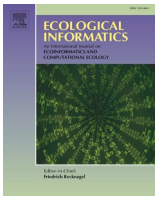
Ecological Informatics 95 (2026) 103768

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Potential of Sentinel-2 time series in capturing boreal forest structural diversity

Audrey Mercier [a], Joel Kostensalo [b], Mari Myllymäki [c], Miina Rautiainen [a,*]

[a] School of Engineering, Aalto University, Espoo, Finland
[b] Natural Resources Institute Finland (Luke), Yliopistokatu 6B, 80100 Joensuu, Finland
[c] Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00790 Helsinki, Finland

ARTICLE INFO

Keywords:
Biodiversity
Canopy height
First echo cover index
Airborne laser scanning
Random forest
Seasonal optical data

ABSTRACT

Forest structural diversity plays a key role in forest ecosystem functions. Capturing the temporal dynamics of forest structural variables could improve the characterization of forest structural diversity. However, the potential of seasonal time series from optical satellite data to predict the structural diversity in boreal forests remains underexplored. This study addresses two research questions: 1) What is the connection between boreal forest structural diversity and its spectral dynamics across seasons? 2) What is the potential of Sentinel-2 time series data for predicting boreal forest structural diversity compared to using a single image? Forest structural variables were derived from airborne LiDAR data and individual tree maps from two study areas in southern Finland. We analyzed the seasonal dynamics of Sentinel-2 reflectance and vegetation indices in relation to the forest structural diversity and predicted structural variables using random forest models with varying numbers of Sentinel-2 image dates. Results show that the diversity of the forest structure affects the seasonal dynamics of Sentinel-2 reflectance data from boreal forests during the spring-summer period. Stands with sparser canopy cover, a lower mean canopy height, higher stem density or higher Shannon index show greater temporal variability in terms of their spectral properties. The prediction accuracy improved for the mean canopy height, first echo cover index, volume of broadleaved trees, Shannon index, and stem density using a time series of Sentinel-2 images rather than a single image. These findings improve the characterization of boreal forest structural diversity through optical satellite time series and encourage further research into their application for monitoring seasonal and inter-annual dynamics in forest ecosystems.

1. Introduction

The boreal forest is the largest terrestrial biome on Earth. It is an important economic resource with a strong tradition of forest management (Moen et al., 2014), playing key roles in ecosystems including carbon storage, nutrient cycling and biodiversity conservation (Girona et al., 2023). An important component of these ecological and management functions is the structure of boreal forests. In ecology, structural diversity refers to the spatial organization and characteristics of living elements within an ecosystem (LaRue et al., 2023). Therefore, forest structural diversity relates to the spatial pattern and the distribution of the characteristics of individual trees and plants within a forest. It can be described by geometric attributes including the tree height, tree diameter, and canopy cover, along with spatial configuration attributes such as the stem density, species diversity, or vertical stratification (Atkins et al., 2023a). The forest structure plays a key role in forest ecosystem functions, e.g., meeting habitat needs for animal species (McElhinny et al., 2005), influencing microclimates (Ehbrecht et al., 2017; Zellweger et al., 2020), and reflecting tree species diversity and forest management practices (Ehbrecht et al., 2017, Hakkenberg et al., 2016). High forest structural diversity is linked to high levels of biodiversity in boreal (Kuuluvainen, 2016), tropical and temperate forests (Coverdale and Davies, 2023). Moreover, the forest structure strongly influences the shortwave radiation regime (Hovi and Rautiainen, 2020), which affects the energy distribution within the ecosystem. Therefore, it is commonly used as a proxy for biodiversity and ecosystem functioning (Gao et al., 2014; Muise et al., 2022). Additionally, it is closely related to the commercial value of forests (Puumalainen et al., 2003). Forest structural variables vary across time, driven by natural forest development, human interventions, post-disturbance

* Corresponding author.
E-mail address: miina.a.rautiainen@aalto.fi (M. Rautiainen).

https://doi.org/10.1016/j.ecoinf.2026.103768
Received 3 October 2025; Received in revised form 8 April 2026; Accepted 9 April 2026
Available online 11 April 2026

regeneration or environmental condition changes on a long-term time scale (Atkins et al., 2023b; Franklin et al., 2002), and by plant phenology on a seasonal scale (Sala et al., 1994). These temporal dynamics may reveal patterns that help characterize structural diversity.

Traditionally, forest structure estimation has been based on field surveys, which though highly accurate are time consuming, and thus are limited in both spatial coverage and temporal frequency. Remote sensing methods offer a suitable alternative for acquiring structural information over large forested areas (Atkins et al., 2023a; Kacic et al., 2023). For this purpose, airborne and terrestrial LiDAR scans are the most accurate methods, providing detailed forest 3D geometry through point clouds (Atkins et al., 2023a). However, acquiring dense LiDAR time series remains difficult due to technology and cost constraints. Openly available data from optical satellite sensors, such as Landsat OLI or Sentinel-2 MSI (S2), could help address these limitations. However, the use of seasonal time series from optical satellite images in understanding forest structure has been underexplored. Currently, only a few studies have addressed how the density and length of the time series of satellite images influences forest structure prediction. Moreover, these studies have focused exclusively on temperate forests and on a single structural variable (Choi et al., 2026; Hemmerling et al., 2021; Zhu and Liu, 2015). In these studies, seasonal time series of S2 and Landsat images have shown higher potential than single images in estimating forest structural variables such as tree species composition (Hemmerling et al., 2021), above-ground biomass (Zhu and Liu, 2015) or canopy height (Choi et al., 2026). To date, we are not aware of similar studies conducted concerning boreal forests, nor studies that would have explored a broader range of variables that would reflect different aspects of structural diversity (e.g. indicators of horizontal, vertical, and compositional dimensions of forest structure). Finally, despite its high temporal resolution (5-day revisit time in Finland) and relatively fine spatial resolution (10 and 20 m spectral bands), the potential of S2 data to explain the relationships between the seasonal reflectance dynamics and structural diversity of forests has not yet been investigated. More specifically, the use of seasonal trajectories of reflectance factors and vegetation indices, rather than single-date imagery has been underexplored. Therefore, in this study, we address two research questions:

(Q1) What is the connection between boreal forest structural diversity and the spectral dynamics across seasons?

(Q2) What is the potential of S2 time series data for predicting boreal forest structural diversity compared to using a single image?

We hypothesize that forest structural diversity influences the temporal dynamics of the spectral properties of boreal forests, and that using a time series of S2 data provides more accurate predictions of forest structural diversity than using a single image. To answer the research questions, we focus on two boreal forest areas located in southern Finland over the spring-summer period, using airborne LiDAR and LiDAR-based tree maps as ground reference data. First, we analyze the temporal dynamics of S2 reflectance factors and spectral vegetation indices in relation to various forest structural variables. Then, we predict the forest structural variables using random forest regression models, varying the density and length of the S2 time series to evaluate how the number of observations and seasonal coverage affect predictive performance of forest structural diversity. These new results provide practical guidance to remote sensing-based forest monitoring workflows.

2. Study area and data

2.1. Study areas

The data were collected from two study areas in southern Finland, which are mainly covered by managed boreal forests. The study areas include Hyytiälä in the west (61°51′N, 24°18′E) and Mikkeli in the east (61°41′N–62°11′N, 26°25′E–27°27′E), covering 18 and 3072 km^2, respectively (Fig. 1). Both areas are dominated by conifers, with Norway spruces (*Picea abies* (L.) Karst.) and Scots pines (*Pinus sylvestris* L.) being the most common, while the broadleaved trees present are mainly birches (*Betula pubescens* Ehrh. and *Betula pendula* Roth). The forests in

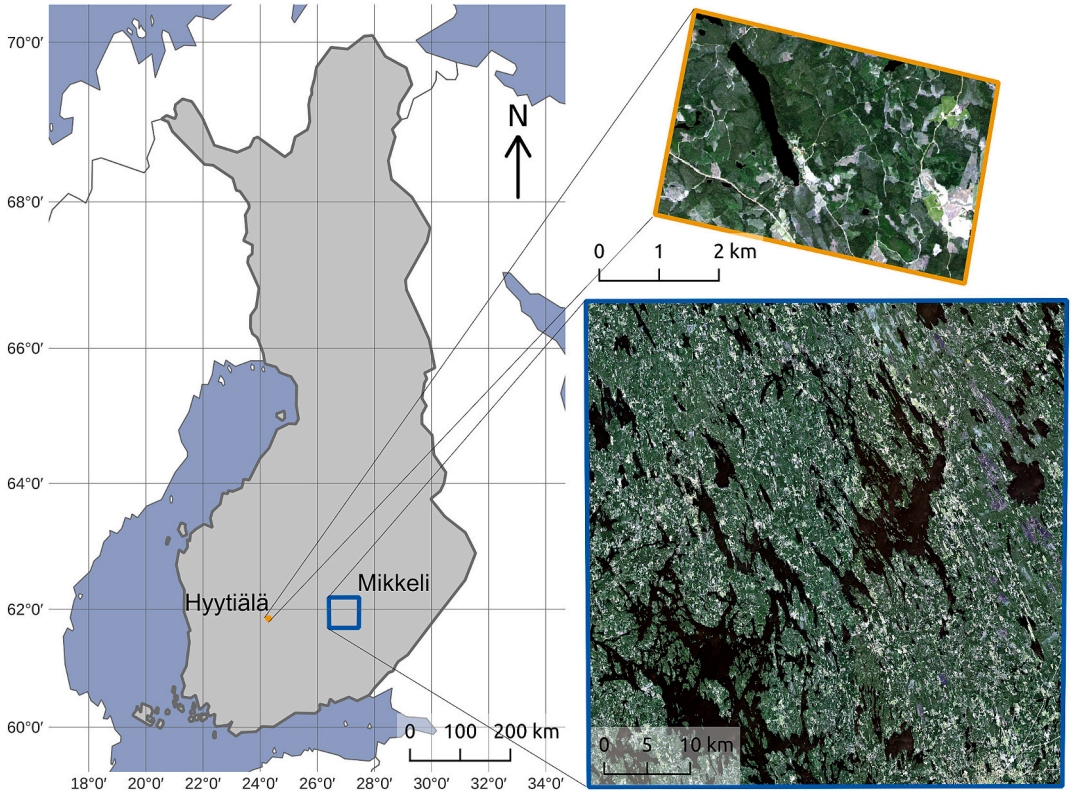


Fig. 1. Location of the two study areas. The Sentinel-2 satellite image dates are May 28, 2019, for Hyytiälä and May 22, 2020, for Mikkeli.

both study areas include all age classes from seedling stands to mature forests and are characterized by even-aged management practices. Previous field inventories indicate that forests in Mikkeli have an average DBH of 17.2 cm (ranging from 3 to 65.2 cm), mean tree height of 16.2 m (1.3–37 m), and stem densities averaging 836 stems ha^{-1} (186–2515 stems ha^{-1}) (Kostensalo et al., 2023). In Hyytiälä, field inventories indicate that trees have a mean DBH of 22 cm (ranging from 7 to 45 cm), an average height of 20 m (6–34 m), and a basal area of 22 m^2 ha^{-1} (4-46 m^2 ha^{-1}) (Hovi et al., 2023). The mean annual temperature is 4.8 °C in Hyytiälä and 4.9 °C in Mikkeli, with a mean annual precipitation of 682 mm and 629 mm, respectively (annual means from 2013 to 2023 recorded at the Juupajoki Hyytiälä and the Mikkeli Airport stations by the Finnish Meteorological Institute). The forest structure of the study areas are detailed in Hovi et al. (2023) for Hyytiälä and in Kostensalo et al. (2023) for Mikkeli. Although the Hyytiälä study area covers a small spatial extent (18 km^2), it includes a broad range of age classes, management regimes, and species compositions (Hovi et al., 2023) making it well suited for analyzing connections between spectral dynamics and forest structural diversity. The inclusion of the larger Mikkeli area (3072 km^2) ensures a regional context analysis. Finally, using two study areas rather than a single area allows to evaluate the robustness of the relationships between seasonal spectral dynamics and forest structural diversity.

2.2. Data

2.2.1. Forest structural variables

For both study areas, six forest structural variables were extracted from airborne LiDAR data and individual tree maps to capture both geometric and compositional aspects of forest structural diversity (Table 1). The limited number of variables aimed at facilitating the interpretability of results and redundancy among dependent variables by limiting strong correlations between them (Appendix A, Fig. A1). Specifically, the LiDAR data were used to compute the mean canopy height (chm_mean) and the first echo cover index (fci). Additionally, the same LiDAR dataset was used to produce maps of individual trees as specified below. These maps provided a basis for deriving four additional forest structural variables: stem density (stem_density, trees/ha), volume of broadleaved trees (broad_vol, m^3/ha), Gini coefficient index of height (gini_h), and the Shannon index (shannon). The selected structural variables are expected to influence the seasonal dynamics of S2 reflectance factors and vegetation indices. Mean canopy height (chm_mean) is partly related to stand age, which is associated with differences in spectral properties (Kuusinen et al., 2015). Canopy cover, quantified by the first echo cover index (fci) and stem density (stem_density) affect the light penetration and thus the forest floor vegetation, which vary during the growing season (Pisek et al., 2015; Rautiainen et al., 2011) and hence contribute to the optical signal (Pisek et al., 2015; Rautiainen et al., 2011). The volume of broadleaved trees (broad_vol) and the Shannon index (shannon) are linked to the specific phenology of the tree species that influence seasonal reflectance dynamics (Heiskanen et al., 2013; Rautiainen and Lukeš, 2015; Richardson et al., 2009). Unlike species proportions, the volume of broadleaved trees integrates both compositional and structural information by accounting for differences in tree size since broadleaved presence is weighted by tree volume. This variable is widely used as a proxy for boreal forest diversity (Kostensalo et al., 2026; Toraño Caicoya et al., 2023). In Finland, the main tree groups are spruce, pine, and broadleaved species (mostly birches); mixed forests including broadleaved trees are therefore considered more diverse. Finally, the height Gini coefficient (gini_h) characterizes vertical structural heterogeneity, which could affect the S2 signal by vertical vegetation masking ground vegetation and its seasonal variations, and more asynchronous phenological development compared to a homogeneous layer.

These six variables capture multiple and complementary aspects of forest structural diversity (Atkins et al., 2023a; LaRue et al., 2019). According to Atkins et al. (2023a), our selection reflects height (chm_mean), cover (fci), stand structure (stem_density, broad_vol), and structural heterogeneity (gini_h and shannon). According to Bakx et al. (2019), our variables would be categorized into height (chm_mean), cover (fci, stem_density), horizontal variability (broad_vol and shannon) and vertical variability (gini_h).

Airborne acquisitions were conducted in Hyytiälä in July 2019 and in Mikkeli in June 2020. In Hyytiälä, the acquisitions were operated by the Flying Laboratory of Imaging Systems FLIS (Hanuš et al., 2023) with a Riegl LMS Q780 waveform-recording sensor (Riegl Gmbh, Austria). The average pulse density was 28 points/m^2 and the average scanning altitude was 1000 m above ground level. In Mikkeli, the LiDAR scans were acquired as part of the national survey supervised by the National Land Survey of Finland and operated by BSF Swissphoto AG. The laser scanner used was a RIEGL VQ-1560i and the average pulse density for the study area was 7.2 points/m^2, while the average scanning altitude was 1525 m above the ground. The Hyytiälä data are openly available (Hovi et al., 2024) and described in detail in Rautiainen et al. (2024). Details of the Mikkeli LIDAR characteristics can be found on the National Land Survey of Finland website (National Land Survey of Finland, 2023) and in Kostensalo et al. (2023).

Individual tree maps were obtained by an approach developed by Kostensalo et al. (2026). The method is based on the previously described airborne LiDAR data, colour-infrared imagery (Mikkeli) or hyperspectral imagery (Hyytiälä), and ground reference measurements. First, the dominant trees were detected using a local-maximum-filter-based individual tree detection (ITD) algorithm, which was optimized using the ground reference data with F1 as the target metric, i.e.,

Table 1
Description of the forest structural variables used in this study.

Variable name	Abbreviation	Description	Unit	Data source
Mean canopy height	chm_mean	Average of the canopy height model	m	canopy height model derived from the LiDAR point clouds
First echo cover index	fci	Number of first echoes from the canopy divided by the total number of first echoes. A standard area-based approach was used, with the echoes above 2 m classified as canopy.	no unit	LiDAR point clouds
Stem density	stem_density	Number of stem per hectare	number of stems/ha	individual tree maps
Volume of broadleaved trees	broad_vol	Volume of broadleaved trees	m^3/ha	individual tree maps
Gini coefficient index of height	Gini_h	Area between the Lorenz-curve and the cumulative distribution of a uniform distribution. Measure of 'height inequality' varies between 0 (all trees have the same height) and 1 (one tall tree and others tiny).	no unit	individual tree maps
Shannon index	shannon	$shannon = -\Sigma\, p_i * \ln(p_i)$ where p_i is the proportional abundance of tree species i relative to the total abundance of all species. Increase with diversity of tree species in a community.	no unit	individual tree maps

penalizing the ITD for undetected trees and false positives equally. Next, the diameter, height, and species type (pine, spruce, broadleaved trees) of each ITD were predicted using models fitted using the ground reference data. Then, the numbers of false detections and missed trees were predicted for each 16 m × 16 m pixel and the corresponding number of ITD trees was eliminated starting from the smallest followed by simulating small trees using a resampling approach. Finally, a 2 m × 2 m forest mask based on land-use maps by National Land Survey Finland was applied. While evaluating the method in Mikkeli, Kostensalo et al. (2026) found the r-squared (r^2) of 0.32, 0.83, 0.23, and 0.20 for stem density, volume of broadleaved trees, Gini coefficient index of the height and Shannon index, respectively. Finally, six forest structural variables were derived on a 16 m × 16 m raster grid georeferenced to ETRS89/TM35FIN (EPSG: 3067).

