is a measure of maturity and higher dbh's can indicate an increased ability to produce offspring.

The model result of a positive influence of *mean tree age* on regeneration probability is as expected. *Mean tree age* can also be a measure of stand maturity and is therefore an indicator of its ability to produce offspring. Lynch et al. (2003) also reported a positive effect of tree age on natural regeneration.

The lower odds for the mixed *stand types* compared to the base level of pure coniferous stands were unexpected. Other studies report a positive relation between species richness and the number of seedlings (Liira et al. 2011). On the other hand, Pretzsch (2014) reported that the crown density is much higher in mixed stands. This reduces available light on the forest floor and therefore hinders natural regeneration. In mixed stands, at least one coniferous and one deciduous species is present in the overstory, which could indicate that increased competition in the crown reduces the occurrence of regeneration. In Fig. 1c, the predicted probabilities for the deciduous pure stands are slightly lower (less than 2%) than those for coniferous pure stands. But the difference is very small and the significance level is very close to the 5% threshold. This pattern was also reflected in the odds ratios. It is therefore reasonable to conclude that probability of finding natural regeneration is highest in pure stands.

The relatively small odds ratios of most variables characterizing the stand (*basal area*, *mean dbh* and *mean tree age*) indicate that the individual impact of single factors is not that important. It instead seems that the mechanism for regeneration is more complex and bound to multiple variables. Hallikainen et al. (2007) reported a similar complexity in the relationship between natural regeneration and different variables. This suggests that natural regeneration is not influenced by single factors alone. The exception is the vertical structure of the stand, where the number of layers within the stand strongly changed the odds of regeneration being present.

Plausibility of our findings for the single-species model

Within the Norway spruce model, the odds ratios of all variables and factor levels had the expected sign in each model.

The odds ratio of the annual precipitation was slightly above one. Given that the odds ratio refers to the change in one mm precipitation, it becomes clear that this variable indicates the optimal precipitation of spruce is above 500 mm (Schütt 2008). The mean tree height serves as an indicator for tree maturation and the growing condition on the given site. An increase of 1 m in height increases the odds of finding regeneration by 2%. The odds ratios of the numbers of stand layers show that the probability increases five times if a second layer is present and by seven times when more than two layers are present. These values are lower than the ones of the model for all species. This may indicate the increased need for light compared to some other species like the European Beech and the Silver Fir. The odds ratios of the factor levels coniferous–deciduous-mixed-stands and deciduous–coniferous-mixed-stands of the variable stand type showed that the spruce regenerates best in pure stands and the regeneration becomes less likely in mixed stands with leading coniferous trees and even less likely in mixed stands with leading deciduous stands. The fact that the odds ratio of deciduous pure stands is higher than the mixed stand can be attributed to the low number of cases within deciduous pure stands ($n = 147$) and the fact that this coefficient is not significant.

The odds ratios of all variables and factor levels of the model for Scots pine had the expected sign, except the coefficient of the factor-level deciduous pure stand. The odds ratios of the variable *stand type* are getting smaller when deciduous species became admixed. The fact that deciduous pure stands had a bigger odds ratio than mixed stands with leading deciduous trees can be attributed to the very low number of cases within deciduous pure stands ($n = 27$). The extreme high values of the odds ratios for 2 layers and selective cutting can be attributed to the ecological pattern of the Scots pine. As a pioneer species, Scots pines reach maturity very early. This means in this context that a second layer can produce offspring very early (Schütt 2008). This increased seed availability increases the probability of natural regeneration. In addition to that, selective cutting creates gaps where direct sunlight is available on the ground and provides the conditions a pioneer species needs. The odds ratio of 1.59 for the *mean temperature in the vegetation period* is the only proxy variable describing the site condition in this model. In middle Europe, the Scots pine is competitive on dry sites with fast water drainage. The mean temperature in the vegetation period is negatively correlated with the precipitation in the vegetation period (correlation coefficient of -0.61). This shows that, within our data, a higher mean temperature in the vegetation period occurs on sites where the precipitation in the same period is low. This explains the high odds ratio as a higher temperature favors the regeneration of Scots pine.

