Generating Tree-Level Harvest Predictions from Forest Inventories with Random Forests

Philipp Kilham [1],* , **Christoph Hartebrodt [2]** and **Gerald Kändler [2]**

[1]   Faculty of Environment and Natural Resources, University of Freiburg, 79110 Freiburg, Germany
[2]   Forest Research Institute Baden-Württemberg (FVA), 79110 Freiburg, Germany;
      christoph.hartebrodt@forst.bwl.de (C.H.); gerald.kaendler@forst.bwl.de (G.K.)
*   Correspondence: philipp-kilham@posteo.de; Tel.: +49-761-401-8312

check for
updates

Abstract: Wood supply predictions from forest inventories involve two steps. First, it is predicted whether harvests occur on a plot in a given time period. Second, for plots on which harvests are predicted to occur, the harvested volume is predicted. This research addresses this second step. For forests with more than one species and/or forests with trees of varying dimensions, overall harvested volume predictions are not satisfactory and more detailed predictions are required. The study focuses on southwest Germany where diverse forest types are found. Predictions are conducted for plots on which harvests occurred in the 2002–2012 period. For each plot, harvest probabilities of sample trees are predicted and used to derive the harvested volume (m^3 over bark in 10 years) per hectare. Random forests (RFs) have become popular prediction models as they define the interactions and relationships of variables in an automatized way. However, their suitability for predicting harvest probabilities for inventory sample trees is questionable and has not yet been examined. Generalized linear mixed models (GLMMs) are suitable in this context as they can account for the nested structure of tree-level data sets (trees nested in plots). It is unclear if RFs can cope with this data structure. This research aims to clarify this question by comparing two RFs—an RF based on conditional inference trees (CTree-RF), and an RF based on classification and regression trees (CART-RF)—with a GLMM. For this purpose, the models were fitted on training data and evaluated on an independent test set. Both RFs achieved better prediction results than the GLMM. Regarding plot-level harvested volumes per ha, they achieved higher variances explained (VEs) and significantly ($p < 0.05$) lower mean absolute residuals when compared to the GLMM. VEs were 0.38 (CTree-RF), 0.37 (CART-RF), and 0.31 (GLMM). Root means squared errors were 138.3, 139.9 and 145.5, respectively. The research demonstrates the suitability and advantages of RFs for predicting harvest decisions on the level of inventory sample trees. RFs can become important components within the generation of business-as-usual wood supply scenarios worldwide as they are able to learn and predict harvest decisions from NFIs in an automatized and self-adapting way. The applied approach is not restricted to specific forests or harvest regimes and delivers detailed species and dimension information for the harvested volumes.

Keywords: tree-level harvest decisions; multilevel data; business-as-usual; random forest

1. Introduction

National forest inventories (NFIs) are conducted worldwide—some using angle-count sample methods [1,2]. Based on these inventories, forest growth and wood supply modeling is used to highlight consequences of potential management decisions and to show future prospects as well as limitations of the status quo [3–5].

With regard to the generation of business-as-usual (BAU) scenarios, there is an increasing trend of learning harvest decisions—and also tree mortality—from past inter-inventory periods using data-driven models [6–8]. Data-driven tree-level harvest decision models must deal with three important issues: (a) the integration of two-step harvest decisions on plot and tree level; (b) the fact that sample trees of the tree-level data set are nested within inventory plots; and (c) the potential presence of nonlinear relationships and interactions.

NFI plots are often not representative for the decision area (e.g., forest stand) in which they are located. Instead, NFI data becomes representative for higher aggregations of inventory plots. For this reason, the fact that no harvests occurred on the plot cannot be taken as evidence that no harvests occurred in the stand [4]. In the following, an NFI plot is considered as harvested if a minimum of one tree is removed from the plot in the period of interest. Two binary outcomes are involved in harvest decision predictions in the context of NFIs: the question of whether a specific inventory plot is harvested, and the question of whether an individual tree on a harvested plot is cut down [9,10]. BAU scenarios for forest growth and wood supply modeling must account for the fact that a considerable number of plots might be spared from harvest interventions within a given time span [6]. Eid [11], for example, stated that assumed thinning operations were often not conducted. One consequence is that the extracted timber is harvested from a lower number of plots treated with higher intensities than average figures would suggest. Neglecting this phenomenon would result in inventory predictions clearly different from BAU [4].

An option for dealing with this issue is to first predict the binary harvest decisions on plot level. Subsequently, tree-level harvest decision models are trained on sample trees from harvested plots. Trees from plots predicted as not-harvested are spared and tree-level prediction models are applied only to plots predicted as harvested. Here, deterministic and random approaches can be differentiated [6]. In a deterministic approach, a clear yes-no decision is applied to each plot for every predicted period. In this context, Kilham et al. [4] demonstrated that the replication of NFI harvest patterns on plot level can be improved with a stratified procedure based on a classification and regression tree (CART) introduced by Breiman et al. [12] and random forest (RF) algorithms [13]. Antón-Fernández and Astrup [6] used harvest probabilities to select plots randomly in numerous repetitions.

Complementing a prediction on plot level with a detailed tree-level prediction—as compared to an overall harvested volume prediction—offers the advantage that more specific information about tree species and dimensions of the harvested timber can be generated. This is of particular importance for predicting harvests for forests with more than one species and/or forests with trees of varying age and dimension. Here, it increases possibilities to derive timber assortments from predictions. As probabilities are learned from the inventory, a tree-level prediction method can replicate different harvest regimes when these are evident in the inventory and indicated by the predictor variables.

Further issues need to be addressed when predicting tree-level harvest decisions. The first is the multilevel structure of the tree-level data set, which results from the arrangement of sample trees within inventory plots. Neglecting this interrelation can result in an underestimation of the variance of variable estimates [9,14]. As a consequence, tree-level predicting attributes cannot be selected based on their significance levels [15].

Grégoire et al. [14] have addressed this issue with generalized linear mixed models (GLMMs). Here, the dependency can be considered in the form of random effects on the model intercept [15]. An alternative approach is the use of copulas as applied in Fortin et al. [15] and Delisle-Boulianne et al. [16]. Inspired by zero-inflated distributions, Manso et al. [9] presented a single mixed-effects model for predicting plot- and tree-level harvest decisions simultaneously. They pointed out that a separate prediction of plot and tree level impedes the covariance estimation between the random effects of the two models.

Another issue is that the underlying relationships linked to harvest decision are not necessarily linear or monotonic. Relationships might be relevant for some subsets of the data, but not others. For example, different harvest regimes apply in different ownership and forest structure settings.

In each of these settings, different indicators might become relevant for determining the harvest probability of a specific sample tree. In linear models, the inclusion of interaction terms requires specifications based on the pre-knowledge of the researcher. Typically, interactions are included on the base level only because the inclusion of high order interactions is too complex to be done manually. Nonlinear recursive partitioning (RP)—which includes CART and RF—offers the advantage that nonlinear, nonmonotonic, and partial relationships as well as interactions can be detected and accommodated by the algorithm in a data-driven way [17,18].

In RP, several approaches were designed to deal with multilevel data structures. These include random effect regression trees [19], exploratory regression analysis with hierarchically clustered data [20], and mixed-effects RF [21]. To date, only few mixed-effects RP models have been extended to generalized versions. Hajjem et al. [22] presented generalized mixed-effects regression trees. The GLMM tree, as introduced by Fokkema et al. [23], is another approach. However, this method is computationally expensive and often produces convergence issues depending on the specific data subset. Consequently, the integration in an RF algorithm in the context of large data sets is difficult.

However, RP methods that do not include random effects should not be discounted prematurely. Previous research suggested that a clustered structure of the data does not affect the prediction accuracy of the algorithms. Thus, it is important that classification errors and variable importance are calculated from an independent test set [17,24,25].

Martin [25] points towards the different implications of multilevel data for different RP procedures. Conditional inference tree (CTree) [26] is a major competitor of CART given an identified bias of the latter towards numeric variables and categorical variables with many categories [18]. Consequently, CTree was found to improve predictions in comparison to CART [27]. However, the CTree algorithm assumes independence between instances and is more likely to select a split in correlated data. In multilevel data sets this can lead to complex trees that can overfit the training data. CART is considered to be more robust against these data correlations but has also been found to exhibit bias in multilevel data sets. Another question is whether the model includes only tree-level attributes or uses a mix of tree- and plot-level variables. In the second case, the splitting procedure in CTrees can be biased. Finally, the mentioned shortcomings might become less relevant within an RF application because CTrees are grown to maximum depth—and CARTs are left unpruned—and the generated complexity is addressed within the averaging procedure [25].

This research focuses on the tree-level harvest prediction. The plot-level harvest decision prediction is excluded by testing models only on NFI plots harvested within the period of interest. Harvest probabilities of sample trees from harvested NFI plots are predicted. For each sample tree, these probabilities are used to define the share of the weight (number of trees per hectare represented by the sample tree) harvested. Harvested volumes per sample tree are calculated from this harvested weight and can be aggregated to derive the respective harvested volumes for each inventory plot.

The principal research question is whether RFs (CTree-RF and CART-RF) are suitable to generate harvested volume predictions for harvested inventory plots. Since GLMMs are suitable for this task, the research question can be addressed by comparing results of RF-based volume predictions to results of GLMM-based volume predictions. The performance of the models is tested on a separate test set not used for training.

Addressing each of the two RF algorithms individually, the research hypothesis is: RF based prediction results are not different from GLMM based prediction results. If this hypothesis cannot be rejected, the respective RF can be used despite the multilevel data structure. The flexibility that comes with the ability of RFs to select variables and define their interaction and relationships in an automatized way would then be an argument for preferencing an RF over a GLMM. This can be important when the model has to be trained frequently and on different data sets (different regions, different periods).

If the first hypothesis is rejected for an RF and the RF delivers results inferior to the GLMM, the use of this RF cannot be recommended. If an RF delivers results superior to the GLMM, this indicates

that wood supply modellings can benefit from the nonparametric nature of the respective RF and its ability to include high-level variable interactions. If the performance of both CTree-RF and CART-RF are equal to or superior to the GLMM, the two RFs can be compared following the same logic.

2. Materials and Methods

2.1. Data

The harvest prediction method was tested on data from the German NFI (referred to below as NFI) for the federal state of Baden-Württemberg located in the southwest of the country. Baden-Württemberg covers an area of around 35,750 km^2, of which approximately 38.4% is considered to be forest [28,29]. A more detailed description of the case can be found in Kilham et al. [4] and MLR (Ministry of Rural Affairs and Consumer Protection Baden-Württemberg) [30]. Models were trained on the period 2002–2012, which corresponds to the most recent inter-NFI period. The aftermath of a storm in 1999, a drought period in 2003, and a storm in 2007 affected the forest in Baden-Württemberg during this period.

Data used were entirely generated from the second (2002) and third (2012) NFIs. NFI data are available online at Thünen Institute [29]. The NFI uses a systematic single-level cluster sampling with permanently marked and remeasured plots. NFI tracts are located on a 2 km × 2 km grid. One NFI tract consists of up to four plots and is officially considered as the primary sampling unit. In one tract, plots are located at the corners of a square with side lengths of 150 m [28,31]. Plots of one tract can be in different forest stands and might be associated with different landholders. For this reason, they were considered as independent research units in this study.

