Forestry: An International Journal of Forest Research, 2024, **97**, 339–348
https://doi.org/10.1093/forestry/cpad053
Advance access publication date 4 November 2023
Original Article

The relationship between forest structure and naturalness in the Finnish national forest inventory

Mari Myllymäki[1,*], Sakari Tuominen[1], Mikko Kuronen[1], Petteri Packalen[1] and Annika Kangas[2]

[1]Natural Resources Institute Finland (Luke), Bioeconomy and Environment Unit, Latokartanonkaari 9, FI-00790 Helsinki, Finland
[2]Natural Resources Institute Finland (Luke), Bioeconomy and Environment Unit, Yliopistokatu 6, FI-80100 Joensuu, Finland
[*]Corresponding author. E-mail: mari.myllymaki@luke.fi

Abstract

There is considerable interest in identifying and locating natural forests as accurately as possible, because they are deemed essential in preventing biodiversity loss. In the boreal region, natural forests contain a substantial amount of dead wood and exhibit considerable variation in tree age, size, and species composition. However, it is difficult to define natural forests in a quantitative manner. This is an issue, for example, in the Finnish national forest inventory. If naturalness could be related to the metrics derived from tree measurements, it would be easier to locate natural forests based on the inventory data. In this study, we investigated the value of metrics computed from tree locations and tree sizes for the characterization of a key aspect of naturalness, namely, structural naturalness as defined in the Finnish national forest inventory. We used L-moments, Gini coefficient, Lorenz asymmetry, and interquartile range to quantify the variations in tree size at the plot level. We summarized the spatial pattern of trees with a spatial aggregation index. We compared the structural metrics, species proportions, and stand age using the classes of structural naturalness described in the Finnish national forest inventory, which have been determined in the field without strict numerical rules. These categories are 'natural', 'near-natural', and 'non-natural'. We found that the forests evaluated as structurally natural had larger variations in tree size and species composition and showed a more clustered spatial pattern of trees on average, although the variation in the structural metrics was considerable in all three classes. In addition, we used the structural metrics to predict naturalness by employing a random forest algorithm. Based on the structural metrics, it was possible to obtain high precision in the classification only if we simultaneously accepted low recall, and vice versa; the link between the inspected metrics and naturalness evaluated in the field was weak. The stand age separated the three classes more clearly and it also improved the classification.

Keywords: forest inventory; Gini coefficient; L-moments; random forest; natural forest; spatial structure

Introduction

Primary forests are rare across Europe, but their importance in the prevention of biodiversity loss is considered high (e.g. Gibson *et al.* 2011, Mackey *et al.* 2015). The terminology with regard to these forests is not fully settled, and they have also been called intact, primaeval, virgin, near-virgin, old-growth, or long-untouched forests (e.g. McRoberts *et al.* 2012, Potapov *et al.* 2017, Sabatini *et al.* 2018). The Food and Agriculture Organization (FAO) collects information on primary forests in their global Forest Resources Assessment and have defined primary forests as 'naturally regenerated forests of native species, where there are no clearly visible indications of human activities, such as logging, road construction or anthropogenic fires, and the ecological processes are not significantly disturbed' (FAO 2015). This definition covers forests that exhibit a high level of naturalness, without implying that these forests were never cleared or disturbed by humans. A similar definition was also adopted by Sabatini *et al.* (2018). The European Union define as main features of natural old forests a considerable amount of dead wood and coarse woody debris, a large variation in tree age, tree height, and species composition, occurrence of trees from previous generations, and a stable microclimate (habitat type '9010: Western Taïga' of Annex I of the Habitats Directive, https://eunis.eea.europa.eu/habitats-annex1-browser.jsp).

According to Morales-Hidalgo *et al.* (2015), 202 out of 234 countries harbour some area of primary forests with the total area reported by these countries estimated at 1277 million ha, which is 32% of the forest area in those countries. The largest areas of primary forests are registered in the Russian Federation, Canada, and Brazil. In the Forest Europe report (2020), 2.2% of the total forest area is considered untouched. Sabatini *et al.* (2018) searched available data sources, such as descriptions in journal articles, existing maps, and questionnaires to experts, and estimated that primary forests cover 1.4 million ha in 32 countries, which is 0.7% of the forest area in Europe.

Because of their importance, it is critical that primary forests are identified and located as accurately as possible. To facilitate the identification and mapping of natural forests, there is an urgent need for a general formulation of the concept of 'forest naturalness' (Brumelis *et al.* 2011). Brumelis *et al.* (2011) identified three dimensions according to which naturalness can be formulated: (i) structure-based, (ii) species-based, and (iii) process-based concepts. Of these, the structural concept is most promising from the viewpoint of locating natural forests using remote

Handling editor: Helen McKay
Received: February 22, 2023. **Revised:** September 27, 2023. **Accepted:** October 10, 2023

sensing-based forest resource data. The structural variables of interest include old trees, variation in tree species composition, occurrence of large deciduous trees, multi-layered and multi-aged tree canopies, dead wood of varying sizes and decay stages, as well as signs of natural disturbances (fire, wind, insects, and fungi) (Brumelis *et al.* 2011). The use of species-based definitions is difficult as identification of species (e.g. bracket fungi) is problematic with field measurements, and impossible with remote sensing methods. Even the identification of the tree species that occur in the area is very uncertain with remote sensing. The process-based concept involves concentrating on the dynamics that have led to the current state.

Many of the metrics available from commonly measured field inventory data, such as occurrence of multiple height layers, are important structural variables, but some relevant variables are typically missing. For instance, age is not usually measured for all trees, and often only a mean age is available, as tree age measurements are too expensive and, moreover, are prohibited on permanent sample plots. Therefore, instead of multiple age layers, it is only possible to address the variability associated with stem diameter. Signs of natural disturbances may also be lacking from the measurements commonly taken in the field.

National forest inventories (NFI) are the most comprehensive and extensive operationally available data sets for the assessment of naturalness (McRoberts *et al.* 2012). In the Finnish NFI, the naturalness of a forest is assessed visually in the field according to three criteria: (i) the structure of the forest, (ii) the amount and composition (continuum) of dead wood, and (iii) signs of human activities. For each variable, the level of naturalness is determined at three levels or classes. According to the 11th Finnish NFI, with data collected in 2009–13, the area of productive forest that fulfilled these criteria was 517 000 ha or 2.5% (Korhonen *et al.* 2017), and according to the 12th Finnish NFI (NFI12), with data collected in 2014–18, the area was about 380 000 ha or 1.9% (Korhonen *et al.* 2021). In the analysis by Sabatini *et al.* (2018), Finland had the greatest proportion of primary forest in Europe (0.9 million hectares, ~3% of the national territory).

The variation across the consecutive NFI may be partly because of the absence of unambiguous definitions for the three criteria that define naturalness, as substantial changes have also been noted in conservation areas. In the Finnish NFI, naturalness based on structure is defined to exhibit, for example, large variation in tree size, as well as a random spatial pattern, but without numerical rules that define what should be interpreted as large variation. This is partly because the internal variation of natural forests (because of site type and dominant species) is considerable. However, ambiguous definitions also give room to subjective considerations, which makes the results harder to interpret and use.

