Map of forest tree species for Poland based on Sentinel-2 data

Ewa Grabska-Szwagrzyk[1], Dirk Tiede[2], Martin Sudmanns[2], Jacek Kozak[1]

[1]Institute of Geography and Spatial Management, Jagiellonian University, Gronostajowa 7, 30-387 Kraków, Poland
[2]Department of Geoinformatics—Z_GIS, University of Salzburg, Schillerstr. 30, 5020 Salzburg, Austria

Correspondence to: Ewa Grabska-Szwagrzyk, ewa2.grabska@uj.edu.pl

Abstract

Accurate information on forest tree species composition is vital for various scientific applications, as well as for forest inventory and management purposes. Country-wide, detailed species maps are a valuable resource for environmental management, conservation, research, and planning. Here, we performed the classification of 16 dominant tree species/genera in Poland using time series of Sentinel-2 imagery. To generate comprehensive spectral-temporal information, we created Sentinel-2 seasonal aggregations known as Spectral-Temporal Metrics (STMs) within Google Earth Engine (GEE). STMs were computed for short periods of 15-30 days during spring, summer, and autumn, covering multi-annual observations from years 2018 to 2021. The *Polish Forest Data Bank* served as reference data, and, to obtain robust samples with pure stands only, it was validated through automated and visual inspection based on very high resolution orthoimagery, resulting in 4500 polygons, serving as training and test data. The forest mask was derived from available land cover datasets in GEE, namely ESA World Cover and Dynamic World. Additionally, we incorporated various topographic and climatic variables from GEE to enhance classification accuracy. The Random Forest algorithm was employed for the classification process, and an area-adjusted accuracy assessment was conducted through cross-validation and test datasets. The results demonstrate that the country-wide forest stand species mapping achieved an accuracy exceeding 80%, however it varies greatly depending on species, region and observation frequency. We provide freely accessible resources including the forest tree species map, training and test data: https://doi.org/10.5281/zenodo.10180469 (Grabska-Szwagrzyk, 2023).

1. Introduction

Information of forest tree species composition is essential for many scientific applications as well as the purposes of the forest inventory and management, such as estimating timber volume, modelling biodiversity, conservation, monitoring of disturbances or carbon and biomass estimation (Hanewinkel et al., 2013; Loiselle et al., 2003; Gillis et al., 2005; Boisvenue and White, 2019). In recent times, use of remote sensing data has greatly improved forest monitoring and management. One such powerful source of data is Sentinel-2 mission, which offers high-resolution and frequent data for mapping tree species. While Sentinel-2 data have been increasingly employed for mapping species composition, most studies focus on smaller regional scales (Immitzer et al., 2016; Puletti et al., 2017; Karasiak et al., 2017; Persson et al., 2018; Grabska et al., 2019; Immitzer et al., 2019; Hościło and Lewandowska, 2019; Bolyn et al., 2018; Grabska et al., 2020; Lechner et al., 2022; Shirazinejad et al., 2022; Axelsson et al., 2021; Wessel et al., 2018; Melnyk et al., 2023) or classify broad forest classes/species

groups over larger regions (Waser et al., 2021; Breidenbach et al., 2021; Schindler et al., 2021; Rüetschi et al., 2021). For larger areas, discrimination of tree species has been performed with the use of Landsat (Turlej et al., 2022; Bonannella et al., 2022). Furthermore, continent-scale studies have utilized high-resolution hyperspectral and field data to develop models for tree species classification, evaluating both general and site-specific models (Marconi et al., 2022). At the national scale, Sentinel-2 time series were successfully used to map seven dominant tree species in Germany (Welle et al., 2022) or map larch plantations in Wales (Punalekar et al., 2021). In studying tree species composition for larger regions, additional environmental variables, for instance topographic predictors, have been found to improve classification accuracy (Waser et al., 2021; Grabska et al., 2020; Ye et al., 2021). Other datasets used as auxiliary variables include climatic variables (Hermosilla et al., 2022), soils (Hemmerling et al., 2021), phenological metrics (Kollert et al., 2021; Hermosilla et al., 2022), spectral indices (Schindler et al., 2021; Ye et al., 2021; Hemmerling et al., 2021; Praticò et al., 2021) and textural metrics (Ye et al., 2021; Hemmerling et al., 2021).

Still, the accurate mapping of forest tree species with remote sensing data remains a challenge (Fassnacht et al., 2023). Particularly, studying species composition in large areas presents significant problems, such as generating good quality predictors from satellite imagery (Grabska et al., 2020). The frequent cloud cover or topographic effects in mountainous regions may limit the number of cloud-free observations or disturb the surface reflectance values (Schindler et al., 2021). Additionally, larger areas exhibit greater environmental variability, including variations in topography, climate, and phenology, which can significantly impact species classification accuracy. The optimal image acquisition dates which are crucial in improved species recognition (Grabska et al., 2019; Immitzer et al., 2019) may substantially differ between regions. Other challenge in large-scale classification is the limited availability of reference data, especially for less common species (Zeug et al., 2018), leading to poorer performance for underrepresented species (Hemmerling et al., 2021; Marconi et al., 2022; Ahlswede et al., 2022). Finally, species classification for large regions requires handling high-volume spatial datasets, which may be difficult to process using locally installed, monolithic software. Google Earth Engine (GEE), the - for research purposes - freely accessible cloud-based platform, enables parallel processing of large spatial datasets (Tamiminia et al., 2020; Gorelick et al., 2017). GEE provides the access to entire, pre-processed Sentinel-2 collections and other environmental datasets as well as tools for processing and classification (Tamiminia et al., 2020). Previous studies have demonstrated the potential and versatility of GEE in forest classification, emphasizing its role in addressing the challenges encountered in large-scale mapping (Forstmaier et al., 2020; Chen et al., 2017; Praticò et al., 2021). Different approaches have been used to produce seamless and cloud-free satellite composites for mapping tree species composition, with multiple studies emphasizing the importance of a multi-temporal approach for accurate tree species classification (Immitzer et al., 2019; Grabska et al., 2019; Hościło and Lewandowska, 2019; Persson et al., 2018; Kollert et al., 2021). However, there are variations in optimal timing for different seasons, and applying a single seamless image composition at a country-wide scale is not feasible. Thus, researchers often employ temporal aggregations such as spectral-temporal metrics (STM) calculated for a season, year, or multi-annual periods.

Here, we present classification of 16 forest tree species/genera for the entire area of Poland. Given the availability of several years of Sentinel-2 imagery, we propose, based on our previous findings (Grabska et al., 2020, 2019), a novel approach that utilizes short-period (15-30 days) seasonal STM using mean values derived from multiple years. This strategy aims to focus on critical periods characterized by dynamic phenological changes while avoiding gaps in imagery that are commonly encountered when using single-year data. We used GEE for pre-processing and classification of the Sentinel-2 time series, along with additional environmental variables.

2. Data and methods

2.1 Study area

Poland's forests cover an area exceeding nine million hectares – 9,265,000 ha according to the Central Statistical Office; (December 31, 2021) or 9,464,000 ha according to the standard adopted for international assessments, taking into account land related to forest management (Zajączkowski et al., 2022). This accounts for approximately 30% of the country's total land area (Figure 1). In terms of ownership, public forests hold the majority share at 80.7% (with 76.9% of forests managed by the State Forests, 2% belonging to National Parks, and 1.8% of communes' properties and others), followed by private forests at around 19.3%. The dominant species is the Scots pine (*Pinus sylvestris*), covering 58.5% of the forested area across all ownership types, according to the National Forest Inventory (NFI) reports (Biuro Urządzania Lasu i Geodezji Leśnej, 2022). The second most prevalent genus is *Quercus*, primarily *Robur* and *Pedunculate* species, accounting for 8.0%. Birch (*Betula pendula*) represents 6.8% and alder species (*Alnus* spp.) 5.7% of tree species. In the mountainous regions in southern Poland, Norway spruce (*Picea abies*), silver fir (*Abies alba*), and common beech (*Fagus sylvatica*) are the most common species, covering 5.3%, 3.3%, and 6.2%, respectively. It is worth noting that European larch (*Larix decidua*) shares are usually not reported

separately but in combination with pine species. Still, larch is also among prevalent species in Poland – based on data from the Polish Forest Data Bank (FDB), the share of larch in Poland's State Forests land property is approximately 2%.