Only trees registered with an angle-count method (basal area factor = 4 m^2 ha^{-1}) with a diameter at breast height (DBH) $\geq$7 cm were used. Moreover, only plots on accessible, productive forest not owned by the federal government were included. The data sets included only sample trees measured in 2002 and 2012, as well as trees measured in 2002 and registered as not standing (harvest or mortality) in 2012. Tree representation factors (weights) from 2002 were used.

Enough data was available to set aside the test set completely, which is preferable to cross-validation or boosting [32]. The training and the test set were created with the createDataPartition function of R's caret package [33]. The function considered the original distribution of the plot-level harvest decision when creating the partitions to reproduce the class distribution in the split [34]. Generally, models were trained on a training set consisting of 75% of the plots. To evaluate model performance, the fitted models were applied to the test set (25% of the plots). Assuming independence between plots, plot- and tree-level test sets were independent from the respective training sets as none of the training set sample trees shared a common plot with a test set sample tree.

The training set and test set included only plots that were harvested according to the NFI in the 2002–2012 period (5684 training set plots with 49,055 sample trees and 1894 test set plots with 15,960 sample trees). An extension of the training set (in the following training set B) was used to generate an additional attribute for the tree-level harvest models (see Chapter 2.5. Plot Harvest Stratum Attribute). This training set included the 5684 harvested plots plus additional 3313 plots not harvested in the 2002–2012 period.

2.2. Response and Predicting Attributes

All considered variables were registered in the NFI or derived from NFI measurements. Training set specifications of response attributes (dependent variables) and predicting attributes (independent variables) are shown in Tables A1 and A2. The binary tree-level attribute harvest decision was the response attribute. A tree was counted as harvested if it was registered in the 2002 NFI and was found to have been removed from the inventory plot in 2012.

Eleven plot-level and 11 tree-level variables were considered as predicting attributes. Comparable to Kilham et al. [4], plot-level attributes were related to plot characteristics (plot standing volume,

average age of trees on the plot, average DBH of trees on the plot, plot type), site characteristics (slope, harvest condition, altitude, site index), ownership characteristics (ownership type), and characteristics related to forest policies (nature protected area, nature park).

Plot standing volume over bark (ob) in m^3 per hectare was calculated for the beginning of the 10-year period across all species. Average age of trees on the plot recorded the average age of all sample trees of the dominant and predominant forest stories (for plenter forest, all sample trees of the plot were considered) at the middle of the period. Average DBH of trees on the plot corresponded to the average DBH of all sample trees of the dominant and predominant forest stories (for plenter forest, all sample trees) at the beginning of the period.

The attribute plot type differentiated between conifer 1, conifer 2, deciduous, mixed 1, and mixed 2. To generate these characteristics, tree species were divided into three groups: 'conifers 1' consisted of species of the genus Picea (*Picea* spp.), and fir (*Abies alba Mill.*); 'conifers 2' referred to other conifers (see Table A3); and 'deciduous' referred to broadleaf species (see Table A3). The differentiation between 'conifer 1' and 'conifer 2' was applied because the group of spruce and fir is of particular market relevance in Baden-Württemberg [30]. Plot type characteristics were then generated by calculating the basal area share of each tree species group per ha from trees of the dominant and predominant forest stories (for plenter forest, all sample trees). From these values, the dominant tree species group was determined for each plot. If the dominant tree species group represented a share of <90%, the plot was considered to be mixed. This threshold is commonly used in the German NFI to differentiate pure from mixed stands [31]. 'Mixed 1' referred to plots where the conifer 1 group represented a share of ≥50%. 'Mixed 2' referred to plots where the conifer 1 group had a share of <50%.

Slope and harvest conditions can have impacts on harvests [6,35]. Slope was recorded in per cent (whole-numbers) and used as a continuous variable. The variable harvest condition had the two classes: 'favorable' and 'unfavorable'. Unfavorable referred to site conditions likely to impede harvest operations (slopes >30%, or machine track distances >150 m, or distances to nearest driveway >1 km). Altitude [36] referred to the elevation of the plot in meters above sea level (a.s.l.). Site index [6] was an attribute that included both plot and site characteristics. It was the average yearly total increment in m^3 ob ha^{-1} over 100 years of the dominant species on the plot and originated from the database of the German forest development and wood supply modeling WEHAM (German abbreviation: Waldentwicklungs- und Holzaufkommensmodellierung) [37].

Ownership type [38,39] differentiated between federal state, community, large private, medium private, and small private properties. The private property size classes refer to the total forest property of a specific owner within Germany—large private: >500 ha; medium private: 5–500 ha; and small private: <5 ha. The binary attributes 'nature protected area' and 'nature park' recorded whether a plot was protected under the respective directive.

Regarding tree-level attributes, tree species were classified according to eight groups including species of the genus Picea, species of the genus Abies (*Abies* spp.), Douglas fir (*Pseudotsuga menziesii (Mirb.)*, species of the genus Pinus and the genus Larix (*Pinus* spp., *Larix* spp.), beech (*Fagus sylvatica* L.), oak (*Quercus petraea (Matt.) Liebl.*, *Quercus robur* L.), ash (*Fraxinus excelsior* L.), and other deciduous species. Species share (%) referred to the basal area share of the sample tree's species per ha. Taking a beech sample tree as an example, a species share of 60% would indicate that beech trees accounted for 60% of the basal area per ha registered at the respective plot. For this attribute, species were not grouped beforehand. F-rDiffDq was proposed by Schelhaas et al. [40] as an estimate of the social position of a tree within its inventory plot and was defined as:

$$F - rDiffDq = \frac{DBH}{DBHq} - 1 \tag{1}$$

where DBHq is the plot level quadratic mean diameter of all sample trees [40].

Tree-level predicting attributes were related to tree type (species, species share), health condition (skidding damage, fungus, beetle, other stem damages, dead), pre-treatment (pruning), dimension (DBH, tree height), and social status (F-rDiffDq). These referred to the condition registered in 2002.

The six variables related to tree health and pre-treatment were binary attributes indicating whether a tree was found to be damaged, infested or treated in 2002. The attribute dead recorded whether the tree had recently died but was still standing and had fine brushwood.

DBH was measured in mm for each sample tree. Tree height was only measured (in dm) for a subsample ($\sim 1/3$) of the trees registered at each plot. The subsamples were selected according to canopy-class and tree species groups. Tree heights for the remaining sample trees were estimated with unit-height-curves (Einheitshöhenkurven) according to Sloboda et al. [41]. The procedure is described in detail in Riedel et al. [31].

2.3. Classification Trees

CART [12] is one of the most popular classification and regression tree (in the following CT) algorithms. CTs recursively partition the space spanned by all predictor variables (feature space) into rectangular areas. In the partitioning process, observations with similar response values are grouped. Once the feature space is fragmented, constant values of the response variable are predicted for each area [18,26]. For the binary response variable harvest decision, harvest probabilities can be derived from the proportion of harvested trees within each partitioned area.

CART recursively produces binary splits using an impurity reduction approach. Every split produces two new groups, each with a majority of either response class (in this research either yes or no). This results in daughter nodes that are purer than their parent nodes. Entropy measures (e.g., Gini index) are used to determine the impurity in each node [18].

One shortcoming is that CTs (both CARTs and CTrees) can overfit the training data. Pruning is used to avoid overfitting in CART. Furthermore, variable selection in CART algorithms is biased towards variables with many possible splits, including both numerical variables and categorical variables with many groups [18,26,42]. To address this, Hothorn et al. [26] introduced the CTree framework, which combines regression trees with the theory of conditional inference procedures. First, the independence of each of the m included predicting variables X_j ($j = 1, \ldots, m$) from the response attribute Y (global null hypothesis H0) is tested with the m partial hypotheses H_0^j:

$$D(Y|X_j) = D(Y) \tag{2}$$

where D is the (conditional) distribution of the response Y (given X_j).

If H0 cannot be rejected at specified level α, the recursion is stopped. Otherwise, the association between the response attribute and each predictor is assessed by using test statistics or p-values which indicate the deviation from H_0^j. The X_j with the strongest relation to Y is selected and the binary split is conducted for this predicting attribute. In CTrees, the use of p-values makes further pruning unnecessary [26,43].

2.4. Random Forest

CTs have two additional important drawbacks. First, though producing good predictions on average, they are considered to be unstable. Even small variations within the training data could result in a different tree structure. Second, if a splitting point is crossed, a small change in the value of the predicting attribute can cause a jump in the prediction [18].

The ensemble learning procedure RF [13] can help to overcome these issues by building a number of CTs on bootstrap samples of the original data. In comparison to Bagging [44], RF adds an additional layer of randomness by using only a subset of the predicting variables in each CT [45]. Weaker variables—usually outplayed by stronger competitors—are allowed to enter the prediction process and might reveal interaction effects that would otherwise remain disregarded [13,18].

Considerations about the different RP methods in the context of multilevel data structures were outlined in the introduction.

RFs are difficult to interpret. Variable importance (VI) measurements can be used to offer some information about which attributes played major roles in the generation of the CTs. VI is usually calculated from the out-of-bag (OOB) sample of each CT and then averaged across all CTs of the RF [18,42].

The original VI—available for CART-RF—assigns higher scores to correlated predictors [46]. For this reason, Strobl et al. [42] suggested a conditional permutation scheme, which reflects only the impact of the concerned attribute in predicting the response. This VI measurement is available for the CTree-RF but not for the CART-RF. It is outlined in Strobl et al. [46].

However, both VI measures assume independence between the training data of a CT and the OOB used to derive the VI. Thus, a multilevel data structure can cause bias in the calculation of VI measures due to correlated OOB samples [17,25]. A solution was presented by Martin [25] within a multilevel exploratory data analysis (MLEDA) framework. However, in this study, VI could be generated from the test set instead because test set trees were considered to be independent from the training set trees.

VI was only generated for the CTree-RFs for the reasons named above. VI results derived from the OOB samples were compared to the results derived from the test set. Following Martin [47], VIs were standardized to facilitate a better comparison.

2.5. Plot Harvest Stratum Attribute

An additional plot-level attribute was generated (plot harvest stratum). This attribute allocated NFI plots to plot-level harvest probability groups. The attribute was generated from training set B. Training set B included the binary response attribute plot-level harvest decision as well as the 11 independent plot-level variables (Tables A4 and A5). For the attribute plot-level harvest decision, the characteristic 'yes' (for the 5,684 harvested plots) indicated that a minimum of one sample tree was removed from the plot between 2002 and 2012.

To create the plot harvest stratum attribute, a CTree was grown with the ctree function of R's party package [26,48] by using the plot-level harvest decision attribute as the dependent variable (Figure A1). CTree was chosen instead of CART to avoid the bias described in the Classification Trees section. The 11 plot-level attributes were used as inputs. The minimum number of plots per node was set to 200. Univariate was used as the test type and the threshold for the p-value was set to 0.001 to avoid overfitting. The fitted CTree was applied to the test set plots to generate the corresponding attribute.

2.6. Harvest Probability Predictions for Sample Trees with Random Forests

For CTree-RF, 100 CTrees were grown from the sample trees of (harvested) training-set plots using the cforest function of R's party package [42,49,50]. For CART-RF, 100 CARTs were grown using the randomForest function of R's randomForest package [45].

Since the inclusion of both plot-level and tree-level attributes can cause bias due to the multilevel data structure [25], both tree-level RFs were tested in two ways. In the first, all 22 plot- and tree-level attributes as well as the plot harvest stratum attribute were used as potential inputs. In this variation, the number of variables per CT was set to five (using the square root of the input attributes is often found to be optimal [42,51]). In the following, the respective variations are referred to as CTree-RF-pt and CART-RF-pt. In the second variation, only the 11 tree-level attributes were available as input variables for the RFs. Here, three variables were used per CT. In the following, the respective variations are referred to as CTree-RF-t and CART-RF-t.