The aim of this study was to determine whether the level of forest structural naturalness, as defined and assessed in the field in the Finnish NFI, varies according to the variables of forest structure computed from NFI field measurements. A better modelled relationship between structural naturalness and the structural variables (tree diameter and height, or other field measurements) will lead to more informative and unambiguous definitions in the NFI. If the relationship between the variables computed from tree measurements and the structural naturalness is vague, then there is a clear need to reformulate both the definitions and the NFI measurements. The need to improve the definitions in the NFI and the potential effect of deficiencies in the currently used definitions on policies is discussed.

Materials

We used tree- and plot-level data from the Finnish NFI (2018–20). The 2018 data were part of NFI12, and the 2019–20 data were part of 13th NFI (NFI13). For each plot, diameter growth at breast height (dbh) was measured for all tally trees, and height for all sample trees (a subset of tally trees). Heights of non-sample trees and upper diameter at 6 m for all trees were predicted as described in Korhonen *et al.* (2021). Stem volume for each tree was predicted using the species-specific three-predictor volume models proposed by Laasasenaho (1982). We only considered such plots in forest land that were completely located within a single stand and contained at least five trees with measured dbh > 4.5 cm and either measured or predicted height. As we only analyzed the structure of standing trees, fallen dead trees or tree stumps were not included. Stand age was defined as the average age of the dominant tree storey (crown layer) of the stand where the plot was located. Age was measured with an age borer on a few trees at breast height (1.3 m), and the total age was obtained by adding species and region-specific age increment values that represent the time it takes for a tree to reach breast height.

Circular plots were used in both NFI12 and NFI13. Trees with dbh ≥ 9.5 cm were measured within a circle of 9 m radius, whereas trees with dbh and ≥ 4.5 < 9.5 cm were measured within a 5.64 m radius in NFI12 or a 4 m radius in NFI13. In our analyses, we ignored trees with dbh < 4.5 cm, which were sampled with a relascope. The unequal inclusion probabilities introduce weightings for trees with a different dbh. The weight assigned to tree k was $w_k \propto 1/r_k{}^2$, where r_k is the radius from which tree k was sampled.

Of the three criteria used to define naturalness in the Finnish NFI (see above), this study concentrated on the structure of the forest, which is divided into three classes as follows:

- Natural (class 0) forests are virgin forests or forests that are close to their natural state. The trees should have random spacing and variable size, and the canopy cover should have several vertical layers. Trees of several generations should be present, and the dominant tree stratum should have achieved at least silvicultural maturity age (i.e. fulfilling the recommended species and site-specific age limit for final cutting, Äijälä *et al.* 2019). In forests close to a natural state, there may be some signs of previous selective cuttings, which should not have markedly changed the tree spacing and species composition. In addition, forests that have regenerated naturally after the occurrence of natural damage (e.g. forest fire or storm) are assigned to class 0 regardless of the stand age, in cases when (i) the damaged stand had reached maturity, and (ii) the damaged trees have not been harvested.
- Near-natural (Class 1) forests have regenerated naturally, but their structure and spacing have been changed by slight thinning or selective cutting.
- Non-natural (Class 2) forests are, in general, evenly spaced and even-aged, and the forest structure is formed by artificial regeneration or cutting. All planted forests belong to this class, even if they are unevenly structured or unmanaged.

For our analyses, we considered data from three parts of Finland, as depicted in Fig. 1. The division was carried out with respect to latitude, because of both the inherent variation in Finnish forests and the amount of natural forests found along the north–south gradient. The northern study region (hereafter called the North region) represents Finnish Lapland, excluding Northern Lapland, which was not measured in 2018–20. The

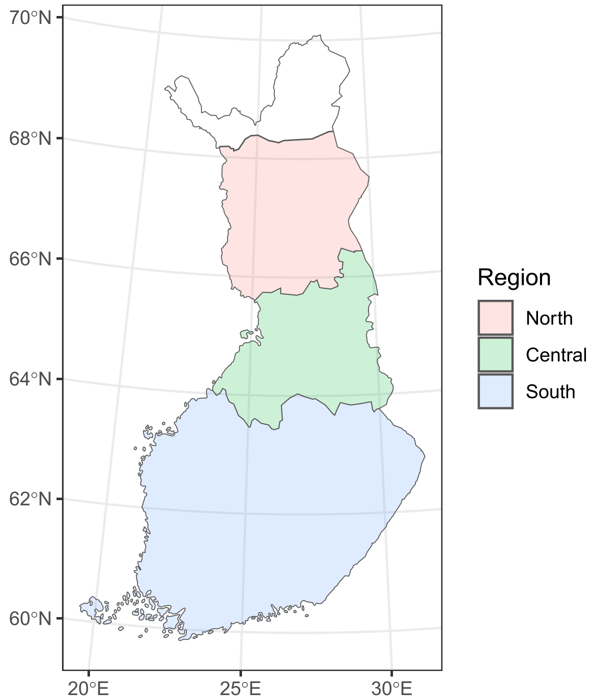

Figure 1. Three geographical regions of Finland utilised in the study. The North region represents Lapland, excluding Northern Lapland, the Central region is composed of North Ostrobothnia and Kainuu, and the South region is the remainder of the country.

Table 1. Summary of the number of plots in the three naturalness classes (0 = natural, 1 = near-natural, 2 = non-natural) in the three regions of Finland shown in Figure 1.

Region	*n*	Naturalness class		
		0	1	2
North	1473	223	113	1137
Central	2335	70	142	2123
South	6748	66	392	6290

central study region (hereafter called the Central region) consists of North Ostrobothnia and Kainuu, and the remainder of Finland was the southern study region (hereafter called the South region). On the basis of a visual assessment of structure, the North region contained the most structurally natural plots (62% of the structurally natural plots included in this study, Table 1) and the Central region contained more than half of the remaining structurally natural plots. Summary of the plots with respect to the other two naturalness criteria can be found in Table S.1 in Supplementary Material. Hereafter naturalness refers to structural naturalness only.

We hypothesized that, in addition to the geographical location, the structural properties of natural and non-natural forests may differ with respect to site conditions and the dominant tree species or species groups. Therefore, we considered the variation because of soil type (mineral land or peatland) and site type (rich or poor sites), as well as the variation because of dominant species. Here, rich site types included forest stands that ranged from mesic to herb-rich sites on mineral land and peatlands that ranged from mesotrophic to eutrophic mires; poor site types included forests that ranged from sub-xeric to barren sites on mineral land and peatlands that ranged from oligotrophic to ombrotrophic mires. With regard to dominant species, we used three groups: (i) conifer-dominated stands where the number of conifers exceeded 67%, (ii) broadleaved-dominated stands with more than 67% broadleaved trees, and (iii) mixed stands.

In the North region, natural forests were most common in mixed stands on mineral land: 31% of mixed stands were natural on poor mineral land, 25% on rich mineral land, as well as 25% of broadleaved stands growing on rich mineral land. However, almost all broadleaved and mixed stands on poor peatlands were non-natural (Table S.2). In the Central region, natural forests were most common in conifer stands on rich peatland (6.5%) and in mixed stands on poor mineral land (6%). In the South region, the proportion of natural forests was greatest in mixed stands on poor mineral land, although the proportion was even lower (3.4%). In the Central and South regions, none of the broadleaved-dominated stands on poor sites were natural, and neither were any of the mixed stands on poor peatlands or broadleaved stands on rich peatlands.