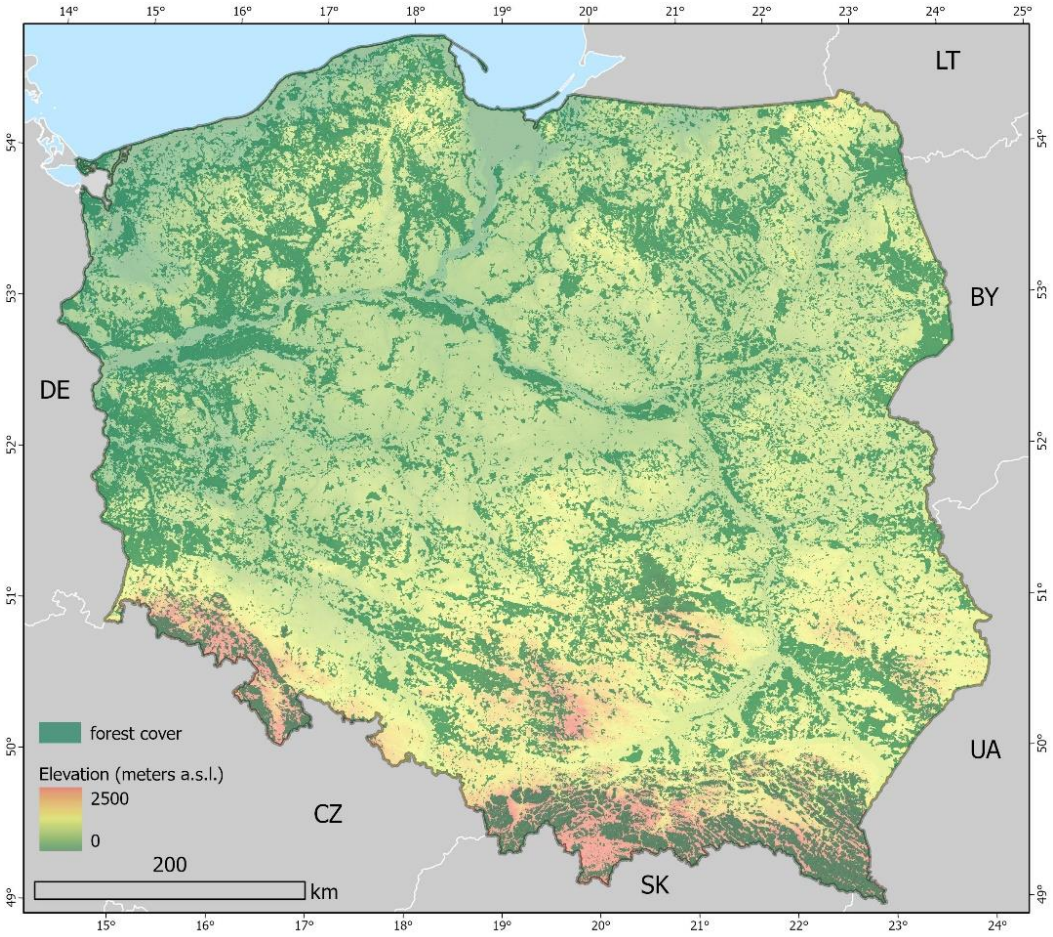


Figure 1 Forest cover in Poland Elevation from EU-DEM; forest mask derived in this study.

2.2. Workflow

We developed an approach to classify 16 tree species in Poland using Sentinel-2 time series within the GEE platform. Polish FDB was used as reference data for training, validation and test samples. We created four seasonal STMs (means) from multi-annual observations (2018-2021), performed pre-processing in GEE, and clipped them to the forest mask derived from existing land cover datasets. Classification involved the Random Forest (RF) classifier with a 10-fold cross-validation technique, and accuracy metrics were computed using test samples. To handle class imbalances, we implemented two strategies: proportional and unproportional allocation. Additionally, we compared accuracy between areas influenced by overlapping and non-overlapping Sentinel-2 orbits.

2.3. Reference data processing

The reference data was gathered from publicly accessible FDB, in which forest management units (forest stands) are represented by polygons. Each polygon contains information on species share expressed by values ranging from 1 to 10, with ten indicating homogenous coverage by a particular species. Nonetheless, the precise spatial distribution of these species within the polygons remains uncertain. In addition, the FDB does not cover private forests.

From FDB, polygons representing pure stands with a single species dominance of 90% or more, and with trees older than 10 years were selected. However, due to frequency of some species in Polish forests, we used other thresholds and additional conditions. Given the large number of reference stands of Scots pine, we randomly chose 10% of pure stands with 100% share of this species, however other pine species uncommon in Poland underwent the same processing procedures as other examined species. On the other hand, due to an insufficient number of reference samples for less common species such as poplar (*Populus* spp.), black locust (*Robinia pseudoacacia*), hornbeam (*Carpinus betulus)*, ash (*Fraxinus excelsior*), maple (*Acer* spp.), lime (*Tillia* sp.), and Douglas fir (*Pseudotsuga menziesii*), additional FDB stands with a 60-80% share of these species were included. The next step involved precise adjustments of reference samples to the actual forest mask derived from two available land cover datasets in GEE, i.e. any samples or their parts falling outside of forest mask were removed. Specifically, we utilized the ESA WorldCover 2021 product ("ESA/WorldCover/v200"; (Zanaga et al., 2022)), selecting only value 10 (i.e., tree cover), and the Dynamic World dataset ("GOOGLE/DYNAMICWORLD/V1", (Brown et al., 2022)) calculated from summer 2021 imagery and aggregated to mean, with a tree probability threshold set at 0.6. Both datasets were employed, as based on our tests, the ESA World Cover product tends to overestimate forests in certain areas, while the DynamicWorld dataset, generated dynamically from available Sentinel-2 observations, may be prone to errors due to frequent cloud cover. In the next step, image segmentation on the Sentinel-2 STM was performed (Harmonized Level-2A data; 'COPERNICUS/S2_SR_HARMONIZED'), utilizing mean values from summer 2021. This segmentation process was carried out using Simple Non-Iterative Clustering (SNIC; (Achanta and Süsstrunk, 2017)) algorithm in GEE limited to the previously selected FDB stands within the forest mask area, with the aim to delineate spectrally homogeneous patches. Segments obtained in this step were intersected with the FDB stands, and for further processing only segments larger than 0.5 hectares that encompassed more than 60% of the stands were selected. Subsequently, the resulting segments were visually checked using very high-resolution orthoimagery.

Finally, 4500 polygons were obtained representing 16 species/genera (Table 1). They were divided into training (2999; corresponding to approx. 400 thousands training pixels) and test polygons (1501). The training data was further divided into training (90%) and validation (10%) and 10-fold cross-validation was employed to calibrate the model. The examples of reference samples for each examined class are illustrated in Figure 2.

Table 1 Classes and species classified in our study with the number of reference polygons and pixels

Class	Species	No. of polygons	No. of pixels	Share of the total reference pixels [%]	Percentage of polygons from stands with 60-80% species share [%]
Pine	*Pinus sylvestris* *Pinus nigra* *Pinus strobus* *Pinus rigida* *Pinus banksiana*	1036	183 768	30.9	0
Oak	*Quercus robur* *Quercus petraea* *Quercus rubro*	512	79 713	13.4	0
Beech	*Fagus sylvatica*	301	43 018	7.2	0
Alder	*Alnus glutinosa* *Alnus incana*	477	37 792	6.4	0
Birch	*Betula pendula* *Betula pubescens*	419	38 744	6.5	0
Larch	*Larix Decidua*	256	25 153	4.2	0
Spruce	*Picea abies*	419	51 615	8.7	0
Fir	*Abies alba*	171	29 319	4.9	0
Hornbeam	*Carpinus betulus*	134	19 376	3.3	66
Poplar	*Populus alba* *Populus tremula* *Populus nigra*	176	22 146	3.7	51
Ash	*Fraxinus Excelsior*	164	16 894	2.8	48
Maple	*Acer pseudoplatanus* *Acer platanoides* *Acer campestre*	122	12 454	2.1	67
Lime	*Tillia cordata* *Tilia platyphyllos*	60	6 567	1.1	67
Douglas fir	*Pseudotsuga menziesii*	124	12 946	2.2	29
Black locust	*Robinia pseudoacacia*	86	9 644	1.6	55
Dwarf mountain pine	*Pinus mugo*	43	5 165	0.9	0