2.7. Harvest Probability Predictions for Sample Trees with Generalized Linear Mixed Model

GLMM extends a generalized linear regression model by incorporating fixed and random effects (mixed effects). The status of a tree i after a harvest intervention on plot g follows a Bernoulli distribution [16]:

$$y_{gi} \sim \text{Bernoulli}(\pi_{gi}) \tag{3}$$

If the tree was harvested, y_{gi} equals one, otherwise its value is zero. π_{gi} is the harvest probability of the tree. The applied GLMM can be described as

$$\text{logit}(\pi_{gi}) = \beta_0 + \beta_1 X_{1gi} + \beta_2 X_{2gi} + \ldots + \beta_p X_{pgi} + b_{0g} \tag{4}$$

where β_0 is the constant (intercept), β_j is the partial regression weight of predictor j, X_{jgi} is the trees i's score of the jth of p predictors in plot g, and b_{0g} is the deviation from β_0 for the random effect group (plot) g with $b_{0g} \sim N(0, \sigma^2_{b_{0g}})$.

The GLMM was fitted on all training set sample trees of harvested plots using the glmer function of R's lme4 package [52]. The model was fitted with a binary logistic regression. Bound optimization by quadratic approximation (BOBYQA, [53]) was used as an optimizer. The Nelder–Mead method [54] was tested but the resulting model failed to converge.

All 22 input variables and the plot harvest stratum attribute were considered. Out of these, only variables with p-values <0.001 were included to avoid overfitting. Numerical variables were scaled and centered. Inventory plots were used as input for the random effects specified on the intercept (sample trees of one plot were in the same random effects group). For discrete attributes with more than two characteristics, characteristics were grouped iteratively when no statistically significant differences ($p > 0.001$) were detected between the individual characteristics or groups of characteristics. Based on this approach, strata of the plot harvest stratum attribute were grouped starting from the terminal nodes of the classification tree. Taking the left side of the tree (Figure A1) as an example, terminal nodes 7 and 8 were grouped first. Subsequently, terminal node 5 was joined to this group. Finally, terminal node 3 was not attached to the group since there was a statistically significant difference between node 3 and the group consisting of nodes 8, 7 and 5. The derived groups were group 1 (5, 7, 8), group 2 (stratum 3 only), group 3 (12, 14, 15, 18, 19, 22, 25, 27, 28, 31, 32, 34, 35) and group 4 (stratum 21 only).

A higher order interaction was included, which replicated the relationship between the plot-level attribute average DBH of trees on the plot and the tree-level attribute F-rDiffDq for the different ownership types and species (groups). The interaction reflects the change in harvest selection in relation to a growing forest stand (Figure A2). On plots with lower average DBH, the probability of a tree being harvested decreases with increasing F-rDiffDq. However, this relationship changes with changing average DBH of trees on the plot. For plots with a larger average DBH, the tree-level harvest probability can also increase with increasing F-rDiffDq. Interactions were clearly different with regard to species (groups) and ownership types (Figure A2).

A stepwise backward variable selection (e.g., as described in Rohner et al. [55]) based on the Akaike information criterion (AIC) [56] did not remove any variable. Nevertheless, average stand age was excluded from the GLMM as its inclusion impeded convergence of the model. Its inclusion also generated the lowest AIC decrease compared to other statistically significant variables. The final GLMM included 15 attributes (Table 1). The random effect of the intercept is shown in Figure A3.

For the GLMM, residual diagnostics were generated with R's DHARMa (diagnostic for hierarchical regression models) package [57,58]. Within the DHARMa approach, 1000 new synthetic data sets were created from the fitted GLMM. Cumulative distributions of simulated values for each observed value and the quantile values corresponding to the observed values were calculated. Data set diagnostics (Figure A4) show that no considerable abnormalities could be detected in the GLMM.

Table 1. Estimates, significance codes, and standard errors (in parentheses) for the generalized linear mixed model (GLMM) used to predict tree-level harvest probabilities. The model was fitted on sample trees from the training set inventory plots on which harvests occurred (49,055 sample trees on 5684 plots; IA: interaction).

Attributes	Tree-Level Harvest Probability	
F-rDiffDq (index, centered and scaled)	−0.245 ***	(0.017)
Tree height (m, centered and scaled)	0.120 ***	(0.027)
Species share (%)	0.133 ***	(0.016)
Skidding damage (yes)	0.324 ***	(0.029)
Fungus (yes)	1.291 ***	(0.328)
Dead (yes)	1.400 ***	(0.246)
Other stem damages (yes)	0.380 ***	(0.045)
Pruned (yes)	−0.419 ***	(0.062)
Species group (other conifers)	−0.545 ***	(0.038)
Species group (oak)	−1.300 ***	(0.070)
Species group (deciduous)	−0.960 ***	(0.041)
Plot standing volume (m^3 over bark ha^{-1}, centered and scaled)	−0.215 ***	(0.024)
Average plot diameter at breast height (DBH, cm, centered and scaled)	0.245 ***	(0.026)
Site index (m^3 over bark ha^{-1}, centered and scaled)	−0.169 ***	(0.021)
Altitude (m above sea level, centered and scaled)	−0.197 ***	(0.019)
Ownership type (medium private)	−0.496 ***	(0.073)
Ownership type (others)	−0.330 ***	(0.056)
Plot harvest stratum (group 1)	0.282 ***	(0.108)
Plot harvest stratum (group 2)	2.336 ***	(0.303)
Plot harvest stratum (group 3)	−0.375 ***	(0.084)
IA: Spruce spp. on state forest—Average DBH of trees on the plot—F-rDiffDq	0.247 ***	(0.048)
IA: Other conifers on state forest—Average DBH of trees on the plot—F-rDiffDq	0.176 ***	(0.053)
IA: Spruce spp. on community forest—Av. DBH of trees on the plot—F-rDiffDq	0.241 ***	(0.037)
IA: Other conifers on community forest—Av. DBH of trees on the plot—F-rDiffDq	0.211 ***	(0.046)
IA: Conifers on large private forest—Av. DBH of trees on the plot—F-rDiffDq	0.267 ***	(0.056)
IA: Spruce spp. on medium priv. forest—Av. DBH of trees on the plot—F-rDiffDq	0.213 ***	(0.064)
IA: Other conifers on medium private—Av. DBH of trees on the plot—F-rDiffDq	0.290 ***	(0.063)
IA: Spruce spp. on small private forest—Av. DBH of trees on the plot—F-rDiffDq	0.489 ***	(0.054)
IA: Other conifers on small private—Average DBH of trees on the plot—F-rDiffDq	0.306 ***	(0.092)
IA: Deciduous—Average DBH of trees on the plot—F-rDiffDq	0.068 ***	(0.018)
Constant (Intercept)	0.450 ***	(0.102)
Observations		49,055

Note: *** $p < 0.001$.

2.8. Prediction

Harvest probabilities for sample trees were derived by applying the harvest models to the respective test set. The probabilities were not converted into yes-no decisions. Instead, a proportion of the standing volume per ha represented by an individual sample tree was predicted to be harvested according to the derived harvest probability of that sample tree. This is similar to the method used in WEHAM where usually only a share of the trees per ha represented by a sample tree are harvested within one harvest intervention [37]. In the NFI, the exact date of a harvest within the ten-year period remains unknown. For this reason, it is assumed that all harvests occurred in the middle of the period [31]. The same assumption was applied in the prediction.

The mid-period (2007) standing volume of a sample tree ($\hat{V}_j$) was expressed as volume in m^3 ob. In the NFI, tree standing volume (V_j) is estimated from DBH, a second diameter measurement (the measurement position of that second diameter on the trunk can vary) and a tree height measurement or estimation. The exact procedure is described in Riedel et al. [31]. For trees measured in 2002 and 2012, $\hat{V}_j$ was estimated as:

$$\hat{V}_j = \frac{V_{j2002} + V_{j2012}}{2} \tag{5}$$

For trees that were recorded as missing in 2012, tree dimension measurements for 2007 were estimated with the Sloboda trend function and $\hat{V}_j$ was derived from these estimates [31]. The harvested volume per sample tree $\hat{H}_j$ (m^3 ob in 10 years) was calculated as:

$$\hat{H}_j = \hat{p}_j \; \hat{V}_j \; w_j \tag{6}$$

where $\hat{p}_j$ is the predicted harvest probability of the sample tree; $\hat{V}_j$ is the mid-period standing volume in m^3 ob of the sample tree; and w_j is the number of trees represented by the sample tree per ha (weight).

2.9. Evaluation Criteria

This study evaluated how well the models replicated the NFI harvest patterns when applied to test set data not used for model training. For the numeric harvested volume predictions, percentage of variance explained (VE) was used as a criterion in the following definition [59]:

$$\text{VE} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{7}$$

where $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ is the residual sum of squares (RSS) derived from measured (y_i) and predicted ($\hat{y}_i$) values and $\sum_{i=1}^{n} (y_i - \overline{y})^2$ is the total sum of squares (TSS), which corresponds to the variation of the measured values [60]. In addition, models were compared by the resulting mean square error (MSE) [61] defined as $1/n \times RSS$. Significance tests were conducted with reference to the test set residuals ($y_i - \hat{y}_i$). Paired Student t-tests were used to analyze the null hypothesis: The arithmetic mean of the absolute residuals of model 1 and model 2 do not differ. Given the large number of test set plots, a test for normal distributions was not required. To be able to compare the p-values among the analyzed models, they were adjusted with R's p.adjust function. When the comparison of two models resulted in a rejection of the null hypothesis, the model with the lower average absolute residuals was identified as the superior model. Effect sizes (*r*) were calculated with the correlation coefficient according to Pearson and evaluated according the classes provided by Cohen [62]. An additional method was the graphical comparison of average harvested volumes (m^3 ob in 10 years) for subsets of the data. The standard errors of the NFI averages were used as critical benchmarks.

3. Results

The RF algorithms were generally superior to the GLMM. The average of the absolute residuals of both RF algorithms that used plot- and tree-level attributes were statistically significantly lower compared to the GLMM (Table 2). However, the corresponding effect sizes were low. No difference was detected between CTree-RF-pt and CART-RF-pt. The RF variations that used plot- and tree-level attributes were statistically significantly superior to the respective variations that did not include plot-level attributes. Neither CTree-RF-t nor CART-RF-t showed statistically significant differences when compared to the GLMM.

Table 2. Testing models for statistically significant differences with reference to absolute residuals of plot-level harvested volume predictions with the paired sample t-test. Null hypothesis: The averages of the residuals of model 1 and model 2 do not differ (RF: random forest; CTree-RF: RF based on conditional inference trees; CART-RF: RF based on classification and regression trees; GLMM: generalized linear mixed model).