The odds ratios in the model for silver fir regeneration had the expected signs, but the magnitude of the odds ratios

for the variable *number of stand layers* was lower than expected compared to the magnitude in the other models for tree species and the overall model because the silver fir is well adapted to dense stands with low available light on the ground. The odds ratio for the *mean tree age* is in line with the other models.

The signs and the magnitude of the odds ratios in the model for European beech were as expected. *Mean tree age* and *mean tree height* are indicators for the maturity and the site quality, respectively, and therefore increase the odds of finding beech seedlings. The positive sign of the odds ratio of *precipitation in the vegetation period* shows the affinity of the European beech for fresh or moist soils (Schütt 2008). The rising odds ratios for the *stand types* with more deciduous wood is understandable, as beech is well adapted to dense stands with low light availability on the ground. The rise of the probability of finding natural regeneration in stands with two or more layers is without surprise.

The signs of the coefficients in the oak model are as expected. Adding a second layer, or more than two layers, results in higher odds for finding natural regeneration. Later stand development phases lead to higher probability of finding emerging regeneration and can be seen as an indicator of stand maturity and an increased ability to produce offspring. The positive influence of the *mean temperature during the vegetation period* shows the ability of oaks to cope with warmer conditions. It also can be seen as an indicator of the vulnerability of oak seedlings to late frost (Schütt 2008).

The model for sycamore maple seedlings behaved as expected. The *number of stand layers* increased the odds of finding natural regeneration. The *stand development phase* increased the odds too. Both are understandable, as the number of stand layers increases the potential available seeds and the stand development phase can be viewed as an indicator for stand maturation. The lower odds for the phase of > 70 cm contrast with the expectation that larger, and therefore older, stands produce more offspring. However, given that the coefficient is not significant, it can be ignored. The raised odds from coniferous–deciduous-mixed stands to deciduous pure stands is straight forward because oak seedlings are dependent on available light on the ground and mixed stands are often denser than pure stands (Pretzsch 2014).

Limitations and applicability of the overall model

Although we included many variables in our initial model, there are still many which are not considered. Hyppönen et al. (2005) stated that differences in the variables included in different models make it difficult to directly compare different studies. For practical application, the correct prediction in 73.51% of the cases with an independent data set is quite high. In the existing literature, few models have had

the same aim as this study, i.e., to predict the probability of natural regeneration occurring in a specific density. Of the studies that did aim to predict the occurrence of natural regeneration, Hallikainen et al. (2007) reached a correct discrimination in 61% of the cases, while Schweiger and Sterba (1997) reported a correct classification rate of 87%. Because of the large number of plots used for the prediction in this study and the distribution of those plots across the whole of Germany, the model can be seen as a general model for the prediction of natural regeneration in Germany.

In order to increase applicability, we calculated a model with variables which can be obtained within a stand without measurement. We included the variables *number of stand layers*, *stand type* and *stand development phase*. The model performs well compared to the single-species models. It had an average AUC 0.8 and an overall correct classification of 0.72. The adaptation to the data as indicated by the goodness-of-fit test was poor. This means that even if the overall predictions are good, there are substantial problems with the model (Hosmer et al. 2013). The Hosmer–Lemeshow test rejected the H0: good fit in 100% of the cases and the Le Chessie test rejected H0 in 15% of all cases. The coefficients were all significantly different from zero and the odds ratios for the factor levels of *number of stand layers* and *stand development phase* indicated an increase in the probability of finding natural regeneration. The exception was the factor level $\text{dbh} > 70$ cm of the variable *stand development phase*, which showed a decrease in the odds. Although the coefficient is significant, there were only a small number of data sets that fell in this category ($n = 273$). This could have led to this discrepancy. The odds ratios of the variable *stand type* showed that pure stands are more likely to contain natural regeneration. The complete model is described in supplement Table 8. Despite the statistical shortcomings of the bad adaptation, this model could be used as a field predictor to estimate the suitability of a stand for natural regeneration.

