

Multitemporal hyperspectral tree species classification in the Białowieża Forest World Heritage site

Aneta Modzelewska¹, Agnieszka Kamińska¹, Fabian Ewald Fassnacht² and Krzysztof Stereńczak¹

¹Department of Geomatics, Forest Research Institute, Sękocin Stary, Braci Leśnej 3 Street, 05-090 Raszyn, Poland

²Department of Geography and Geoecology, Karlsruhe Institute of Technology, Kaiserstraße 12 76131 Karlsruhe, Germany

*Corresponding author: Tel: +48 22 7150 327; E-mail: a.modzelewska@ibles.waw.pl

Received 28 July 2020

Tree species composition maps derived from hyperspectral data have been found to be accurate but it is still unclear whether an optimal time window exists to acquire the images. Trees in temperate forests are subject to phenological changes that are species-specific and can have an impact on species recognition. Our study examined the performance of a multitemporal hyperspectral dataset to classify tree species in the Polish part of the Białowieża Forest. We classified seven tree species including spruce (*Picea abies* (L.) H.Karst), pine (*Pinus sylvestris* L.), alder (*Alnus glutinosa* Gaertn.), oak (*Quercus robur* L.), birch (*Betula pendula* Roth), hornbeam (*Carpinus betulus* L.) and linden (*Tilia cordata* Mill.), using Support Vector Machines. We compared the results for three data acquisitions—early and late summer (2–4 July and 24–27 August), and autumn (1–2 October) as well as a classification based on an image stack containing all three acquisitions. Furthermore, the sizes (height and crown diameter) of misclassified and correctly classified trees of the same species were compared. The early summer acquisition reached the highest accuracies with an Overall Accuracy (OA) of 83–94 per cent and Kappa (κ) of 0.80–0.92. The classification based on the stacked multitemporal dataset resulted in slightly higher accuracies (84–94 per cent OA and 0.81–0.92 κ). For some species, e.g. birch and oak, tree size differed notably for correctly and incorrectly classified trees. We conclude that implementing multitemporal hyperspectral data can improve the classification result as compared with a single acquisition. However, the obtained accuracy of the multitemporal image stack was in our case comparable to the best single-date classification and investing more time in identifying regionally optimal acquisition windows may be worthwhile as long hyperspectral acquisitions are still sparse.

Introduction

Information on forest tree species composition is relevant for forest management and nature conservation (Fassnacht *et al.*, 2016). Hyperspectral data are an efficient remotely sensed data source to map tree species (e.g. Ghosh *et al.*, 2014; Trier *et al.*, 2018). Most earlier studies using hyperspectral data to classify tree species focused on a single acquisition, typically acquired during the regional peak of the vegetation season (the period after spring green-up, Kern *et al.*, 2020, but before leaves started ageing). However, in temperate forests, seasonal changes influence both tree spectra and data quality (e.g. due to different amounts of shadow and varying illumination conditions). Trees' phenology, including spring leaf-out, flowering and senescence are species-specific processes that can influence species recognition. Hence, seasonal impacts should be considered when planning flight campaigns. However, the question whether an optimal time window to acquire hyperspectral data exists is still unclear and may also differ regionally. Trees' reflective properties change during the growing season. This is mostly visible in

changes related to leaves and their colour: these changes are predominantly driven by the leaf pigment composition which varies during the year following temperature and photoperiod changes (Delpierre *et al.*, 2016; Keskitalo *et al.*, 2005). In the temperate zone of the Northern hemisphere, trees' annual cycle consists of winter dormancy, spring budburst and leaf out or needle elongation, florescence, fruit and cone maturation and, for deciduous trees autumn tinting and leaf abscission, for coniferous—preparation for dormancy (Delpierre *et al.*, 2016).

In autumn, leaves start the complex process of senescence, which consists of biochemical, physiological and metabolic changes (Kim *et al.*, 2016). It starts with leaf tinting caused by the degradation of chlorophyll (Delpierre *et al.*, 2016). Differences in the timing of leaf senescence and drop out may be large for tree species. Most of the temperate lowland tree species start senescence in the second part of October, e.g. linden (*Tilia cordata* Mill.), oak (*Quercus robur* L.), birch (*Betula pendula* Roth) or hornbeam (*Carpinus betulus* L.), while alder (*Alnus glutinosa* (L.) Gaertn.) only colours in the second part of November and

December (Panchen *et al.*, 2015). Leaf senescence is related to the leaves' age; however, environmental factors or stresses may advance the process (Kim *et al.*, 2016).