Model 1 [1] (Mean Absolute Residuals [2])		Model 2 [1] (Mean Absolute Residuals [2])		Result	Adjusted *p*-Value [3]		Effect Size (*r*) [4]
CTree-RF-pt	(95.7)	CTree-RF-t	(98.6)	Model 1 superior	0.009	**	0.06
CTree-RF-pt	(95.7)	CART-RF-pt	(95.8)	Model 1 ≈ Model 2	0.904	n	
CTree-RF-pt	(95.7)	CART-RF-t	(101.9)	Model 1 superior	0.001	**	0.12
CTree-RF-pt	(95.7)	GLMM	(98.8)	Model 1 superior	0.007	**	0.06
CTree-RF-t	(98.6)	CART-RF-pt	(95.8)	Model 2 superior	0.020	*	0.05
CTree-RF-t	(98.6)	CART-RF-t	(101.9)	Model 1 superior	0.049	*	0.12
CTree-RF-t	(98.6)	GLMM	(98.8)	Model 1 ≈ Model 2	0.884	n	
CART-RF-pt	(95.8)	CART-RF-t	(101.9)	Model 1 superior	0.001	**	0.12
CART-RF-pt	(95.8)	GLMM	(98.8)	Model 1 superior	0.014	*	0.05
CART-RF-t	(101.9)	GLMM	(98.8)	Model 1 ≈ Model 2	0.077	n	

[1] pt: version using plot- and tree-level attributes; t: version using tree-level attributes only; [2] in m^3 over bark per ha in 10 years; [3] Note: ** $p < 0.01$; * $p < 0.05$; n $p \geq 0.05$ (no statistical significance); [4] 0.1 = low; 0.3 = medium; 0.5 = high [62].

Highest overall VE was achieved with the CTree-RF-pt (Table 3). VE was calculated with regard to the harvested volume (m^3 ob in 10 years) on plot level. The highest VE values were obtained with the CTree-RF-pt and CART-RF-pt. VE values produced by the GLMM were lower. For both RF algorithms, VE was reduced when plot-level attributes were not considered for the prediction. This reduction was comparably strong for CART-RF.

Table 3. Harvested volume predictions on plot level from random forest (RF) based on conditional inference trees (CTree-RF), RF based on classification and regression trees (CART-RF), and generalized linear mixed models (GLMMs) (1894 plots with 15,960 sample trees).

| Tree-Level Method [1]	VE [2]	RMSE [3]
CTree-RF-pt	0.380	138.27
CTree-RF-t	0.327	144.15
CART-RF-pt	0.366	139.88
CART RF-t	0.177	159.42
GLMM	0.314	145.47

[1] pt: variation using plot- and tree-level attributes; t: version using tree-level attributes only; [2] VE: Variance explained; [3] RMSE: Root mean squared error.

The GLMM deviated more from the NFI value than the other methods with regard to average harvested volumes. The species shown in Figure 1, Figure 2, and Figure A5 correspond to: spruce (*Picea abies (L.) Karst.*), fir (*Abies alba Mill.*), pine (*Pinus sylvestris* L.), Douglas fir (*Pseudotsuga menziesii (Mirb.) Franco*), beech (*Fagus sylvatica* L.), oak (*Quercus petraea (Matt.) Liebl.*, *Quercus robur* L.), ash (*Fraxinus excelsior* L.), and maple (*Acer pseudoplatanus* L.). NFIs are only representative for larger aggregations of inventory plots. Figure 1 compares predicted test set harvested volumes for individual tree species and tree dimensions. The GLMM underestimated beech harvests (in particular in the DBH range 20–40 cm) to a greater extent than the other models. Figure 2 shows average harvested volume predictions for test set plots. The displayed average volumes refer to the average of plots included in the respective data subsample (N). For most subsets, predictions of all models remain within the standard error of the NFI. GLMM average harvested volume predictions for spruce in higher altitudes, and for plots located within nature parks, were above the NFI mean plus the standard error. Furthermore, GLMM estimates for total average beech harvested volumes were below the NFI mean minus the standard error. Such underestimations for beech were found for plots located in state forest, plots at higher altitudes, plots with favorable harvest conditions, and plots not located in nature parks. Average beech harvested volume predictions of the two RF models were comparably low for the state forest—though they remained within the NFI standard error. Maple was included in the group other deciduous in the tree-level models. Predictions for average harvested volumes of maple

from small private forests (all three models), sites with unfavorable site conditions (all three models), sites in higher altitudes (all three models), and from nature parks (RFs) were also above the NFI mean plus standard error. Predictions for total average harvested volumes of maple were above the NFI mean plus standard error when averages were calculated only for plots on which the respective species (group) occurred (Figure A5). Other examples of average harvested volume prediction above the NFI mean plus standard error in Figure 2 were fir in higher altitudes (CTree-RF), pine in higher altitudes (RFs), ash in higher altitudes (all three models), and Douglas fir from small private forests (all three models).
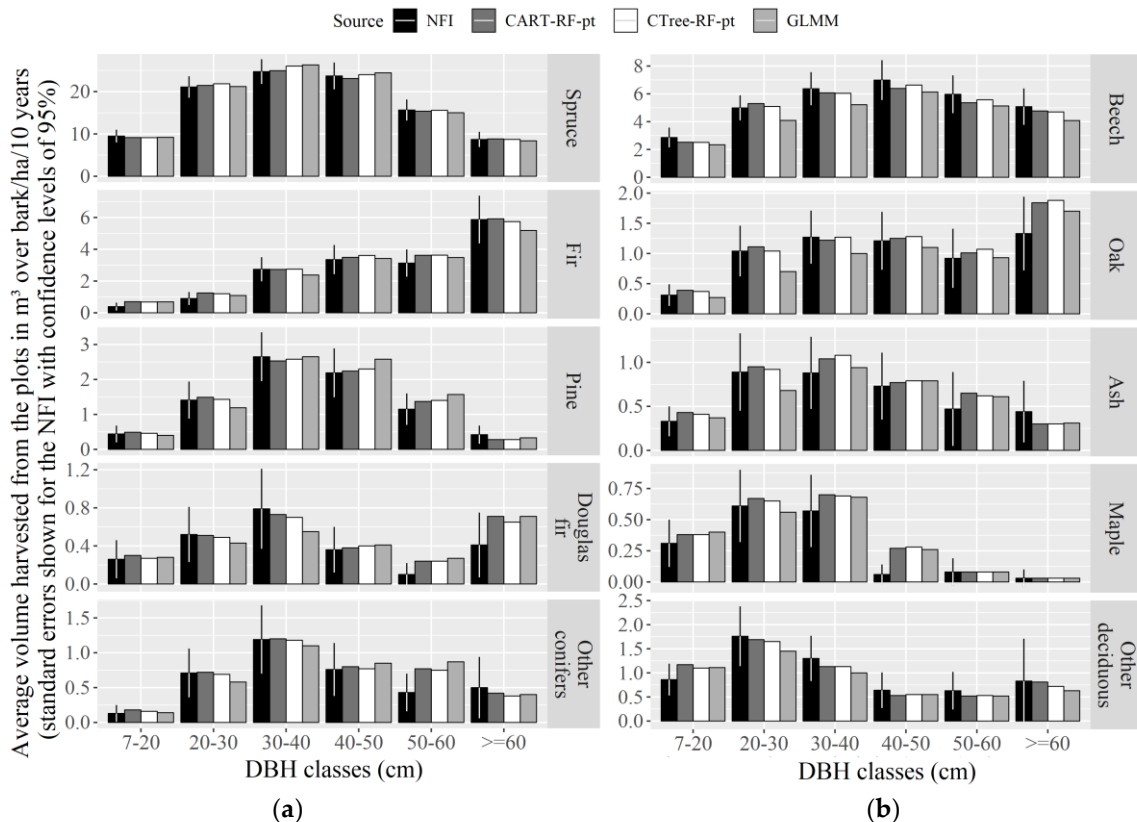
Figure 1. Comparing predicted average harvested volumes to national forest inventory (NFI) measurements on species level for the period 2002–2012. (**a**) Conifer species. (**b**) Deciduous species. Harvested volumes refer to the average of the 1,894 test set plots. The reference year for the diameter at breast height (DBH) classes and harvested volumes was 2007 (CART: Classification and regression tree; CTree: Conditional inference tree; RF: Random forest; pt: variation that used plot- and tree-level attributes; GLMM: Generalized linear mixed model).
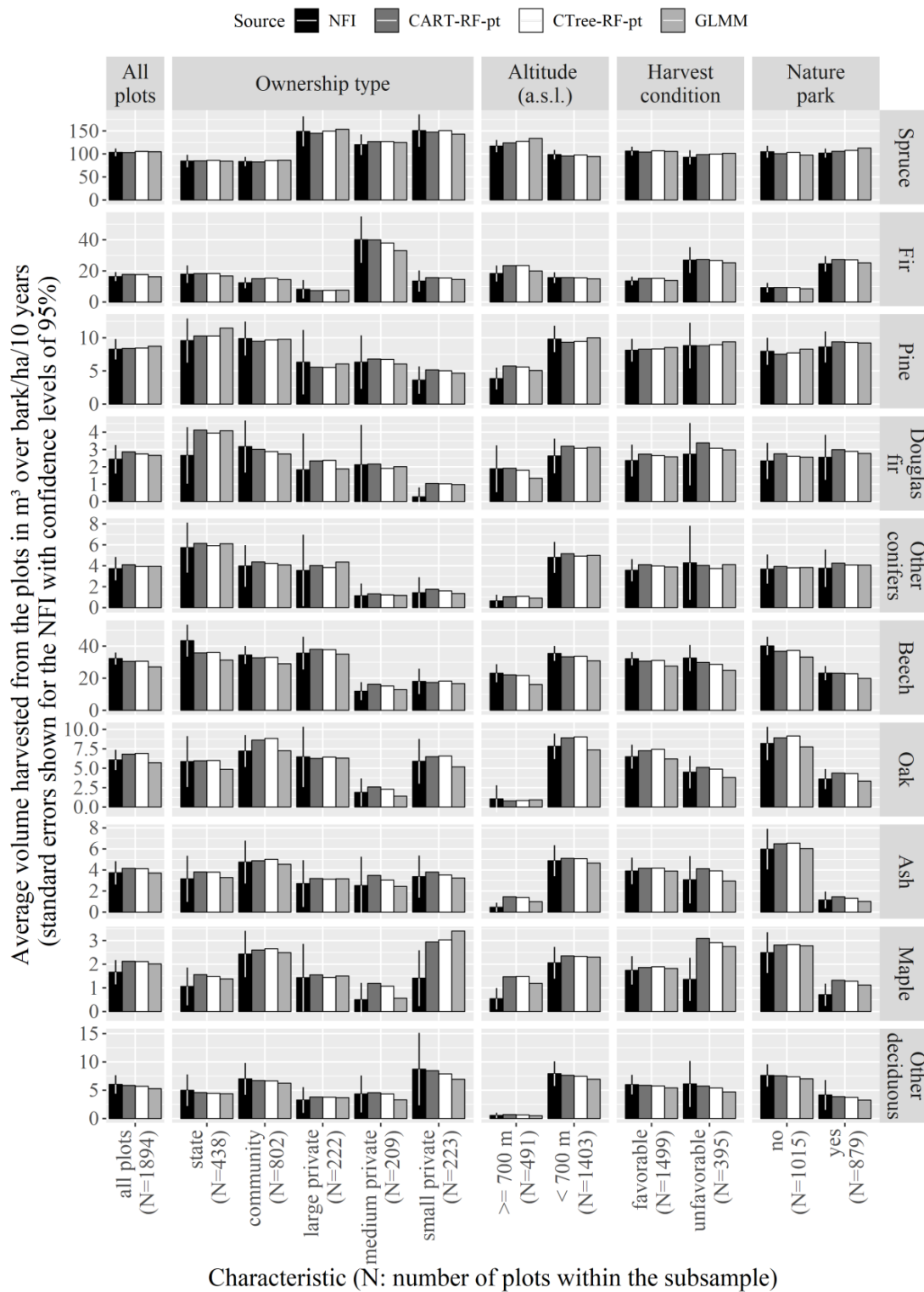
Figure 2. Comparing 2002–2012 predicted average harvested volumes (tree-level models only) for plots on which harvests occurred according to the national forest inventory (NFI) for species (groups) and selected data subsets. Harvested volumes refer to the average of the plots included in the respective test set subsample (N). The reference year for the diameter at breast height (DBH) classes and harvested volumes was 2007 (CART: Classification and regression tree; CTree: Conditional inference tree; RF: Random forest; pt: variation that used plot- and tree-level attributes; GLMM: Generalized linear mixed model; a.s.l.: above sea level).