Another potential shortcoming is that the model is not able to quantify the impact of former management changes as this information is not available. It is, however, able to show the consequences of future management decisions because they would change one or more variables. For example, basal area, mean tree age, mean tree diameter, mean tree height, number of stand layers and stand type are all related to the management. If rotations or the intensity of silvicultural operations would change, stand characteristics would change. Similarly, if tree species would change, we would get other stand types. The effect of changed management in the future could be derived easily from the odds ratios. One could see which changes are necessary to increase the likeliness of emerging regeneration. With this knowledge, one could determine the effects of changed management. For instance, if the goal is natural regeneration of a stand, one could see that decreasing the basal area would lead to a

higher probability of emerging natural regeneration. Another example would be as simple as to wait for higher ages. This too would lead to higher probabilities of regeneration. As our data set does include plots with and without regeneration, the model is able to predict the absence of regeneration too, as the probability would be estimated below the given threshold and the model would predict the absence of natural regeneration.

Further limitations of our overall model as well as for the single-species models are listed in Supplement 1.

Comparison of the models

The main difference between the overall model and the models fitted for individual species is the number of variables. The largest single-species model is about 40% smaller than the overall model, making the species models easier to apply because less data are needed. Another advantage of the single-species models is that the used variables can be measured in the field or are usually known by forest personnel. In particular, the variable *ownership structure* cannot be estimated within the stand. The same holds true for the variable *slope*. However, the reduction in model size and the easy to estimate variables bring a decreased ability to predict the presence of natural regeneration. Depending on the tree species, the mean out-of-sample AUC is up to 16% lower. The percentage of out-of-sample overall correct specification went down to 63% which is 10% points lower than the overall model. The difference to a random guess (i.e., a correct specification in 50% of the cases) became smaller with some single-species models.

A possibly even more important factor is that the single-species models are not well adapted to the whole range of possible ecological habitats. For instance, there are only two federal states where the number of complete cases for Scots pine exceeded 100. For sycamore maple, only 4 out of 16 federal states had more than 100 complete cases. This makes it hard to apply those models in other federal states, even if the conditions there would fit the requirements of those species. Only the models for oaks and the European beech had sufficient complete cases in almost all federal states. This is in line with their large ecological amplitude. The models for Norway spruce and Silver fir are adopted to data in those federal states to which these species are limited due to their climatic requirements.

These two points favor the use of the overall model, whenever the information necessary for the overall model is available, and the question is if a stand is suitable for natural regeneration in its current state. Within the mentioned limitations, some of the single-species models can be applied whenever it should be determined if a specific species should be capable to produce natural regeneration in a specific stand. The models for European beech are almost as

good as the overall model. Although the predictive abilities of the model for Scots pine are equal to those of the overall model, its adaptation is specific to environmental conditions in Brandenburg and Lower Saxony.

Conclusion

In almost 75% of the cases, our overall model correctly predicted the occurrence of natural regeneration. The area under the receiver-operating-characteristic curve shows very good discrimination. Therefore, we conclude that such a model can satisfactorily predict natural regeneration. Larsen et al. (1997) suggest that such models may be used to evaluate the potential success of natural regeneration in a given stand, while practitioners could use it to evaluate how specific silvicultural changes in the stand structure could improve regeneration success.

The single-species models are partially suitable for the prediction. Those for European beech and Scots pine are as good as the overall model. The remaining models can be used for prediction, but the models for oaks and the silver fir are much worse than the overall model. They can be used, but the results should be treated with caution.