Some earlier studies examined the influence of phenology on tree species recognition. Wolter *et al.* (1995) and Mickelson *et al.* (1998) applied Landsat time series from different seasons and compared the results of classifications for single-date images with the results for stacked multitemporal ones. Wolter *et al.* (1995) obtained a classification accuracy of 80.1 per cent for the layered multitemporal image, whereas it was not possible to classify dominant species of the research area using single date images. In the work of Mickelson *et al.* (1998), spring and autumn data were found potentially more useful to classify forest types than the summer ones, however, combining the three datasets resulted in the best accuracy (OA: 78.9 per cent). Both studies compared images from different years. The authors consider different acquisition years as a disadvantage and assume that intra-annual data could give better results (Wolter *et al.*, 1995). Intra-annual datasets were analysed in Key *et al.* (2001) and Hill *et al.* (2010) who used multispectral airborne data to distinguish several temperate deciduous species. Key *et al.* (2001) acquired CIR and RGB images nine times in 1997 (from May to October) and Hill *et al.* (2010)—five times in 2003 (from March to October). Both studies found the image from October (with accuracy exceeding 70 per cent) as the best due to the peak of autumn colours. Hill *et al.*, (2010) prove in their work that combining multitemporal datasets increases the ability to map temperate deciduous tree species (combining three images gave 84 per cent accuracy, around 10 per cent higher than the best single one). In contrary, Key *et al.*, (2001) did not report improved performances when combining images (selecting one single classification during the optimal period gave better results than combined datasets) but claimed that if it is impossible to acquire data at the optimal time of the year, applying multitemporal images may be a solution.

Most studies comparing tree species classifications from data collected during several time points in a year used multispectral data (Hill *et al.*, 2010; Key *et al.*, 2001). Comparisons based on multiple hyperspectral datasets are rare. Nevertheless, some recent studies address this topic (e.g. Richter *et al.*, 2016; Tagliabue *et al.*, 2016; Voss and Sugumaran, 2008). All but one of these studies used data acquired in different years. Richter *et al.* (2016) and Voss and Sugumaran (2008) applied two hyperspectral datasets acquired in summer and autumn with a 2 years gap. When comparing datasets, Richter *et al.* (2016) obtained slightly better results for the summer dataset, but Voss and Sugumaran (2008) did not find notable differences (overall accuracy (OA) equalled 56 per cent in autumn and 57 per cent in summer). Meaningful improvement was observed when the classification was accomplished using the two stacked datasets (Richter *et al.*, 2016). Forzieri *et al.* (2012) applied datasets acquired in winter and autumn with a 1-year time gap. They performed a land cover classification including seven classes of tree species. October data were found slightly more useful than the winter ones (87 per cent compared to 83 per cent). A study by Tagliabue *et al.* (2016) was the only one to use an intra-annual dataset, acquired in spring and autumn. They classified five tree species: hornbeam, linden, pine, sessile and pedunculate oak. The classification performed on spring data gave a higher accuracy (59.5

per cent) than autumn (48 per cent), but clearly, the highest results were obtained for the combined dataset: 74.4 per cent (Tagliabue *et al.*, 2016). Generally, summer data were found to give slightly better results than the autumn ones, when hyperspectral data were applied (e.g. Richter *et al.*, 2016; Tagliabue *et al.*, 2016), but spring or autumn data were found more efficient when multispectral data were used (Hill *et al.*, 2010; Key *et al.*, 2001).