125

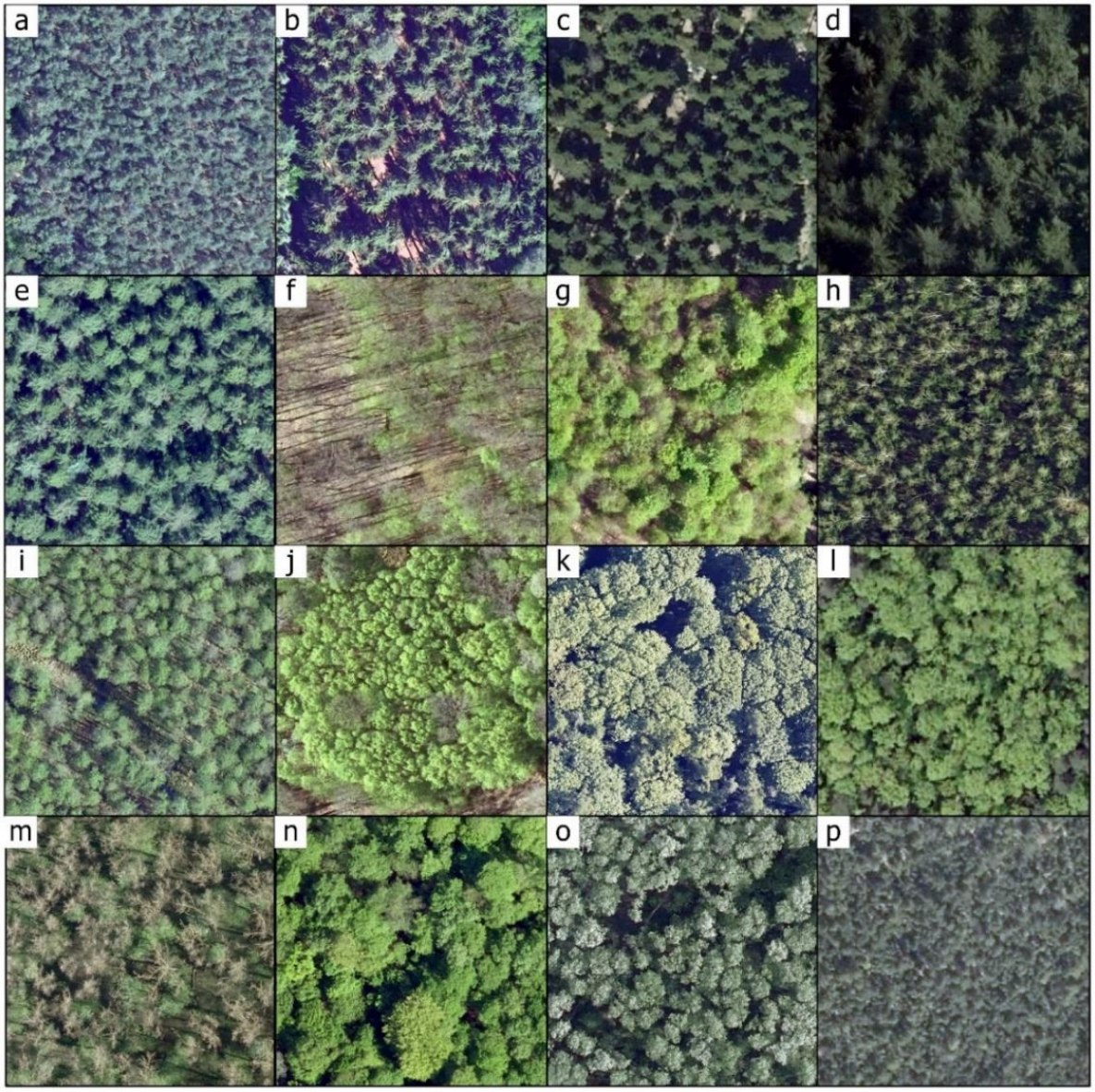


Figure 2 Examples of reference samples for each analysed tree species/genera shown in (spring) very high resolution orthoimagery: a) pine; b) spruce; c) fir; d) Douglas fir; e) larch; f) oak; g) beech; h) birch; i) alder; j) hornbeam; k) maple ; l) ash; m) poplar; n) lime; o) black locust; p) dwarf mountain pine. Orthoimagery is openly available at Polish Geoportal (https://mapy.geoportal.gov.pl/; Head Office of Geodesy and Cartography).

2.4. Satellite imagery processing and additional variables

Regarding satellite imagery predictors, numerous studies have demonstrated the significance of a multi-temporal approach in accurately distinguishing tree species (Immitzer et al., 2019; Grabska et al., 2019; Hościło and Lewandowska, 2019; Persson et al., 2018; Kollert et al., 2021). For instance, our previous study on species classification in a smaller area highlighted the

optimal timing for distinguishing forests tree species in temperate zones, which varies during the spring and autumn seasons (Grabska et al., 2019). At a national scale, however, applying a single seamless image composition for the entire growing season is impractical. While seasonal STMs can provide important phenological information (Müller et al., 2015), areas with frequent cloud-cover may still experience difficulties in acquiring high-quality observations for all needed temporal time steps (Grabska et al., 2020). Different approaches to calculate Sentinel-2 based STMs were employed, such as utilizing seasonal metrics calculated over two to four months (Praticò et al., 2021) or testing long-term, seasonal, and monthly composites (Nasiri et al., 2023).

Here, we employed seasonal Sentinel-2 (L2A) Spectral-Temporal Metrics (STM) calculated in GEE for four periods: (1) the second half of April, (2) May, (3) June/July, and (4) October from the years 2018-2021. For each period, one seasonal STM from multi-annual observations was calculated. The specific periods for each season and year are provided in Table 2. They were selected based on findings from our previous studies (Grabska et al., 2019, 2020; Grabska-Szwagrzyk and Tymińska-Czabańska, 2023). The spring imagery was chosen to capture the greening-up phase, while autumn imagery was selected to represent the period when leaves undergo colour changes. Furthermore, we decided to include two spring STMs, one early and one late spring, as our previous study revealed significant differences among deciduous species in this period. For instance, on a smaller site, there was an 8-18 day gap between early leafing species like larch and birch, and late leafing species like alder and oak (Grabska-Szwagrzyk and Tymińska-Czabańska, 2023). Moreover, we included a summer STM, as it represents a relatively stable and certain period and allows to utilize a greater number of images. In the previous study on forest tree species classification in the Polish Carpathians, bands from July STM were among the most important variables (Grabska et al., 2020). The dates were slightly modified due to meteorological conditions in particular years and therefore phenology variations, as well as missing observations in some cases (Table 2).

All available Sentinel-2 images from the Harmonized Level-2A collection captured during these periods and with cloud cover below 40% were pre-processed, including cloud, cloud shadow and dark pixel masking based on the Sentinel-2 cloud probability dataset (based on the Sentinel2-cloud detector, see https://github.com/sentinel-hub/sentinel2-cloud-detector) also available in GEE ('COPERNICUS/S2_CLOUD_PROBABILITY'). The number of clear observations for each period varied largely due to cloud cover as well as overlapping Sentinel-2 orbits (Figure 3).

Table 2 Periods of Sentinel-2 imagery used in analysis.

Name	2018	2019	2020	2021
Early spring	04/15 – 05/10*	04/20 – 05/10	04/20 – 05/10	05/05 – 05.25
Late spring	05/10 – 05/30	05/15 – 06/05	05/15 - 06/05	05/25 – 06/15
Summer	06/10 – 07/10	06/10 – 07/10	06/10 – 07/10	06/10 – 07/10
Autumn	10/25 – 11/10	10/20 – 11/05	10/20 – 11/05	10/25 – 11/10

*Period increased due to not enough observations

The pre-processed imagery was then clipped to match the actual forest mask, ensuring that only relevant areas were considered for analysis. In addition, the Normalized Difference Vegetation Index (NDVI) was calculated to mitigate the potential impact of disturbances on the obtained results and remove recent clear cuts, ensuring that only areas with healthy vegetation were considered. Specifically, based on tests, the pixels with NDVI values below 0.6 from the summer 2021 STM were excluded from the analysis (Figure 4). Final step employed calculating mean reflectance values for each pixel, and for each specific season, based on the seamless Sentinel-2 imagery.

Additional variables for classification included environmental datasets available in GEE. They included: elevation data (reprocessed 30m SRTM data: "NASA/NASADEM_HGT/001"), WorldClim variables ("WORLDCLIM/V1/BIO"): temperature and precipitation (bio1, bio12, bio17), soils ("OpenLandMap/SOL/SOL_GRTGROUP_USDA-SOILTAX_C/v01") and Terra Climate ('IDAHO_EPSCOR/TERRACLIMATE') maximum air temperature for 2018 (see Table A1 in the appendix).

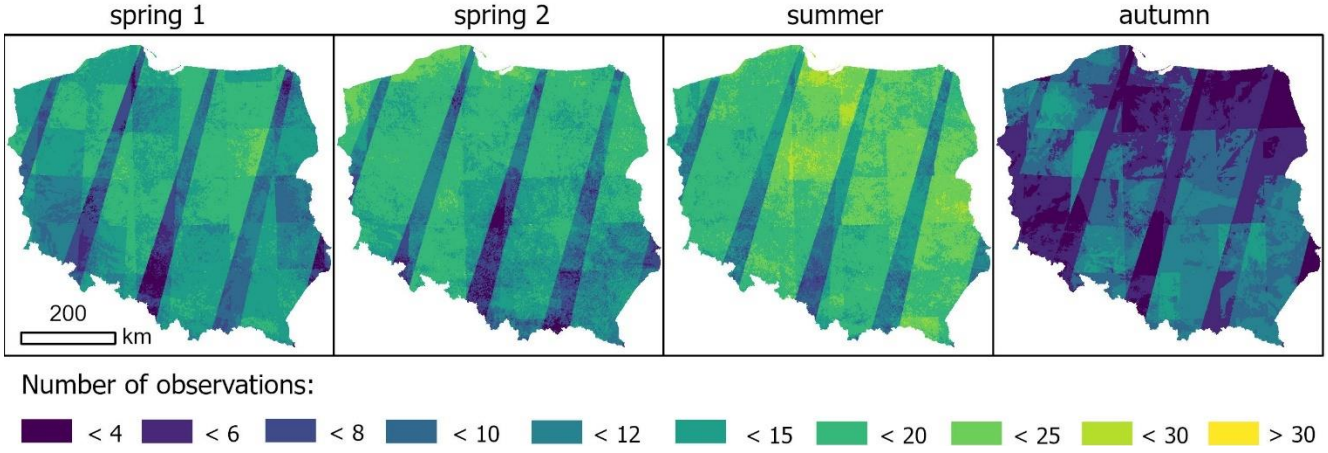


Figure 3 Number of cloud-free observations in the analysed periods combined for all years.