The VI comparison did not indicate that the correlation of OOB samples had a strong impact on the VI. The overall importance of predicting attributes within the models offers a hint as to how far harvest decisions are influenced by the specific impact factors. The comparison between the OOB VIs and test set VIs of the CTree-RFs did not reveal large differences (Figure 3). Most obvious is the change in the ranking of F-rDiffDq (Figure 3a).
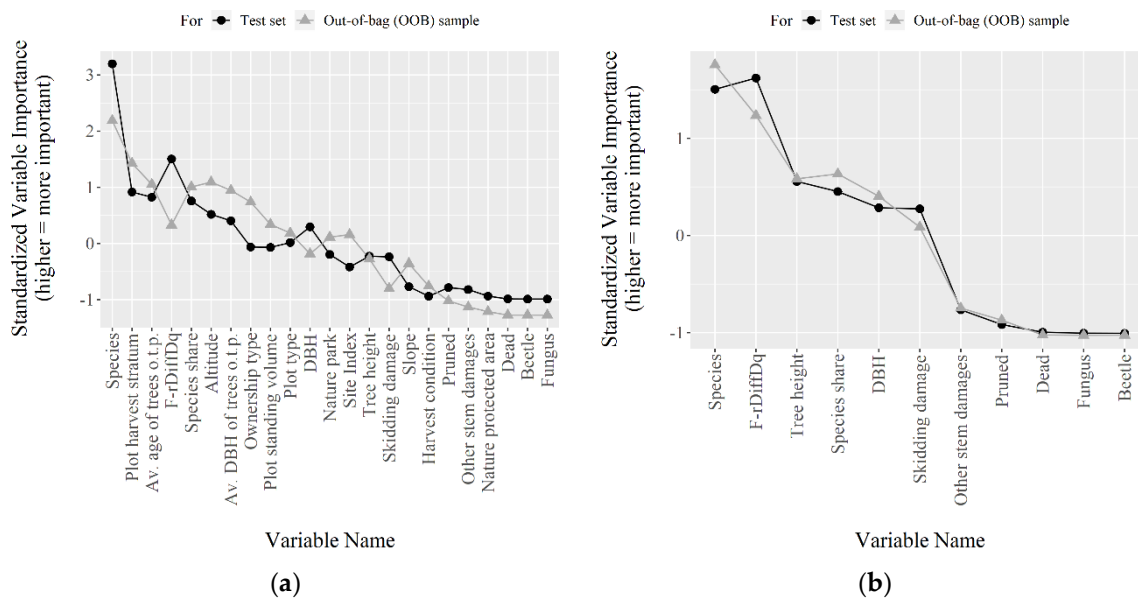
Figure 3. Standardized variable importance (VI) as permutation importance for CTree-RF in (**a**) the variation that used plot- and tree-level attributes as well as strata and (**b**) the variation that used tree-level attributes and strata only (DBH: Diameter at breast height, Av.: average, o.t.p.: on the plot).

4. Discussion

A major advantage of the RF based tree-level harvest models is that model fitting can be automatized. To the knowledge of the authors, this research is the first attempt to predict detailed BAU harvest decisions on the level of sample trees from NFIs with modern machine learning algorithms. Previous studies restricted predictions to inventory plots or larger areas [4,6,35] or used generalized linear models (GLM), GLMM or a copula approach to predict harvest probabilities of sample trees [8–10,15,16]. The advantage of RF over these methods is that variable selection, interactions, and relationships are learned by the algorithm from the training data in an automatized way. Therefore, no presumptions about harvest regimes (e.g., thinning, single-tree selection harvest or clear cut) under specific stand or ownership conditions were necessary. Furthermore, the RF-based harvest models can be adapted easily to other regions, inventory designs, or harvest practices.

RFs are suited to predicting tree-level harvests in the given inventory design. Concerning average absolute residuals, all tested RFs resulted in equal or superior performances when compared to the GLMM. Since a GLMM can account for the multilevel data structure found in NFIs [14–16], the respective RFs can also be considered suitable in this regard.

RFs improved predictions in comparison to the GLMM. Compared to a GLMM, the tested RF algorithms that used plot- and tree-level attributes showed a better prediction performance. With regard to overall test set results, they achieved higher VEs and resulted in significantly ($p < 0.05$) lower average absolute residuals. One reason for the superior overall performance of CTree-RF-pt and CART-RF-pt was that they predicted harvested volumes of spruce and beech more accurately than the GLMM. Among other tree species, spruce and beech accounted for the highest, and second highest harvested volumes respectively (Figure 2). The findings demonstrate that RFs can help to improve the quality of predictions on the level of NFI sample trees.

The RFs performed better when both plot-level and tree-level attributes were used as variable inputs. Regarding average absolute residuals, CTree-RF-pt and CART-RF-pt were significantly ($p < 0.05$) superior to the respective variations that did not use plot-level attributes. The VE of CART-RF dropped considerably when plot-level attributes were not used.

When comparing CTree-RF-pt and CART-RF-pt, neither of the two algorithms could be clearly identified as the better performer. No significant differences were found between the two models regarding absolute residuals and the VEs were comparably close. This does not necessarily mean that the CTs of CTree-RF were not biased as a consequence of the multilevel structure [25]. In that sense, the integration of random effects into generalized RP methods [22,23] might help to further improve the RF based tree-level harvest prediction. Other studies have found that prediction accuracies in RF with multilevel data could be further improved if bagging subsets were resampled at the cluster level [24,25]. This effect was not tested here as this would involve a manipulation of the original CART-RF and CTree-RF algorithms.

The models could not capture the full effects of tree damages on harvest decisions. Delisle-Boulianne et al. [16] found that a classification system based on defects could be used to improve tree-level harvest models. Ledermann [8] also found tree damage to be one important indicator for tree-level harvest selections. This study included tree damages in the RF and GLMM models. However, the corresponding variables were ranked relatively low in terms of VI, though one reason for this was certainly the low number of trees recorded as being infected by beetle or fungus. Furthermore, the scenario could not capture new infestations and damages that occurred within the 10-year period. So, it is likely that tree damages played a larger role in harvest decisions. However, since this relation was not evident in the NFI data, it could not be captured by the models.

Tree species had an important impact on harvest decisions. Species was found to be the number one tree-level indicator according to the VI. In the GLMM, species share—which also had a relatively high VI—was found to be positively correlated with harvest probability. The fact that tree species that were a minority on a plot were less likely to be harvested might reflect a stronger diversification of forest stands. The diversification of stands is frequently discussed in the context of climate change, biodiversity, sustainable forestry, and hazard control [63]. With a focus on forest conversion periods towards uneven-aged stands, Ledermann [8] identified stem quality and tree species as important estimators of tree-level harvest selection.

Tree dimension was an important indicator for tree-level harvest probabilities, which is consistent with other studies [8,10,64,65]. In this context, the F-rDiffDq transformation of the DBH was found to be more effective than the pure DBH.

Distance-dependent attributes can be of relevance but were not included in this research. Delisle-Boulianne et al. [16] found that tree-level harvest probabilities were influenced by the status of neighboring trees. These attributes were not included in this study due to small plot sizes and potential bias due to edge-effects.

The general restrictions related to data availability and sampling outlined in Kilham et al. [4] were also relevant for this study. This also includes that the effect of disturbances from storms and dry periods could not be deducted from the harvests. Thus, a future BAU scenario based on the described methods and data would assume the weather conditions and disturbances of the time period 2002–2012 for every predicted 10-year period.

In general, predicting harvest decisions for plots and trees of large-scale inventories is difficult. In these inventory designs, individual sample plots are not representative of the decision area—e.g., forest stand—in which they are located [4]. However, harvest intensities and the decision to harvest a tree might be influenced more strongly by the overall condition of the stand. At the same time, information about decision makers that can be connected to NFI plots is scarce. In this research, owner-related information was restricted to rough ownership types and property size classes. Under these circumstances, highly accurate harvest predictions on the level of inventory plots cannot be expected. This is expressed in the rather low VEs shown in Table 3. However, NFIs—and therefore

also NFI-based forest growth and wood supply modeling—are designed to generate representative data on a larger scale [4,66]. In this context, the models were able to replicate observed harvest patterns for aggregations of NFI plots relatively well on a comparably detailed scale.

To be able to concentrate on plot-level clusters, other potential levels of the data structure were consciously excluded from this study. However, inventory plots themselves can be considered as being embedded in multiple hierarchical structures. This includes the fact that several inventory plots are managed by one owner or manager, and several owners are serviced by one forest official and profit from the same local infrastructure. Also, plots within one inventory tract were considered to be independent, which might be true in some cases but not in all. Including these effects in RP harvest prediction procedures is an interesting field of future research.

To make use of the tree-level prediction within a BAU scenario, it is necessary to first predict harvest decisions for individual NFI plots. If plot-level harvest probabilities are predicted with RF algorithms as well, the entire model combination could benefit from the RF-specific advantages (automatized training, nonlinear relationships, high-level interactions). Here, it is essential to convert plot-level harvest probabilities into binary decisions with an approach (e.g., randomized procedure [6], deterministic cut-off, or deterministic stratified procedure [4]) that suits the requirements and goals of the respective scenario. For readers interested in the results of a joint plot-level and tree-level prediction, a randomized plot-level prediction based on CTree-RF was combined with the three presented tree-level models (CTree-RF-pt, CART-RF-pt, and GLMM). The prediction process is described in Table S1. Results are shown in Table S2 and Figures S1–S3. When combined with tree growth and mortality models, such a harvest model combination can be used to generate future scenarios—e.g., by applying the fitted models to 2012 NFI data to predict the period 2012–2022 and to the predicted 2022 NFI data to predict the period 2022–2032.

5. Conclusions

To a statistically significant degree, the performances of both analyzed RF algorithms were superior to that of the GLMM. With reference to harvested volume prediction on the level of inventory plots, both RFs achieved higher overall VEs (0.38 for CTree-RF-pt, and 0.37 for CART-RF-pt) in comparison to the GLMM (0.31). The RFs resulted in statistically significant but low ($r < 0.10$) reductions of average absolut residuals when compared to the GLMM. Predictions for the frequently harvested species spruce and beech could be improved. The RFs need to be trained on a combination of plot-level and tree-level attributes. For CART-RF this is even more essential, since the performance of the algorithm was found to drop considerably when the plot-level attributes were not considered.

Predicting the harvested volumes on the basis of sample trees has the advantage that more detailed information about species and dimension of the harvested timber can be derived. For regions that include mixed species forests with diverse tree dimensions and for stakeholders interested in specific material availabilities and properties (e.g., within the wood processing industry or wood-based bioeconomy) this is of particular importance.