In addition to the influence of phenology, other properties may affect the ability of a classification workflow to classify individual trees to the correct species. In the current state-of-the-art, very few studies investigated the properties of misclassified trees. One of the few exceptions is the study of Leckie *et al.* (2005) who examined the classification performance in old-growth stands, taking into account shadowed, unhealthy and odd trees. In all of these groups, higher error rates were observed as compared with the other classes of trees (Leckie *et al.*, 2005). Korpela *et al.* (2014) assumed that misclassification of individual trees may be traced back to erroneously calibrated reflectance or directional reflectance anisotropy. However, they found that up to 70 per cent of reflectance variance in the near-infrared is caused directly by the tree crown structures of the individual trees. It can furthermore be assumed that smaller trees are generally harder to classify as their species-specific spectral signal may be altered by neighbouring tree crowns or by the background signal from understorey vegetation and soil. This problem is likely to be more pronounced at a coarser spatial grain. Hence, in this study, we examined whether tree height and crown area differ between correctly and incorrectly classified tree individuals. Tree height determines if a tree is visible in the upper canopy captured by the remote sensing images. Crown area determines what fraction of a pixel is occupied by the tree. Some earlier studies investigated the possible influence of tree traits on the classification success (e.g. Ghosh *et al.*, 2014; Jones *et al.*, 2010). Height features were not found helpful in species recognition (Ghosh *et al.*, 2014), however, smaller tree crowns may actually be partly hidden below the taller ones, which can result in misclassification of those. Similarly, crown features are known to be species-dependent (Ørka *et al.*, 2009) and the relation between the size of the average tree and the applied pixel size was found to influence the classification result (Fassnacht *et al.*, 2016). The best accuracies were obtained when the size of a pixel was close to the size of the average crown (Ghosh *et al.*, 2014), which can differ in differently managed forests (Modzelewska *et al.*, 2020). Despite the large numbers of studies focusing on tree species classification and mapping, a numerical assessment of the size of misclassified trees has to our knowledge—not been conducted, yet. However, such analyses are key for an improved understanding of which factors have to be addressed in the future for further improving our workflows to classify tree species.

Furthermore, the majority of tree species classification studies working with multitemporal data, both multi- or hyper-spectral data, compared data from different years, which was considered disadvantageous. Some authors underline the need for intra-annual dataset comparisons. To the best of our knowledge, only three studies (Hill *et al.*, 2010; Key *et al.*, 2001; Tagliabue *et al.*, 2016) investigated intra-annual images with only one of them using intra-annual hyperspectral datasets.

Thus, the aims of this study are: (1) to examine which of the investigated hyperspectral images acquired across 1 year (2015) in three different stages of trees' phenological cycle (July, August, October) is optimal for tree species classification; (2) to examine whether multitemporal hyperspectral datasets improve the classification performance; (3) to examine whether tree height and crown area differ between correctly and incorrectly classified trees.

Materials and methods

Study area

The study area is the Białowieża Forest (BF), in Poland (Figure 1A). The BF is located in Poland and Belarus in the temperate climate zone, with Atlantic and continental influences. Around one-third of the BF lays in Poland (Figure 1B). It is a vast and highly heterogeneous, partially protected forest area. The BF consists of a mosaic of forest communities including coniferous and mixed coniferous forests, wet deciduous forests, rich mesic deciduous stands and early successional stands with birch (*Betula* spp.) and aspen (*Populus tremula* L.) (Jędrzejewska and Jędrzejewski, 1998). According to Faliński (1986), the dominant tree species are oaks (*Quercus robur* L. and *Quercus petraea* (Matt.) Liebl.), Norway spruce (*Picea abies* (L.) H.Karst), Scots pine (*Pinus sylvestris* L.), European hornbeam (*Carpinus betulus* L.), followed by Black alder (*Alnus glutinosa* (L.) Gaertn.), small-leaved linden (*Tilia cordata* Mill.), Norway maple (*Acer platanoides* L.), birch (*Betula pendula* Roth and *B. pubescens* Ehrh.) and European ash (*Fraxinus excelsior* L.); however, the latter's share has drastically decreased in recent years due to ash decline (Miścicki, 2016) and the share of Norway maple was below 2 per cent according to recent inventory data and we hence decided to not consider these two species.

Remote sensing data

Hyperspectral data applied in this study were acquired in three flight campaigns held in 2015: the first, on 2–4 July (early summer), the second, on 24–27 August (late summer), the third, on 1–2 October (autumn). All the remote sensing and reference data used in the study were acquired in the framework of the LIFE+ ForBioSensing 'Comprehensive monitoring of stand dynamics in Białowieża Forest supported with remote sensing techniques' project (<http://forbiosensing.eu/>). From each campaign (each consisting of 40 stripes), data of three flight stripes were used as the basis for our study (covering 75.5 km²). We restricted the experiment to this smaller area in order to save processing time and built upon earlier tree species mapping studies which already examined the complete dataset but for monotemporal acquisitions (Modzelewska et al., 2020). The selected stripes are representative for the whole area in terms of species composition and abundant reference information is available for these areas (Figure 1C). Images were acquired using HySpex VNIR-1800 and SWIR-384 cameras designed by Norsk Electro Optics company. At the selected flight height, the HySpex VNIR-1800 takes images in 182 bands with 2.5 m of spatial resolution. Its spectral range covers 0.4–1.0 µm, while HySpex SWIR-384 operates in the spectral range of 1.0–2.5 µm with 5 m spatial resolution in 288 bands (<https://www.hyspex.no/>) (Norsk