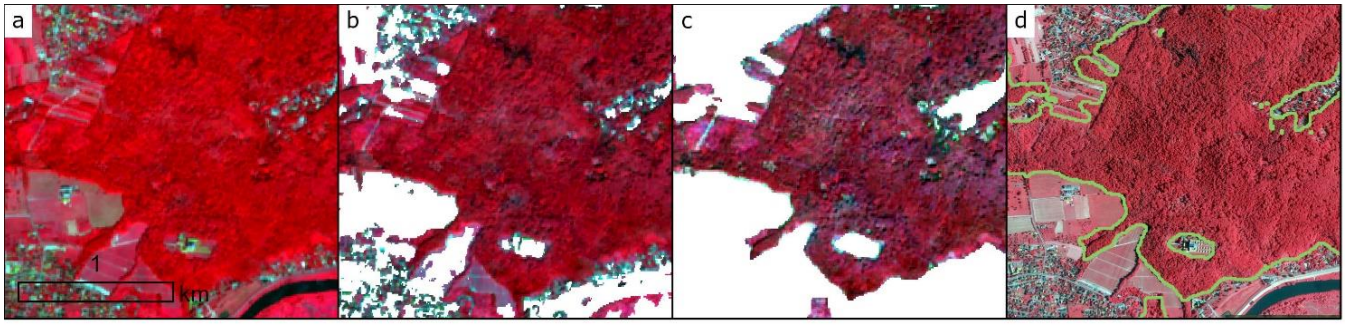


Figure 4 Procedure for obtaining forest mask used in this study on the example of part of Kraków, southern Poland: a) Sentinel-2 image (NIR, VIS R, VIS G); b) Sentinel-2 clipped to ESA World Cover v200 dataset extracted tree cover class; 3) Sentinel-2 clipped to ESA World Cover, Dynamic World and NDVI thresholding; 4) high resolution CIR orthoimage with borders of the calculated forest mask.

2.5. Classification and accuracy assessment.

Classification for the entire area of Poland was performed using approx. 400,000 sample pixels, employing a 10-fold cross-validation technique. RF classifier (Breiman, 2001) was used within the GEE, with the number of trees set to 200. This algorithm was chosen because it is reported to be insensitive to overfitting and outliers in training samples (Belgiu and Drăgu, 2016). Moreover, RF is commonly used in vegetation mapping studies for large areas (Rüetschi et al., 2021; Hermosilla et al., 2022). Among the classification algorithms available in GEE, RF has been reported to be less computationally intensive than SVM (Bonannella et al., 2022) and to outperform other algorithms (Praticò et al., 2021). Accuracy assessment included estimation of area-adjusted confusion matrices, producer's accuracy (PA), user's accuracy (UA), F1-score, which is a weighted harmonic mean of UA and PA, and overall accuracy (OA). For this task, 1501 test polygons (see section 2.3) were utilized. To ensure the robustness of accuracy assessment, a stratified random sampling approach based on species was adopted, as recommended by Olofsson et al. (2014) and based on our previous research (Grabska et al., 2020). Furthermore, we tested the unproportional allocation approach which is commonly employed when dealing with substantial class imbalances (Marconi et al., 2022; Maxwell et al., 2018; Jackson and Adam, 2021).

In recognition of class imbalance, a two-fold strategy was implemented. The first approach involved proportional allocation, while the second approach involved an unproportional dataset. The sample size for less common species was increased through oversampling, whereas undersampling was employed to the most common class, *Pinus*. In both approaches, the size of the sample was approximately 20,000 pixels (see Table A2 in the appendix) and a minimum sampling distance of 20 meters was used. Finally, regarding the significant differences in number of observations between Sentinel-2 orbit overlapping and non-overlapping areas, further analyses were conducted to evaluate the impact of observation frequency on accuracy. This included the calculation of OA separately for overlapping and non-overlapping areas in both sampling approaches.

3. Results and discussion

3.1. Overall accuracy of the tree species maps and variable importance

On average, the classification process yielded high OA, achieving values of approximately 80% or higher. Employing a 10-fold cross-validation, the average OA was equal to 83.3%, ranging between 79.3 and 84.9%. Subsequently, the species map with the best performance in terms of OA from the initial step was validated with approximately 20,000 pixels in two approaches: proportional and unproportional. The proportional approach demonstrated an OA of 89.6%, while in the unproportional approach a lower accuracy of 84% was achieved. This decline in accuracy when transitioning from proportional to unproportional samples allocation is reasonable, as more samples represent less-common species, usually underperforming the most common ones.

OA varied between regions with overlapping and non-overlapping Sentinel-2 orbits. The following OAs were obtained: 86.7% for not-overlapping areas and 90.1% for overlapping area using proportional allocation, and 83.8% and 84.1% using

unproportional allocation respectively. Although the difference using unproportional allocation seems to be low, limited number of clear observations may increase the uncertainty of estimations (Schindler et al., 2021). In studies which utilize Landsat imagery, the number of clear observations plays a vital role in classification accuracy improvement (Turlej et al., 2022). Furthermore, in mapping large areas accuracy metrics are not expected to be uniform in space due to high species and environmental diversity. Examples of selected regions with low and high accuracies are illustrated in Figure **5**. Numerous environmental and forest-related factors can impact the results. For example, heterogeneous forest structure with high diversity in age and species (Figure **5**A) result in misclassifications and require further examination and addressing. Also, misclassification occurs more often in the mountainous areas, particularly in the Carpathian forests due to higher species and environmental diversity and topography effects (Figure **5**B). High accuracy is observed in areas featuring a combination of various species but composing pure stands with a similar forest structure (Figure **5**C) as well as in locations where dense black locust stands are present (Figure **5**D).

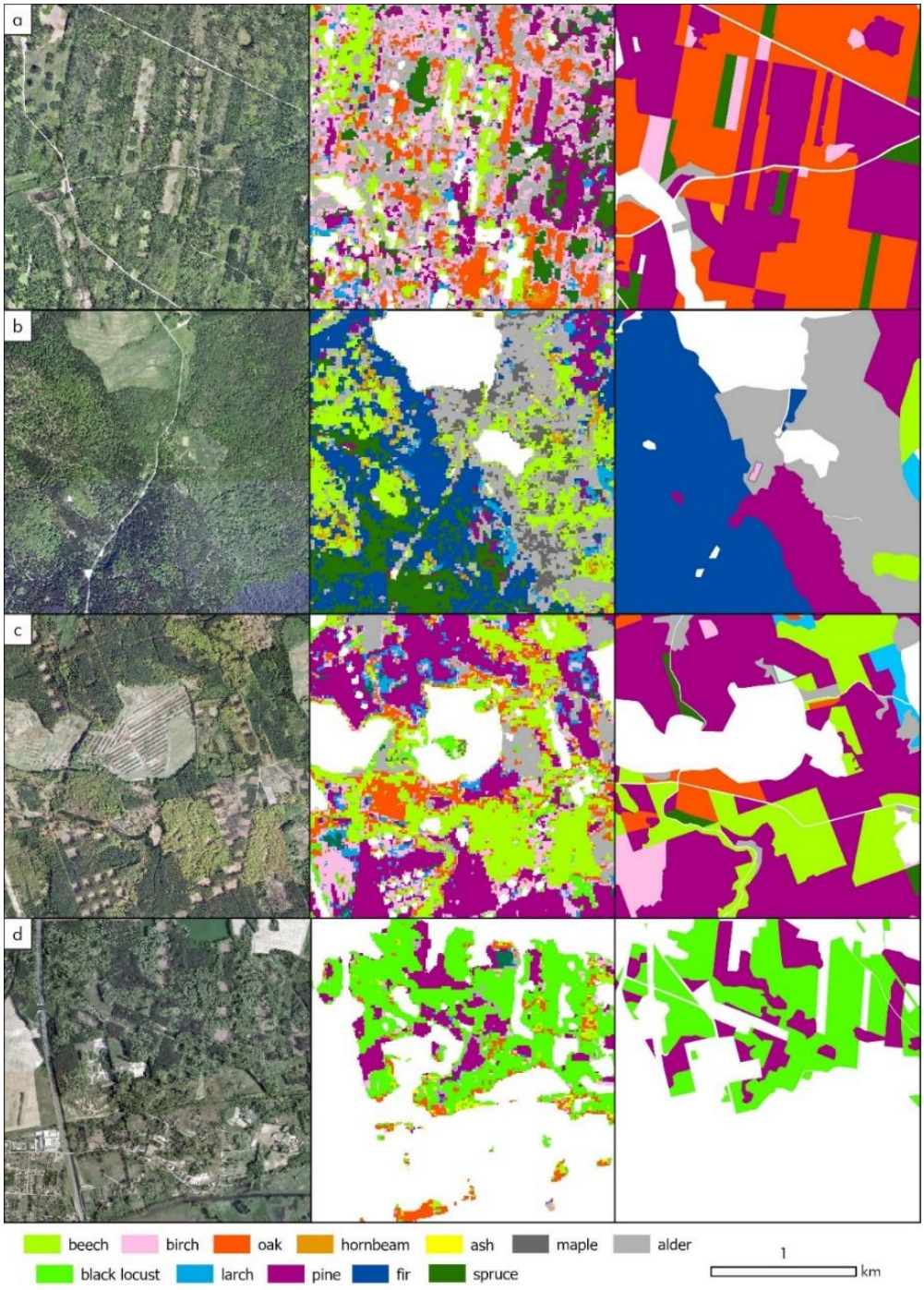
The variable importance analysis (see Figure A1 in the appendix) revealed the highest contributions from environmental variables such as maximum temperature, annual precipitation, mean annual temperatures, and elevation, similar to findings in other studies for large areas (Hermosilla et al., 2022). Among the periods used to calculate STMs, bands from autumn appeared to be the highest ranked, followed by early spring bands. Notably, visible, red-edge and SWIR bands showed stronger importance. On the other hand, the soils dataset exhibited notably lower importance compared to other predictors, despite previous reports indicating soils as more significant than climatic variables in temperate tree species distribution (Walthert and Meier, 2017). However, it is important to note that these findings may vary across regions and be scale-dependent, and more detailed soil information could enhance the accuracy of the results.