We also investigated the number of stands according to their development class (Table S.3). The development class describes the developmental phase of the growing stock in relation to the expected rotation determined in the field (Tomppo *et al.* 2011). Natural forests were mainly assigned to the development classes that represented young thinning stands, advanced thinning stands or mature stands (Table S.3). Here a young thinning stand indicates that the first commercial thinning has not yet been performed. In the advanced thinning stands, the growing stock is older and the trunk size is larger. The growing stock of a mature stand is either old or is considered sufficiently large for a regeneration cutting from a management point of view.

Methods

We compared the natural (0), near-natural (1), and non-natural (2) plots using several indices that characterized the structure of the stand based on the size and location of the trees. To characterize the shape of the size distributions, we used L-moments (mean, L-scale, L-skewness, L-kurtosis), the Gini coefficient, the Lorenz asymmetry (LA) coefficient, and the interquartile range (IQR). We considered these indices for dbh (cm), basal area (dm^2), height (m), and volume (dm^3) of the trees in the sample plot. To summarize the spatial pattern of the trees, we used a traditional aggregation index based on nearest neighbour distances (Clark and Evans, 1954). We further inspected stand age, total volumes, and species proportions, as specified below.

The IQR is the simplest of the structural variables tested as it is the difference between the low (25%) and high (75%) quartiles of the variable of interest. Because the trees of different sizes had different inclusion probabilities, we used a weighted version of IQR from the R (R Core Team 2021) package DescTools (Signorell *et al.* 2022).

The main advantage of L-moments over conventional moments is that they are more robust to outliers in the data, and they enable more secure inferences about the underlying probability distributions based on small sample sizes (Hosking 1990). The L-moments defined by Hosking (1990) are

$$L_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathrm{E}X_{r-k:r} \tag{1}$$

where $X_{r-k:r}$ is the $(r-k)$th-order statistic of the variable X in a sample of size r. The first four L-moments are

$$L_1 = \mathrm{E}X \tag{2}$$

$$L_2 = (\mathrm{E}X_{2:2} - \mathrm{E}X_{1:2})/2 \tag{3}$$

$$L_3 = (\mathrm{E}X_{3:3} - 2\mathrm{E}X_{2:3} + \mathrm{E}X_{1:3})/3 \tag{4}$$

$$L_4 = (\mathrm{E}X_{4:4} - 3\mathrm{E}X_{3:4} + 3\mathrm{E}X_{2:4} - \mathrm{E}X_{1:4})/4 \tag{5}$$

L_1 is more conventionally the mean and L_2 is the scale. The ratio $\frac{L_3}{L_2}$ is L-skewness, and the ratio $\frac{L_4}{L_2}$ is the L-kurtosis. According to Hosking (1990), the rth sample moment of $x_1, \ldots, x_n$ can be estimated by

$$l_r = \binom{n}{r}^{-1} \sum \cdots \sum\nolimits_{1 \le i_1 < i_2 < \cdots < i_r \le n} r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{(i_{r-k})}$$

where $x_{(i)}$ is the ith-order statistics of x. Because of unequal inclusion probabilities, we used the following weighted version of (6) as the estimator of the rth sample moment

$$l_r = \left(\sum_{1 \le i_1 < i_2 < \cdots < i_r \le n} \prod_{k=1}^{r} w_{(i_{r-k})} \right)^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_r \le n} \left(r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{(i_{r-k})} \right) \left(\prod_{k=1}^{r} w_{(i_{r-k})} \right)$$

where $w_{(i)}$ are the weights associated with the order statistics $x_{(i)}$. Here, we weighted each r-tuple of statistics $x_{(i)}$ by the product of their weights.

For a continuous variable X, the Gini index is defined through the distribution function $F(x)$ and its expectation μ

$$G = \frac{1}{\mu} \int_0^{\infty} F(x)\,(1 - F(x))\,dx \tag{7}$$

The Gini coefficient is a measure of inequality. For non-negative random variables, it is related to the L-moments by $G = L_2/L_1$. We used the R package dineq (Schulenberg 2018) for the calculation of the weighted Gini coefficient. Interpretation of the Gini coefficient has been discussed, for example, in Wittebolle *et al.* (2009) and Valbuena *et al.* (2012).

We further calculated the LA coefficient according to Valbuena *et al.* (2013). Let d_{QMD} be the weighted quadratic mean diameter in a plot with n trees with dbh d_k, $k = 1, \ldots, n$, and associated weightings w_k, i.e.

$$d_{\mathrm{QMD}} = \sqrt{\sum_{k=1}^{n} w_k d_k^2 \Big/ \sum_{k=1}^{n} w_k} \tag{8}$$

Furthermore, let

$$D(x_{\mathrm{QMD}}) = \frac{\sum_{k=1}^{n} \mathbf{1}(d_k > d_{\mathrm{QMD}})\, w_k}{\sum_{k=1}^{n} w_k},$$
$$M(x_{\mathrm{QMD}}) = \frac{\sum_{k=1}^{n} \mathbf{1}(d_k > d_{\mathrm{QMD}})\, w_k \pi d_k^2}{\sum_{k=1}^{n} w_k \pi d_k^2} \tag{9}$$

be the proportions of the basal area and the stem density of the trees with $d_k > d_{\mathrm{QMD}}$ from the totals. Then, LA is the average of these two proportions

$$\mathrm{LA} = \left(M(x_{\mathrm{QMD}}) + D(x_{\mathrm{QMD}}) \right)/2 \tag{10}$$

We used the aggregation index R (Clark and Evans 1954) as a measure of clustering or regularity of the pattern of the trees. We calculated this index for trees with dbh $\geq$ 9.5 cm that were observed within the 9 m radius plots. This was carried out as we assumed that the spatial distribution of larger trees would be more indicative of naturalness, and on the other hand, we are also not aware of a weighted estimator of R. Index R is the ratio of the observed mean nearest neighbour distance in the tree pattern to that expected for a Poisson point process of the same intensity. We estimated the R index using a Kaplan–Meier type of edge correction as implemented in the R library spatstat (Baddeley *et al.* 2015). In theory, the aggregation index can obtain values between 0 and 2.1491. A value $R > 1$ suggests regularity, whereas $R < 1$ suggests clustering. We chose the index R, which is based on nearest neighbour distances, because it is not possible to evaluate clustering or regularity at large inter-tree distances based on the rather small NFI sample plots. In total, there were 307 plots (0 natural, 9 near-natural, and 298 non-natural) where R could not be estimated with the chosen edge correction. These cases were omitted from the distribution graphs in the Results section. For the random forest (RF) algorithm, their value was set to 1, which corresponds to the case of complete spatial randomness.

We first investigated the distributions of the different indices within the naturalness classes and site conditions. Second, we used the RF algorithm (Breiman 2001) as implemented in the R package randomForest (Liaw and Wiener 2002) to investigate how well the indices, together with site conditions, could assign forest stands to the naturalness classes. In addition to the indices, as explanatory variables in the RF algorithm, we added: the total basal area in the plot; total basal area and proportions of small (dbh < 15 cm) and large (dbh > 30 cm) trees; proportions of various species groups (spruce, pines, conifers, broadleaved); and optionally stand age. Then, we separately carried out the classification with the RF algorithm and the selected variables for the three regions shown in Fig. 1. For comparison, we also tested it for the whole country, without any regional information. Our dependent variable had only two classes: (i) we combined the near-natural forests either with the non-natural forests and tried to predict the natural forests, or (ii) we combined it with the natural forests in an attempt to predict 'natural or near-natural' forests. We randomly selected 75% of the plots to train the RF algorithm and the remaining 25% were used for validation. The data were split separately for both response classes. We report the average results from a 100 train-test splits of the data. We used the variable selection using random forests (VSURF) (Genuer *et al.* 2015) to evaluate the importance of the variables. Namely, for each variable, we computed the number of train-test splits where that variable was selected by VSURF.