2.2.2. Temporal variables derived from Sentinel-2 data

S2 time series were downloaded from the Copernicus Data Space Ecosystem (Copernicus Data Space Ecosystem, 2024). Level-2A products were selected as they provide atmospherically corrected surface reflectance, which we will refer to as reflectance factors in this paper. Time series acquired during the same year as the LiDAR campaigns (2019 for Hyytiälä and 2020 for Mikkeli) were downloaded, removing images with full cloud cover or snow. This process resulted in images from six dates in Hyytiälä (2019: Apr-28, May-18, Jun-30, Jul-25, Jul-27, Aug-29) and five dates in Mikkeli (2020: May-22, Jun-26, Aug-20, Aug-25, Sep-14). Details on the bands used in this study are provided in Section 3 and Table 2. First, the 20 m bands were resampled to fit the 10 m bands, then cloud masks provided with the S2 Level-2A products were applied on the full time series, so the analysis was conducted on the pixels free from clouds throughout the time series. The two tiles covering the Mikkeli study area were mosaicked, with the western tile partly overlapping the eastern tile (i.e., 20 km overlap). Finally, 14 vegetation indices (Appendix A, Fig. A3) were computed based on their potential to study vegetation. As the indices were highly correlated, we reduced them to three indices based on the Pearson correlation coefficient (Appendix A, Fig. A3). We retained the normalized difference vegetation index (NDVI) (Rouse et al., 1973), the weighted difference vegetation index (WDVI) (Clevers, 1988) and the soil brightness index (BI) (Escadafal et al., 1989). NDVI, which ranges from −1 to +1, is sensitive to the amount of photosynthetically active vegetation (eq. 1). WDVI was originally developed to estimate the leaf area index of vegetation (eq. 2). It is a near-infrared reflectance index corrected for soil influence with an unlimited range, which increases with the leaf area index, but it is highly sensitive to atmospheric conditions (Qi et al., 1994). The BI increases with soil brightness and is influenced by soil properties. It is used in this study as it is inversely related to vegetation abundance (eq. 3). The spectral vegetation indices were computed using the following equations:

$$\mathrm{NDVI} = \frac{(NIR - RED)}{(NIR + RED)} = \frac{(B8 - B4)}{(B8 + B4)} \quad (1)$$

$$\mathrm{WDVI} = NIR - 0.5 \times RED = B8 - 0.5 \times B4 \quad (2)$$

$$\mathrm{BI} = \sqrt{\frac{RED^2 + GREEN^2}{2}} = \sqrt{\frac{B4^2 + B3^2}{2}} \quad (3)$$

where GREEN, RED, and NIR correspond to the reflectance factors in the green, red, and near-infrared S2 bands, respectively. B3, B4, and B8 indicate the specific S2 bands used in this study and are detailed in Table 2.

The reflectance factors and vegetation indices were resampled using the same area, spatial resolution (16 m × 16 m) and coordinate system (EPSG: 3067) as the maps of forest structural variables.

Table 2
Sentinel-2 bands used in our study. Bands used for further analysis (Q2) are denoted with an asterisk (*).

Sentinel-2 band number	Sentinel-2 band name	Central wavelength (nm)	Original spatial resolution (m)
B2	Blue	493	10
B3*	Green	560	10
B4	Red	665	10
B7	Vegetation red-edge 3 (RE3)	783	20
B8A	Narrow near-infrared (NIR)	833	20
B8*	Near-infrared (NIR)	865	10
B11	Shortwave infrared 1 (SWIR1)	1614	20
B12*	Shortwave infrared 2 (SWIR2)	2202	20

2.2.3. Forest mask derived from MS-NFI products

A forest mask was created using the Multi-source National Forest Inventory (MS-NFI) Raster Maps of 2021 (National Land Survey of Finland, 2023). The raster maps of the stand mean height and canopy cover were downloaded from the LUKE platform (Natural Ressources Institute Finland, 2024). Their coordinate system is ETRS89 / TM35FIN (EPSG: 3067) and their spatial resolution is 16 m × 16 m. The forested areas were defined as height > 5 m and canopy cover >15%. The forest mask was applied to the maps of the forest structural variables and S2 reflectance factors and spectral vegetation indices.

3. Methods

3.1. Connection between boreal forest structural diversity and their spectral dynamics across seasons (Q1)

The first part of the methodological framework was designed to address the first research question (Q1): What is the connection between boreal forest structural diversity and the spectral dynamics across seasons? Characterization of forest structural diversity and analysis of the temporal dynamics of S2 reflectance factors and spectral vegetation indices in relation to the six forest structural variables were conducted.

First, forest structural diversity within each study area was characterized using a principal component analysis (PCA) based on the six forest structural variables. The PCA was computed using the *factoextra* and *FactoMineR* R packages (Husson et al., 2016; Kassambara and Mundt, 2020) to compare the forest structure of the two study areas and to visualize the variability and correlations among the forest structural variables. The PCA results were not used in later steps of the methodology.

Second, the relationships between S2 reflectance factors and the stage of the growing season were analyzed by plotting the temporal dynamics of the spectral bands in pixels dominated by broadleaved and coniferous trees in parallel to the growing degree days (GDD). Eight spectral bands were selected from the ten S2 bands at 10 and 20 m, excluding perfectly correlated bands (i.e. the Pearson coefficient = 1) (Appendix A, Fig. A2). Pixels dominated by broadleaves were defined as those where the volume of broadleaved trees constituted more than 75% of the total volume, while coniferous dominated pixels had less than 25%. For each tree type dominated pixel and S2 date, the average of the reflectance factors was calculated.

To relate seasonal dynamics of reflectance factors and vegetation indices to phase of the growing season, GDD were calculated from daily weather observations obtained from the Finnish Meteorological Institute's open data portal (Finnish Meteorological Institute, 2024). The Juupajoki Hyytiälä weather station (61°85′N, 24°29′E) was used for the Hyytiälä study area and the Mikkeli airport station (61°69′N, 27°20′E) for the Mikkeli study area. Daily minimum and maximum temperatures

(tmax and tmin) were downloaded from January to end of September (2019 for Hyytiälä, 2020 for Mikkeli). They correspond respectively to the lowest and highest temperature during two 12-h periods (from 8 p.m. the previous evening to 8 p.m. the evening of the day of interest, and from 9 p.m. to 9 p.m. during summertime). The GDD were calculated using the GDD function of the *climatrends* R package (de Sousa et al., 2023) using the default formula: GDD = ((tmax + tmin)/2) − tbase. The base temperature (tbase) was defined as 5 °C, which is suitable and commonly used for trees in cold environments (Prentice et al., 1992).

Finally, the influence of forest structural diversity on the temporal dynamics of S2 reflectance factors and vegetation indices was assessed by analyzing mean seasonal profiles stratified by forest structural variable classes. For each forest structural variable, three classes were defined using, equal intervals from minimum to maximum values. For example, in the Hyytiälä study area, the class boundaries of the mean canopy height were 0.05, 6.94, 13.81, and 20.71 m (interval = 6.89 m) and the ones for the fci were 0, 0.33, 0.66, and 0.99 (interval = 0.33). In Mikkeli, the class boundaries for the mean canopy height were 2.01, 12.71, 23.40, and 34.10 m (interval = 10.70 m) and the ones for the fci were 0.05, 0.37, 0.68 and 1.00 (interval = 0.32). S2 pixels corresponding to each class were extracted, and for each S2 date, the mean reflectance factor or vegetation index and its date-wise 95% confidence interval were computed.

3.2. The potential of S2 time series data for predicting boreal forest structural diversity compared to using a single image (Q2)

The second part of the methodological framework was designed to address the second research question (Q2): What is the potential of S2 time series data for predicting boreal forest structural diversity compared to using a single image? The potential of S2 time series data to predict the diversity of forest structure was analyzed by modeling the six forest structural variables using random forest regressions and varying numbers of dates of S2 images used as predictors.

For each study area, the independent training and test datasets were defined by dividing them into two spatial parts. First, 70% of the forested pixels were allocated to the training dataset by sequentially selecting pixels from the first row of the forest mask until the required proportion was reached. This resulted in training areas of 12 km^2 in Hyytiäla, and 2055 km^2 in Mikkeli. The remaining 30% of forested pixels corresponding to 6 km^2 in Hyytiäla, and 973 km^2 in Mikkeli were allocated to the testing datasets and were located in the southern part of the study areas. To reduce the influence of the imbalanced distributions of forest structural variable values, the training dataset was stratified. For each forest structural variables, three classes with equal intervals, as defined in the previous section, were used. From the 70% preselected training datasets, pixels were randomly selected with the number of sampled pixels per class equal to the smallest class size (Table 3). The test dataset was not stratified and correspond to the preselected 30% of forested pixels. We stratified the training dataset by randomly selecting pixels in each class to ensure that all classes of the forest structural variables were balanced, allowing the model to learn from rare and common conditions. In contrast, the test dataset was not stratified to reflect the natural distribution of the forested pixels and provide a realistic evaluation of model performance. As a result of this stratification strategy, the number of training samples varies across variables (Table 3), with broad_vol in both study areas and Gini_h at Hyytiälä having limited sample sizes. Results for these variables should be considered exploratory. For the remaining variables, sample sizes are sufficient to support robust model training.

Table 3
Number (nb) of pixels per class in the training dataset (divided in three classes with equal interval) for each forest structural variable.

Variable	Hyytiäla (nb pixels/class)	Mikkeli (nb pixels/class)
chm_mean	667	31,108
fci	1609	141,893
stem_density	338	77,573
broad_vol	11	41
Gini_h	10	11,848
shannon	3358	215,288

Random forest regressions were trained using the *ranger* R package (Wright, 2015) using 500 trees. Input variables included three S2 spectral bands and the three spectral vegetation indices. The green (B3), near infrared (B8), and shortwave infrared (B12) bands were selected from the ten available S2 bands at 10 m and 20 m spatial resolution based on their low correlations, in order to limit redundancy and minimize the processing time (Appendix A, Fig. A2). The number of S2 dates used as independent variables in the random forest was systematically varied to evaluate the influence of time series length and density on the forest structure prediction. All possible time series length and date combinations were tested, corresponding to a number of S2 dates from 1 to 6 for Hyytiälä and 1 to 5 for Mikkeli. In Hyytiälä, 1, 2, 3, 4, 5, and 6 input dates correspond to 6, 15, 20, 15, 6, and 1 combinations. In Mikkeli, 1, 2, 3, 4, and 5 input dates correspond to 5, 10, 10, 5, and 1 combinations.

Model performance was evaluated on the independent test dataset using the coefficient of determination (r^2), root mean squared error (RMSE), and bias. The RMSE represents the average magnitude of the squared errors between predicted and test values. The r^2 measures the proportion of the variance in the dependent variable (forest structural variable) that can be explained by the independent variables (S2 reflectance factors and spectral vegetation indices). Bias indicates whether the prediction model consistently underestimates or overestimates the test values, represented by negative and positive values, respectively.

4. Results

4.1. Connection between boreal forest structural diversity and their spectral dynamics across seasons (Q1)

The PCA results for Mikkeli and Hyytiälä show that the forest structure varied smoothly without a clear separation into groups (Fig. 2). The spread of forested pixels was relatively uniform along the PCA axes, with no dominant direction. Thus, variability in the forest structure was influenced by multiple forest structural variables. The Hyytiälä site partly overlapped with the Mikkeli site (which is 171 times larger in area), showing a tendency toward lower canopy heights and higher stem densities. The mean canopy height was positively correlated with the Gini coefficient index for the height, while the Shannon index, volume of broadleaved trees, and the first echo cover index were also positively correlated with each other. The stem density was slightly negatively correlated with the mean canopy height.

In both study areas, the temporal dynamics of the reflectance factors in the S2 bands and GDD show similar patterns with broadleaved forests exhibiting higher variability than coniferous ones in terms of the S2 temporal dynamics (Fig. 3). Additionally, the GDD was closely related to the temporal variation in certain spectral bands. Specifically, reflectance factors in the red-edge and near-infrared bands increased with the GDD from May to July, while reflectance factors in the red and shortwave-infrared bands showed negative correlations with the GDD. Since the GDD is linked to the tree phenology, the dynamics in the reflectance factors in the different S2 spectral bands portray the phenological phase of the forest in Hyytiälä and Mikkeli.

The temporal dynamics of reflectance factors and spectral vegetation indices were influenced by the forest structural diversity (Figs. 4 and 5). Reflectance factors in the green, NIR, and SWIR bands, as well as the BI and WDVI indices, decreased with the mean canopy height and the first echo cover index, while the NDVI showed the opposite trend. Additionally, forests with a low canopy height and low cover exhibited the strongest temporal variations for all S2 reflectance factors and spectral

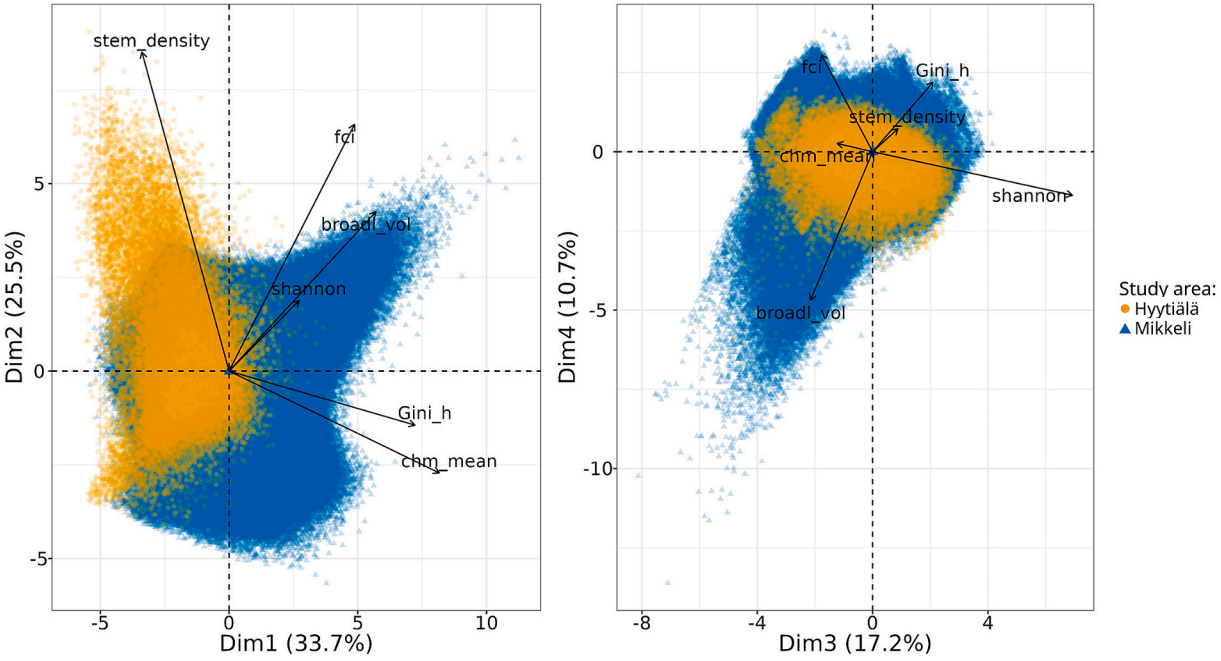


Fig. 2. Results of the principal component analysis with forest structural variables from Hyytiälä (orange) and Mikkeli (blue) as input. The first and second principal components are plotted on the left subfigure and the third and fourth on the right. The six forest structural variables correspond to broad_vol = volume of broadleaved trees (m^3/ha), chm_mean = mean canopy height (m), fci = first echo cover index (not unit), gini_h = Gini coefficient index of height (no unit), Shannon index (no unit) and stem_density = stem density (trees/ha). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

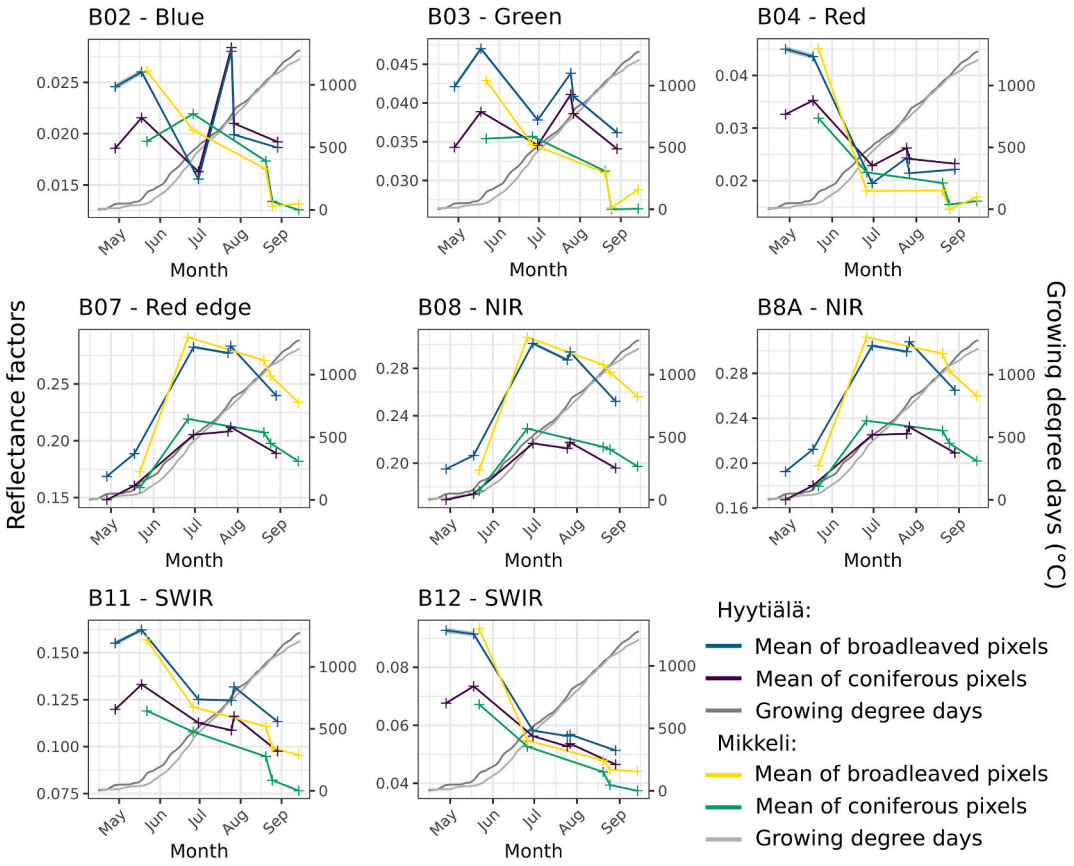


Fig. 3. Temporal dynamics of the reflectance factors for broadleaved and coniferous dominated pixels for the Sentinel-2 spectral bands and corresponding growing degree days at Hyytiälä (2019) and Mikkeli (2020).