Elektro Optikk (n.d.). Similar spatial resolution was previously found sufficient to classify tree species (e.g. Ghosh et al., 2014) and was also a compromise between data quality and acquisition cost. The images from the two cameras were combined into 5 m resolution images in 451 spectral bands by the data provider. Pre-processing, including PARGE geometric correction based on GPS/IMU data and atmospheric correction using the MODTRAN5 model, were executed by the data provider (MGGP Aero Sp. z o.o.).

Airborne Laser Scanning data were acquired using a full-waveform Riegl LMS-Q680i scanner on 2–5 July 2015. The data were collected with a footprint size of a laser beam equal to 0.25 m and a maximum scan angle of $\pm 30^\circ$. The average density of the resulting point cloud is 6 pts./m². ALS data were used to generate a digital terrain model (DTM) and a digital surface model (DSM). On the base of a canopy height model (CHM) derived by subtracting the DTM from the DSM a segmentation of individual tree crowns was carried out (Kamińska et al., 2018; Stereńczak, 2013; Stereńczak et al., 2017). The segmentation method is based on the assumption that larger trees in the upper layer of the stand have larger crowns. Therefore, for the CHM model with a fixed resolution, taller trees were smoothed with a larger kernel window and smaller trees with a smaller kernel window. In total, three groups of trees were defined in terms of their height for coniferous and deciduous species. After the smoothing, a local maxima approach was used to identify the tops of trees and a subsequent region growing approach (starting from the identified treetops, neighbouring cells are continuously examined according to user-defined rules to decide for each pixel whether it belongs to the tree crown or not—the crowns 'grow' until no further pixels fulfil the rules) led to the final tree crown segments. After the crown delineation, some typical errors were corrected with post-processing routines in order to improve the segmentation results. The process is described in detail in previous studies (Stereńczak et al., 2017; Stereńczak et al., 2020). Evaluation of the results shows that the method works well for dominant trees in the sample (Stereńczak et al., 2017). In the case of the classifications based on hyperspectral data conducted here, only dominant trees were considered, since the images cannot provide spectral information of below canopy vegetation. We hence consider the results of the segmentation approach sufficient for the purposes of this study.

The tree segments were only used for extracting the training data and for characterizing the misclassified trees. The classification map was derived at the pixel level.

Data Reduction and Feature Selection

In order to reduce the high dimensionality of the hyperspectral data, a Minimum Noise Fraction (Green et al., 1988) rotation was applied after mosaicking the three hyperspectral stripes (stripes were mosaiced in ENVI with use of its Georeferenced Mosaicking Tool). Based on a visual interpretation of the MNF bands and the ranked eigenvalues, a total of 15 features were deemed coherent and informative for use in the final classification. The typical approach is to first choose MNF bands with eigenvalues higher than 1 as bands with lower ones are usually noisy (Vinceh and Arfania, 2017). Rechecking all bands with eigenvalues higher than 1, we found that the information in