We used different metrics as measures of goodness of predictions. All of these metrics can be derived from the numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as well as the true numbers of positives (P) and negatives (N) with P + N giving the total number of observations. We inspected the rather classical metrics; overall accuracy $\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{P}+\mathrm{N}}$, Cohen's kappa and F_1-score, as well as precision and recall. The precision (or user's accuracy) is the proportion of correct classification in the group that was classified as positive, $\frac{TP}{TP+FP}$, i.e. the proportion of natural plots amongst the plots classified as natural. The recall (or producer's accuracy) is the proportion of correct classification in the group of real positives in the data, $\frac{TP}{TP+FN}$, i.e. the proportion of natural plots classified as natural. The F_1-score is the harmonic mean of precision and recall, with 1 indicating perfect precision and recall, and 0 being the lowest possible value.

We inspected the above metrics for the default class predictions given by the randomForest package. For more accurate analyses, the RF algorithm also provides the class probabilities. By default, a prediction is classified to the target group (natural) if the class probability is larger than the cut-off value of 0.5. However, in an imbalanced situation, the choice of the cut-off is not obvious and a powerful solution is instead to consider all possible cut-offs (e.g. Saito and Rehmsmeier 2015). Thus, we computed the precision and recall for the cut-offs of unique class probabilities between 0 and 1, each cut-off providing different TP, TN, FP, and FN values and thus different precision and recall values.

Results

The Gini coefficient of basal area ($G < 0.5$) suggests a slight tendency of non-natural plots to be even-aged, and the tendency of the natural and near-natural plots to be irregular or have a reverse J distribution type ($G \approx 0.5$, according to Valbuena *et al.* 2016) (Fig. S.1 in Supplementary Material). In the North and Central regions (see Fig. 1), the near-natural group seemed to lie between the other two groups, whereas in the South region, the Gini coefficient values in the natural and near-natural groups seemed to coincide, and the non-natural group was more distinct with lower Gini values. However, the range of the Gini coefficients associated with dbh, basal area, height, and volume in each group overlapped notably (Fig. S.1). This means that it was possible to distinguish only the most even-aged plots as non-natural, but most plots were somewhat irregular, which made identification based on the Gini coefficient very uncertain.

The IQR distributions in the natural, near-natural, and non-natural NFI sample plots overlapped in a similar manner to those of the Gini coefficient (Fig. S.2). There was a slight tendency in the natural plots to have a higher IQR, although each group exhibited a wide range of values.

The L_1 moments indicated that the largest mean dbh, basal area, height, and volume values occurred slightly more often in the natural plots in all regions (Fig. S.3). The differences between the groups were clearer in the L_2 moment (L-scale): the variation was larger in the natural plots (Fig. 2). The near-natural group seemed to lie within the other two groups, even though the distributions of L_2 in some cases almost coincided with the natural and near-natural groups. With regard to L-skewness, the groups differed only slightly in the North region (Fig. S.4), and the differences in L-kurtosis distributions were negligible (Fig. S.5). Overall, the ability of L-moments (L_2) to separate the classes was comparable to that of the Gini coefficient and the IQR.

The distributions of the LA coefficients were similar in all naturalness groups in all three regions (Fig. S.6). However, the aggregation index R again showed some patterning, even though the R values also had considerable variability in each group (Fig. S.7). In any case, the non-natural plots showed a slight tendency towards regular spatial arrangement of trees and the natural plots towards a clustered spatial arrangement. In the Central and South regions, the natural and near-natural groups were somewhat similar with respect to their R value distributions.

We hypothesized that the lack of structural variation between the groups was partly because of the site type, which represents the different fertility classes that support the various types of tree populations. As such, the poorest sites were often pine-dominated and the rich sites usually had a more variable tree species composition. Therefore, we further inspected the Gini coefficient with respect to the dominant species groups (Fig. S.8) and the soil (mineral or peatland) and site (poor or rich) types (Fig. S.9). The naturalness groups seemed to differ most clearly in the broadleaved dominated stands located in the North and South regions. However, there were only three natural broadleaved dominated stands in the Central region, and four in the South region, in comparison to 37 in the North (Table S.2). With regard to soil and site types, the differences between the groups seemed more prominent in the mineral soil.

The natural forests were assumed to exhibit large variation in tree species composition, in addition to structural variation, but the observed differences between the natural, near-natural, and non-natural forests in the NFI plots was also small in this respect. The plots where the recorded tree species were solely conifer (0% broadleaved) had a slightly greater probability of being non-natural than the others, but in other respects there were no clear differences between the groups (Fig. S.10). We also inspected the number of trees with dbh < 15 and > 30 cm and observed that their proportions were somewhat greater in natural than in non-natural forests (Fig. S.11). This was also reflected by the IQR and associated variation (L-scale), which were somewhat larger in the natural plots.

The differences between the naturalness groups were more prominent with respect to stand age (Fig. 3). The natural forests were older, and across all three regions, about 23% of natural forests were older than 200 years, whereas the corresponding proportions in near-natural and non-natural forests were less than 3% and 0.2%, respectively. This meant that 72% of forests older than 200 years were natural. Furthermore, the proportions of natural, near-natural, and non-natural forests greater than 150 years were 60%, 13%, and 1%, respectively.

We then used the RF algorithm to classify the plots as natural (or natural or near-natural) using all the above-mentioned indices, as well as site variables. We carried out the same analyses with and without stand age. Predicting a plot as natural if its class probability was larger than the default cut-off 0.5, the overall accuracy was always rather high (>0.82 in all cases, see Table S.4), as the proportion of the natural plots was rather low. However, Cohen's kappa and the F_1-score were rather low: in the North region, for example, the F_1-score varied between 0.41 and 0.72, whereas in the Central and South regions, it was close to 0 when predicting the natural group without stand age, while it increased to between 0.03 and 0.24 when stand age was included or when the near-natural plots and natural plots were combined.

The trade-off between recall and precision for all possible cut-offs is shown in Fig. 4. As with the basic metrics, classification worked better in the North region than in other regions (higher precision for a given recall) and also worked better if the group to be predicted consisted of both natural and near-natural plots (see solid and dashed lines in Fig. 4). Furthermore, the inclusion of stand age in the set of predictions clearly improved the results (see red and blue lines in Fig. 4).