This research demonstrated for the first time that RFs are suitable to predict harvest probabilities of sample trees despite the multi-layered data structures that occur in NFIs. Researchers working in the field of forest growth and wood supply projections can benefit from this finding in two ways. First, the RFs were found to improve the quality of predictions, which is linked to the ability of the algorithms to integrate nonlinearity and high order variable interactions. Second, RFs can learn and predict BAU harvest decisions from NFIs in an automatized and self-adapting way, which is a key advantage over common approaches in this field. Regression models require extensive fitting procedures and pre-knowledge about variable interactions and nonlinearity. In RF-based procedures, model fitting can be automatized, thereby facilitating a flexible application. This could be of particular interest for scientists who seek to develop inventory-based prediction applications of use for practitioners or researchers.

Appendix A

Table A1. Characteristics of continuous training set variables (5684 plots with 49,055 sample trees).

Level	Attribute	Range	Mean	SD
Plot-level	Plot-level standing volume (m^3 over bark/ha)	7.1–2137.7	433.3	234.7
Plot-level	Average DBH of trees on the plot (cm)	7.0–100.4	33.1	12.5
Plot-level	Average age of trees on the plot (years)	15.0–380.0	81.3	35.8
Plot-level	Slope (%)	0–77.0	11.1	9.7
Plot-level	Site Index (m^3 over bark/ha)	2.0–21.0	10.4	4.1
Plot-level	Altitude (m above sea level)	94.0–1370.0	554.7	214.0
Tree-level	Species share (%)	0.0–1.0	0.7	0.3
Tree-level	F-rDiffDq (index)	−0.9–2.3	0.1	0.3
Tree-level	Diameter at breast height (DBH, cm)	7–144.1	36.2	15.6
Tree-level	Tree height (m)	3.2–49.7	26.1	7.4

Table A2. Characteristics of discrete training set variables (5684 plots with 49,055 sample trees).

Level	Attribute	Characteristics
Plot-level	Nature protected area	yes (126); no (5,558)
Plot-level	Nature park	yes (2553); no (3,131)
Plot-level	Harvest condition	favorable (4471); unfavorable (1,213)
Plot-level	Ownership type	state (1353); community (2377); large private (640); medium private (636); small private (678)
Plot-level	Plot type	conifer 1 (2004); conifer 2 (193); deciduous (1260); mixed 1 (1025); mixed 2 (1202)
Tree-level	Harvest decision	yes (17,991); no (31,064)
Tree-level	Dead	yes (108); no (48,947)
Tree-level	Skidding damage	yes (10,016); no (39,039)
Tree-level	Other stem damage	yes (3238); no (45,817)
Tree-level	Beetle	yes (80); no (48,975)
Tree-level	Fungus	yes (56); no (48,999)
Tree-level	Pruned	yes (2814); no (46,241)
Tree-level	Species (groups)	spruce (24,109); fir (4614); pine (4117); Douglas fir (1511); beech (8654); oak (2484); ash (1101); other deciduous (2465)

Table A3. Species included in the plot type attribute characteristics. Only species with more than 30 individuals in the training set are named explicitly. The species (groups) were ranked according to their frequency in the training set.

Group	Species
Conifer 1	*Picea abies* L. Karst; *Abies alba* Mill.; other *Picea* spp.
Conifer 2	*Pinus sylvestris* L.; *Pseudotsuga menziesii (Mirb.)*; *Larix decidua* L.; *Larix kaempferi (Lamb.) Carrière*; *Pinus nigra J.F.Arnold*; other conifers
Deciduous	*Fagus sylvatica* L.; *Quercus petraea (Matt.) Liebl.*; *Fraxinus excelsior* L.; *Quercus robur* L.; *Acer pseudoplatanus* L.; *Carpinus betulus* L.; *Tilia* spp.; *Alnus glutinosa* (L.) Gaertn.; *Quercus rubra* L.; *Prunus avium* L.; *Betula pendula* Roth; *Populus nigra* L.; *Acer platanoides* L.; *Robinia pseudoacacia* L.; *Castanea sativa* Mill.; *Salix* spp.; *Acer campestre* L.; *Ulmus* spp.; other deciduous

Table A4. Characteristics of continuous variables of training set B (8997 plots).

Level	Attribute	Range	Mean	SD
Plot-level	Plot-level standing volume (m^3 over bark/ha)	6.3–2137.7	388.5	238.2
Plot-level	Average DBH of trees on the plot (cm)	7.0–123.6	32.5	13.1
Plot-level	Average age of trees on the plot (years)	15.0–380.0	80.2	37.1
Plot-level	Slope (%)	0–77.0	12.0	10.4
Plot-level	Site Index (m^3 over bark/ha)	2.0–21.0	10.0	4.2
Plot-level	Altitude (m above sea level)	90.0–1370.0	557.3	220.4

Table A5. Characteristics of discrete variables of training set B (8997 plots).

| Level	Attribute	Characteristics
Plot-level	Harvest decision	yes (5684); no (3313)
Plot-level	Nature protected area	yes (296); no (8701)
Plot-level	Nature park	yes (4189); no (4808)
Plot-level	Harvest condition	favorable (6715); unfavorable (2282)
Plot-level	Ownership type	state (2116); community (3570); large private (949); medium private (1069); small private (1293)
Plot-level	Plot type	conifer 1 (2833); conifer 2 (339); deciduous (2390); mixed 1 (1477); mixed 2 (1958)
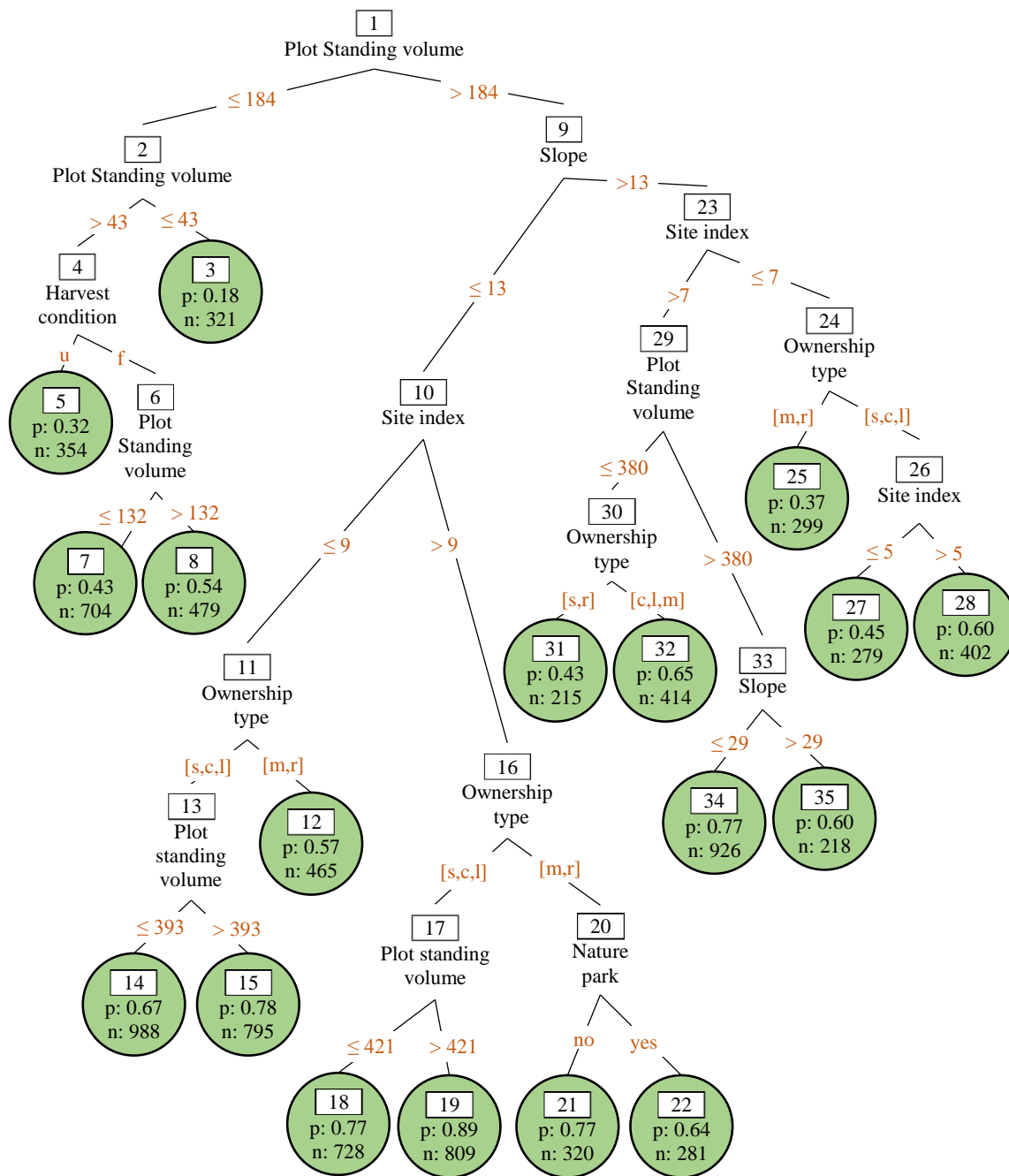
Figure A1. Classification tree based on the conditional inference tree (CTree) algorithm used to classify plots into strata to generate the plot harvest stratum attribute (p: harvest probability; n: number of training set plots; Ownership type as s: state; c: community; l: large private; m: medium private; r: small private; Harvest condition as f: favorable; u: unfavorable; Plot standing volume in m^3 over bark (ob) ha^{-1}; Site index as average yearly total increment in m^3 ob ha^{-1} over 100 years; Slope in %).
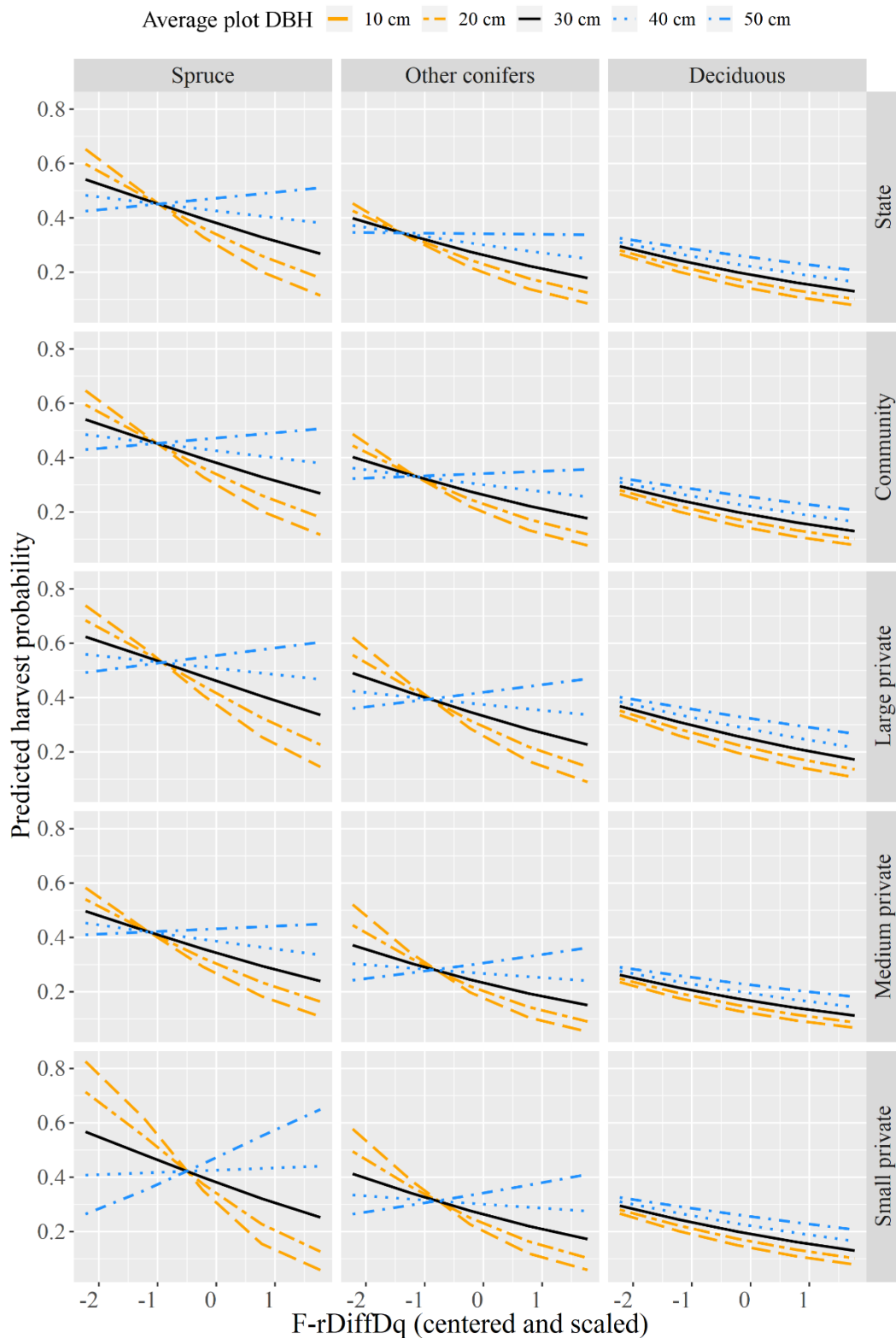
Figure A2. Effect of the interaction between F-rDiffDq, average diameter at breast height (DBH) of trees on the plot, (grouped) species, and ownership type on tree-level harvest probability predictions within the generalized linear mixed model (GLMM). For the visualization, values of all other variables were constant (unpruned; no damages; favorable harvest conditions; plot harvest stratum attribute group 3; training set mean values for numeric attributes). Spruce refers to *Picea* spp.
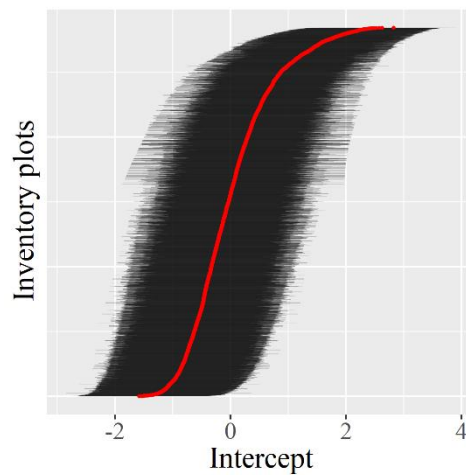
Figure A3. Random effects (red dots) with a variance of 0.97 and error bars (black horizontal lines) of the inventory plots within the generalized linear mixed model (GLMM). The individual inventory plots—sorted from the highest positive to the lowest negative shift—are shown on the y-axis.