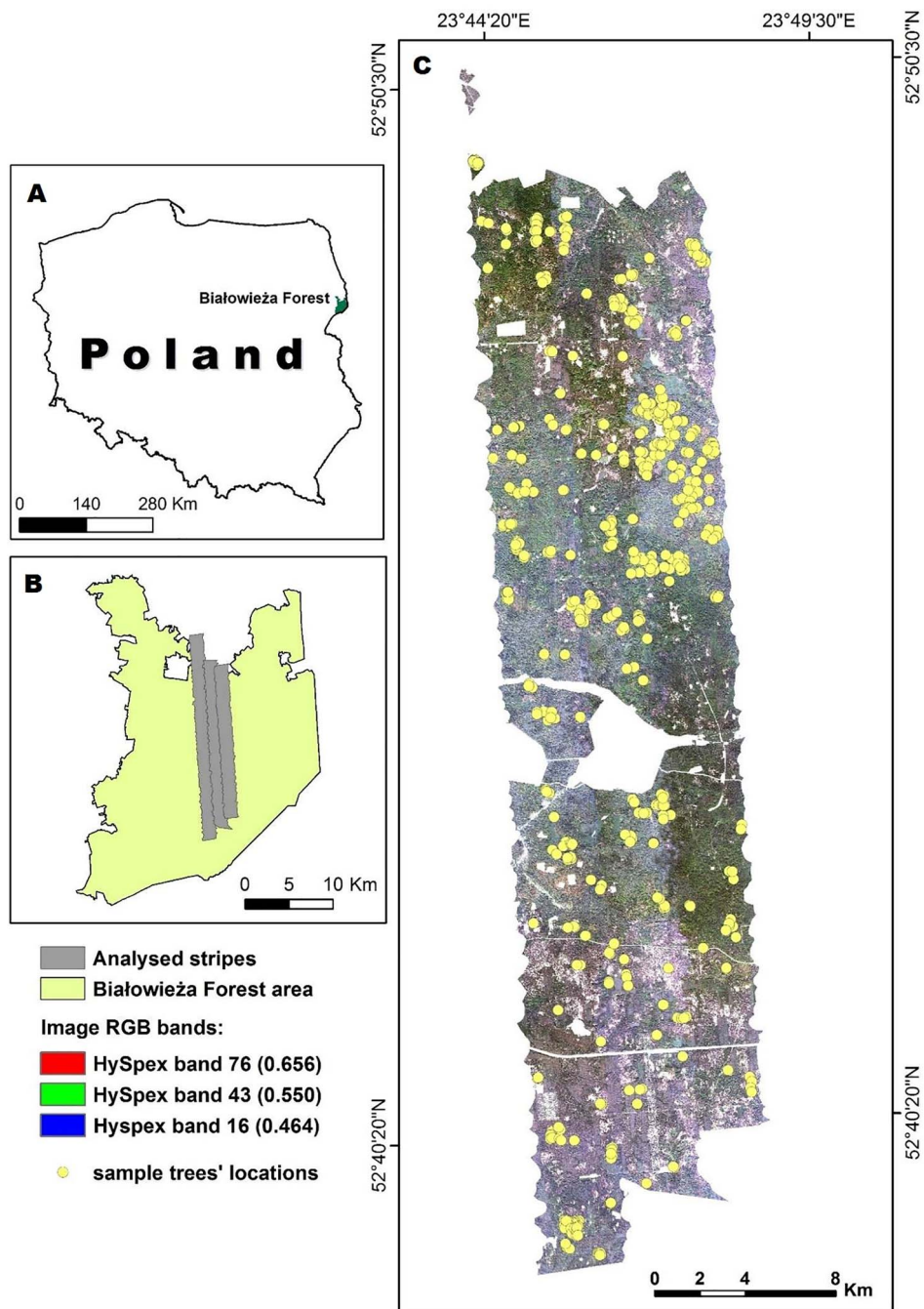


Figure 1 Study area. The location of BF in Poland (A), stripes selected for investigation (B) and RGB composite of hyperspectral data with reference data plotted (C).

some bands was dominated by stripping or illumination effects and those bands were excluded. The final set of selected bands contained the following MNF bands: 1, 3, 5, 6–17 (the same bands for each season).

Reference data

The reference data were acquired during a field campaign conducted between July and September 2015 (simultaneously with

the flight missions). The inventory plots surveyed in the field campaign covered the research area in a regular grid of 500 m. In each 500 m² inventory plot the positions of all trees were recorded and several tree metrics and characteristics were collected (i.a. dbh, height, crown height, species). A visual assessment of whether the tree can possibly be seen from above was also conducted. Based on the locations of the trees assessed as possibly visible in remote sensing images, we visually chose a set of sample pixels for the classification. To ensure the correct identification