Figure 4 is indicative of how the results could be used: let us assume that in the North region, one would like to find 75% of the natural plots (recall 0.75). This could be reached, but only if we accept that about 38% of plots classified as natural were non-natural (precision 0.62). To find 75% of the natural or near-natural

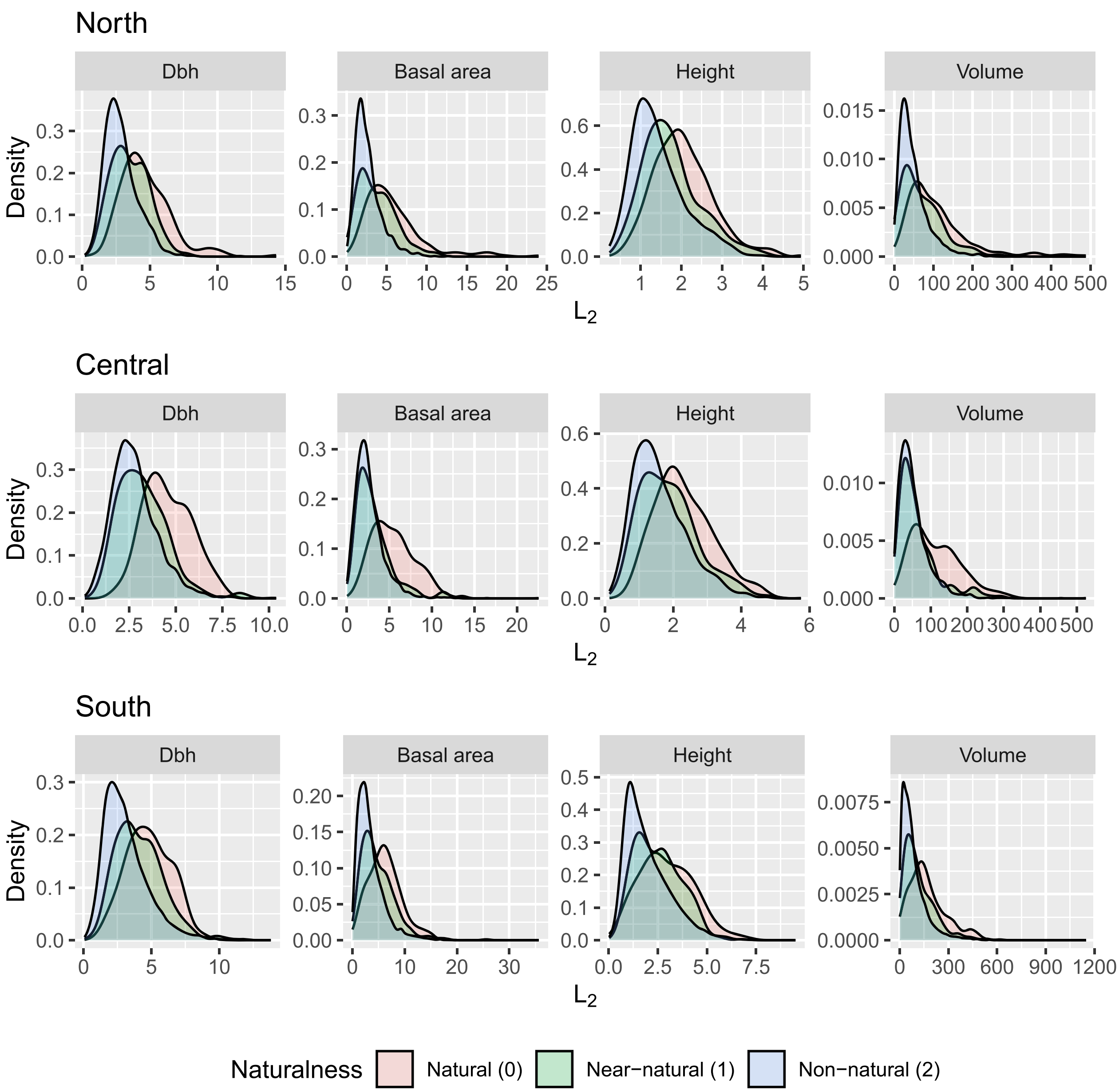

Figure 2. The second L-moment in the NFI plots in the three naturalness groups and regions shown in Figure 1.

plots, the proportion of non-natural plots assigned to the target group (natural or near-natural) was lower (25%). The proportion of forests predicted as natural (19% or 23% when near-natural was also included) was close to the true proportion (15% or 23%, Table 1). The classification worked less successfully in the Central and South regions. For example, in the South region, combining the near-natural and natural plots, using stand age (the best case), and classifying 50% of the natural plots as natural (recall 0.5) meant accepting that 69% of the plots classified as natural were non-natural (precision 0.31) and predicting 11% of forests to this category even though the correct proportion was 6.8% (Table 1).

The most important indices in the VSURF classifications are shown in Fig. 5 when the predicted class consisted of natural and near-natural plots. Stand age was clearly the most important predictor in all cases. It was chosen by VSURF in each of the 100 train-test splits. In the South region and for the whole of Finland, no other indices were chosen more than five times. In the North region, the proportion of pines and the total basal area of trees with dbh $>$ 30 cm were the top predictors after stand age. In addition, the proportion of spruce and the L_2-moment of the dbh and basal area distributions were important predictors as they were chosen more than 25 times. In the Central region, the total basal area of the plot, number of trees with dbh $<$ 15 cm, and the aggregation index R were found to be the most important predictors after stand age.

Discussion

The distributions of the NFI plot structural variables of natural, near-natural, and non-natural forests were surprisingly similar, irrespective of the variables tested. The natural plots had greater variability in tree size, greater proportions of broadleaved trees, and more clustered pattern of trees on average, but the variability was high in all groups leading to large overlaps in the value ranges

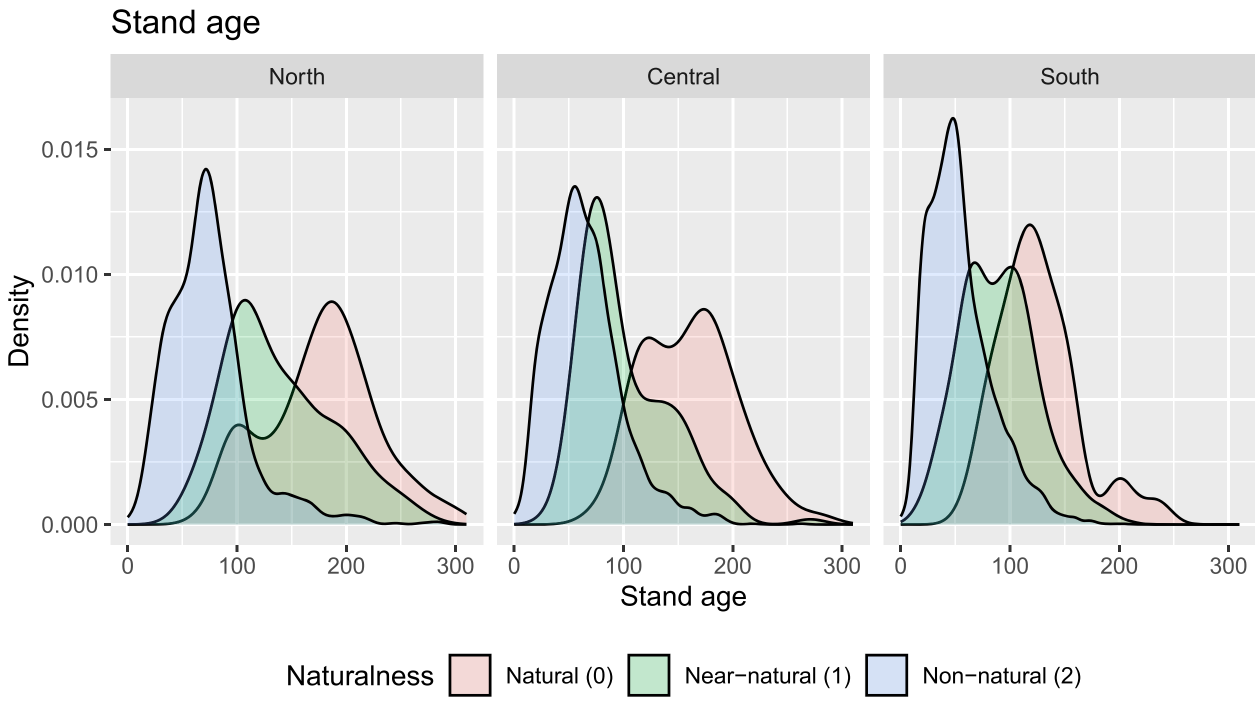

Figure 3. Stand age (years) in the three naturalness groups and regions shown in Figure 1.

of the three classes. The groups were, however, notably different with respect to stand age, with the natural forests much older. In classifying the forests as natural and non-natural (as defined in the Finnish NFI) through the NFI tree measurements carried out in the field, there was a clear trade-off between recall and precision: it was possible to find a substantial proportion of the natural (and near-natural) forests (high recall), but only when we accepted that many non-natural plots also had to be assigned to the natural class (low precision). In the North region, the classification results using the RF algorithm were sufficiently good to be considered operationally useful, particularly when stand age was also used as a predictor. In contrast, classification in the Central and South regions was poor.