3.2. Tree species distribution and accuracy

The obtained map of forest tree species/genera reveals the share and spatial distribution of forests in Poland. Pine-dominated stands are the most common, accounting for 47.5% of the total forest cover in the country. Several other common species prevalent across Polish forests are birch occupying 11.7% of the forested areas, along with alder at 9%, beech at 8.1%, and oak at 7.2%. Other common species include spruce (3.7%) and fir (2.8%), predominantly occurring in mountainous areas in the southern Poland. Additionally, larch-dominated stands are relatively common (3.6%), along with ash (1.7%), hornbeam (1.1%) and poplar (1%). Several other species each hold a share of less than 1% in the overall forest composition, including Douglas fir, maple, and black locust. Lastly, lime and dwarf mountain pine have a more marginal presence in the obtained map.

The comparison of the results with official statistics shows some discrepancies. Firstly, the share of pine in our map is underestimated by more than 10 percentage points, which may result from several factors. One possible reason is the misclassification of pine as spruce or other coniferous trees, which accounts for 0.65% of the reference data, particularly in mountainous regions. Additionally, the share of pine is decreasing in recent years due to shifts in forest management practices,

such as the transition from monocultures to stands with more diversified species composition (Tomaś and Jagodziński, 2019). Furthermore, pine has been susceptible to disturbances in recent years, which may have led to misclassifications (Hemmerling et al., 2021). Another species with share lower in our map than reported is spruce (3.6% vs 5.3%), which, in recent years is exposed to significant disturbances and dieback, particularly in the Western Carpathian mountains and Białowieża forest (Grodzki, 2010; Bałazy, 2020; Kamińska et al., 2021). Consequently, the share of spruce is also decreasing. On the other hand, certain species like alder and birch are seemingly more common than in the official reports. The larger share of birch may be attributed to the fact that this species are common on abandoned agricultural land, it is also regarded as pioneer and successional species (Hynynen et al., 2010). The area analysed in our study might include former agricultural lands where forest succession takes place, a process that is very common in different parts of Poland (Shahbandeh et al., 2022; Kolecka et al., 2017; Zgłobicki et al., 2020; Majchrowska, 2013). Abandoned areas with forest succession, however, are not included in the official reporting for forests. Also, while very young forests have been excluded from our analysis, the visual inspection indicates frequent misclassifications of younger stands covered with broad-leaved trees as alder, which may be one of the reasons for its overestimation.



255

Figure 5 Examples of classification (middle) compared with high-resolution orthoimagery (left) and dominant species from Forest Data Bank (right): a) Czarna Białostocka forest district, NE Poland lowlands; b) Baligród forest district; SE Poland, Bieszczady mountains; c) Kłodawa forest district, NW Poland lowlands; d) Sulechów forest district; W Poland lowlands. Orthoimagery is openly available at Polish Geoportal (https://mapy.geoportal.gov.pl/; Head Office of Geodesy and Cartography).

In terms of species accuracy, the most abundant species in Poland, pine, was classified with the highest accuracy, exceeding 90% F1-score (Figure 6). Other species demonstrating F1-score of 80% or higher included dwarf mountain pine, alder, beech, fir, spruce, oak and larch. With the exception of dwarf pine mountain, these species are common in forests of Poland. On the other hand, the classification of poplar, Douglas fir, maple, lime, hornbeam, and ash revealed relatively poor accuracy levels, below 60%. Surprisingly, rare species such as black locust achieved high classification accuracy around 75%. Confusion matrix reveals the frequent misclassifications (Table 3). Typically, broad-leaved species such as ash, hornbeam and lime are misclassified – ash and lime as oak and hornbeam as oak and beech species; while coniferous Douglas fir – as pine. Similarly, in the study of Hemmerling et al. (2021), less common species with relatively high accuracy was black locust. This is a result of its unique spectral-temporal properties, as usually it leaves out later than other broad-leaved and is characterized by flowering in late spring (Rusňák et al., 2022; Somodi et al., 2012). It is promising result, taking into account the invasiveness of this non-native species in Europe (Richardson and Rejmánek, 2011). The visual inspection also indicates that frequent misclassifications include younger stands, such as oak, misclassified as other broad-leaved species, e.g. alder. Importantly, the age structure within the examined species differs largely (based on FDB), with average values between approx. 50 years old for birch, larch and alder; around 70 for spruce and pine and above 80 for beech, oak and fir. Furthermore, the species classification in young forests, characterized by the distinguished spectral characteristics than the mature ones, is challenging. Finally, not all species occurring in Poland were classified.

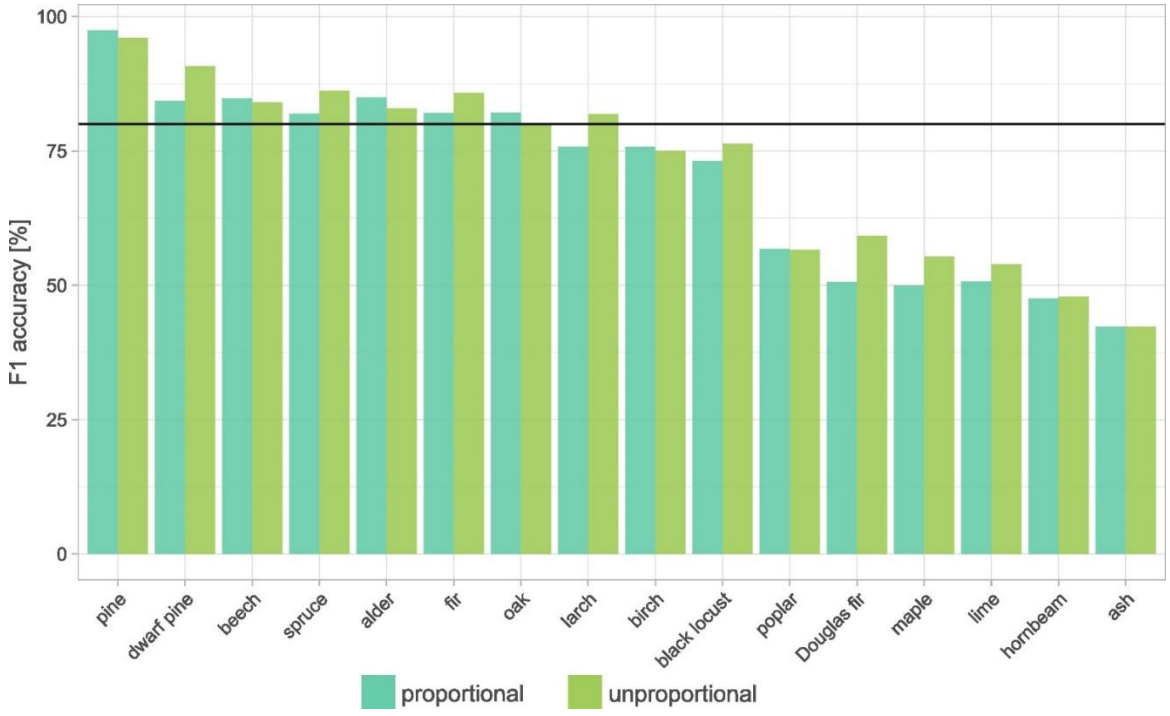


Figure 6 F1-score for 16 analysed species in two approaches: proportional samples allocation and unproportional allocation with down sampling of pine and oversampling of other classes.