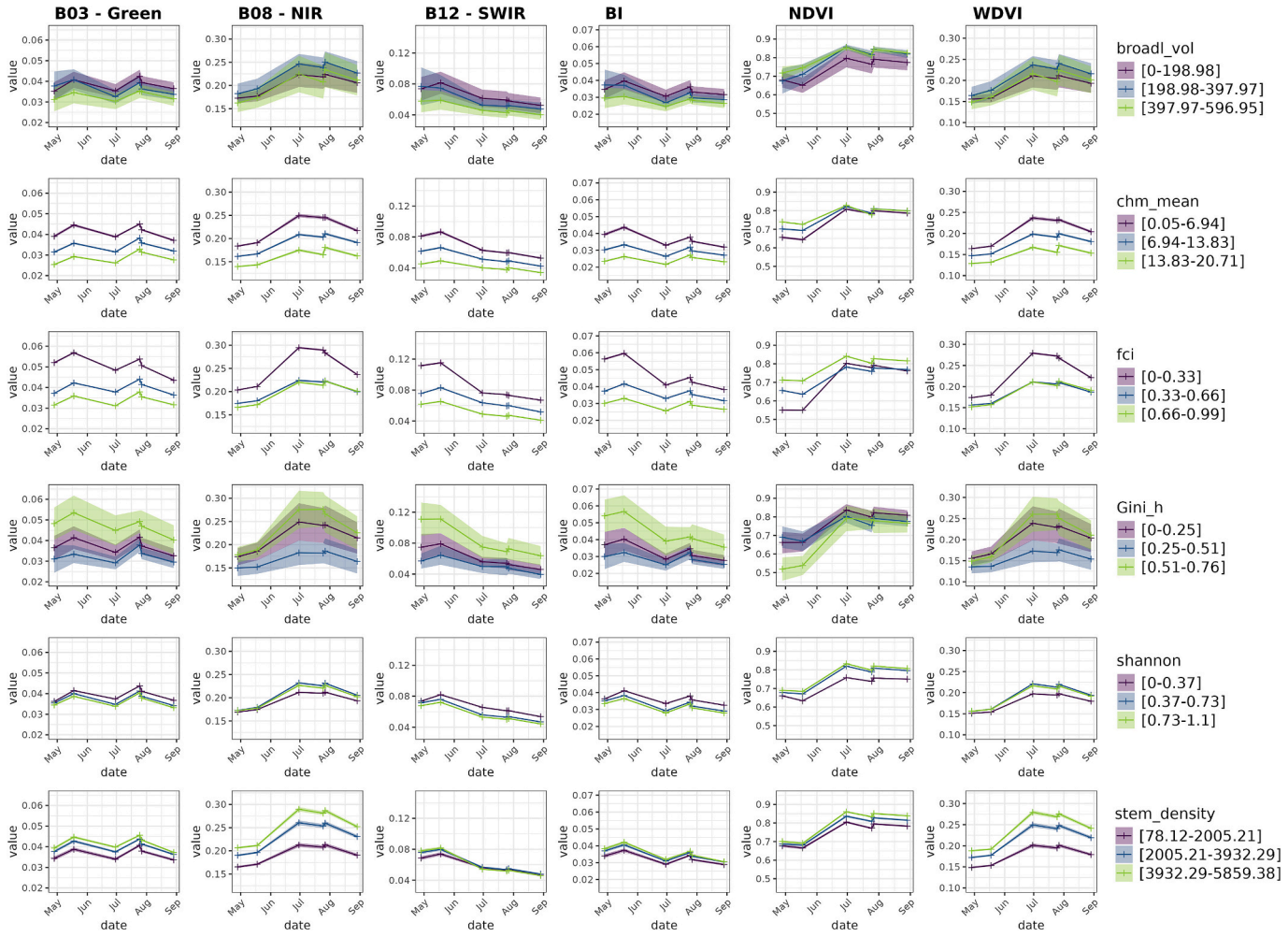


Fig. 4. Temporal dynamics of reflectance factors for three Sentinel-2 spectral bands (B03 = green, B08 = near infrared, B12 = shortwave-infrared) and three spectral vegetation indices (BI = brightness index, NDVI = normalized difference vegetation index, WDVI = weighted difference vegetation index) per class of forest structural variables at Hyytiälä. The six forest structural variables correspond to broad_vol = volume of broadleaved trees (m^3/ha), chm_mean = mean canopy height (m), fci = first echo cover index (not unit), gini_h = Gini coefficient index of height (no unit), Shannon index (no unit) and stem_density = stem density (trees/ha). The lines and bands correspond to the mean and confidence intervals of reflectance factors in the grouped pixels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

vegetation indices. In contrast, a medium to high stem density and a higher Shannon index (species diversity) resulted in stronger temporal variations in the S2 reflectance factors and spectral vegetation indices. The volume of broadleaved trees did not influence the seasonal reflectance dynamics at Hyytiälä, whereas at Mikkeli, the reflectance factors in the NIR band and WDVI were the only ones showing prominent differences between the three structural classes. However, the broadleaved tree volume varied greatly at Mikkeli (broad_vol = [0, 1352]) compared to Hyytiälä (broad_vol = [0, 597]). The highest values of the Gini coefficient index for height led to larger variations in the green and SWIR reflectance factors, and BI and NDVI at both sites, while the NIR band and WDVI behaved differently at Mikkeli and Hyytiälä.

Overall, the NIR band and WDVI showed the strongest differences between forest structural classes across all variables. The BI showed little variation for broad_vol, stem density and Shannon index but varied clearly with the fci and chm_mean. For fci and chm_mean, all reflectance factors and vegetation indices exhibited marked variations for all forest structural variables. In contrast, for stem density, Shannon index, and height Gini coefficient, differences were mainly observed in the NIR band and WDVI.

4.2. The potential of S2 time series data for predicting boreal forest structural diversity compared to using a single image (Q2)

Overall, increasing the number of S2 images (i.e., dates) enhanced the accuracy of the forest structure predictions (Fig. 6). The accuracy metrics indicate that the mean canopy height, first echo cover index, volume of broadleaved trees, Shannon index, and stem density were better predicted using a time series of S2 images rather than using a single image. Specifically, the mean canopy height and first echo cover index achieved the highest prediction accuracy with four dates, the Shannon index and stem density with three dates, and volume of broadleaved trees with four dates for Mikkeli, although no improvement was observed for Hyytiälä. Among these variables, the first echo cover index and mean canopy height were the best predicted, with maximum r^2 values of 0.66 and 0.75, and 0.54 and 0.67, respectively. However, the precision of the Gini coefficient for height prediction remained very low ($r^2 < 0.1$), regardless of the size of the S2 time series.

The most optimal dates differed depending on the predicted forest structural variable (Table 4). Overall, the months leading to the highest accuracy for Hyytiälä were June, July, and August, while for Mikkeli, they were May and June. The dates of July 25 or 27 for Hyytiälä, and August 20 or 25 for Mikkeli were among the last to be added as inputs in

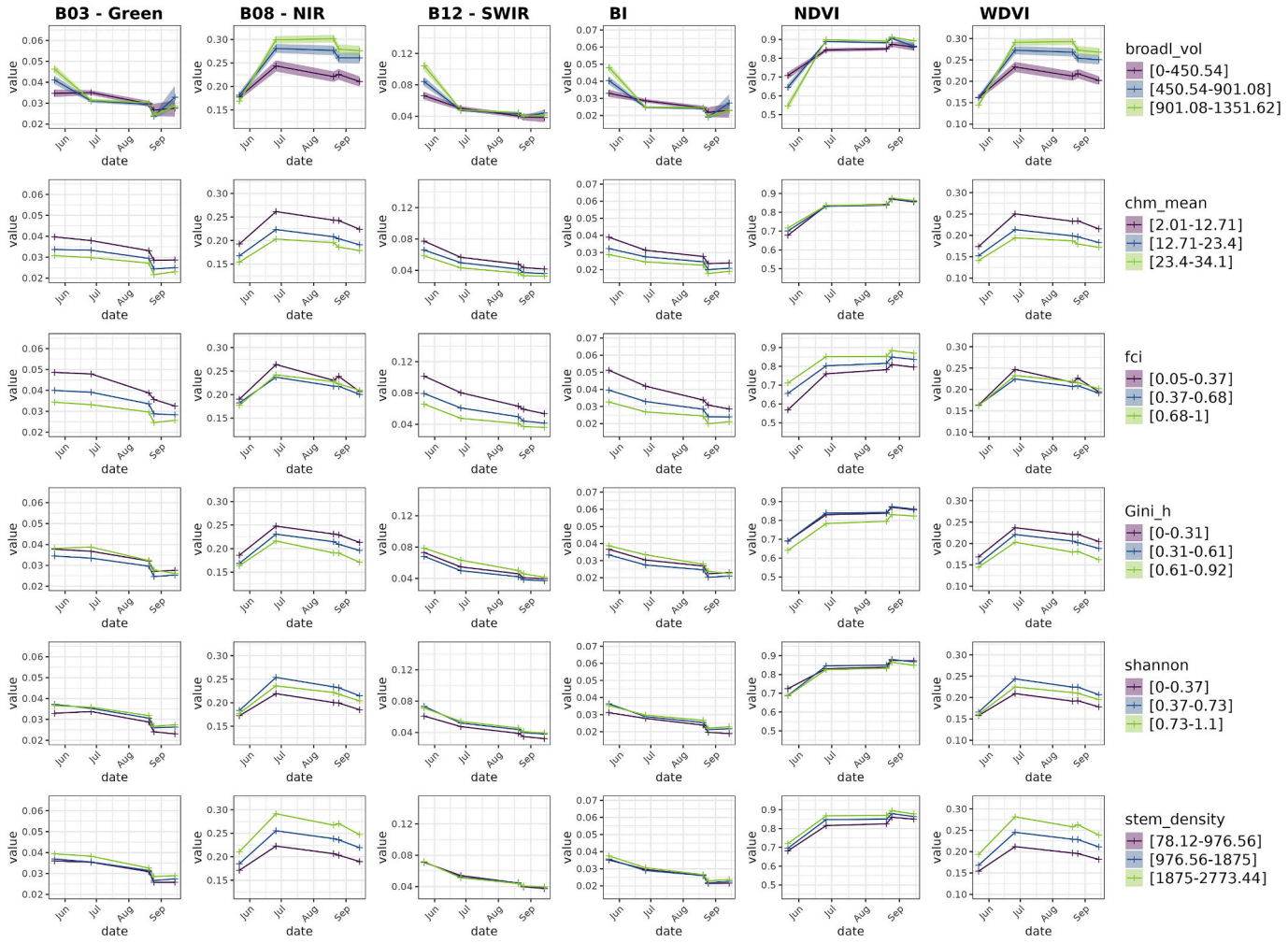


Fig. 5. Temporal dynamics of the three Sentinel-2 spectral bands (B03 = green, B08 = near-infrared, B12 = shortwave-infrared) and the three derived indices (BI = brightness index, NDVI = normalized difference vegetation index, WDVI = weighted difference vegetation index) per class of forest structural variables at Mikkeli. The six forest structural variables correspond to broad_vol = volume of broadleaved trees (m^3/ha), chm_mean = mean canopy height (m), fci = first echo cover index (not unit), gini_h = Gini coefficient index of height (no unit), Shannon index (no unit) and stem_density = stem density (trees/ha). The lines and bands correspond to the mean and confidence intervals of reflectance factors in the grouped pixels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the time series. Adding one of these dates, when a July or August date was already included, did not improve the prediction accuracy. As the first echo cover index was the most accurately predicted variable, we produced maps of it for both study sites (Figs. 7 and 8). The first echo cover index was overestimated for areas where the canopy cover was very sparse (low observed fci), and patterns of underestimation appeared on some road edges at the Hyytiälä site, apart from which the errors were randomly distributed over Hyytiälä (Fig. 7) and Mikkeli (Fig. 8). Maps of other forest structural variables with optimal date combinations are given in Appendix B (Figs. B1, B2, B3, B4 and B5 for Hyytiäla and Figs. B6, B7, B8, B9 and B10 for Mikkeli).

5. Discussion

In this study, we have shown that the forest structural diversity in boreal forests affects the seasonal reflectance dynamics of forests observed in S2 satellite data. Boreal forests with the lowest groups of canopy height and cover, and the highest groups of stem density and Shannon index (related to species diversity) showed more temporal variation in reflectance factors and vegetation indices. Broadleaved forests, indicated by the highest volumes of broadleaved trees, showed more variation in reflectance factors compared to coniferous forests (Figs. 4 and 5) as also found in previous studies (Heiskanen et al., 2013; Rautiainen and Lukeš, 2015; Richardson et al., 2009). The stronger temporal variations of NIR reflectance factors observed in broadleaved stands compared to coniferous stands has been attributed to a rapid increase in canopy leaf area index between May and July (Rautiainen et al., 2012; Rautiainen and Lukeš, 2015). The boreal forest understory also contributes to the total stand reflectance, and its seasonal dynamics are stronger in broadleaved than coniferous stands, with the understory NDVI increasing strongly in May–June and decreasing in the fall (Pisek et al., 2015). Stands with a low first echo cover index, corresponding to sparse canopy cover, exhibited stronger seasonal reflectance dynamics. For example, in early May, the NDVI values were lower for stands with a sparse canopy compared to those with a medium or dense canopy, but from mid-May to July, the NDVI increased until it matched the levels of denser canopy stands. This temporal behavior likely reflects the sensitivity of the remotely sensed spectral data to forest floor (Eriksson et al., 2006; Hovi et al., 2023; Rautiainen and Lukeš, 2015; Rautiainen and Stenberg, 2005). In southern boreal forests, the forest floor contributes more to the S2 data in areas with low canopy cover, though the contribution changes throughout the season as the forest ground vegetation grows actively from May to July (Rautiainen et al., 2011). Similarly, the mean canopy height exhibited a similar effect to the first echo