Because of the small amount of natural plots in the prediction data set, it can be preferable to adjust the cut-off value to the user's needs. On one hand, the greater the number of non-natural plots that are classified as natural (high FP, low precision), then the larger the proportion of natural forests found (high recall). On the other hand, if the number of FP is required to be small, then only a small proportion of natural plots are found (low recall). Thus, the users can choose between (i) finding a large proportion of natural forests with the increased cost of checking a lot of FP in the field, or (ii) not undertaking field checks with the cost of not finding all the natural forests. For example, in the North region, the requirement to find 75% of natural plots was possible, but only if we accepted that about 38% of the plots classified as natural could be non-natural (or near-natural). The percentage of forests predicted as natural was 19%, close to the true proportion, 15% (Table 1).

It is notable that Uotila *et al.* (2001) observed statistically significant differences between natural and managed forests in regard to both age and diameter distributions, but not with regard to the proportions of broadleaved trees. Their plots were 900–2500 m^2 in size, in comparison to our 254 m^2. One obvious reason for our results may be the small size of the NFI plots. It is possible that a larger plot size would enhance the performance of the classification (e.g. Häbel *et al.* 2019, and references therein). In general, increasing the radius of the NFI plots is not feasible because of the cost (e.g. Henttonen and Kangas 2015, Häbel *et al.* 2019), but it could be possible to improve the classification, for instance, by measuring exceptionally large trees from a larger area, such as a 20-m radius plot. In contrast, stand age was evaluated at the stand level not at the plot level, which may have facilitated its importance in the RF algorithm classification. Naturalness was also evaluated at the stand level in the field.

Another possibility is that naturalness is mainly assessed in the field based on information other than the spatial pattern of the trees and tree size distribution. In particular, the occurrence of very old trees, or trees of previous generations, is likely to be regarded as an important signs of naturalness by the field group. While tree-level age information is not currently measured, this is reflected by stand age in our study. Moreover, features such as scars because of fire or other natural hazards may have triggered the field groups to assess the plot as natural. Information on such characteristics was not available in this study, as they are currently not recorded in NFI field measurements; only the classification is recorded. Therefore, in the future, it is important to improve data collection in such a way that the properties that the field group use in assessing naturalness is duly recorded and used in the future development of measurement guidelines. This requires obtaining feedback from the field groups on the metrics that they have utilized in their assessment, as well as the inclusion of such indicators in future field measurements. Moreover, it would be useful if the recorded information also included details as to whether (or not) the classification was affected by, for example, trees from the previous generation outside the plot or by signs of non-naturalness, such as the existence of roads in the vicinity of the plot.

Previous studies have noted that the capture of a spatially clustered structure is affected by the selected minimum stem

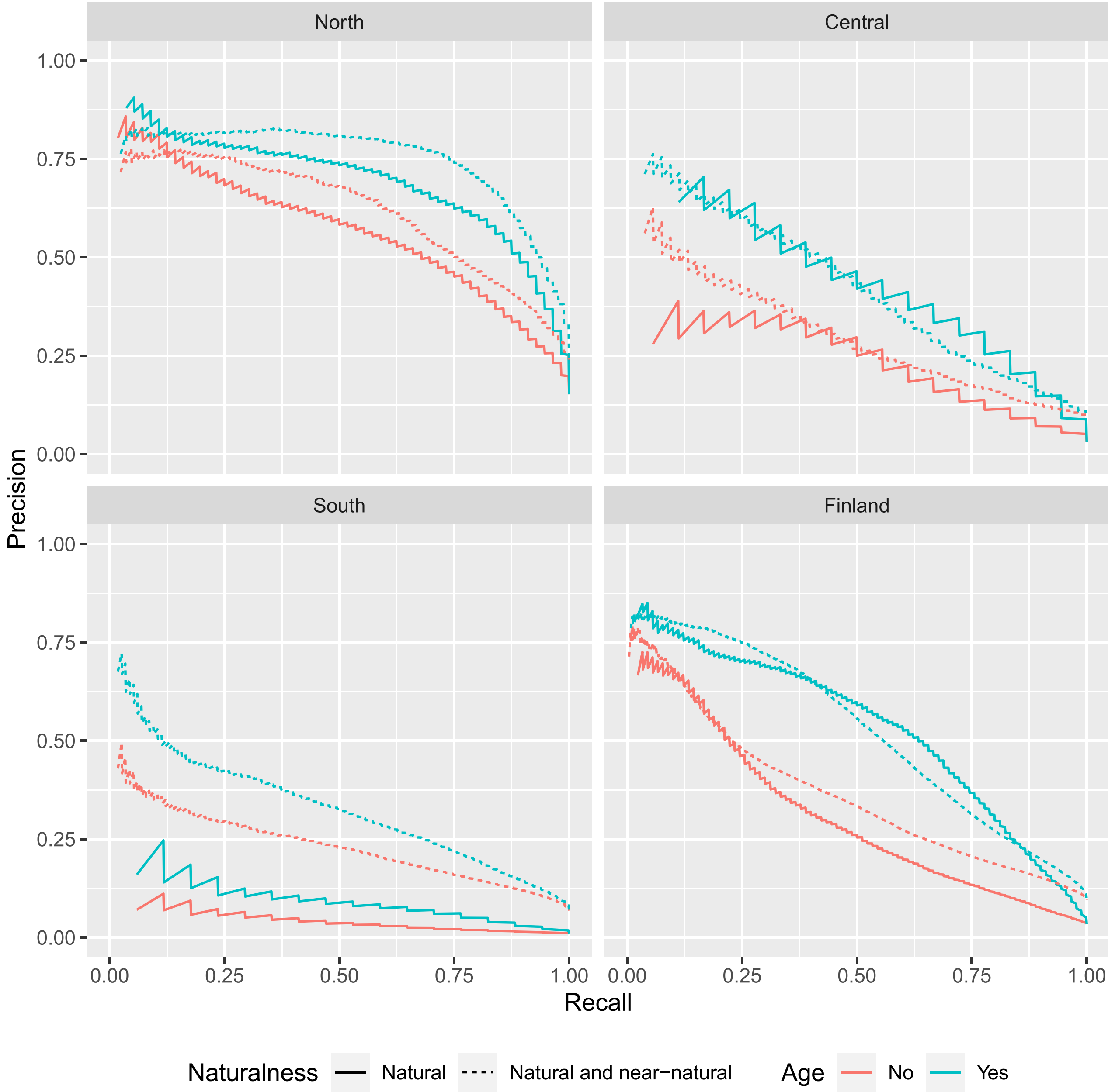

Figure 4. The average precision-recall curve of the RF algorithm classifications in the 100 test sets in the three regions shown in Figure 1 and for the whole of Finland, when the predicted class consisted of (i) natural (solid lines), (ii) natural and near-natural plots (dashed lines), and when stand age was either included or excluded in the set of explanatory variables.