Table 3 Area-adjusted confusion matrix for the unproportional samples allocation (populated by estimated proportions of area).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Beech (1)	7.51	0.12	0.28	0.00	0.46	0.01	0.08	0.13	0.02	0.01	0.17	0.01	0.08	0.01	0.02	0.00
Birch (2)	0.07	7.14	0.35	0.02	0.07	0.02	0.01	0.01	0.02	0.21	0.29	0.05	0.34	0.01	0.01	0.00
Oak (3)	0.85	1.16	10.77	0.00	0.81	0.01	0.21	0.16	0.27	0.13	0.21	0.00	0.40	0.00	0.10	0.00
Douglas f. (4)	0.00	0.01	0.00	0.61	0.00	0.25	0.01	0.00	0.00	0.01	0.01	0.12	0.00	0.04	0.00	0.00
Hornbeam (5)	0.27	0.03	0.10	0.00	1.12	0.00	0.03	0.14	0.15	0.00	0.09	0.00	0.02	0.00	0.00	0.00
Fir (6)	0.03	0.00	0.00	0.08	0.01	5.50	0.00	0.00	0.00	0.01	0.01	0.13	0.00	0.18	0.00	0.00
Ash (7)	0.03	0.03	0.06	0.00	0.06	0.00	0.46	0.07	0.08	0.00	0.34	0.00	0.02	0.00	0.04	0.00
Maple (8)	0.03	0.00	0.01	0.00	0.01	0.00	0.01	0.42	0.01	0.00	0.05	0.00	0.01	0.01	0.00	0.00
Lime (9)	0.00	0.00	0.01	0.00	0.02	0.00	0.01	0.01	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Larch (10)	0.03	0.54	0.02	0.01	0.01	0.07	0.00	0.00	0.01	3.57	0.03	0.16	0.04	0.05	0.00	0.00
Alder (11)	0.15	0.90	0.13	0.00	0.10	0.01	0.15	0.03	0.03	0.01	7.29	0.04	0.06	0.03	0.08	0.00
Pine (12)	0.00	0.28	0.01	0.22	0.00	0.14	0.00	0.00	0.00	0.20	0.03	29.56	0.02	0.44	0.02	0.06
Poplar (13)	0.00	0.16	0.06	0.00	0.05	0.00	0.02	0.01	0.06	0.01	0.02	0.00	0.91	0.00	0.00	0.00
Spruce (14)	0.01	0.00	0.00	0.08	0.00	0.90	0.00	0.00	0.00	0.03	0.00	0.51	0.00	7.43	0.00	0.08
Black l. (15)	0.00	0.07	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.01	0.00	0.66	0.00
Dwarf p. (16)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.68
Prop (ref)	8.98	10.42	11.82	1.02	2.70	6.89	0.97	0.97	1.03	4.19	8.58	30.57	1.91	8.20	0.93	0.82
Prop (map)	8.88	8.60	15.09	1.05	1.96	5.93	1.18	0.54	0.43	4.53	9.01	30.97	1.29	9.04	0.80	0.69
PA	83.6	68.5	91.1	60.3	41.4	79.9	47.0	43.0	38.3	85.3	85.0	96.7	47.4	90.6	70.8	83.6
UA	84.6	83.0	71.4	58.2	57.0	92.8	38.5	77.7	91.5	78.8	80.9	95.4	70.3	82.2	82.9	99.2

3.3 Limitations in large-area species mapping and proposed solutions

In the country-wide or other large-extent mapping cases, there are several challenges and limitations. Larger regions are often characterized by higher diversity of species and environmental conditions. Certain species occur only in spatially limited areas – for example, in Poland, Silver fir is typical for the mountain areas only, while oaks and hornbeams tend to occur more often in the lowlands. In addition, due to the variability in meteorological conditions, the optimal period for classification of specific

species may differ largely among regions, particularly during the spring when processes of leaf unfolding take place, and autumn while leaf coloring occur. Furthermore, these optimal periods may vary from year-to-year due to variations in spring temperatures and other meteorological conditions (Grabska-Szwagrzyk and Tymińska-Czabańska, 2023). Future research should also consider specific periods of imagery acquisition when aimed to distinguish different species, i.e. covering periods when particular species exhibit the highest phenological variations. It would be profitable to use multiple autumn (e.g. early and late autumn) STMs, however it is very challenging due to insufficient number of clear observations during this time of the year.

One solution may be the division of the study area into smaller regions – in the country-wide or other large-extent mapping of species composition, the subdivision to smaller parts may play an important role, also due to computational power; similarly as in the study from Pazúr et al. (2022) or Hermosilla et al. (2022). However, another question arises how to define the optimal borders of smaller regions to achieve higher accuracy of the obtained map, which is rarely discussed in studies focused on remote sensing-based classification.

Another methodological challenge is the underrepresentation of clear observations in some regions. In this study, we employed short-period seasonal STMs from Sentinel-2 time series rather than one seasonal mean, as the information from specific periods of growing season is crucial in distinguishing species. In calculation of seasonal means, multi-annual observations were used, still, for some regions the underrepresentation of clear observations occurs. It may have significant impact on map accuracy in regions of lower observation frequency. In the case of Poland, it is particularly observed in the places where two orbits do not overlap, specifically for autumn (Figure 3). This issue should be addressed in studies on species classification for larger regions using Sentinel-2 or similar satellite constellations.

As a result of abovementioned factors, the design of robust training, test and validation datasets is challenging. Finally, in certain regions such as privately-owned forests or lands not officially reported as forests (e.g., successional forests that have emerged on previously abandoned agricultural lands), there is no reference data available. These areas tend to exhibit greater complexity, making the task of assessing classification accuracy particularly demanding.

4. Conclusions

We have obtained the first national-scale forest tree species map for Poland, achieving an accuracy exceeding 80%. This was accomplished through a novel approach that involved the calculation of Sentinel-2 seasonal STMs spanning multiple years. The resulting map is an important dataset for both forest management and the scientific community, facilitating tasks like modeling biodiversity and monitoring non-native and invasive species. It can enhance our understanding of forest ecosystems and support more informed and precise forestry and conservation effort. Unlike other existing data sources, such as the FDB, which primarily provide information about the share of species within forest stands, this new map offers a view of tree species

distribution at a finer scale. Furthermore, our map provides a unique advantage over traditional forest inventories like NFI, which offers point-based data rather than continuous spatial representation of species distribution.

5. Data availability

We provide freely accessible resources including the forest tree species map, training and validation data: https://doi.org/10.5281/zenodo.10180469 (Grabska-Szwagrzyk, 2023). The map can be explored online: https://ee-aweaksbarg.projects.earthengine.app/view/speciesmappl

Competing interests

The contact author has declared that none of the authors has any competing interests.

References

Achanta, R. and Süsstrunk, S.: Superpixels and polygons using simple non-iterative clustering, Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, 2017-Janua, 4895–4904, https://doi.org/10.1109/CVPR.2017.520, 2017.

Ahlswede, S., Schulz, C., Gava, C., Helber, P., Bischke, B., Förster, M., Arias, F., Hees, J., Demir, B., and Kleinschmit, B.: TreeSatAI Benchmark Archive : A multi-sensor , multi-label dataset for tree species classification in remote sensing, 1–22, 2022.

Axelsson, A., Lindberg, E., Reese, H., and Olsson, H.: Tree species classification using Sentinel-2 imagery and Bayesian inference, Int. J. Appl. Earth Obs. Geoinf., 100, 102318, https://doi.org/10.1016/j.jag.2021.102318, 2021.

Bałazy, R.: Forest dieback process in the Polish mountains in the past and nowadays – literature review on selected topics, Folia For. Pol. Ser. A, 62, 184–198, https://doi.org/10.2478/ffp-2020-0018, 2020.

Belgiu, M. and Drăgu, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS J. Photogramm. Remote Sens., 114, 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

Biuro Urządzania Lasu i Geodezji Leśnej: Wielkoobszarowa inwentaryzacja stanu lasów. Wyniki za okres 2017-2021, 2022.

Boisvenue, C. and White, J. C.: Information needs of next-generation forest carbon models: Opportunities for remote sensing science, Remote Sens., 11, https://doi.org/10.3390/rs11040463, 2019.

Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., and Bonnet, S.: Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery, 22, 2018.

Bonannella, C., Hengl, T., Heisig, J., Parente, L., Wright, M. N., Herold, M., and de Bruin, S.: Forest tree species distribution for Europe 2000-2020: mapping potential and realized distributions using spatiotemporal Machine Learning, Preprint, https://doi.org/10.7717/peerj.13728, 2022.

Breidenbach, J., Waser, L. T., Debella-Gilo, M., Schumacher, J., Rahlf, J., Hauglin, M., Puliti, S., and Astrup, R.: National mapping and estimation of forest area by dominant tree species using Sentinel-2 data, Can. J. For. Res., 51, 365–379,

https://doi.org/10.1139/cjfr-2020-0170, 2021.

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

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., and Tait, A. M.: Dynamic World, Near real-time global 10 m land use land cover mapping, Sci. Data, 9, 1–17, https://doi.org/10.1038/s41597-022-01307-4, 2022.

Chen, B., Xiao, X., Li, X., Pan, L., Doughty, R., Ma, J., Dong, J., Qin, Y., Zhao, B., Wu, Z., Sun, R., Lan, G., Xie, G., Clinton, N., and Giri, C.: A mangrove forest map of China in 2015: Analysis of time series Landsat 7/8 and Sentinel-1A imagery in Google Earth Engine cloud computing platform, ISPRS J. Photogramm. Remote Sens., 131, 104–120, https://doi.org/10.1016/j.isprsjprs.2017.07.011, 2017.

Fassnacht, F. E., White, J. C., Wulder, M. A., and Næsset, E.: Remote sensing in forestry : current challenges , considerations and directions, 1–27, 2023.

Forstmaier, A., Shekhar, A., and Chen, J.: Mapping of Eucalyptus in Natura 2000 areas using Sentinel 2 imagery and artificial neural networks, Remote Sens., 12, 4–6, https://doi.org/10.3390/rs12142176, 2020.