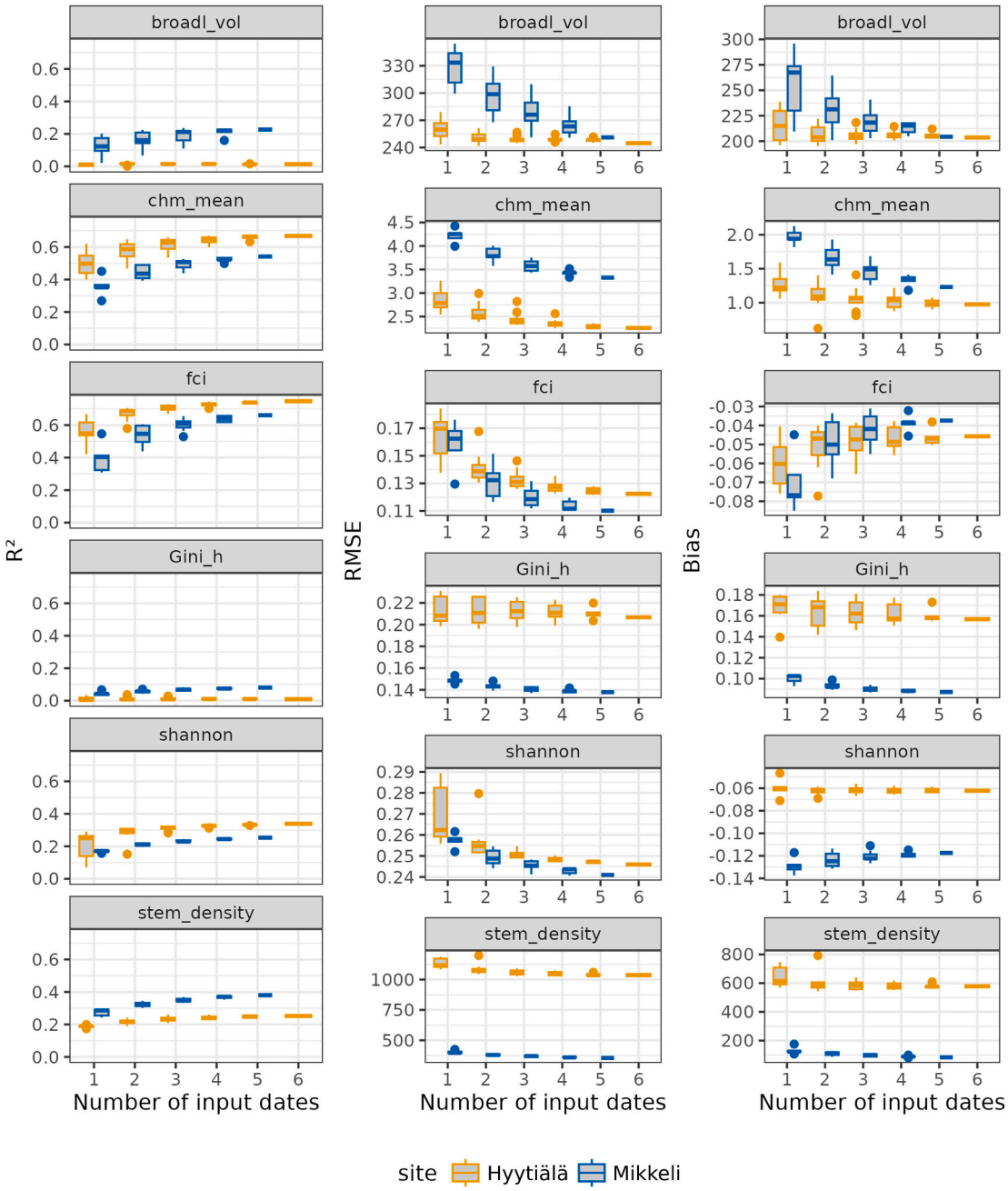


Fig. 6. Accuracy results (R^2, RMSE and bias) of the random forest regressions to predict forest structural variables according to the number of input Sentinel-2 dates. The six forest structural variables predicted were the volume of broadleaved trees (broad_vol, m^3/ha), mean canopy height (chm_mean, m), first echo cover index (fci, not unit), Gini coefficient index of height (gini_h, no unit), Shannon index (shannon, no unit) and stem density (stem_density, trees/ha).

cover index, with low values resulting in a higher temporal variation in the spectral data from S2. The canopy height was positively correlated with the first echo cover index (Appendix A, Fig. A1), indicating that low stands were associated with sparse canopies. Therefore, ground vegetation might again explain the observed seasonality of the S2 reflectance factors. Inversely, the reflectance factors and spectral vegetation indices increased with the stem density, with low variations in the SWIR reflectance factors and BI. The stem density was negatively correlated with the mean canopy height (Appendix A, Fig. A1), so pixels with a high stem density could correspond to young forests. In Finland, thinning is a common silvicultural practice which aims to improve the timber quality and tree growth by reducing tree competition (Mielikäinen and Hynynen, 2003; Pukkala et al., 2014). As a result, the stem density in managed forest stands decreases over time, through successive thinning (Pukkala et al., 2014). Since younger stands exhibit higher stem density, the high NIR reflectance factors observed with the highest stem density are consistent with the findings by Kuusinen et al. (2015), which show that NIR reflectance for pine and spruce declines rapidly in young stands and stabilizes around the time the stand is 50 years old. Finally, the Shannon index, increasing with tree species diversity, led to increased and larger temporal variations in the NIR reflectance factors, NDVI and WDVI, but contradictory results between Hyytiälä and Mikkeli for the green and SWIR reflectance factors, and BI. Since the Shannon index was computed using only three tree species groups (spruce, pine and broadleaved trees), and spruce and pine dominate both landscapes, high Shannon index values would suggest more broadleaved trees. As shown by comparing the spectral signatures of broadleaved and coniferous forests, the differences are particularly high in the NIR domain, with

Table 4
Sentinel-2 date combinations achieving the highest R^2 for predicting forest structural variables. For the Hyytiälä site, the Sentinel-2 dates used in this study were Apr-28, May-18, Jun-30, Jul-25, Jul-27, Aug-29 from the 2019 year. For the Mikkeli site, the dates used in this study were May-22, Jun-26, Aug-20, Aug-25, Sep-14 from the 2020 year.

Response variable	Best single date		Best 2-date combination		Best 3-date combination		Best 4-date combination		Best 5-date combination
	Hyytiälä	Mikkeli	Hyytiälä	Mikkeli	Hyytiälä	Mikkeli	Hyytiälä	Mikkeli	Hyytiälä
broad_vol	Jul-25	Jun-26	Jun-30 Jul-27	Jun-26 Aug-20	Apr-28 Jun-30 Jul-27	May-22 Jun-26 Aug-20	Apr-28 May-18 Jun-30 Jul-27	May-22 Jun-26 Aug-20 Sep-14	Apr-28 May-18 Jun-30 Jul-25 Jul-27
chm mean	Jul-25	Jun-26	Jul-25 Aug-29	Jun-26 Sep-14	Jun-30 Jul-25 Aug-29	May-22 Jun-26 Sep-14	May-18 Jun-30 Jul-25 Aug-29	May-22 Jun-26 Aug-25 Sep-14	May-18 Jun-30 Jul-25 Jul-27 Aug-29
fci	Jul-25	Jun-26	Jul-25 Aug-29	May-22 Jun-26	May-18 Jul-25 Jul-27	May-22 Jun-26 Aug-20	May-18 Jun-30 Jul-25 Aug-29	May-22 Jun-26 Aug-20 Sep-14	May-18 Jun-30 Jul-25 Jul-27 Aug-29
gini_h	Jul-25	May-22	Jun-30 Jul-25	May-22 Jun-26	Jun-30 Jul-25 Jul-27	May-22 Jun-26 Sep-14	Jun-30 Jul-25 Jul-27 Aug-29	May-22 Jun-26 Aug-25 Sep-14	May-18 Jun-30 Jul-25 Jul-27 Aug-29
shannon	Jun-30	Sep-14	Jun-30 Aug-29	May-22 Sep-14	Jun-30 Jul-25 Aug-29	May-22 Jun-26 Sep-14	May-18 Jun-30 Jul-25 Aug-29	May-22 Jun-26 Aug-20 Sep-14	May-18 Jun-30 Jul-25 Jul-27 Aug-29
stem_density	Jul-25	Jun-26	Apr-28 Jul-25	May-22 Jun-26	Apr-28 Jul-25 Aug-29	May-22 Jun-26 Aug-20	Apr-28 Jul-25 Jul-27 Aug-29	May-22 Jun-26 Aug-20 Sep-14	Apr-28 May-18 Jul-25 Jul-27 Aug-29

higher values for broadleaved dominated stands compared to conifers. The Gini coefficient of height showed different S2 temporal dynamics depending on the site, however the sample size between Hyytiälä and Mikkeli differ strongly. Nevertheless, Gini coefficient was poorly predicted from the S2 data on both study areas, suggesting a lack of sensitivity of spectral data in the S2's bands to the heterogeneity of tree height in boreal forests at fine spatial scale. While field inventories and airborne LiDAR data provide more accurate estimations on forest vertical structure, their low temporal frequency, due to technology and cost constraint, limits the analysis of seasonal structural dynamics.

Regarding the potential of time series for predicting forest structural diversity, we found more accurate predictions when the number of S2 images (i.e., dates) was increased. More specifically, the predictions for the mean canopy height, first echo cover index, volume of broadleaved trees, Shannon index, and stem density were more accurate using a time series of S2 images rather than when using a single image during the growing season. We observed a higher prediction accuracy for the first echo cover index and mean canopy height compared to other forest structural variables with a comparable training sample size (i.e. stem density and shannon). Consistent with our findings, Choi et al. (2026) showed that S2 time series improved canopy height prediction, with performance increasing up to four acquisitions in nine temperate forests of the United States (broadleaf, coniferous, and mixed) and performed best in forests with low species diversity. The higher performance of optical time series was attributed to species-specific phenological and reflectance dynamics that constrain reflectance-height relationships. Complementarily, our results suggest that S2 sensitivity to canopy height partly reflects canopy cover dynamics. It should be noted that the first echo cover index and mean canopy height were derived directly from the airborne LiDAR point clouds, while the others were based on proxy tree maps. The uncertainty of the variables derived from the tree maps might lead to over or underestimations of the true forest structure, while first echo cover index and mean canopy height are closer to the field truth. Capturing the full temporal range of the growing season in the satellite images achieved the highest prediction accuracy. Our study demonstrates that the accuracy was higher when images from May to end of August or September for the Mikkeli site, and from end of April or mid-May to end of August for the Hyytiälä site, were used. Remotely sensed spectral data can capture the seasonality of trees (Kobayashi et al., 2018) and forest floor variations (Pisek et al., 2015; Rautiainen et al., 2011). In boreal forests, the changes in phenology and the structure of the understory impact its spectral properties during the peak growing period (Pisek et al., 2015; Rautiainen et al., 2011), the most fertile forest type (herb-rich) showing the largest temporal dynamics. Moreover, increasing the number of S2 acquisition dates helps to better manage meteorological conditions by reducing cloud cover and accounting for variations in atmospheric conditions. The best dates for estimating the forest structure from the S2 data varied from one forest structural variable to another. This suggests that different variables impact the S2 data at different stages of the growing season. To predict forest structural diversity from a single S2 date during the spring-summer seasons, we recommend choosing a date between the end of May and August. Indeed, when using a single date to predict forest structural variables, June, July, and August achieved the highest R^2 for the Hyytiälä site, while May and June were highlighted for the Mikkeli site. Finally, increasing the density of the S2 time series in the middle of the summer does not necessarily lead to improved prediction accuracy of forest structural diversity, as July 25 and 27, and August 20 and 25 did not appear together on the most accurate date combinations. Denser time series could be more valuable in the spring, when the leaf area index and understory reflectance of boreal forests change rapidly, before stabilizing in mid-summer (Pisek et al., 2015; Rautiainen et al., 2012).

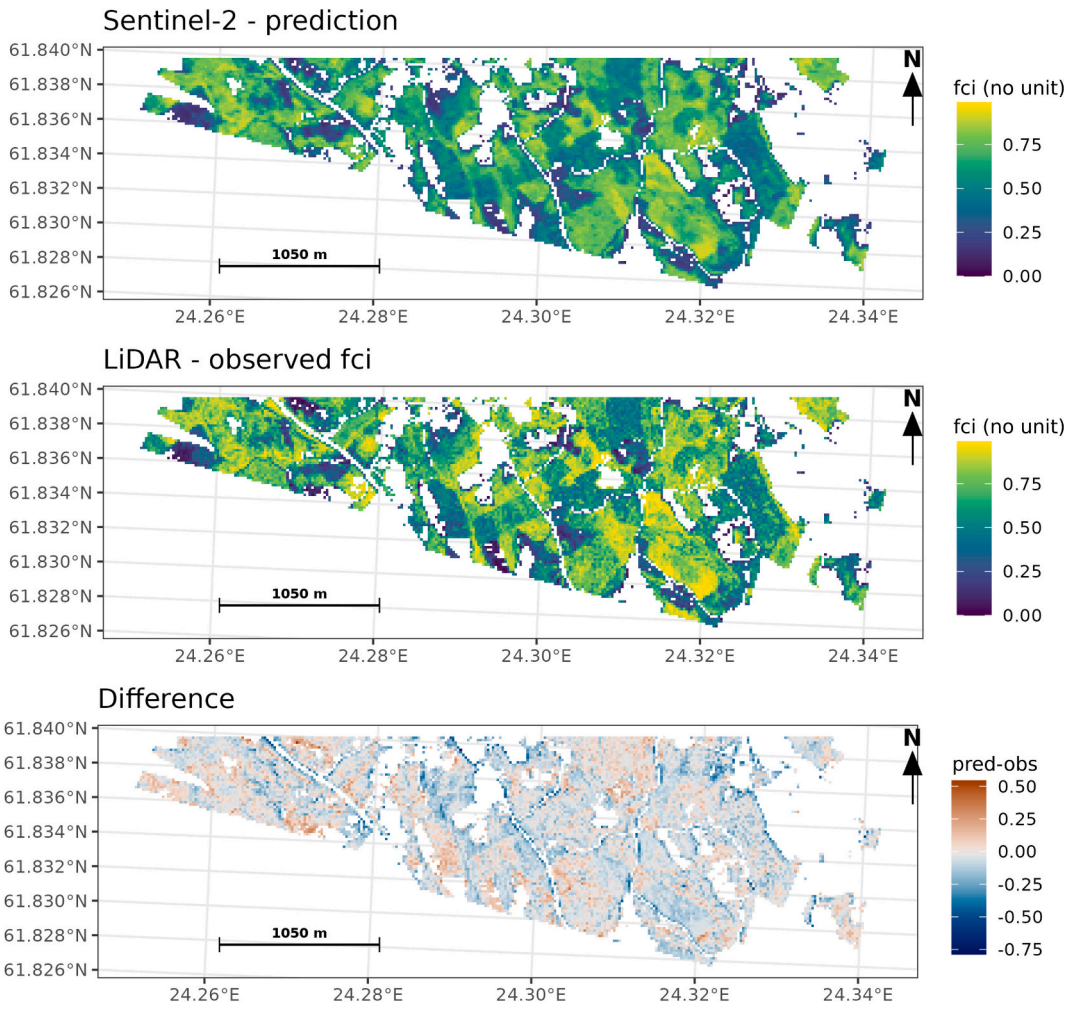


Fig. 7. The most accurate map of any combination of dates and forest structural variables for the Hyytiälä site. The forest echo cover index (fci) was predicted using five Sentinel-2 dates: (2019-05-18, 2019-06-30, 2019-07-25, 2019-07-27, 2019-08-29) The resulting R^2, RMSE and bias were 0.75, 0.12 and − 0.05, respectively.

Overall, our results demonstrate that seasonal dynamics of S2 data contribute to characterizing structural diversity of boreal forests, through variations linked to canopy openness changes, understory development, and tree species phenology, as discussed throughout this section.

In the future, one should explore the potential of the S2 time series to monitor forest structure dynamics during the growing season more closely, through frequent field measurements of the tree canopy structure. This study focuses on seasonal variations occurring during the growing season in boreal forests, but investigating the potential of S2 on long-term time scales would be interesting too, for example, to study inter-annual dynamics of the boreal forest structural diversity. Lastly, further research should focus on the combined effects of forest structural variables. For example, factors such as the stem density or canopy cover could influence the prediction of other metrics, such as the canopy height.

6. Conclusion

Our results demonstrated that in boreal forests, the diversity of the forest structure affects the seasonal dynamics of S2 data. Overall, the NIR band and WDVI index showed the strongest sensitivity to forest structural diversity, while canopy height (chm_mean) and cover (fci) affected the full range of reflectance factors and spectral vegetation indices. This visual interpretation provides initial insights that could be further assessed using quantitative modeling. Additionally, increasing the number of S2 dates can enhance the characterization of the forest structural diversity. Stands with sparser canopy cover (first echo cover index), lower mean canopy height, higher stem density or higher tree species diversity (Shannon index) showed greater temporal variability in S2 reflectance factors during the spring-summer period. The mean canopy height and canopy cover (first echo cover index) were the most accurately predicted, with optimum performance achieved using four S2 dates. The Gini coefficient of height showed poor predictability, emphasizing the limited sensitivity of S2 data to fine-scale heterogeneity in canopy height. Additionally, variables derived from proxy tree maps (Gini coefficient of height, volume of broadleaved trees, Shannon index and stem density) were less accurate than LiDAR-derived ones (first echo cover index, mean canopy height). The best periods for predicting forest structural diversity using S2 vary by site and variable. This highlights the context dependency of these optimal periods and the need for further investigation using multi-year data and additional forest types to understand how the S2 dates influence prediction performance. Overall, these findings enhance the characterization of boreal forest structural diversity using optical satellite time series and provide practical guidance to remote sensing-based forest monitoring workflows. Future research should confirm the robustness of our findings across broader spatial extents and various forest types. Future studies could go further exploring S2 time series for monitoring seasonal and inter-annual dynamics in forest ecosystems and investigating how the interplay of forest structural variables, such as the stem density, height, and canopy cover, affect their own prediction accuracy.