Although the model uses a wide variety of variables for its prediction, it is still applicable for practitioners because the variables included in the models that describe the forest stand are commonly reported in forest inventories. The climate variables are commonly known by professional foresters. Furthermore, we have shown that there are some key stand and site characteristics that strongly influence the occurrence of natural regeneration. Among the influence of the existing stand and the site, which is frequently reported in the literature (e.g. Hallikainen et al. 2007; Hyppönen et al. 2005; Klopčič and Boncina 2012; Lynch et al. 2003; Miina and Saksa 2013), we have shown the importance of *ownership structure*. As discussed above, the influence of *ownership structure* is likely to be linked to different management. Especially in German state forest management, there is a strong commitment to seminatural silviculture (Bavarian State Forests 2008; Landesbetrieb Forst Baden-Württemberg 2014).

Within the context of seminatural silviculture, forest management must consider aspects of game management. As stated in the silvicultural and hunting rules of several state forests (e.g., Faltl 2011; Staatsbetrieb Sachsenforst 2014; ThüringenForst 2015), hunting is an integral part of seminatural silviculture. It is necessary that the game population is kept at a level that allows natural regeneration to emerge. Repeated ungulate browsing can lead to dieback of seedlings and saplings and hinder natural regeneration (Harmer 2001; Kuijper et al. 2010). This is particularly the case with seminatural silviculture, because it creates stands with admixed

species such as fir, beech, oak, ash and maple, which deer selectively browse (Hothorn et al. 2008). Reduced deer densities can therefore increase seedling survival (Gaston et al. 2008; Tremblay et al. 2007). Adaptive hunting strategies can improve conditions for tree regeneration in a very short time period (see Hothorn and Müller 2010), making it likely that state forests have already reduced ungulate densities to a level that allows for successful regeneration.

Although one could argue that different silvicultural management leads to the higher probability of finding natural regeneration in state-owned forests, it seems reasonable to conclude that the hunting regime in state forests, as part of the overall management, is also responsible for the higher probability of occurrence of natural regeneration.

Our model describes the probability of natural regeneration based on the condition of the stand and site characteristics. Ungulate browsing is not considered as a variable in this model. Nevertheless, in cases where the model predicts a high potential (high probability) of natural regeneration occurring, yet no regeneration is found in the given stand, it is likely that ungulate browsing is limiting natural regeneration. With this theoretical link, our model can be used as an indicator to identify areas where high densities of ungulates may be hindering natural regeneration. Therefore, our model cannot only evaluate the suitability of a forest stand for natural regeneration, but it can also help to objectively assess the impact of ungulate browsing on natural regeneration.

Acknowledgements Supported by the German Federal Agency for Nature Conservation (BfN) with resources of the Federal Programme for biological diversity (Bundesprogramm Biologische Vielfalt) from the German Federal Ministry for Environment, Nature Conservation, Construction and Nuclear Safety (BMUB). We also thank Elizabeth Gosling and Chelsea Jones for language editing.

References

- Backhaus K, Erichson B, Weiber R, Plinke W (2016) Discriminant analysis. In: Backhaus K, Erichson B, Plinke W, Weiber R (eds) *Multivariate Analysemethoden*. Springer, Berlin, pp 215–282. doi:10.1007/978-3-662-46076-4_5 (in German)
- Bauer W, Braun W, Braunger M, Dörr T, Ebinger T, Göckel C, Karopka, M, Mann P, Morell M, Schmid R, Thumm H, Wieners M (2009) Seedlings and planting. Ministry for Nutrition and Countryside (in German)
- Bavarian Ministry for Nutrition, Agriculture and Forestry (2010) Planting and nursing forest stands (in German)
- Bavarian State Forests (2008) Silvicultural principles of the Bavarian State Forest (in German)
- Bitterlich W (1984) The relascope idea. Relative measurements in forestry. Commonwealth Agricultural Bureaux, Farnham Royal, Slough
- Burschel P, Huss J (2003) Compendium of silviculture. A guideline for academic studies and practice, 3rd edn. Ulmer, Stuttgart (in German)