Gillis, M. D., Omule, A. Y., and Brierley, T.: Monitoring Canada's forests: The national forest inventory, For. Chron., 81, 214–221, https://doi.org/10.5558/tfc81214-2, 2005.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sens. Environ., 202, 18–27, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Grabska-Szwagrzyk, E.: National-scale tree species/genera map for Poland from Sentinel-2 time series, https://doi.org/10.5281/zenodo.10180469, 2023.

Grabska-Szwagrzyk, E. and Tymińska-Czabańska, L.: Sentinel-2 time series : a promising tool in monitoring temperate species spring phenology, Forestry, 1–15, 2023.

Grabska, E., Hostert, P., Pflugmacher, D., and Ostapowicz, K.: Forest stand species mapping using the sentinel-2 time series, Remote Sens., 11, 1–24, https://doi.org/10.3390/rs11101197, 2019.

Grabska, E., Frantz, D., and Ostapowicz, K.: Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians, Remote Sens. Environ., 251, 112103, https://doi.org/10.1016/j.rse.2020.112103, 2020.

Grodzki, W.: The decline of Norway spruce, Beskydy, 3, 19–26, 2010.

Hanewinkel, M., Cullmann, D. A., Schelhaas, M. J., Nabuurs, G. J., and Zimmermann, N. E.: Climate change may cause severe loss in the economic value of European forest land, Nat. Clim. Chang., 3, 203–207, https://doi.org/10.1038/nclimate1687, 2013.

Hemmerling, J., Pflugmacher, D., and Hostert, P.: 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, 2021.

Hermosilla, T., Bastyr, A., Coops, N. C., White, J. C., and Wulder, M. A.: Mapping the presence and distribution of tree

species in Canada's forested ecosystems, Remote Sens. Environ., 282, 113276, https://doi.org/10.1016/j.rse.2022.113276, 2022.

Hościło, A. and Lewandowska, A.: Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data, Remote Sens., 11, 929, https://doi.org/10.3390/rs11080929, 2019.

Hynynen, J., Niemistö, P., Viherä-Aarnio, A., Brunner, A., Hein, S., and Velling, P.: Silviculture of birch (Betula pendula Roth and Betula pubescens Ehrh.) in Northern Europe, Forestry, 83, 103–119, https://doi.org/10.1093/forestry/cpp035, 2010.

Immitzer, M., Vuolo, F., and Atzberger, C.: First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe, Remote Sens., 8, 166, https://doi.org/10.3390/rs8030166, 2016.

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., and Atzberger, C.: Optimal input features for tree species classification in Central Europe based on multi-temporal Sentinel-2 data, Remote Sens., 11, https://doi.org/10.3390/rs11222599, 2019.

Jackson, C. M. and Adam, E.: Machine learning classification of endangered tree species in a tropical submontane forest using worldview-2 multispectral satellite imagery and imbalanced dataset, Remote Sens., 13, https://doi.org/10.3390/rs13244970, 2021.

Kamińska, A., Lisiewicz, M., Kraszewski, B., and Stereńczak, K.: Mass outbreaks and factors related to the spatial dynamics of spruce bark beetle (Ips typographus) dieback considering diverse management regimes in the Białowieża forest, For. Ecol. Manage., 498, https://doi.org/10.1016/j.foreco.2021.119530, 2021.

Karasiak, N., Sheeren, D., Fauvel, M., Willm, J., Dejoux, J.-F., and Monteil, C.: Mapping Tree Species of Forests in Southwest France using Sentinel-2 Image Time Series, 1–4, 2017.

Kolecka, N., Kozak, J., Kaim, D., Dobosz, M., Ostafin, K., Ostapowicz, K., Wężyk, P., and Price, B.: Understanding farmland abandonment in the Polish Carpathians, Appl. Geogr., 88, 62–72, https://doi.org/10.1016/j.apgeog.2017.09.002, 2017.

Kollert, A., Bremer, M., Löw, M., and Rutzinger, M.: Exploring the potential of land surface phenology and seasonal cloud free composites of one year of Sentinel-2 imagery for tree species mapping in a mountainous region, Int. J. Appl. Earth Obs. Geoinf., 94, 102208, https://doi.org/10.1016/j.jag.2020.102208, 2021.

Lechner, M., Dostálová, A., Hollaus, M., Atzberger, C., and Immitzer, M.: Combination of Sentinel-1 and Sentinel-2 Data for Tree Species Classification in a Central European Biosphere Reserve, Remote Sens., 14, 1–16, https://doi.org/10.3390/rs14112687, 2022.

Loiselle, B. A., Howell, C. A., Graham, C. H., Goerck, J. M., Brooks, T., Smith, K. G., and Williams, P. H.: Avoiding Pitfalls of Using Species Distribution Models in Conservation Planning, Conserv. Biol., 17, 1591–1600, https://doi.org/10.1111/j.1523-1739.2003.00233.x, 2003.

Majchrowska, A.: Abandonment of agricultural land in central Poland and its ecological role, Ekol. Bratislava, 32, 320–327, https://doi.org/10.2478/eko-2013-0028, 2013.

Marconi, S., Weinstein, B. G., Zou, S., Bohlman, S. A., Zare, A., Singh, A., Stewart, D., Harmon, I., Steinkraus, A., and White, E. P.: Continental-scale hyperspectral tree species classification in the United States National Ecological Observatory Network,

Remote Sens. Environ., 282, 113264, https://doi.org/10.1016/j.rse.2022.113264, 2022.

Maxwell, A. E., Warner, T. A., and Fang, F.: Implementation of machine-learning classification in remote sensing: An applied review, Int. J. Remote Sens., 39, 2784–2817, https://doi.org/10.1080/01431161.2018.1433343, 2018.

Melnyk, O., Manko, P., and Brunn, A.: Remote sensing methods for estimating tree species of forests in the Volyn region, Ukraine, Front. For. Glob. Chang., 6, https://doi.org/10.3389/ffgc.2023.1041882, 2023.

Müller, H., Rufin, P., Griffiths, P., Barros Siqueira, A. J., and Hostert, P.: Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape, Remote Sens. Environ., 156, 490–499, https://doi.org/10.1016/j.rse.2014.10.014, 2015.

Nasiri, V., Beloiu, M., Asghar, A., Griess, V. C., Maftei, C., and Waser, L. T.: International Journal of Applied Earth Observations and Geoinformation Mapping tree species composition in a Caspian temperate mixed forest based on spectral-temporal metrics and machine learning, Int. J. Appl. Earth Obs. Geoinf., 116, 103154, https://doi.org/10.1016/j.jag.2022.103154, 2023.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sens. Environ., 148, 42–57, https://doi.org/10.1016/j.rse.2014.02.015, 2014.

Pazúr, R., Huber, N., Weber, D., Ginzler, C., and Price, B.: A national extent map of cropland and grassland for Switzerland based on Sentinel-2 data, Earth Syst. Sci. Data, 14, 295–305, https://doi.org/10.5194/essd-14-295-2022, 2022.

Persson, M., Lindberg, E., and Reese, H.: Tree Species Classification with Multi-Temporal Sentinel-2 Data, Remote Sens., 10, 1794, https://doi.org/10.3390/rs10111794, 2018.

Praticò, S., Solano, F., Di Fazio, S., and Modica, G.: Machine learning classification of mediterranean forest habitats in google earth engine based on seasonal sentinel-2 time-series and input image composition optimisation, Remote Sens., 13, 1–28, https://doi.org/10.3390/rs13040586, 2021.

Puletti, N., Chianucci, F., and Castaldi, C.: Use of Sentinel-2 for forest classification in Mediterranean environments, Ann. Silvic. Res., 0, 1–7, https://doi.org/10.12899/asr-1463, 2017.

Punalekar, S. M., Planque, C., Lucas, R. M., Evans, D., Correia, V., Owers, C. J., Poslajko, P., Bunting, P., and Chognard, S.: National scale mapping of larch plantations for Wales using the Sentinel-2 data archive, For. Ecol. Manage., 501, 119679, https://doi.org/10.1016/j.foreco.2021.119679, 2021.

Richardson, D. M. and Rejmánek, M.: Trees and shrubs as invasive alien species - a global review, Divers. Distrib., 17, 788–809, https://doi.org/10.1111/j.1472-4642.2011.00782.x, 2011.

Rüetschi, M., Weber, D., Koch, T. L., Waser, L. T., Small, D., and Ginzler, C.: Countrywide mapping of shrub forest using multi-sensor data and bias correction techniques, Int. J. Appl. Earth Obs. Geoinf., 105, 102613, https://doi.org/10.1016/j.jag.2021.102613, 2021.

Rusňák, T., Halabuk, A., Halada, L., Hilbert, H., and Gerhátová, K.: Detection of Invasive Black Locust (Robinia pseudoacacia) in Small Woody Features Using Spatiotemporal Compositing of Sentinel-2 Data, Remote Sens., 14,

https://doi.org/10.3390/rs14040971, 2022.

Schindler, J., Dymond, J. R., Wiser, S. K., and Shepherd, J. D.: Method for national mapping spatial extent of southern beech forest using temporal spectral signatures, Int. J. Appl. Earth Obs. Geoinf., 102, 102408, https://doi.org/10.1016/j.jag.2021.102408, 2021.