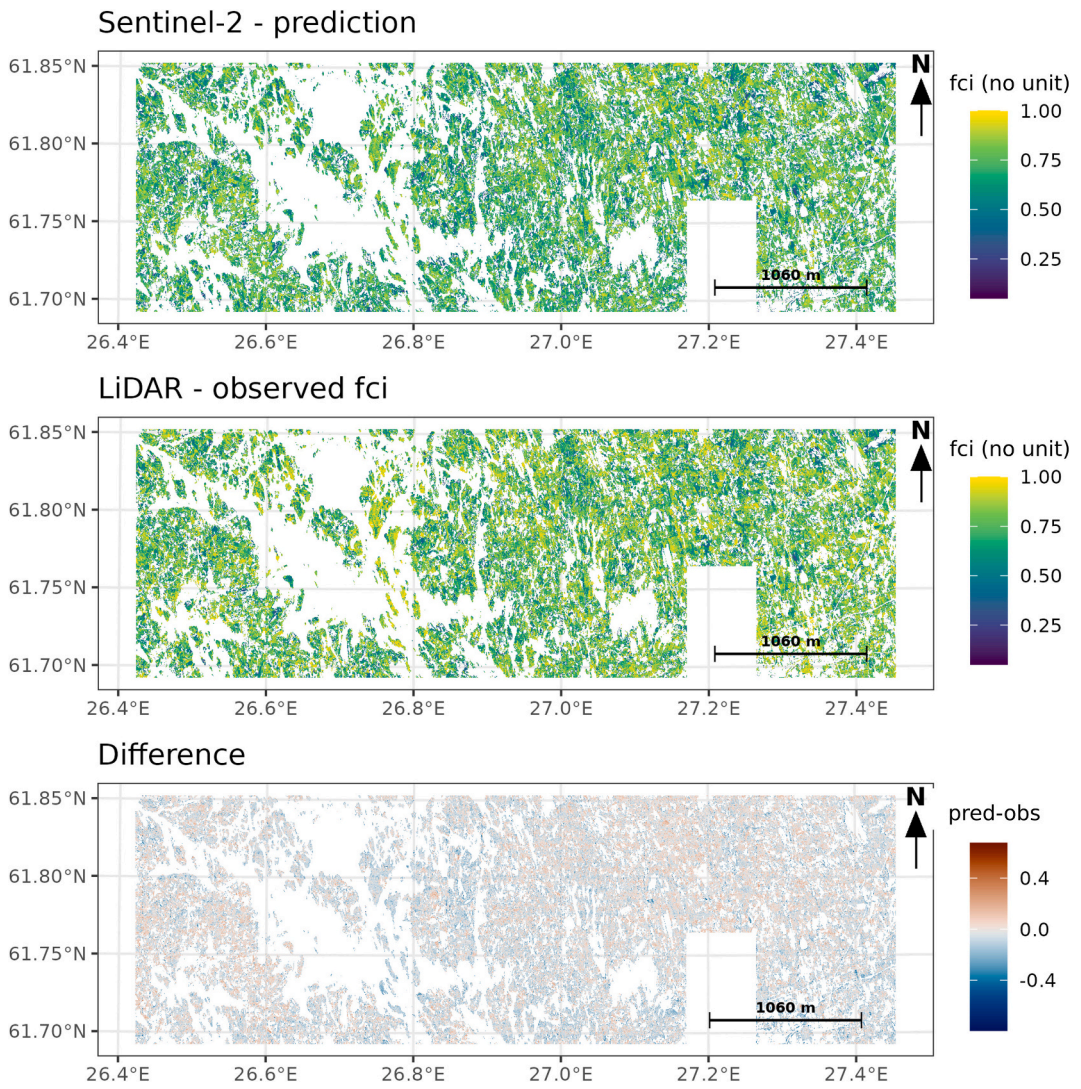


Fig. 8. The most accurate map of any combination of dates and forest structural variables for the Mikkeli site. The forest echo cover index (fci) was predicted using five Sentinel-2 dates (2020-05-22, 2020-06-26, 2020-08-20, 2020-08-25, 2020-09-14). The corresponding R^2, RMSE and bias were 0.66, 0.11 and – 0.04, respectively.

CRediT authorship contribution statement

Audrey Mercier: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Joel Kostensalo:** Writing – review & editing, Software, Investigation, Data curation. **Mari Myllymäki:** Writing – review & editing, Investigation, Funding acquisition. **Miina Rautiainen:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dr. Aarne Hovi for scientific collaboration on the LiDAR data processing and data visualizations. This study was mainly funded by the European Union – NextGenerationEU as part of the Research Council of Finland project ARTISDIG (decision numbers 348152 and 348154). JK and MM did their work under the Research Council of Finland's flagship ecosystem for Forest-Human-Machine Interplay—Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) (Grant number 357909). JK was also financially supported by the Research Council of Finland (Grant number 361209).

Appendix A. Appendices

A.1. Appendix A

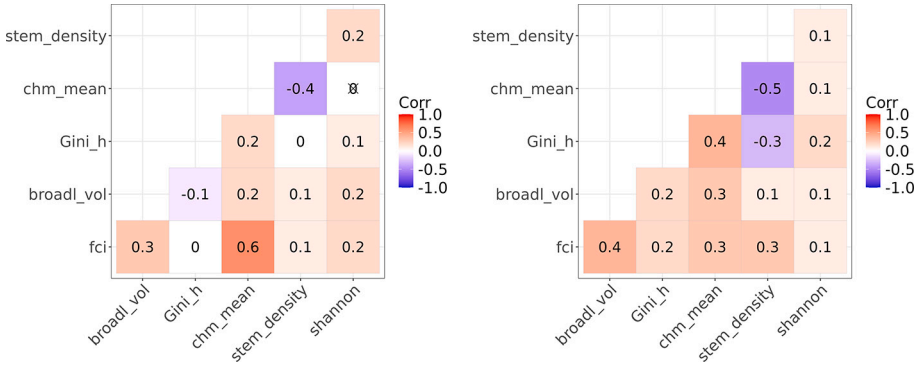


Fig. A1. Analysis of the strength and direction of relationships between the six forest structural variables through correlation matrices using the Pearson correlation coefficient.

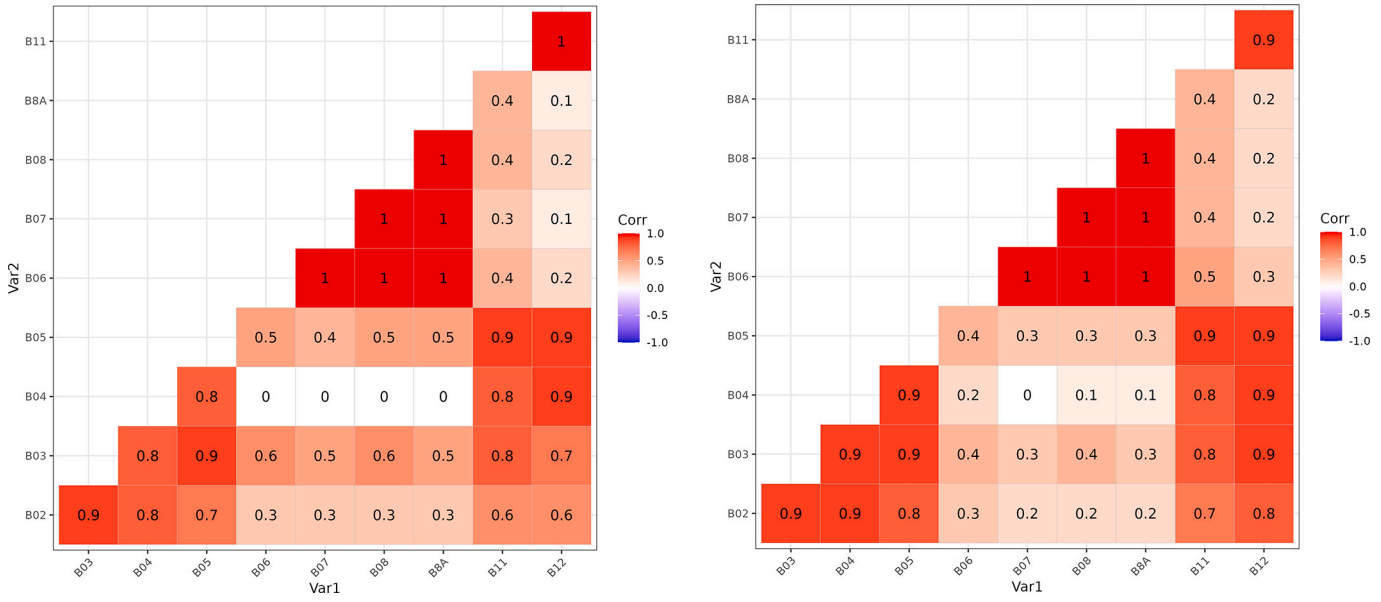


Fig. A2. Analysis of the strength and direction of relationships between the 10 Sentinel-2 spectral bands through correlation matrices using the Pearson correlation coefficient. Based on these figures, eight S2 bands were used in the exploratory part (B02, B03, B04, B07, B08, B8A, B11 and B12) removing the perfectly correlated bands (B05 and B06). In the prediction part, three bands were used to minimize the processing time. The B03, B08 and B12 were selected based on their low correlations between each other.

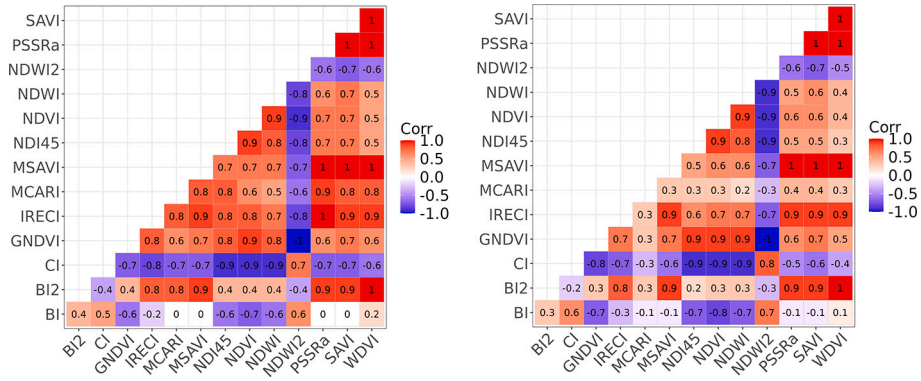


Fig. A3. Analysis of the strength and direction of relationships between the 14 Sentinel-2 vegetation indices through correlation matrices using the Pearson correlation coefficient. Based on these figures, three S2 indices (BI, NDVI, WDVI) were selected in our study.

A.2. Appendix B

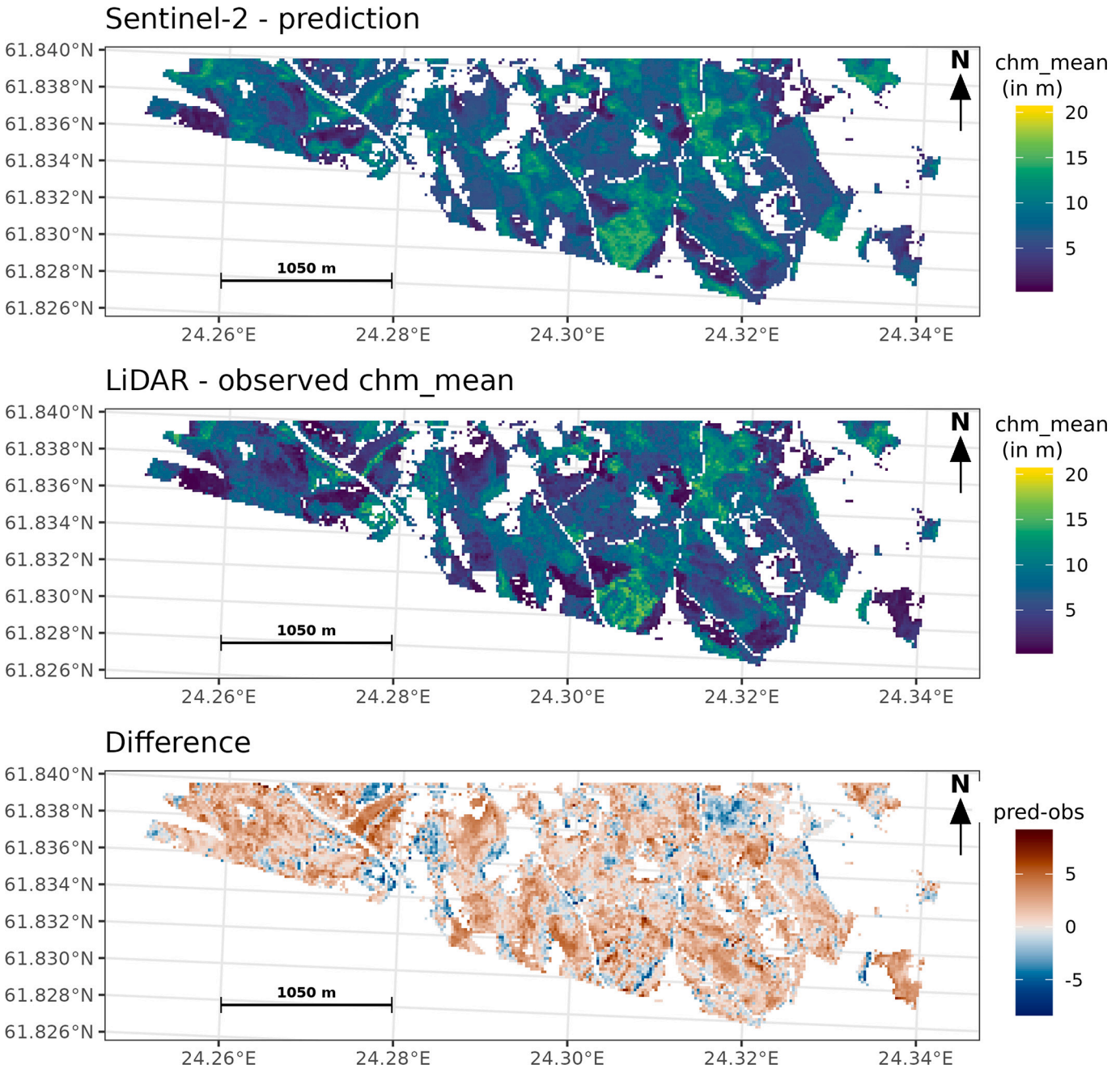


Fig. B1. Among 63 combinations of Sentinel-2 dates used as independent variables in random forest models in Hyytiälä, the most accurate mean canopy height (chm_mean, in m) map was produced using five input Sentinel-2 dates (2019-05-18, 2019-06-30, 2019-07-25, 2019-07-27, 2019-08-29). The corresponding R^2, RMSE and bias were 0.67, 2.28 and 1.04.

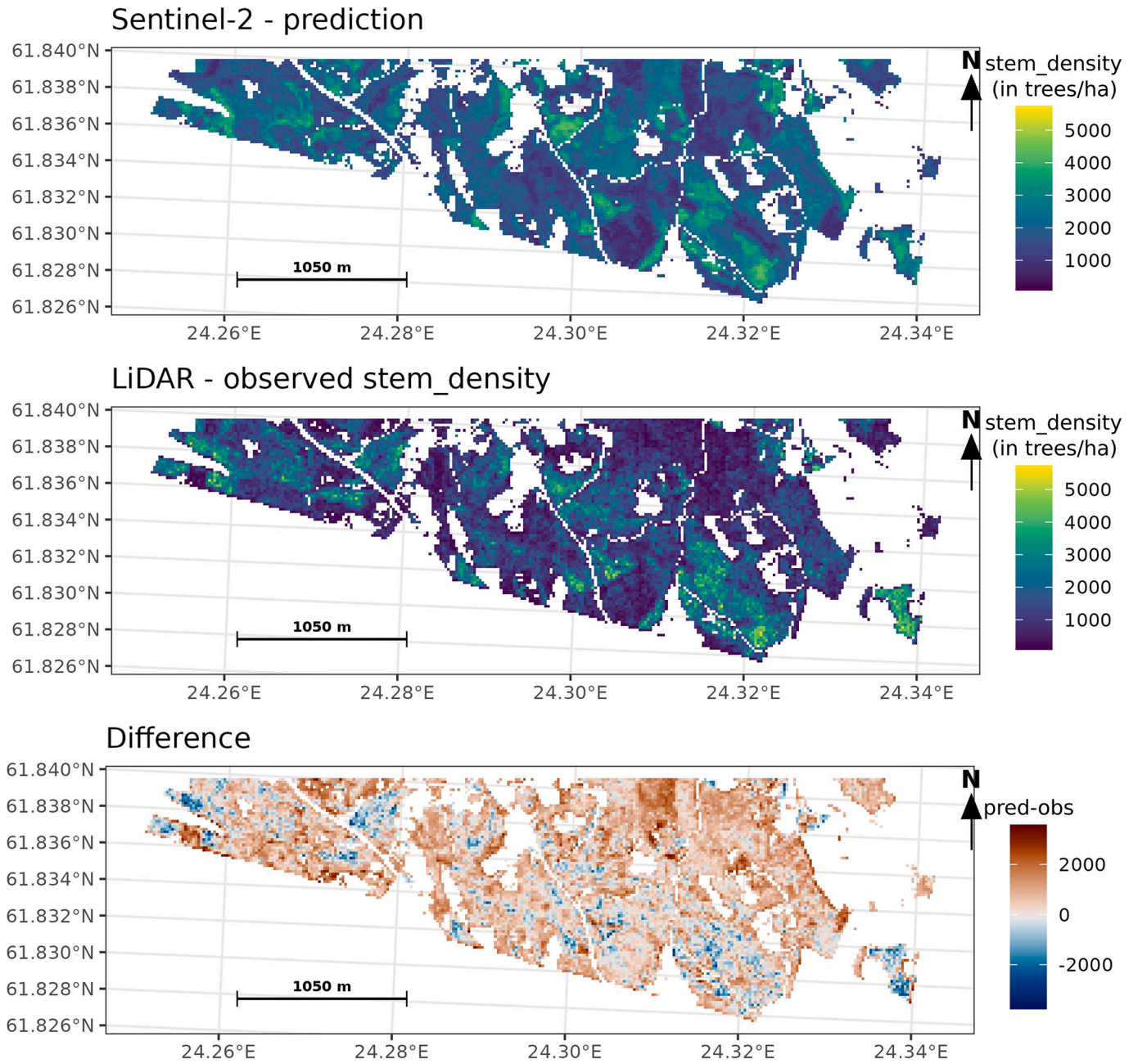


Fig. B2. Among 63 combinations of Sentinel-2 dates used as independent variables in random forest models in Hyytiälä, the most accurate stem density (stem_-density, in trees/ha) map was produced using three input Sentinel-2 dates (2019-04-28, 2019-07-25, 2019-08-29). The corresponding R^2, RMSE and bias were 0.26, 1030.69 and 572.24.