- Cramér H (1999) *Mathematical methods of statistics*. Nineteenth print. Princeton University Press, Princeton (Princeton landmarks in mathematics and physics)
- Dodson EK, Root HT (2013) Conifer regeneration following stand-replacing wildfire varies along an elevation gradient in a ponderosa pine forest, Oregon, USA. *For Ecol Manag* 302:163–170. doi:10.1016/j.foreco.2013.03.050
- Dreiseitl S, Ohno-Machado L (2002) Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform* 35(5–6):352–359. doi:10.1016/S1532-0464(03)00034-0
- Esri Inc (2014) ArcGIS. Esri Inc., Redlands
- Faltl W (2011) Principles of hunting in the Bavarian State Forest. Bayerische Staatsforsten AöR, Regensburg (in German)
- Ferguson DE (1996) Modeling natural regeneration establishment in the northern Rocky Mountains of the U.S.A.. In: Conference on modelling regeneration success and early growth of forest stands, pp 30–40
- Ferguson DE, Stage AR, Boyd RJ (1986) Predicting regeneration in the grand fir-cedar-hemlock ecosystem of the northern Rocky Mountains, vol 26. Forest science. Monograph. Society of American Foresters, Washington
- Fortin Mathieu, DeBlois J (2007) Modeling tree recruitment with zero-inflated models: the example of hardwood stands in Southern Quebec, Canada. *For Sci* 53(4):529–539
- Gaston AJ, Sharpe ST, Stockton S, Golumbia TE, Martin JL (2008) Reduction in deer numbers on Reef Island and SGang Gwaay: progress, results, and vegetation changes. In: Gaston AJ, Golumbia TE, Martin J-L, Sharpe ST (eds) *Lessons from the Islands: proceedings from the Research Group on Introduced Species 2002 symposium*
- German Weather Service (2016) Climate data. ftp://ftp-cdc.dwd.de/pub/CDC/grids_germany/monthly/. 11 Oct 2016
- German Weather Service (2017) The phenological clock. <http://www.dwd.de/DE/klimaumwelt/klimaueberwachung/phaenologie/produkte/phaenouhr/phaenouhr.html?nn=575800>. 23 Feb 2017
- Hallikainen V, Hyppönen M, Hyvönen J, Niemelä J (2007) Establishment and height development of harvested and naturally regenerated Scots pine near the timberline in North-East Finnish Lapland. *Silva Fenn*. doi:10.14214/sf.308
- Harmer R (2001) The effect of plant competition and simulated summer browsing by deer on tree regeneration. *J Appl Ecol* 38(5):1094–1103. doi:10.1046/j.1365-2664.2001.00664.x
- Harrell F (2015) rms: regression modeling strategies. <https://CRAN.R-project.org/package=rms>. 11 Oct 2016
- Hartwich R, Behrens J, Eckelmann W, Haase G, Richter A, Roeschmann G, Schmidt R (1998) Soil map of the Federal Republic of Germany 1:1000000(BÜK 1000). Hannover
- Hosmer DW, Lemeshow S, Sturdivant RX (2013) *Applied logistic regression*, 3rd edn. Wiley series in probability and statistics
- Hothorn T, Müller J (2010) Large-scale reduction of ungulate browsing by managed sport hunting. *For Ecol Manag* 260(9):1416–1423
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biom J* 50(3):346–363. doi:10.1002/bimj.200810425
- Hyppönen M, Alenius V, Valkonen S (2005) Models for the establishment and height development of naturally regenerated *Pinus sylvestris* in Finnish Lapland. *Scand J For Res* 20(4):347–357. doi:10.1080/02827580510036391
- Jarvis A, Reuter HI, Nelson A, Guevara E (2006) Hole-filled seamless srtm data V3. version: International Centre for Tropical Agriculture (CIAT). Online verfügbar unter <http://srtm.csi.cgiar.org>
- Jonášová M, Vávrová E, Cudlín P (2010) Western Carpathian mountain spruce forest after a windthrow: natural regeneration in cleared and uncleared areas. *For Ecol Manag* 259(6):1127–1134. doi:10.1016/j.foreco.2009.12.027