diameter (Häbel et al. 2019), and the larger the minimum diameter the more likely the stands will be regular. In this study, when considering the spatial structure, we ignored the smallest trees that might be clustered in all types of forests and instead calculated the spatial aggregation index for trees with a minimum diameter of 9.5 cm. The applied index had some success in predicting naturalness in the Central region, as measured by VSURF. Assessment of the effect of the spatial patterning could also benefit from larger plot sizes. Recording a (visually assessed) classification of the NFI plots as random, clustered, or regular stands during field measurements might also give further insights into the potential importance of spatial point pattern variables in assessing naturalness of forests in the field.

A final explanation for the relatively low success of structural metrics in detecting field assessed naturalness may lay in its subjective assessment. It is possible that the field group may be looking at different structural metrics in different types of stands. It may also be that different field groups pay attention to slightly different aspects in the forest. Recording all the criteria that affect the classification made in the field would help to understand the dependencies between classification, stand qualities, and structural metrics, as well as the combined effects of the different characteristics. For instance, spatial randomness combined with large tree volumes might be more important than either variable on its own. It is possible that the large variation in previous NFI results is—at least partly—related to subjectivity, and such analysis would also help in characterizing and quantifying this subjectivity. It could be also valuable, though costly, to study the variability between field groups by measuring the same plots by more than one group. This information could be used to provide more detailed guidelines for the field groups, in order to have a consistent classification.

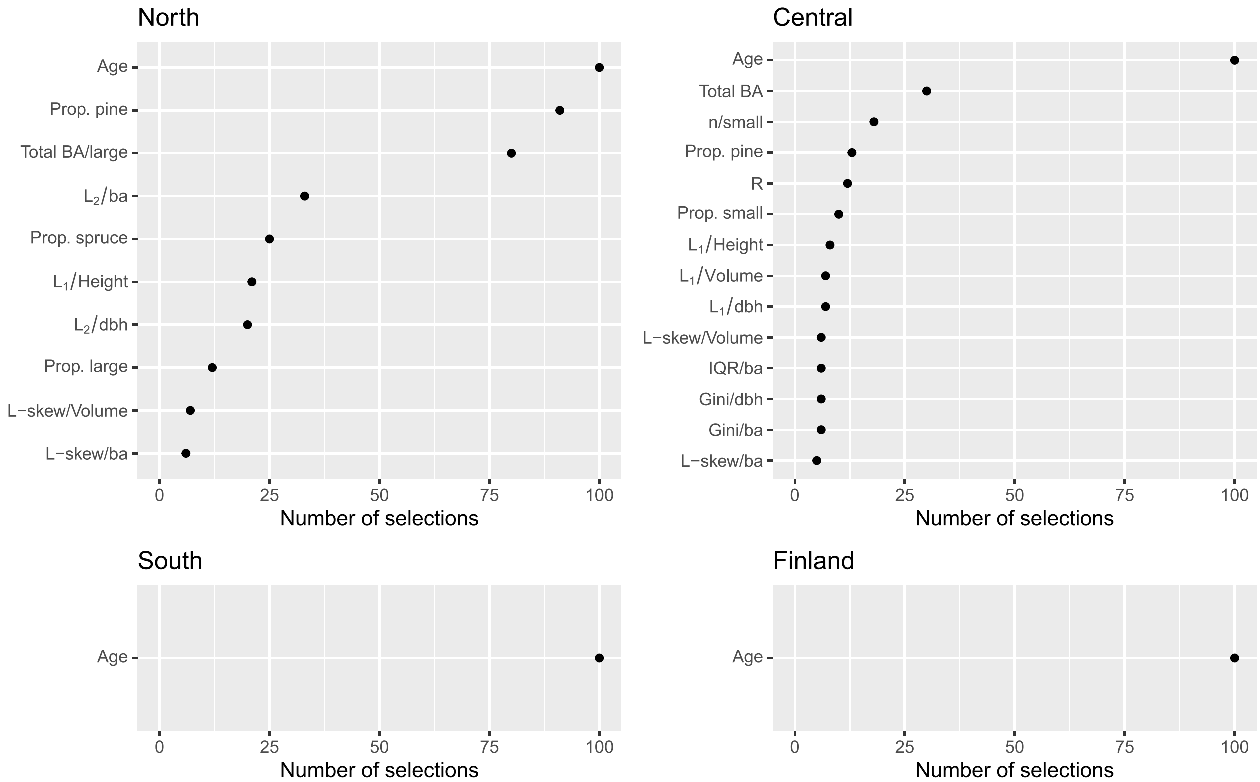

Figure 5. The number of times a variable was selected by VSURF in the 100 train sets in the three regions shown in Figure 1, as well as for the whole of Finland when predicting the group 'Natural or near-natural'. Variables selected less than 5 times were not included. Here 'ba' stands for the basal area of a tree and 'BA' is the stand basal area.

There is considerable interest in predicting forest naturalness or structural metrics from remote sensing data (e.g. Ørka and Naesset 2022), in order to help in local decision-making with regard to the prevention of biodiversity loss. Our results show that even if we were able to map, for instance, the Gini coefficient accurately, the produced map would not directly represent naturalness. From a forest management point of view, however, the possibility of accurately locating forests with the largest structural variability may be valuable as such. This information could be used, for instance, in selecting such forest stands that can be restored to a natural class given sufficient time. Classification based on age might be more related to actual naturalness, but unfortunately stand age is a difficult variable to predict from remote sensing (e.g. Maltamo *et al.* 2020, Ørka and Naesset 2022), particularly in old forests, as the development of height and volume of the stands stabilizes with increasing age. Maltamo *et al.* (2020), for instance, removed all stands older than 100 years from their data set as it was impossible to predict the age in these stands with airborne laser scanning data. On the other hand, remote sensing data can contain other relevant information, e.g. with regard to canopy cover and gap-based indicators (e.g. Häbel *et al.* 2021), which may proof useful to separate natural and non-natural forests.

Conclusion

Our analysis shows that it is a considerable challenge to deduce structural naturalness as defined in the Finnish NFI, based on indices of tree sizes and tree locations only. Trees in natural forests were on average more clustered and exhibited a larger variation in size than non-natural forests, but the distributions of all the studied structural metrics overlapped considerably. The small sample plot size may partly explain these results. Stand age was a clearly better indicator of naturalness. Our results revealed that it is necessary to improve the way that naturalness is recorded in the field, both from the perspective of more reliable mapping of natural forests and from the perspective of improving the quality of NFI naturalness assessments.

Author contributions

Mari Myllymäki (Conceptualization, Formal analysis, Writing—review & editing), Sakari Tuominen (Conceptualization, Writing—review & editing), Mikko Kuronen (Conceptualization, Formal analysis, Writing—review & editing), Petteri Packalen (Conceptualization, Writing—review & editing), and Annika Kangas (Conceptualization, Writing—original draft, Writing—review & editing)

Supplementary data

Supplementary data are available at *Forestry Journal* online.