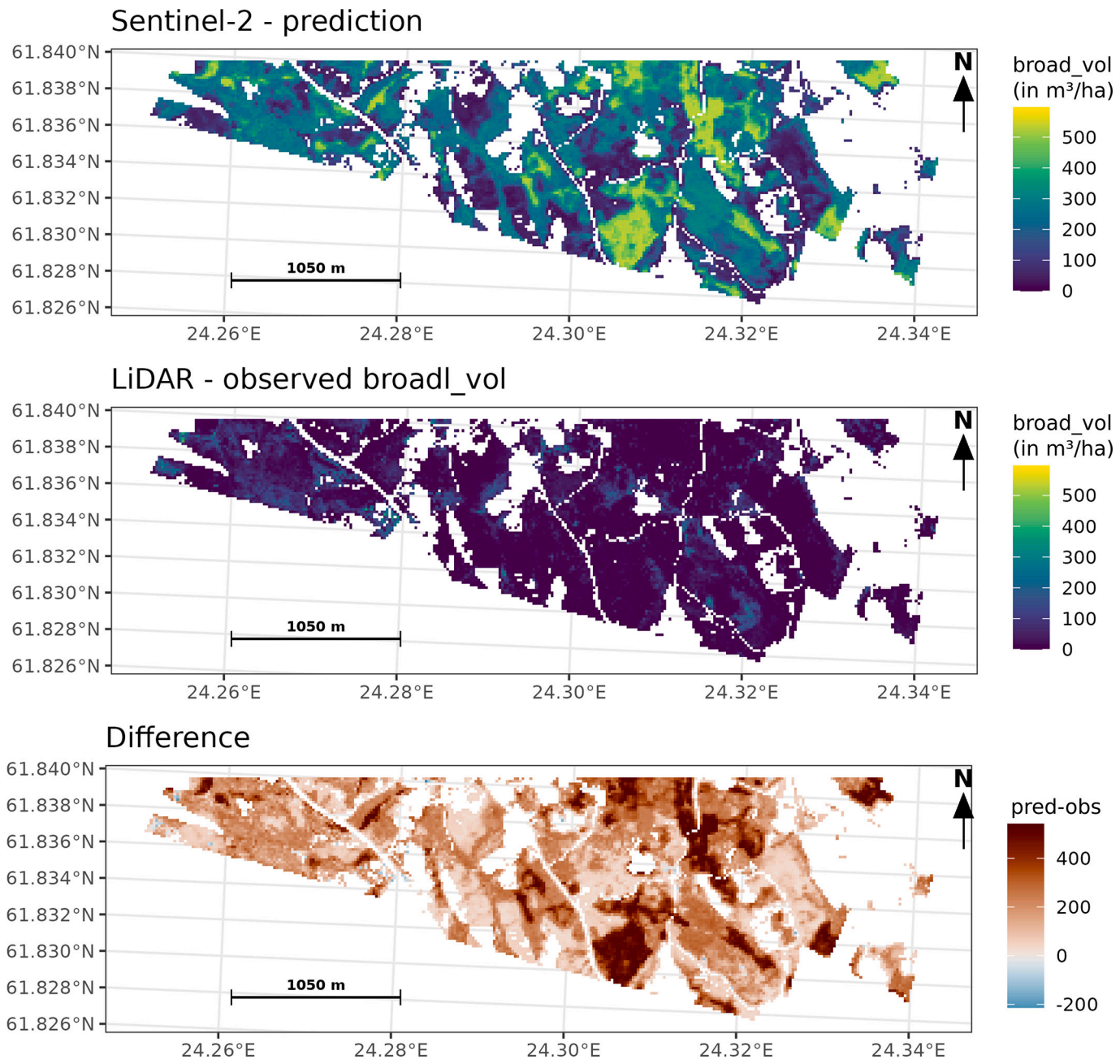


Fig. B3. Among 63 combinations of Sentinel-2 dates used as independent variables in random forest models in Hyytiälä, the most accurate volume of broadleaved trees (broad_vol, in m^3/ha) map was produced using. Three Sentinel-2 dates (2019-04-28, 2019-06-30, 2019-07-27). The corresponding R^2, RMSE and bias were 0.02, 247.76 and 206.49.

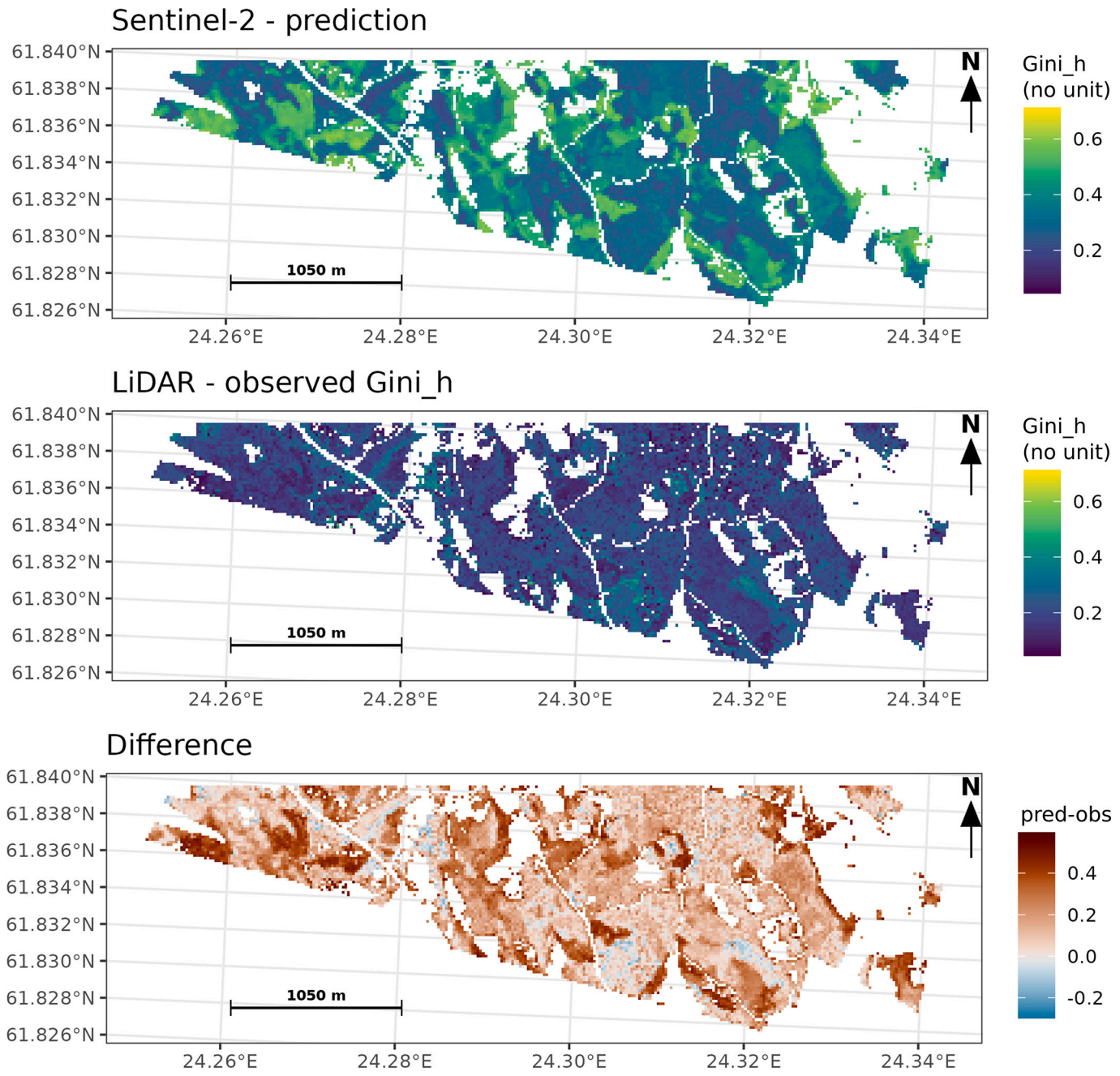


Fig. B4. Among 63 combinations of Sentinel-2 dates used as independent variables in random forest models in Hyytiälä, the most accurate Gini coefficient index of height (Gini_h, no unit) map was produced using two Sentinel-2 dates (2019-06-30, 2019-07-25). The corresponding R^2, RMSE and bias were 0.04, 0.23 and 0.18.

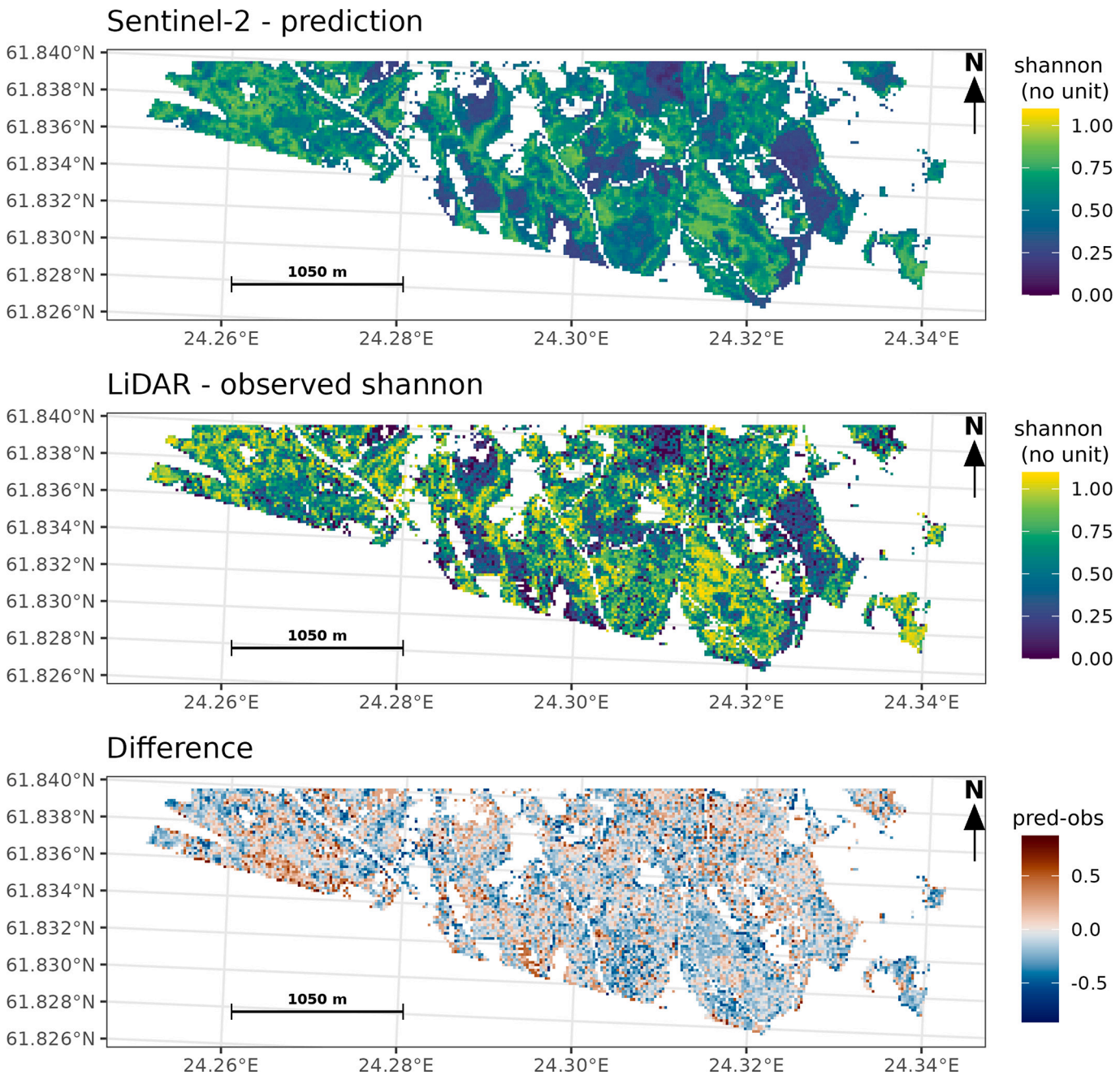


Fig. B5. Among 63 combinations of Sentinel-2 dates used as independent variables in random forest models in Hyytiälä, the most accurate Shannon index (shannon, no unit) map was produced using six Sentinel-2 dates (2019_04_28, 2019-05-18, 2019-06-30, 2019-07-25, 2019-07-27, 2019-08-29). The corresponding R^2, RMSE and bias were 0.34, 0.25 and − 0.06.

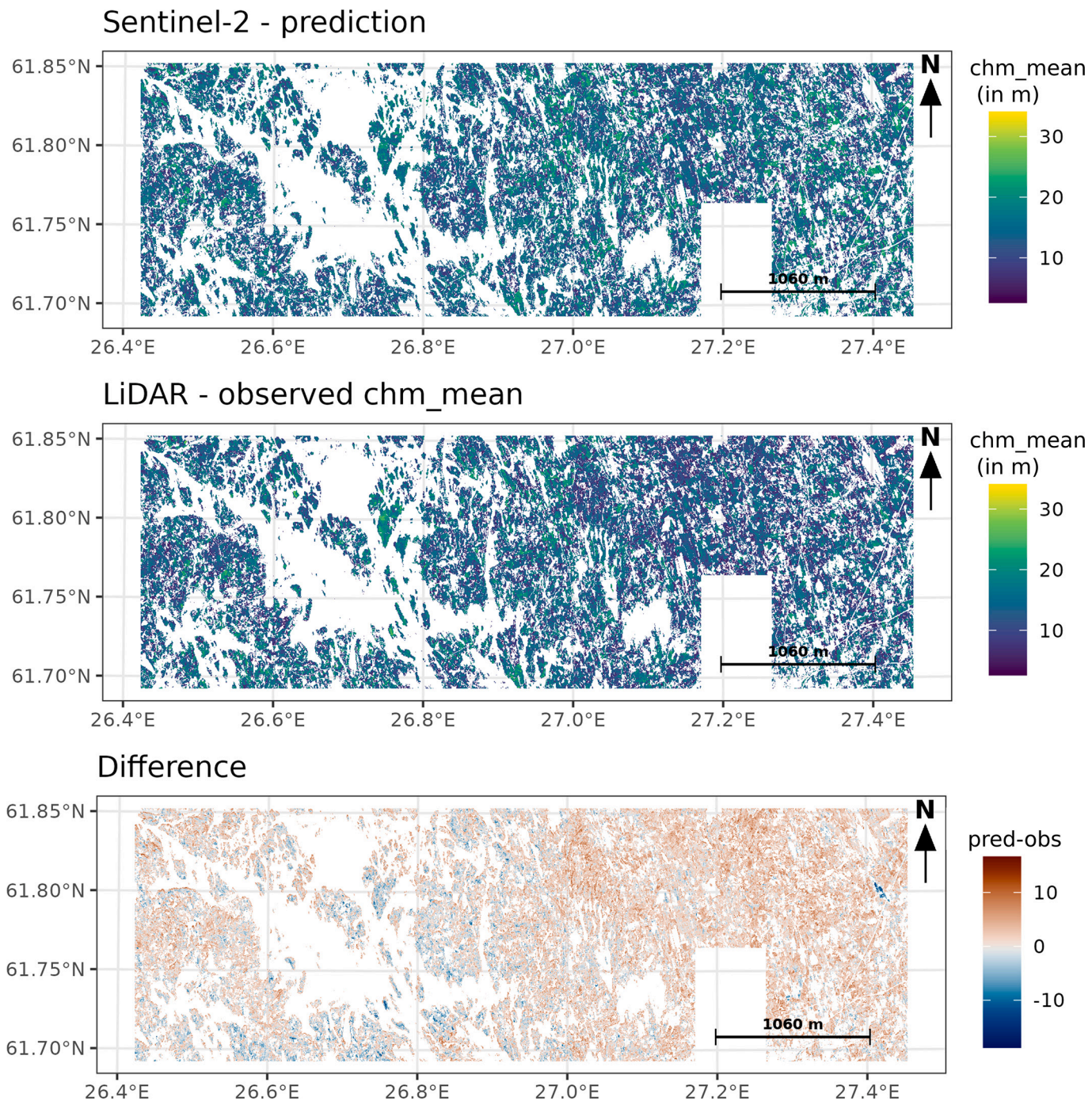


Fig. B6. Among 31 combinations of Sentinel-2 dates used as independent variables in random forest models in Mikkeli, the most accurate mean canopy height (chm_mean, in m) map was produced using five Sentinel-2 dates (2020-05-22, 2020-06-26, 2020-08-20, 2020-08-25, 2020-09-14). The corresponding R^2, RMSE and bias were 0.54, 3.33 and 1.23.