- Kleinbaum DG, Klein M (2010) *Logistic regression: a self-learning text*, vol Statistics for biology and health, 3rd edn. Springer, New York
- Klopčič M, Boncina A (2012) Recruitment of tree species in mixed selection and irregular shelterwood forest stands. *Ann For Sci* 69(8):915–925. doi:10.1007/s13595-012-0224-1
- Klopčič M, Poljanec A, Boncina A (2012) Modelling natural recruitment of European beech (*Fagus sylvatica* L.). *For Ecol Manag* 284:142–151. doi:10.1016/j.foreco.2012.07.049
- Kuijper DP, Jędrzejewska B, Brzeziecki B, Churski M, Jędrzejewski W, Żybura H (2010) Fluctuating ungulate density shapes tree recruitment in natural stands of the Białowieża Primeval Forest, Poland. *J Veg Sci* 21(6):1082–1098. doi:10.1111/j.1654-1103.2010.01217.x
- Landesbetrieb Forst Baden-Württemberg (2014): Principles of statewide types of forest development. Landesbetrieb Forst Baden-Württemberg. http://www.forstbw.de/fileadmin/forstbw_infothek/forstbw_praxis/wet/ForstBW_Waldentwicklung_web.pdf (in Germany). 11 Oct 2016
- Larsen DR, Metzger MA, Johnson PS (1997) Oak regeneration and overstory density in the Missouri Ozarks. *Can J For Res* 27(6):869–875. doi:10.1139/x97-010
- Le Chessie S, van Houwelingen JC (1991) A Goodness-of-Fit Test for Binary Regression Models, Based on Smoothing Methods. *Biometrics* 47(4):1267–1282. doi:10.2307/2532385
- Lele SR, Keim JL, Solymos P (2015) ResourceSelection: resource selection (probability) functions for use-availability. <https://CRAN.R-project.org/package=ResourceSelection/>. 11 Oct 2016
- Lemeshow S, Hosmer DW (1982) A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 115(1):92–106
- Lexerød NL (2005) Recruitment models for different tree species in Norway. *Forest Ecol Manag* 206(1–3):91–108. doi:10.1016/j.foreco.2004.11.001
- Lexerød N, Eid T (2005) Recruitment models for Norway spruce, Scots pine, birch and other broadleaves in young growth forests in Norway. *Silva Fenn*. doi:10.14214/sf.376
- Liira J, Sepp T, Kohv K (2011) The ecology of tree regeneration in mature and old forests: combined knowledge for sustainable forest management. *J For Res Jpn* 16(3):184–193. doi:10.1007/s10310-011-0257-6
- Lucas-Borja ME, Candel-Pérez D, Morote FGA, Onkelinx T, Tíscar PA, Balandier P (2016) *Pinus nigra* Arn ssp *salzmannii* seedling recruitment is affected by stand basal area, shrub cover and climate interactions. *Ann For Sci* 73(3):649–656. doi:10.1007/s13595-016-0550-9
- Lynch TB, Nkouka J, Huebschmann MM, Guldin JM (2003) Maximum likelihood estimation for predicting the probability of obtaining variable shortleaf pine regeneration densities. *For Sci* 49(4):577–584
- Miina J, Saksala T (2013) Predicting establishment of tree seedlings in regeneration areas of *Picea abies* in Southern Finland. *Balt For* 19(2):187–200
- Nilsson U, Gemmel P, Johansson U, Karlsson M, Welanders T (2002) Natural regeneration of Norway spruce, Scots pine and birch under Norway spruce shelterwoods of varying densities on a mesic-dry site in southern Sweden. *For Ecol Manag* 161(1–3):133–145. doi:10.1016/S0378-1127(01)00497-2
- Nyland RD (2007) *Silviculture: concepts and applications*, 2nd edn. Waveland Press, Long Grove
- Osman KT (2013) *Soils: principles, properties and management*. Springer, Dordrecht
- Polley H (2011) Survey instructions for the 3rd National Forest Inventory (2011–2012). Federal Ministry for Nutrition, Agriculture and Consumer Protection, Bonn