(a)　　　　　　　　　　　　　　　　　　　　　　　(b)
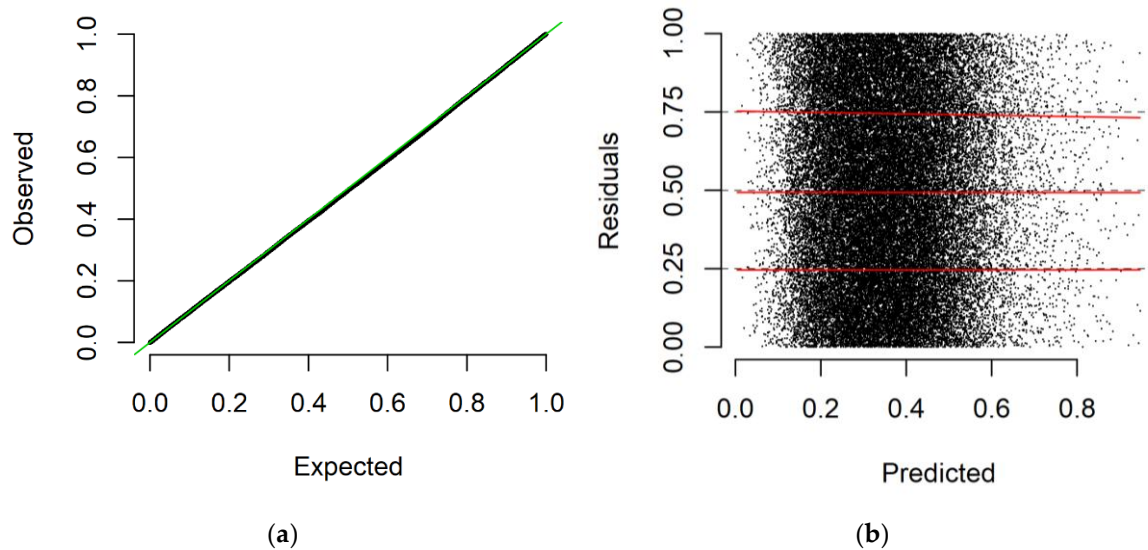
Figure A4. (**a**) Quantile-quantile plot (QQ-plot) for the generalized linear mixed model (GLMM). The QQ-plot can be used to detect overall deviations from the expected distribution (green line). (**b**) GLMM residuals in relation to predicted values. In a well-fitted model, the results of the quantile regression (red lines) should be straight, horizontal, and at y-values of 0.25, 0.5 and 0.75 as indicated by the black dashed lines [58]. The two plots were generated from a synthetic data set within the diagnostic for hierarchical regression models (DHARMa) approach.
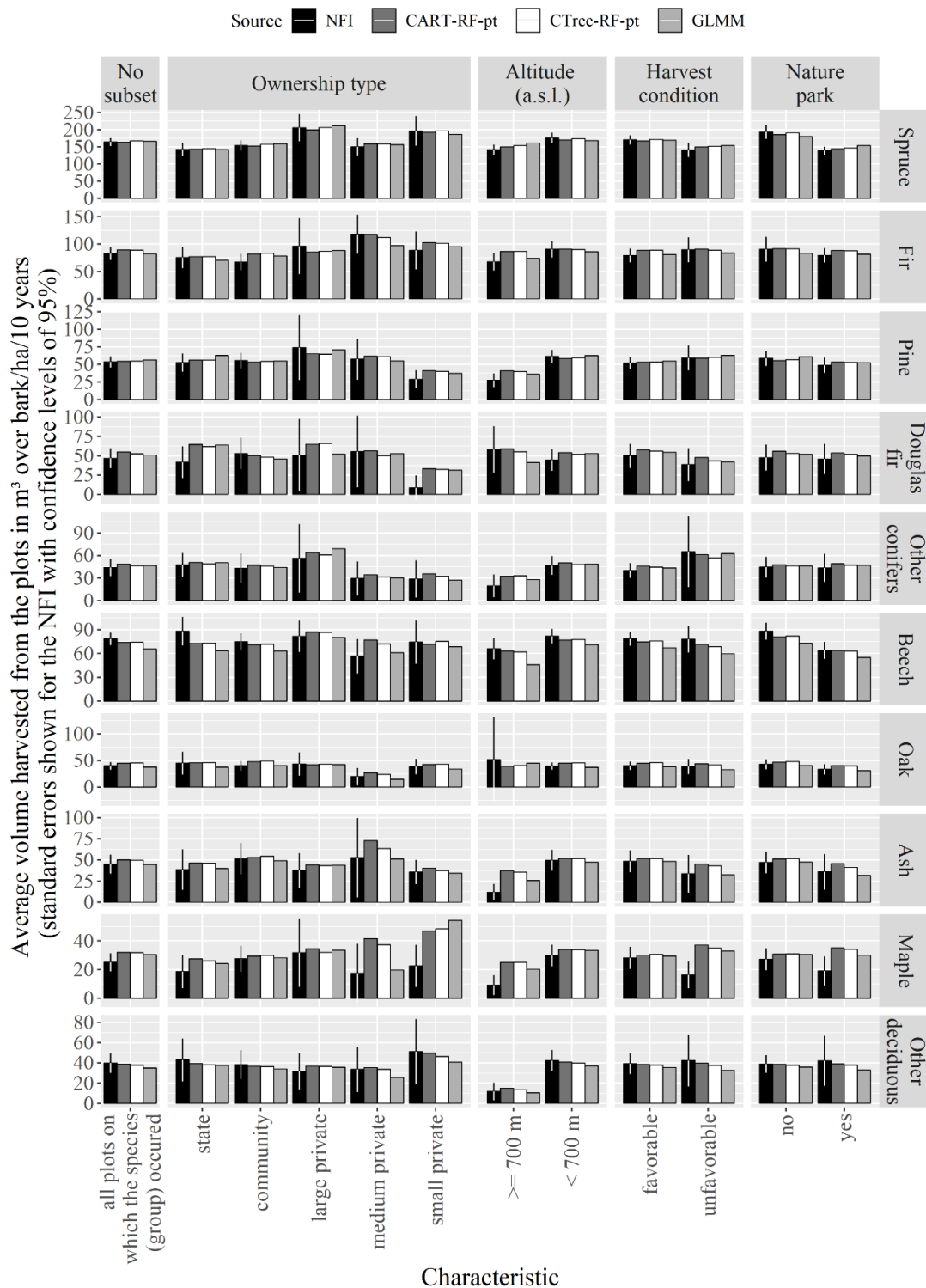
Figure A5. Comparing 2002–2012 predicted average harvested volumes (tree-level models only) from plots on which harvests occurred according to the national forest inventory (NFI) for species (groups) and selected data subsets. Harvested volumes refer to the average of the plots included in the respective test set subsample on which the species (group) occurred. The reference year for the diameter at breast height (DBH) classes and harvested volumes was 2007 (CART: Classification and regression tree; CTree: Conditional inference tree; RF: Random forest; pt: variation that used plot- and tree-level attributes; GLMM: Generalized linear mixed model; a.s.l.: above sea level).