Shahbandeh, M., Kaim, D., and Kozak, J.: The Substantial Increase of Forest Cover in Central Poland Following Extensive Land Abandonment: Szydłowiec County Case Study, Remote Sens., 14, https://doi.org/10.3390/rs14163852, 2022.

Shirazinejad, G., Javad Valadan Zoej, M., and Latifi, H.: Applying multidate Sentinel-2 data for forest-type classification in complex broadleaf forest stands, Forestry, 95, 363–379, https://doi.org/10.1093/forestry/cpac001, 2022.

Somodi, I., Čarni, A., Ribeiro, D., and Podobnikar, T.: Recognition of the invasive species Robinia pseudacacia from combined remote sensing and GIS sources, Biol. Conserv., 150, 59–67, https://doi.org/10.1016/j.biocon.2012.02.014, 2012.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., and Brisco, B.: Google Earth Engine for geo-big data applications: A meta-analysis and systematic review, ISPRS J. Photogramm. Remote Sens., 164, 152–170, https://doi.org/10.1016/j.isprsjprs.2020.04.001, 2020.

Tomaś, Ł. and Jagodziński, A. M.: Przebudowa drzewostanów, Mag. Pol. Akad. Nauk, 3–4, 94–97, 2019.

Turlej, K., Ozdogan, M., and Radeloff, V. C.: Mapping forest types over large areas with Landsat imagery partially affected by clouds and SLC gaps, Int. J. Appl. Earth Obs. Geoinf., 107, 102689, https://doi.org/10.1016/j.jag.2022.102689, 2022.

Walthert, L. and Meier, E. S.: Tree species distribution in temperate forests is more influenced by soil than by climate, Ecol. Evol., 7, 9473–9484, https://doi.org/10.1002/ece3.3436, 2017.

Waser, L. T., Rüetschi, M., Psomas, A., Small, D., and Rehush, N.: Mapping dominant leaf type based on combined Sentinel-1/-2 data – Challenges for mountainous countries, ISPRS J. Photogramm. Remote Sens., 180, 209–226, https://doi.org/10.1016/j.isprsjprs.2021.08.017, 2021.

Welle, T., Aschenbrenner, L., Kuonath, K., Kirmaier, S., and Franke, J.: Mapping Dominant Tree Species of German Forests, Remote Sens., 14, https://doi.org/10.3390/rs14143330, 2022.

Wessel, M., Brandmeier, M., and Tiede, D.: Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data, Remote Sens., 10, https://doi.org/10.3390/rs10091419, 2018.

Ye, N., Morgenroth, J., Xu, C., and Chen, N.: Indigenous forest classification in New Zealand – A comparison of classifiers and sensors, Int. J. Appl. Earth Obs. Geoinf., 102, 102395, https://doi.org/10.1016/j.jag.2021.102395, 2021.

Zajączkowski, G., Jabłoński, M., Jabłoński, T., Sikora, K., Kowalska, A., Małachowska, J., and Piwnicki, J.: Raport o stanie lasów w Polsce 2021, 1–23 pp., 2022.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., and Arino, O.: ESA WorldCover 10 m 2021 v200, https://doi.org/doi:10.5281/zenodo.7254221, 2022.

Zeug, G., Geltendorf, T., Immitzer, M., and Atzberger, C.: Machbarkeitsstudie zur Nutzung von Satellitenfernerkundungsdaten (Copernicus) für Zwecke der Ableitung ökologischer Belastungsgrenzen und der Verifizierung von Indikatoren der Deutschen

Anpassungsstrategie an den Klimawandel, 2018.

Zgłobicki, W., Karczmarczuk, K., and Baran-Zgłobicka, B.: Intensity and driving forces of land abandonment in eastern Poland, Appl. Sci., 10, https://doi.org/10.3390/app10103500, 2020.

Appendix

Table A1: Variables used for classification**.**

Group	Predictor
Early spring	Blue
	Green
	Red
	RE1
	RE2
	RE3
	NIR1
	NIR2
	SWIR1
	SWIR2
Late spring	Blue
	Green
	Red
	RE1
	RE2
	RE3
	NIR1
	NIR2
	SWIR1
	SWIR2
Summer	Blue
	Green
	Red
	RE1
	RE2
	RE3
	NIR1
	NIR2

	SWIR1
	SWIR2
Autumn	Blue
	Green
	Red
	RE1
	RE2
	RE3
	NIR1
	NIR2
	SWIR1
	SWIR2
Topography	Elevation
Climate	Annual mean temperature (bio01)
	Annual precipitation (bio12)
	Precipitation of driest quarter (bio17)
	Maximum temperature in spring 2018
Soils	Soils

Table A2: Number of test pixels for accuracy assessment in two approaches – proportional and not proportional.

	Estimated proportions	**Proportional**	**Not proportional**
Pinus	59%	11 800	5900
Quercus	8%	1600	2400
Betula	6.8%	1360	2040
Fagus	6.2%	1240	1860
Alnus	5.7%	1140	1710
Picea	5.3%	1060	1600
Abies	3.3%	660	1320
Larix	2%	400	800
Carpinus	1.3%	260	520
Populus	1%	200	400
Fraxinus	<1%	100	200
Pseudotsuga	<1%	100	200
Acer	<1%	100	200
Robinia pseudoacaccia	<1%	100	200

Tilia	<1%	100	200
Pinus mugo	<1%	100	200

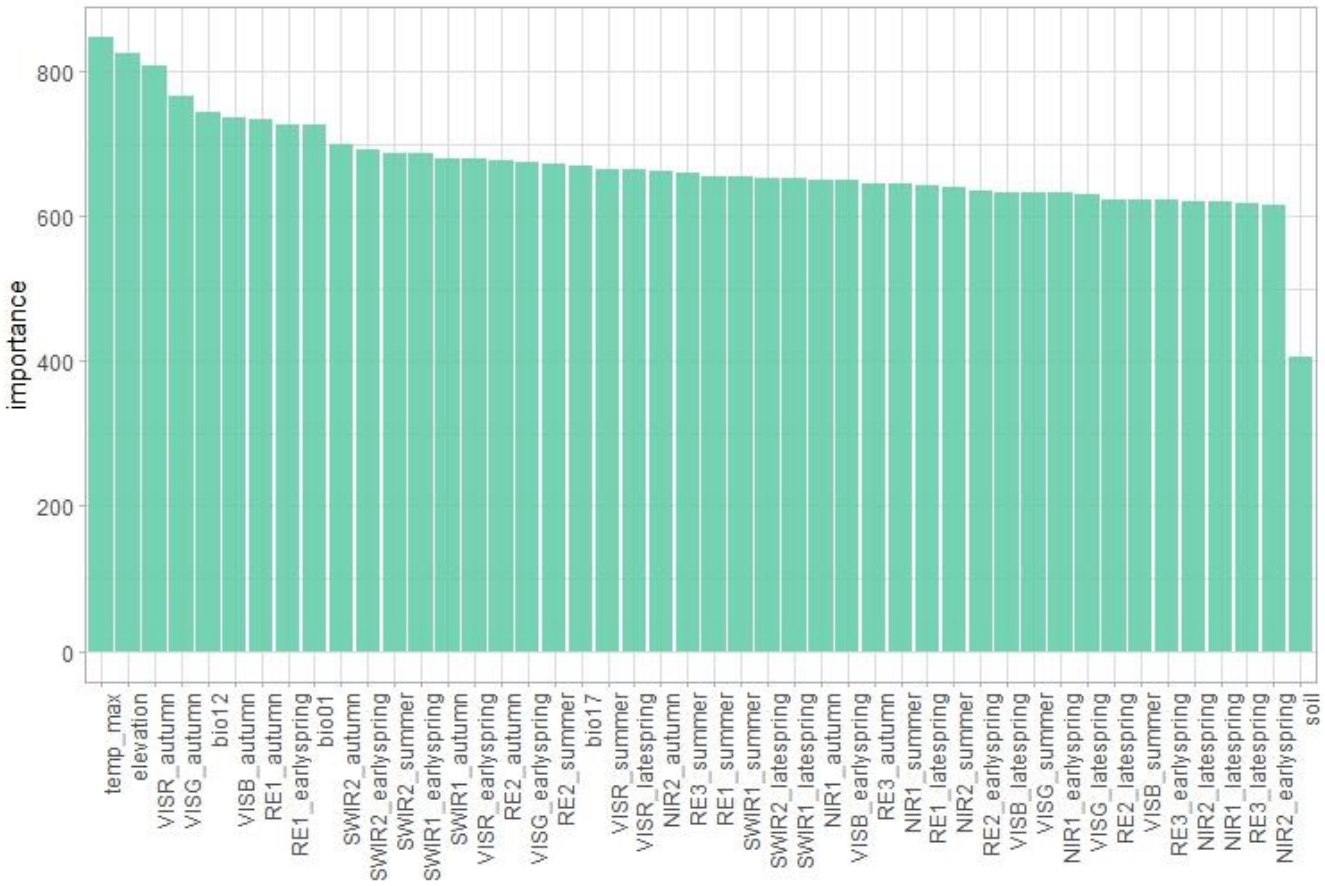


Figure A1. Importance of variables used in classification.