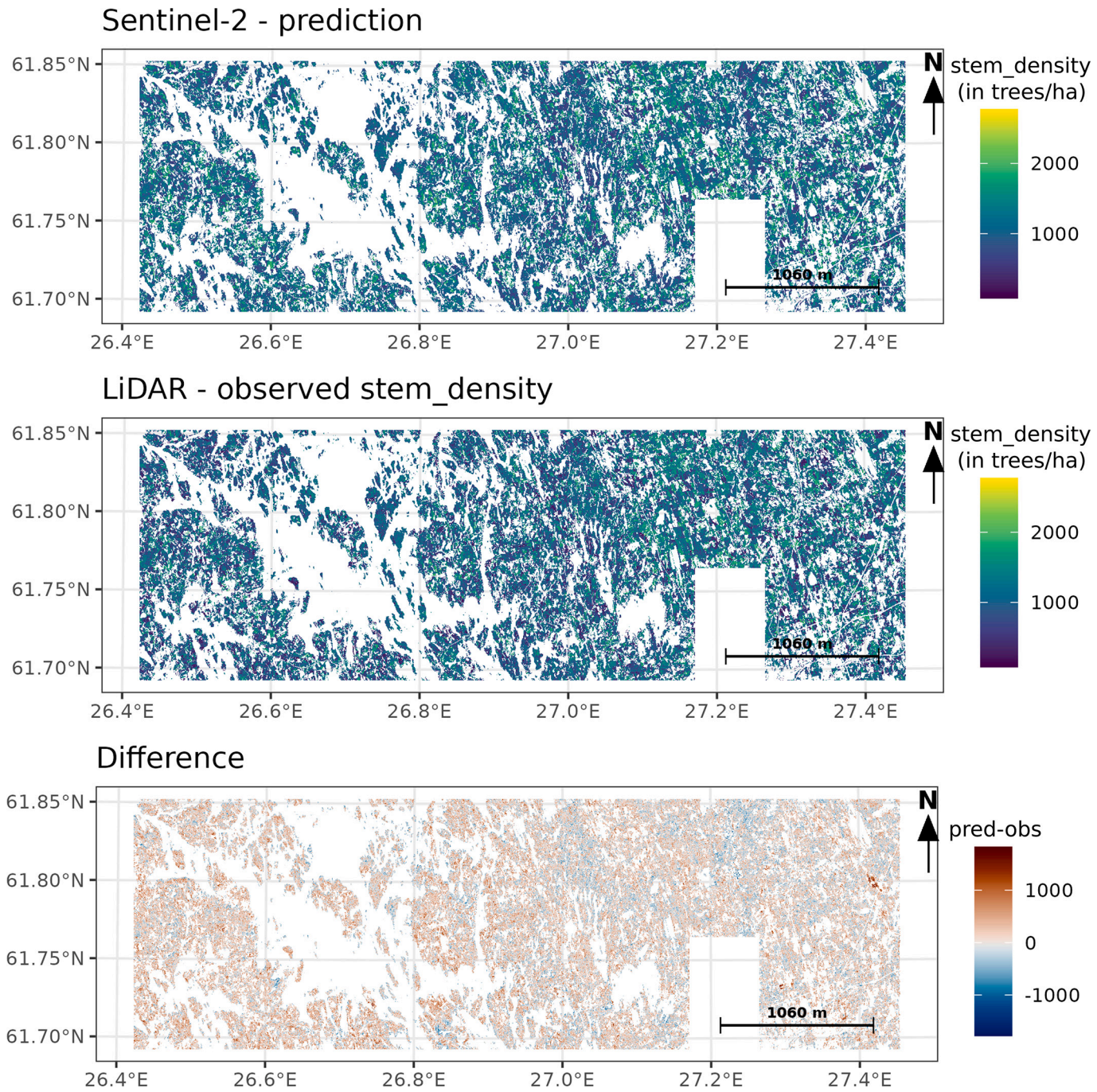


Fig. B7. Among 31 combinations of Sentinel-2 dates used as independent variables in random forest models in Mikkeli, the most accurate stem density (stem_density, in trees/ha) map was produced using five Sentinel-2 dates (2020-05-22, 2020-06-26, 2020-08-20, 2020-08-25, 2020-09-14). The corresponding R^2, RMSE and bias were 0.38, 355.32 and 83.55.

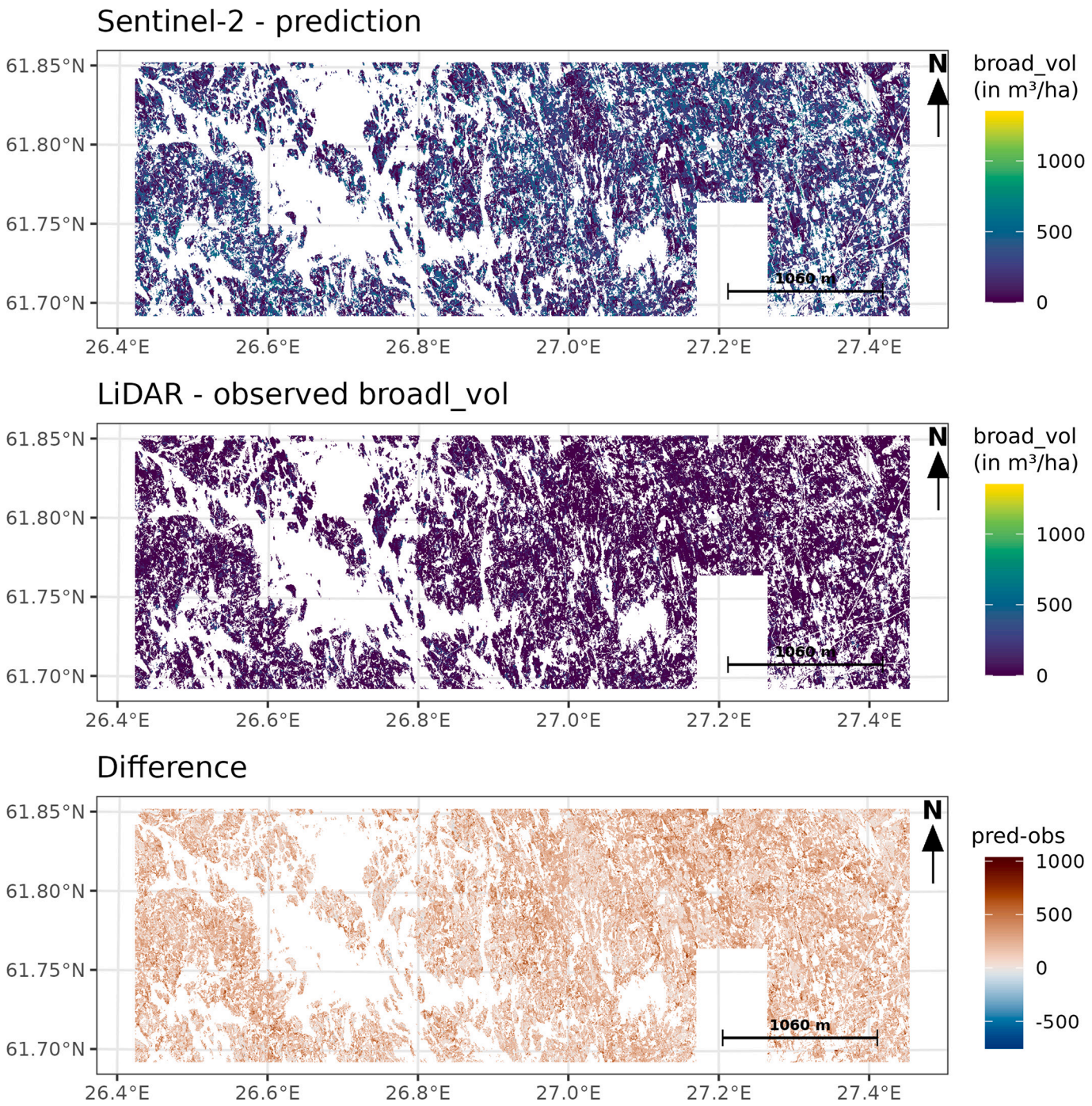


Fig. B8. Among 31 combinations of Sentinel-2 dates used as independent variables in random forest models in Mikkeli, the most accurate volume of broadleaved trees (broad_vol, in m^3/ha) map was produced using three Sentinel-2 dates (2020-05-22, 2020-06-26, 2020-08-20). The corresponding R^2, RMSE and bias were 0.24, 251.32 and 202.91.

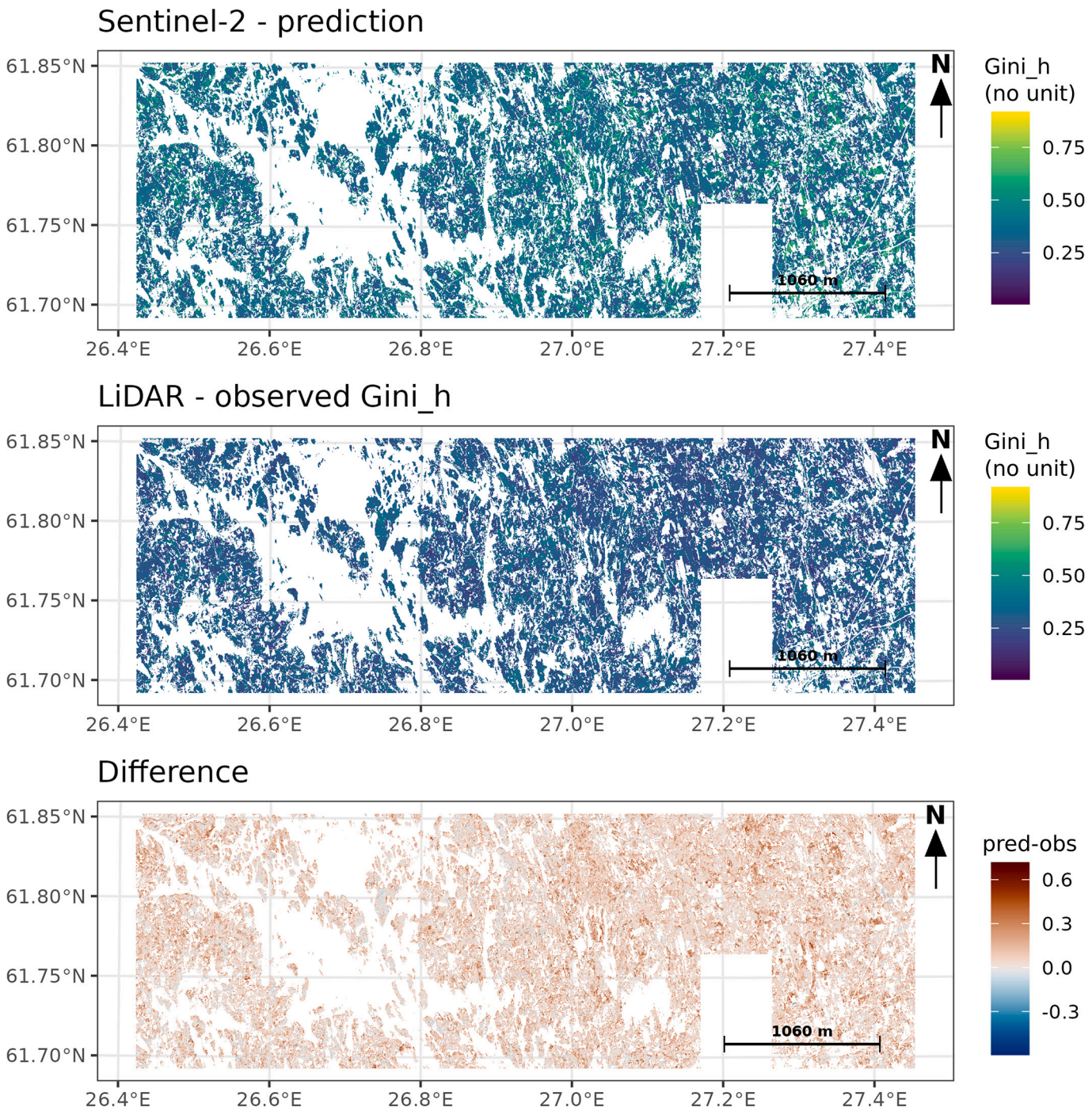


Fig. B9. Among 31 combinations of Sentinel-2 dates used as independent variables in random forest models in Mikkeli, the most accurate Gini coefficient index of height (Gini_h, no unit) map was produced using four Sentinel-2 dates (2020-05-22, 2020-06-26, 2020-08-25, 2020-09-14). The corresponding R^2, RMSE and bias were 0.08, 0.14 and 0.09.

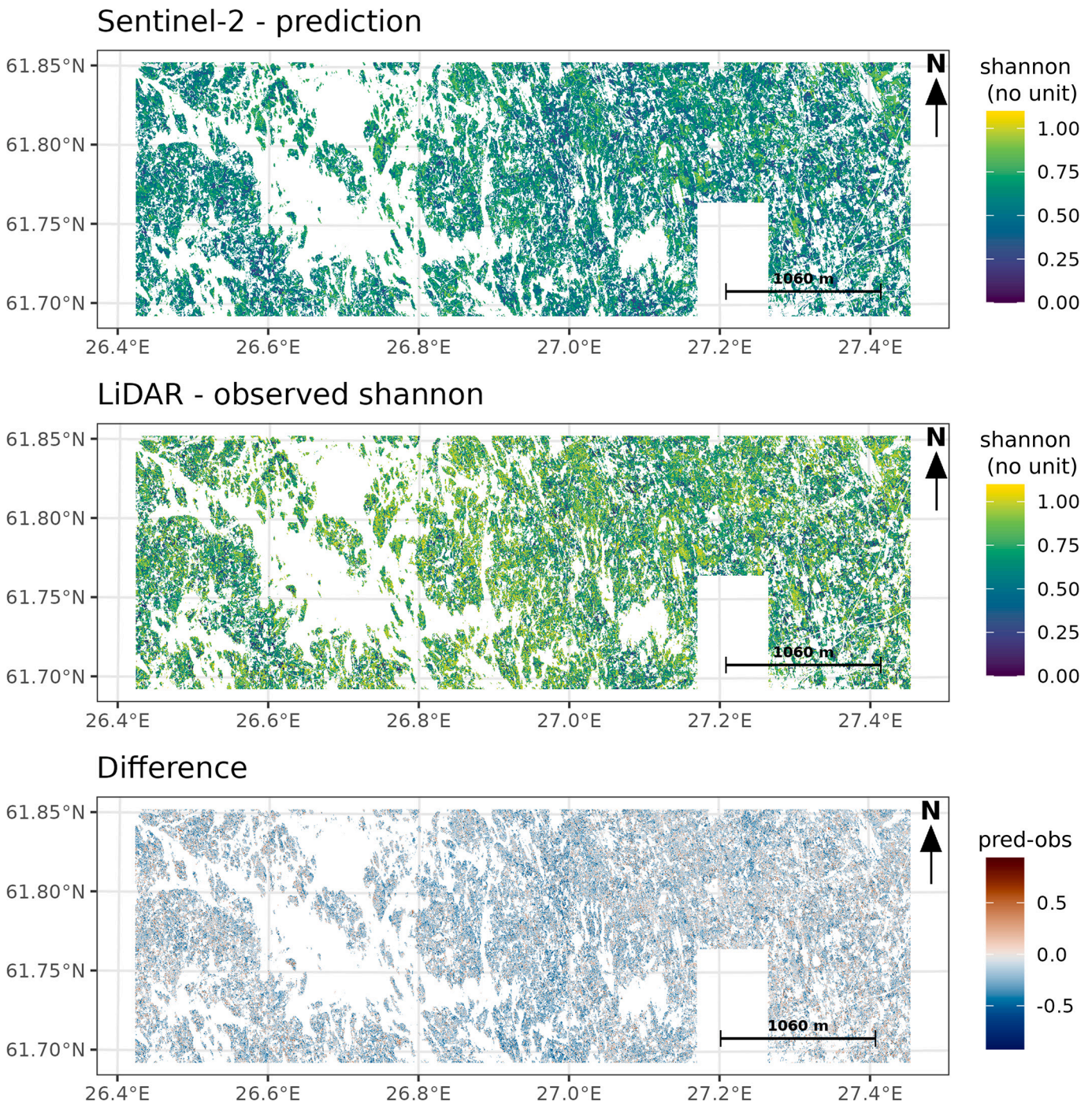


Fig. B10. Among 31 combinations of Sentinel-2 dates used as independent variables in random forest models in Mikkeli, the most accurate Shannon index (shannon, no unit) map was produced using five Sentinel-2 dates (2020-05-22, 2020-06-26, 2020-08-20, 2020-08-25, 2020-09-14). The corresponding R^2, RMSE and bias were 0.25, 0.24 and − 0.12.