References

1. Fischer, C.; Gasparini, P.; Nylander, M.; Redmond, J.; Hernandez, L.; Brändli, U.-B.; Pastor, A.; Rizzo, M.; Alberdi, I. Joining Criteria for Harmonizing European Forest Available for Wood Supply Estimates. Case Studies from National Forest Inventories. *Forests* **2016**, *7*, 104. [CrossRef]
2. Tomppo, E.; Schadauer, K.; McRoberts, R.E.; Gschwantner, T.; Gabler, K.; Ståhl, G. Introduction. In *National Forest Inventories*; Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E., Eds.; Springer: Dordrecht, The Netherlands, 2010; pp. 1–18.
3. Barreiro, S.; Schelhaas, M.-J.; Kändler, G.; Antón-Fernández, C.; Colin, A.; Bontemps, J.-D.; Alberdi, I.; Condés, S.; Dumitru, M.; Ferezliev, A.; et al. Overview of methods and tools for evaluating future woody biomass availability in European countries. *Ann. For. Sci.* **2016**, *73*, 823–837. [CrossRef]
4. Kilham, P.; Kändler, G.; Hartebrodt, C.; Stelzer, A.-S.; Schraml, U. Designing Wood Supply Scenarios from Forest Inventories with Stratified Predictions. *Forests* **2018**, *9*, 77. [CrossRef]
5. Rock, J.; Gerber, K.; Klatt, S.; Oehmichen, K. The WEHAM 2012 "baseline scenario": Center line or guardrail? *Forstarchiv* **2016**, *87*, 66–69. [CrossRef]
6. Antón-Fernández, C.; Astrup, R. Empirical harvest models and their use in regional business-as-usual scenarios of timber supply and carbon stock development. *Scand. J. For. Res.* **2012**, *27*, 379–392. [CrossRef]
7. Fridman, J.; Ståhl, G. A Three-step Approach for Modelling Tree Mortality in Swedish Forests. *Scand. J. For. Res.* **2001**, *16*, 455–466. [CrossRef]
8. Ledermann, T. Using logistic regression to model tree selection preferences for harvesting in forests in conversion. In *Continuous Cover Forestry*; Gadow, K., Nagel, J., Saborowski, J., Eds.; Springer: Dordrecht, The Netherlands, 2002; pp. 203–216.
9. Manso, R.; Ningre, F.; Fortin, M. Simultaneous Prediction of Plot-Level and Tree-Level Harvest Occurrences with Correlated Random Effects. *For. Sci.* **2018**, *36*, 2994. [CrossRef]
10. Thurnher, C.; Klopf, M.; Hasenauer, H. Forests in transition: A harvesting model for uneven-aged mixed species forests in Austria. *Forestry* **2011**, *84*, 517–526. [CrossRef]
11. Eid, T. Testing a large-scale forestry scenario model by means of successive inventories on a forest property. *Silva Fenn.* **2004**, *38*, 305–317. [CrossRef]
12. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Taylor & Franxis Group, LLC: Boca Raton, FL, USA, 1984; p. 368. ISBN 978-0-412-04841-8.
13. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
14. Grégoire, T.G.; Schabenberger, O.; Barrett, J.P. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can. J. For. Res.* **1995**, *25*, 137–156. [CrossRef]
15. Fortin, M.; Delisle-Boulianne, S.; Pothier, D. Considering Spatial Correlations Between Binary Response Variables in Forestry: An Example Applied to Tree Harvest Modeling. *For. Sci.* **2013**, *59*, 253–260. [CrossRef]
16. Delisle-Boulianne, S.; Fortin, M.; Achim, A.; Pothier, D. Modelling stem selection in northern hardwood stands: Assessing the effects of tree vigour and spatial correlations using a copula approach. *Forestry* **2014**, *87*, 607–617. [CrossRef]
17. Finch, H. Recursive Partitioning in the Presence of Multilevel Data. *Gen. Linear Model J.* **2015**, *41*, 30–44.
18. Strobl, C.; Malley, J.; Tutz, G. An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. *Psychol. Methods* **2009**, *14*, 323–348. [CrossRef] [PubMed]
19. Sela, R.J.; Simonoff, J.S. RE-EM trees: A data mining approach for longitudinal and clustered data. *Mach. Learn.* **2012**, *86*, 169–207. [CrossRef]
20. Miller, P.; Mcartor, D.; Lubke, G. Metboost: Exploratory Regression Analysis with Hierarchically Clustered Data. *Stat.ML.* preprint. arXiv:1702.03994.
21. Hajjem, A.; Bellavance, F.; Larocque, D. Mixed-effects random forest for clustered data. *J. Stat. Comput. Simul.* **2014**, *84*, 1313–1328. [CrossRef]
22. Hajjem, A.; Larocque, D.; Bellavance, F. Generalized mixed effects regression trees. *Stat. Probab. Lett.* **2017**, *126*, 114–118. [CrossRef]
23. Fokkema, M.; Smits, N.; Zeileis, A.; Hothorn, T.; Kelderman, H. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behav. Res. Methods* **2017**. [CrossRef]

24. Karpievitch, Y.V.; Hill, E.G.; Leclerc, A.P.; Dabney, A.R.; Almeida, J.S. An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. *PLoS ONE* **2009**, *4*, e7087. [CrossRef]

25. Martin, D.P. Efficiently Exploring Multilevel Data with Recursive Partitioning. Ph.D. Thesis, University of Virginia, Charlottesville, VA, USA, 2015.

26. Hothorn, T.; Hornik, K.; Zeileis, A. Unbiased Recursive Partitioning: A Conditional Inference Framework. *J. Comput. Graph. Stat.* **2006**, *15*, 651–674. [CrossRef]

27. Fu, W.; Simonoff, J.S. Unbiased regression trees for longitudinal and clustered data. *Comput. Stat. Data Anal.* **2015**, *88*, 53–74. [CrossRef]

28. Schmitz, F.; Polley, H.; Hennig, P.; Dunger, K.; Schwitzgebel, F. *Die zweite Bundeswaldinventur BWI 2: Inventur und Auswertungsmethoden. The Second National Forest Inventory NFI 2: Inventory and Evaluation Methods*; BMELV: Braunschweig, Germany, 2008; p. 85.

29. Thünen Institut. Dritte Bundeswaldinventur. 2012. Available online: http://www.webcitation.org/72ru1KjDW (accessed on 2 October 2018).

30. Ministerium für Ländlichen Raum und Verbraucherschutz Baden-Württemberg. Clusterstudie Forst und Holz Baden-Württemberg: Analyse der Spezifischen Wettbewerbssituation des Clusters Forst und Holz und Ableitung von Handlungsempfehlungen. Available online: http://www.webcitation.org/72rslkLVu (accessed on 2 October 2018).

31. Riedel, T.; Hennig, P.; Kroiher, F.; Polley, H.; Schmitz, F.; Schwitzgebel, F. *Die dritte Bundeswaldinventur BWI 2012. Inventur- und Auswertungsmethoden*; BMEL: Berlin, Germany, 2017; p. 124.

32. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2nd ed.; Corrected at 7th Printing 2013; Springer: New York, NY, USA, 2013; p. 745, ISBN 978-0-387-84857-0.

33. Kuhn, M. Caret: Classification and Regression Training. Available online: http://www.webcitation.org/72ruazu3v (accessed on 2 October 2018).

34. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*. [CrossRef]

35. Verkerk, P.J.; Levers, C.; Kuemmerle, T.; Lindner, M.; Valbuena, R.; Verburg, P.H.; Zudin, S. Mapping wood production in European forests. *For. Ecol. Manag.* **2015**, *357*, 228–238. [CrossRef]

36. Thürig, E.; Schelhaas, M.-J. Evaluation of a large-scale forest scenario model in heterogeneous forests: A case study for Switzerland. *Can. J. For. Res.* **2006**, *36*, 671–683. [CrossRef]

37. FVA. *WEHAM 2012. Modelle und Algorithmen*; FVA: Freiburg, Germany, unpublished.

38. Polyakov, M.; Wear, D.N.; Huggett, R.N. Harvest choice and timber supply models for forest forecasting. *For. Sci.* **2010**, *56*, 344–355.

39. Rinaldi, F.; Jonsson, R.; Sallnäs, O.; Trubins, R. Behavioral Modelling in a Decision Support System. *Forests* **2015**, *6*, 311–327. [CrossRef]

40. Schelhaas, M.-J.; Hengeveld, G.M.; Heidema, N.; Thürig, E.; Rohner, B.; Vacchiano, G.; Vayreda, J.; Redmond, J.; Socha, J.; Fridman, J.; et al. Species-specific, pan-European diameter increment models based on data of 2.3 million trees. *For. Ecosyst.* **2018**, *5*, 716. [CrossRef]

41. Sloboda, V.B.; Gaffrey, D.; Matsumura, N. Regionale und lokale Systeme von Höhenkurven für gleichaltrige Waldbestände. *Allg. Forst Jagdztg.* **1993**, *164*, 225–228.

42. Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **2008**, *9*, 307. [CrossRef]

43. Hothorn, T.; Seidbold, H.; Zeileis, A. Package 'Partykit': A Toolkit for Recursive Partytioning. Available online: http://www.webcitation.org/72rumdgV2 (accessed on 2 October 2018).

44. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

45. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.

46. Strobl, C.; Hothorn, T.; Zeileis, A. Party on!: A New, Conditional Variable-Importance Measure for Random Forests Available in the party Package. *R J.* **2009**, *1*, 14–17.

47. Martin, D. Mleda: Multilevel Exploratory Data Analysis. Available online: http://www.webcitation.org/72ruRxsA0 (accessed on 2 October 2018).

48. R Core Team. *R: A Language and Environment for Statistical*, version 3.4.3; R Foundation for Statistical Computing: Vienna, Austria, 2017.

49. Hothorn, T.; Bühlmann, P.; Dudoit, S.; Molinaro, A.; van der Laan, M.J. Survival ensembles. *Biostatistics* **2006**, *7*, 355–373. [CrossRef]

50. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [CrossRef] [PubMed]

51. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J.C.; Sheridan, R.P.; Feuston, B.P. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947–1958. [CrossRef] [PubMed]

52. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*. [CrossRef]

53. Powell, M.J.D. *The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives*; Cambridge NA Report NA2009/06; University of Cambridge: Cambridge, UK, 2009; pp. 26–46.

54. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [CrossRef]

55. Rohner, B.; Waldner, P.; Lischke, H.; Ferretti, M.; Thürig, E. Predicting individual-tree growth of central European tree species as a function of site, stand, management, nutrient, and climate effects. *Eur. J. For. Res.* **2018**, *137*, 29–44. [CrossRef]

56. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. *Akaike Information Criterion Statistics*; KTK Scient. Publ: Tokyo, Japan, 1986; p. 290, ISBN 9027722536.

57. Hartig, F. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models. Available online: http://www.webcitation.org/72ruusQZJ (accessed on 2 October 2018).

58. r-project. DHARMa Vignette. Available online: http://www.webcitation.org/72rtsdcBU (accessed on 2 October 2018).

59. Faraway, J.J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*; CRC Press: Boca Raton, FL, USA, 2016; p. 399. ISBN 1498720986.

60. Renaud, O.; Victoria-Feser, M.-P. A robust coefficient of determination for regression. *J. Stat. Plan. Inference* **2010**, *140*, 1852–1862. [CrossRef]

61. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998; p. 590, ISBN 9780387985022.

62. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]

63. Kändler, G. The design of the second German national forest inventory. In Proceedings of the Eighth Annual Forest Inventory and Analysis Symposium, Monterey, CA, USA, 16–19 October 2006; McRoberts, R.E., Reams, G.A., van Duesen, P.C., McWilliams, W.H., Eds.; U.S. Department of Agriculture: Washington, DC, USA, 2006; pp. 19–24.

64. Spiecker, H. Silvicultural management in maintaining biodiversity and resistance of forests in Europe—Temperate zone. *J. Environ. Manag.* **2003**, *67*, 55–65. [CrossRef]

65. Fortin, M. Population-averaged predictions with generalized linear mixed-effects models in forestry: An estimator based on Gauss−Hermite quadrature. *Can. J. For. Res.* **2013**, *43*, 129–138. [CrossRef]

66. Fortin, M. Using a segmented logistic model to predict trees to be harvested in forest growth forecasts. *For. Syst.* **2014**, *23*, 139. [CrossRef]