- Polley H, Hennig P, Kroihner F, Marks A, Riedel T, Schmid U, Schwitzgebel F, Stauber T (2015) The forests in Germany: Selected results of the Third National Forest Inventory. Berlin
- Pretzsch H (2014) Canopy space filling and tree crown morphology in mixed-species stands compared with monocultures. *For Ecol Manag* 327:251–264. doi:10.1016/j.foreco.2014.04.027
- Príncipe A, Nunes A, Pinho P, do Rosário L, Correia O, Branquinho C (2014) Modeling the long-term natural regeneration potential of woodlands in semiarid regions to guide restoration efforts. *Eur J For Res* 133(4):757–767. doi:10.1007/s10342-014-0787-5
- R Core Team (2016) R: A language and environment for statistical computing, Vienna, Austria. <https://www.R-project.org/>
- Rencher AC (2002) Methods of multivariate analysis, Wiley series in probability and mathematical statistics, 2nd edn. Wiley, New York
- Schütt P (2008) Lexikon der Nadelbäume. Die große Enzyklopädie mit über 800 Farbfotos unter Mitwirkung von 30 Experten. Lizenzausg. für Nikol Verl.-Ges. Wiley, Hamburg
- Schweiger J, Sterba H (1997) A model describing natural regeneration recruitment of Norway spruce (*Picea abies* (L.) Karst.) in Austria. *For Ecol Manag* 97(2):107–118. doi:10.1016/S0378-1127(97)00092-3
- Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCr: visualizing classifier performance in R. *Bioinformatics* 21(20):7881
- Staatsbetrieb Sachsenforst (2014) Hunting in SaxonyForst. Staatsbetrieb Sachsenforst, Pirna (in Germany)
- Stange F (2007) Soil types of the surface soil of Germany 1:1.000.000. <https://produktcenter.bgr.de/terraCatalog/OpenSearch.do?search=45E74823-B-4BB7-866E-5E84DA81301F%5C&type=/Query/OpenSearch.do>
- Stewart JD, Landhäuser SM, Stadt KJ, Lieffers VJ (2001) Predicting natural regeneration of white spruce in boreal mixedwood understories. *For Chron* 77(6):1006–1013. doi:10.5558/tfc771006-6
- Tegelmark DO (1998) Site factors as multivariate predictors of the success of natural regeneration in Scots pine forests. *Forest Ecol Manag* 109(1–3):231–239. doi:10.1016/S0378-1127(98)00255-2
- Thünen-Institut (2012) Third Federal Forest Inventory—Results database. ThünenInstitut. <https://bwi.info/>. 11 Oct 2016
- ThüringenForst (2015) Our forest with game: hunting ThüringenForst. ThüringenForst, Erfurt (in Germany)
- Tremblay JP, Huot J, Potvin F (2007) Density-related effects of deer browsing on the regeneration dynamics of boreal forests. *J Appl Ecol* 44(3):552–562
- Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 49(11):1225–1231. doi:10.1016/S0895-4356(96)00002-9
- Vanclay JK (1992) Modelling regeneration and recruitment in a tropical rain forest. *Can J For Res* 22(9):1235–1248. doi:10.1139/x92-165
- Youden WJ (1950) Index for rating diagnostic tests. *Cancer* 3(1):32–35. doi:10.1002/1097-0142(1950)3:1<32:AID-CNCR2820030106>3.0.CO;2-3