Data availability

The Hyytiälä airborne LiDAR data are openly available (Hovi et al., 2024) and described in detail in Rautiainen et al. (2024). Sentinel-2 data are downloadable on the Copernicus datahub (https://browser.dataspace.copernicus.eu/). The Hyytiälä and Mikkeli tree maps are available on request due to security and private land ownership considerations. The scripts associated to this study are available on Zenodo: https://zenodo.org/records/18378868?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjM2OWRhMmE3LWU5YWItNGNiMy04MDFjLTU5ZWYwNDM2Y2Q4MyIsImRhdGEiOnt9LCJyYW5kb20iOiIzYzA5MjJlMTBmOTQ4MWQzODI0NmIzNjNjMjkwZjQ3OSJ9.IMF5MCMd9HXb9oCvHfPxOe3rd7lh250PIPhVXpZpRc1hhLjlQfjIaKxkqt8fE45r322uty9gj64SgHny6ZQ2PA

References

Atkins, J.W., Bhatt, P., Carrasco, L., Francis, E., Garabedian, J.E., Hakkenberg, C.R., Hardiman, B.S., Jung, J., Koirala, A., LaRue, E.A., Oh, S., Shao, Gang, Shao, Guofan, Shugart, H.H., Spiers, A., Stovall, A.E.L., Surasinghe, T.D., Tai, X., Zhai, L., Zhang, T.,

Krause, K., 2023a. Integrating forest structural diversity measurement into ecological research. Ecosphere 14, e4633. https://doi.org/10.1002/ecs2.4633.
Atkins, J.W., Shiklomanov, A., Mathes, K.C., Bond-Lamberty, B., Gough, C.M., 2023b. Effects of forest structural and compositional change on forest microclimates across a gradient of disturbance severity. Agric. For. Meteorol. 339, 109566. https://doi.org/10.1016/j.agrformet.2023.1095667.
Bakx, T.R.M., Koma, Z., Seijmonsbergen, A.C., Kissling, W.D., 2019. Use and categorization of light detection and ranging vegetation metrics in avian diversity and species distribution research. Divers. Distrib. 25, 1045–1059. https://doi.org/10.1111/ddi.12915.
Choi, C., Ryu, Y., Dechant, B., Yang, T., Wan, L., Jeong, S., Wu, J., 2026. Species information in multi-temporal sentinel-2 data improves forest canopy height estimation. Agric. For. Meteorol. 382, 111114. https://doi.org/10.1016/j.agrformet.2026.111114.
Clevers, J.G.P.W., 1988. The derivation of a simplified reflectance model for the estimation of leaf area index. Remote Sens. Environ. 25, 53–69. https://doi.org/10.1016/0034-4257(88)90041-7.
Copernicus Data Space Ecosystem, 2024. Copernicus Data Space Ecosystem | Europe's eyes on Earth [WWW Document]. (accessed 6.12.24). URL. https://dataspace.copernicus.eu/.
Coverdale, T.C., Davies, A.B., 2023. Unravelling the relationship between plant diversity and vegetation structural complexity: a review and theoretical framework. J. Ecol. 111, 1378–1395. https://doi.org/10.1111/1365-2745.14068.
de Sousa, K., van Etten, J., Solberg, S.Ø., de Sousa, M.K., 2023. Package 'climatrends' https://doi.org/10.32614/CRAN.package.climatrends.
Ehbrecht, M., Schall, P., Ammer, C., Seidel, D., 2017. Quantifying stand structural complexity and its relationship with forest management, tree species diversity and microclimate. Agric. For. Meteorol. 242, 1–9. https://doi.org/10.1016/j.agrformet.2017.04.012.
Eriksson, H.M., Eklundh, L., Kuusk, A., Nilson, T., 2006. Impact of understory vegetation on forest canopy reflectance and remotely sensed LAI estimates. Remote Sens. Environ. 103, 408–418. https://doi.org/10.1016/j.rse.2006.04.005.
Escadafal, R., Girard, M.-C., Courault, D., 1989. Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data. Remote Sens. Environ. 27, 37–46. https://doi.org/10.1016/0034-4257(89)90035-7.
Finnish Meteorological Institute, 2024. Download Observations - Finnish Meteorological Institute [WWW Document]. (accessed 4.12.24) URL. https://en.ilmatieteenlaitos.fi/download-observations.
Franklin, J.F., Spies, T.A., Pelt, R.V., Carey, A.B., Thornburgh, D.A., Berg, D.R., Lindenmayer, D.B., Harmon, M.E., Keeton, W.S., Shaw, D.C., Bible, K., Chen, J., 2002. Disturbances and structural development of natural forest ecosystems with silvicultural implications, using Douglas-fir forests as an example. For. Ecol. Manage. 155, 399–423. https://doi.org/10.1016/S0378-1127(01)00575-8.
Gao, T., Hedblom, M., Emilsson, T., Nielsen, A.B., 2014. The role of forest stand structure as biodiversity indicator. For. Ecol. Manage. 330, 82–93. https://doi.org/10.1016/j.foreco.2014.07.007.
Girona, M.M., Morin, H., Gauthier, S., Bergeron, Y., 2023. Boreal Forests in the Face of Climate Change: Sustainable Management. Springer Nature. https://doi.org/10.1007/978-3-031-15988-6.
Hakkenberg, C.R., Song, C., Peet, R.K., White, P.S., 2016. Forest structure as a predictor of tree species diversity in the North Carolina piedmont. J. Veg. Sci. 27, 1151–1163. https://doi.org/10.1111/jvs.12451.
Hanuš, J., Slezák, L., Fabiánek, T., Fajmon, L., Hanousek, T., Janoutová, R., Kopkáně, D., Novotný, J., Pavelka, K., Pikl, M., Zemek, F., Homolová, L., 2023. Flying laboratory of imaging systems: fusion of airborne hyperspectral and laser scanning for ecosystem research. Remote Sens. 15, 3130. https://doi.org/10.3390/rs15123130.
Heiskanen, J., Rautiainen, M., Stenberg, P., Mõttus, M., Vesanto, V.-H., 2013. Sensitivity of narrowband vegetation indices to boreal forest LAI, reflectance seasonality and species composition. ISPRS J. Photogramm. Remote Sens. 78, 1–14. https://doi.org/10.1016/j.isprsjprs.2013.01.001.
Hemmerling, J., Pflugmacher, D., Hostert, P., 2021. Mapping temperate forest tree species using dense sentinel-2 time series. Remote Sens. Environ. 267, 112743. https://doi.org/10.1016/j.rse.2021.112743.
Hovi, A., Rautiainen, M., 2020. Spectral composition of shortwave radiation transmitted by forest canopies. Trees 34, 1499–1506. https://doi.org/10.1007/s00468-020-02005-7.
Hovi, A., Schraik, D., Kuusinen, N., Fabiánek, T., Hanuš, J., Homolová, L., Juola, J., Lukeš, P., Rautiainen, M., 2023. Synergistic use of multi- and hyperspectral remote sensing data and airborne LiDAR to retrieve forest floor reflectance. Remote Sens. Environ. 293, 113610. https://doi.org/10.1016/j.rse.2023.113610.
Hovi, A., Schraik, D., Hanuš, J., Lukeš, P., Lhotáková, Z., Homolová, L., Rautiainen, M., 2024. A Spectral-Structural Characterization of European Temperate, Hemiboreal And Boreal Forests: Airborne Data. Aalto University. https://doi.org/10.23729/c6da63dd-f527-4ec9-8401-57c14f77d19f.
Husson, F., Josse, J., Le, S., Mazet, J., Husson, M.F., 2016. Package 'factominer.'. R Package 96, 698. https://doi.org/10.32614/CRAN.package.FactoMineR.
Kacic, P., Thonfeld, F., Gessner, U., Kuenzer, C., 2023. Forest structure characterization in Germany: novel products and analysis based on GEDI, sentinel-1 and sentinel-2 data. Remote Sens. 15, 1969. https://doi.org/10.3390/rs15081969.
Kassambara, A., Mundt, F., 2020. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. https://doi.org/10.32614/CRAN.package.factoextra.
Kobayashi, H., Nagai, S., Kim, Y., Yang, W., Ikeda, K., Ikawa, H., Nagano, H., Suzuki, R., 2018. In situ observations reveal how spectral reflectance responds to growing season phenology of an open evergreen forest in Alaska. Remote Sens. 10, 1071. https://doi.org/10.3390/rs10071071.
Kostensalo, J., Mehtätalo, L., Tuominen, S., Packalen, P., Myllymäki, M., 2023. Recreating structurally realistic tree maps with airborne laser scanning and ground measurements. Remote Sens. Environ. 298, 113782. https://doi.org/10.1016/j.rse.2023.113782.
Kostensalo, J., Packalen, P., Kuronen, M., Mehtätalo, L., Tuominen, S., Myllymäki, M., 2026. Large-scale tree-level mapping of forest structure including species type with remote sensing data and ground measurements. Remote Sens. Environ. 334, 115223. https://doi.org/10.1016/j.rse.2025.115223.
Kuuluvainen, T., 2016. Conceptual models of forest dynamics in environmental education and management: keep it as simple as possible, but no simpler. For. Ecosyst. 3, 18. https://doi.org/10.1186/s40663-016-0075-6.
Kuusinen, N., Stenberg, P., Tomppo, E., Bernier, P., Berninger, F., 2015. Variation in understory and canopy reflectance during stand development in finnish coniferous forests. Can. J. For. Res. 45, 1077–1085. https://doi.org/10.1139/cjfr-2014-0538.
LaRue, E.A., Hardiman, B.S., Elliott, J.M., Fei, S., 2019. Structural diversity as a predictor of ecosystem function. Environ. Res. Lett. 14, 11401. https://doi.org/10.1088/1748-9326/ab49b.
LaRue, E.A., Fahey, R.T., Alveshere, B.C., Atkins, J.W., Bhatt, P., Buma, B., Chen, A., Cousins, S., Elliott, J.M., Elmore, A.J., Hakkenberg, C.R., Hardiman, B.S., Johnson, J. S., Kashian, D.M., Koirala, A., Papeş, M., St Hilaire, J.B., Surasinghe, T.D., Zambrano, J., Zhai, L., Fei, S., 2023. A theoretical framework for the ecological role of three-dimensional structural diversity. Front. Ecol. Environ. 21, 4–13. https://doi.org/10.1002/fee.2587.
McElhinny, C., Gibbons, P., Brack, C., Bauhus, J., 2005. Forest and woodland stand structural complexity: its definition and measurement. For. Ecol. Manage. 218, 1–24. https://doi.org/10.1016/j.foreco.2005.08.034.
Mielikäinen, K., Hynynen, J., 2003. Silvicultural management in maintaining biodiversity and resistance of forests in Europe–Boreal zone: case Finland. J. Environ. Manage. 67, 47–54. https://doi.org/10.1016/S0301-4797(02)00187-1.
Moen, J., Rist, L., Bishop, K., Chapin III, F.S., Ellison, D., Kuuluvainen, T., Petersson, H., Puettmann, K.J., Rayner, J., Warkentin, I.G., Bradshaw, C.J.A., 2014. Eye on the taiga: removing global policy impediments to safeguard the boreal forest. Conserv. Lett. 7, 408–418. https://doi.org/10.1111/conl.12098.
Muise, E.R., Coops, N.C., Hermosilla, T., Ban, S.S., 2022. Assessing representation of remote sensing derived forest structure and land cover across a network of protected areas. Ecol. Appl. 32, e2603. https://doi.org/10.1002/eap.2603.
National Land Survey of Finland, 2023. Laser Scanning Data 5 p [WWW Document]. (accessed 1.8.25) URL. https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/laser-scanning-data-5-p.
Natural Ressources Institute Finland, 2024. Luke - Data Download Service [WWW Document]. (accessed 4.10.24) URL. https://kartta.luke.fi/.
Pisek, J., Rautiainen, M., Nikopensius, M., Raabe, K., 2015. Estimation of seasonal dynamics of understory NDVI in northern forests using MODIS BRDF data: semi-empirical versus physically-based approach. Remote Sens. Environ. 163, 42–47. https://doi.org/10.1016/j.rse.2015.03.003.
Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., Solomon, A.M., 1992. Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. J. Biogeogr. 19, 117–134. https://doi.org/10.2307/2845499.
Pukkala, T., Lähde, E., Laiho, O., 2014. Stand management optimization – the role of simplifications. For. Ecosyst. 1, 3. https://doi.org/10.1186/2197-5620-1-3.
Puumalainen, J., Kennedy, P., Folving, S., 2003. Monitoring forest biodiversity: a European perspective with reference to temperate and boreal forest zone. J. Environ. Manage. 67, 5–14. https://doi.org/10.1016/S0301-4797(02)00183-4.
Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1.
Rautiainen, M., Lukeš, P., 2015. Spectral contribution of understory to forest reflectance in a boreal site: an analysis of EO-1 hyperion data. Remote Sens. Environ. 171, 98–104. https://doi.org/10.1016/j.rse.2015.10.009.
Rautiainen, M., Stenberg, P., 2005. Application of photon recollision probability in coniferous canopy reflectance simulations. Remote Sens. Environ. 96, 98–107. https://doi.org/10.1016/j.rse.2005.02.009.
Rautiainen, M., Mõttus, M., Heiskanen, J., Akujärvi, A., Majasalmi, T., Stenberg, P., 2011. Seasonal reflectance dynamics of common understory types in a northern European boreal forest. Remote Sens. Environ. 115, 3020–3028. https://doi.org/10.1016/j.rse.2011.06.005.
Rautiainen, M., Heiskanen, J., Korhonen, L., 2012. Seasonal changes in canopy leaf area index and MODIS vegetation products for a boreal forest site in Central Finland. Boreal Environ. Res. 17, 72–84.
Rautiainen, M., Hovi, A., Schraik, D., Hanuš, J., Lukeš, P., Lhotáková, Z., Homolová, L., 2024. A spectral–structural characterization of European temperate, hemiboreal, and boreal forests. Earth Syst. Sci. Data 16, 5069–5087. https://doi.org/10.5194/essd-16-5069-2024.
Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Jenkins, J.P., Ollinger, S.V., 2009. Near-surface remote sensing of spatial and temporal variation in canopy phenology. Ecol. Appl. 19, 1417–1428. https://doi.org/10.1890/08-2022.1.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring Vegetation Systems in the Great Plains with ERTS. In: Third Earth Resources Technology Satellite-1 Symposium: The Proceedings of a Symposium Held by Goddard Space Flight Center at Washington, DC On. Presented at the Third ERTS Symposium. NASA, p. 309.
Sala, A., Sabaté, S., Gracia, C., Tenhunen, J.D., 1994. Canopy structure within a quercus ilex forested watershed: variations due to location, phenological development, and water availability. Trees 8, 254–261. https://doi.org/10.1007/BF00196629.

Toraño Caicoya, A., Vergarechea, M., Blattert, C., Klein, J., Eyvindson, K., Burgas, D., Snäll, T., Mönkkönen, M., Astrup, R., Di Fulvio, F., Forsell, N., Hartikainen, M., Uhl, E., Poschenrieder, W., Antón-Fernández, C., 2023. What drives forest multifunctionality in central and northern europe? Exploring the interplay of management, climate, and policies. Ecosyst. Serv. 64, 101575. https://doi.org/10.1016/j.ecoser.2023.101575.

Wright, M.N., 2015. ranger: A Fast Implementation of Random Forests. https://doi.org/10.32614/CRAN.package.ranger.

Zellweger, F., De Frenne, P., Lenoir, J., Vangansbeke, P., Verheyen, K., Bernhardt-Römermann, M., Baeten, L., Hédl, R., Berki, I., Brunet, J., Van Calster, H., Chudomelová, M., Decocq, G., Dirnböck, T., Durak, T., Heinken, T., Jaroszewicz, B., Kopecký, M., Máliš, F., Macek, M., Malicki, M., Naaf, T., Nagel, T.A., Ortmann-Ajkai, A., Petřík, P., Pielech, R., Reczyńska, K., Schmidt, W., Standovár, T., Świerkosz, K., Teleki, B., Vild, O., Wulf, M., Coomes, D., 2020. Forest microclimate dynamics drive plant responses to warming. Science 368, 772–775. https://doi.org/10.1126/science.aba6880.

Zhu, X., Liu, D., 2015. Improving forest aboveground biomass estimation using seasonal landsat NDVI time-series. ISPRS J. Photogramm. Remote Sens. 102, 222–231. https://doi.org/10.1016/j.isprsjprs.2014.08.014.