Conflict of interest

None declared.

Funding

This work was supported by funding from the Ministry of Agriculture and Forestry in Finland [Research grant VN/22626/2020 (PLUS)] and the Research Council of Finland's flagship programme, the Forest-Human-Machine Interplay (UNITE) flagship (Grant numbers 337655 and 357909) and by the European

Union—NextGenerationEU in the Research Council of Finland's project (Grant number 348154) under UNITE.

Data availability

The authors do not have permission to share the original NFI data. The processed data used in this article will be shared on reasonable request to the corresponding author with permission of Natural Resources Institute Finland (Luke).

References

Äijälä O, Koistinen A, Sved J, Vanhatalo K, Väisänen P. (eds.) *Metsänhoidon Suositukset*. Tapion julkaisuja, 2019. https://tapio.fi/wp-content/uploads/2020/09/Metsanhoidon_suositukset_Tapio_2019.pdf.

Baddeley A, Rubak E, Turner R. *Spatial Point Patterns: Methodology and Applications with R*. London: Chapman and Hall/CRC Press, 2015.

Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32. https://doi.org/10.1023/A:1010933404324.

Brumelis G, Jonsson BG, Kouki J. *et al.* Forest naturalness in northern Europe: perspectives on processes, structures and species diversity. *Silva Fennica* 2011;**45**:807–21. https://doi.org/10.14214/sf.446.

Clark P, Evans F. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 1954;**35**:445–53. https://doi.org/10.2307/1931034.

FAO. Global Forest Resources Assessment 2015. Terms and definitions. In: *Forest Resources Assessment Working Paper 180*. Rome: FAO, 2015; 36.

Forest Europe. *State of Europe's Forests 2020*. 2020. https://foresteurope.org/wp-content/uploads/2016/08/SoEF_2020.pdf (Accessed 30 October 2023)

Genuer R, Poggi JM, Tuleau-Malot C. VSURF: an R package for variable selection using random forests. *R J* 2015;**7**:19–33. https://doi.org/10.32614/RJ-2015-018.

Gibson L, Lee TM, Koh LP. *et al.* Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature* 2011;**478**:378–81. https://doi.org/10.1038/nature10425.

Henttonen HM, Kangas A. Optimal plot design in a multipurpose forest inventory. *For Ecosyst* 2015;**2**:31. https://doi.org/10.1186/s40663-015-0055-2.

Hosking JRM. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J R Stat Soc B Methodol* 1990;**52**:105–24. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x.

Häbel H, Kuronen M, Henttonen HM. *et al.* Effect of spatial structure of forests on the accuracy and costs of plot-level forest resource estimation. *For Ecosyst* 2019;**6**:8. https://doi.org/10.1186/s40663-019-0167-1.

Häbel H, Balazs A, Myllymäki M. Spatial analysis of airborne laser scanning point clouds for predicting forest structure. *Math Comp For Nat Resour Sci* 2021;**13**:15–28. http://mcfns.net/index.php/Journal/article/view/13.2/2021.

Korhonen KT, Ihalainen A, Ahola A, Heikkinen J, Henttonen HM, Hotanen J-P, Nevalainen S, Pitkänen J, Strandström M, Viiri H. Suomen metsät 2009–2013 ja niiden kehitys 1921–2013. (Forests of Finland 2008–2013 and their development in 1921–2013). Luonnonvara- ja biotalouden tutkimus 2017;**59**. http://urn.fi/URN:ISBN:978-952-326-467-0.

Korhonen KT, Ahola A, Heikkinen J. *et al.* Forests of Finland 2014–2018 and their development 1921–2018. *Silva Fennica* 2021;**55**:49. doi.org/10.14214/sf.10662.

Laasenaho, J. Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae* 1982, 108. http://urn.fi/URN:ISBN:951-40-0589-9.

Liaw A, Wiener M. Classification and regression by random Forest. *R News* 2002;**2**:18–22.

Mackey B, DellaSala DA, Kormos C. *et al.* Policy options for the world's primary forests in multilateral environmental agreements. *Conserv Lett* 2015;**8**:139–47. https://doi.org/10.1111/conl.12120.

Maltamo M, Kinnunen H, Kangas A. *et al.* Predicting stand age in managed forests using National Forest Inventory field data and airborne laser scanning. *For Ecosyst* 2020;**7**:44. https://doi.org/10.1186/s40663-020-00254-z.

McRoberts RE, Winter S, Chirici G. *et al.* Assessing forest naturalness. *For Sci* 2012;**58**:294–309. https://doi.org/10.5849/forsci.10-075.

Morales-Hidalgo D, Oswalt SN, Somanathan E. Status and trends in global primary forest, protected areas, and areas designated for conservation of biodiversity from the Global Forest Resources Assessment 2015. *For Ecol Manage* 2015;**352**:68–77. https://doi.org/10.1016/j.foreco.2015.06.011.

Potapov P, Hansen MC, Laestadius L. *et al.* The last frontiers of wilderness: tracking loss of intact forest landscapes from 2000 to 2013. *Sci Adv* 2017;**3**:e1600821. https://doi.org/10.1126/sciadv.1600821.

R Core Team 2021 *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Ørka H-OJ-P, Næsset MC. A framework for a forest ecological base map – an example from Norway. *Ecol Indic* 2022;**136**:108636. https://doi.org/10.1016/j.ecolind.2022.108636.

Sabatini FM, Burrascano S, Keeton WS. *et al.* Where are Europe's last primary forests? *Divers Distrib* 2018;**24**:1426–39. https://doi.org/10.1111/ddi.12778.

Saito, T. and Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE* 2015;**10**:3. https://doi.org/10.1371/journal.pone.0118432.

Schulenberg R. Dineq: decomposition of (income) inequality. R package version 0.1.0., 2018. https://CRAN.R-project.org/package=dineq.

Signorell A, Aho K, Alfons A. *et al.* DescTools: tools for descriptive statistics. R package version 0.99.46, 2022. https://CRAN.R-project.org/package=DescTools

Tomppo E, Heikkinen J, Henttonen HM. *et al.* *Designing and Conducting a Forest Inventory - Case: 9th National Forest Inventory of Finland*. Dordrecht: Springer, 2011.

Uotila A, Maltamo M, Uuttera J. *et al.* Stand structure in semi-natural and managed forests in eastern Finland and Russian Karelia. *Ecol Bull* 2001;**49**:149–58.

Valbuena R, Packalen P, Martín-Fernández S. *et al.* Diversity and equitability ordering profiles applied to the study of forest structure. *For Ecol Manage* 2012;**276**:185–95. https://doi.org/10.1016/j.foreco.2012.03.036.

Valbuena R, Packalen P, Mehtätalo L. *et al.* Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. *Can J For Res* 2013;**43**:1063–74. https://doi.org/10.1139/cjfr-2013-0147.

Valbuena R, Eerikäinen K, Packalen P. *et al.* Gini coefficient predictions from airborne lidar remote sensing display the effect of management intensity on forest structure. *Ecol Indic* 2016;**60**:574–85. https://doi.org/10.1016/j.ecolind.2015.08.001.

Wittebolle, L., Marzorati, M., Clement, L. *et al.* Initial community evenness favours functionality under selective stress. *Nature* 2009;**458**:623–6. https://doi.org/10.1038/nature07840.