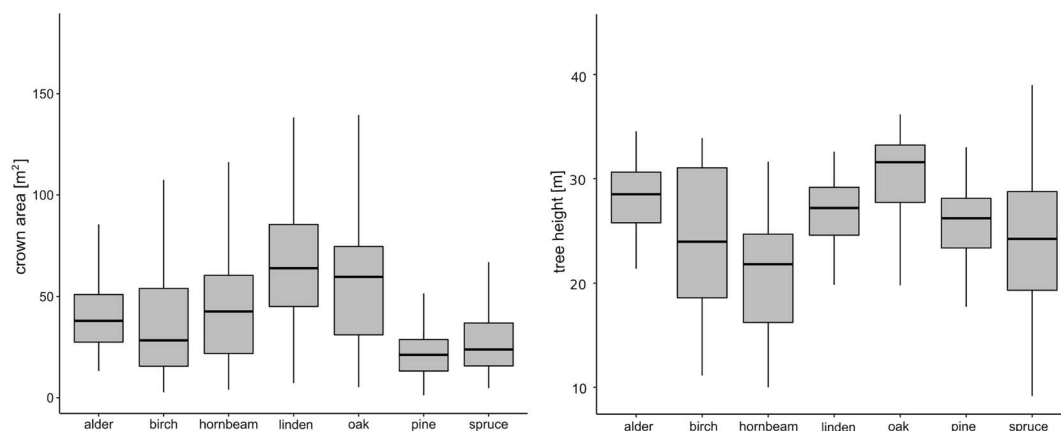


Figure 2 The ranges of analysed tree traits values (crown area, tree height) for each of the tree species presented in box-whisker plots.

of the location of the trees in the hyperspectral images, we additionally used an orthophoto-map (0.5 m spatial resolution). Seven dominant tree species in the BF were considered (the number of samples in parenthesis): birch (*Betula pendula* Roth, 90), oak (*Quercus robur* L., 63), hornbeam (*Carpinus betulus* L., 50), linden (*Tilia cordata* Mill., 94), alder (*Alnus glutinosa* Gaertn., 91), pine (*Pinus sylvestris* L., 101) and spruce (*Picea abies* (L.) H. Karst, 118). These species account for an estimated 98 per cent of all trees in the study area. The number of samples for each tree species corresponds to its overall frequency in the study area (more common species were sampled more frequently).

Classification strategy

The classification was conducted using Support Vector Machines (SVM). A grid search to determine optimal parameters for the SVM was accomplished separately for each of the time periods and the stacked image. We applied a fivefold-cross-validation with 25 repetitions. The classification was conducted in R-project (R Development Core Team, 2015) using the 'caret' package.

The SVM classifier (Vapnik, 1998) is a supervised, non-parametric statistical learning technique, which aims at finding an optimal hyperplane for solving a class separation problem (Mountrakis et al., 2011; Pal and Mather, 2005). The optimal hyperplane is the one that maximizes the distance between the closest training samples and the separating hyperplane (Melgani and Bruzzone, 2004). If a linear separability of the two classes is not possible, the data are mapped into a higher dimensional feature space via a kernel function (kernel-trick), leading to non-linearity in the original feature space (Fassnacht et al., 2014). A radial basis function kernel with parameters: C (representing a trade-off between complexity and the proportion of non-separable samples, thus controlling for over and underfitting in the model) and σ (controlling the shape of the hyperplane) were used in this study. The applied SVM algorithm is implemented in R via the library 'e1071' (Meyer, 2012). We optimized the C and σ parameters via the mentioned grid search for each classified image.

Accuracy assessment

The accuracy of the classifications was assessed using an error matrix and the following indices: OA, Kappa coefficient (κ),

producer's accuracy (PA) and user's accuracy (UA) (Cohen, 1960; Story and Congalton, 1986). To avoid overfitting of the classification model, independent validation was conducted by dividing the available reference data into two sets: one for training the classification model (70 per cent), and the other for testing the performance (30 per cent). The procedure was repeated 100 times and mean classification accuracy indices were obtained. Additionally, the McNemar test was used to determine whether the difference between the two classification results was statistically significant (level of significance $\alpha = 0.05$).

Examination of the misclassified trees

During the iterative classification process, we extracted and investigated trees that were not classified correctly. A tree was classified as 'incorrect' if it was misclassified more than five times in the 100 repetitions (5 per cent). For this task, we investigated segments representing individual trees. Groups of segments representing misclassified and correctly classified trees were compared in terms of their size expressed in two tree traits:

1. tree height (based on CHM data)—defined as the highest CHM pixel value inside the tree segment,
2. crown area derived from CHM (based on single tree detection)—defined as the horizontal area of the crown segment.

Figure 2 summarizes the range of values for the analysed tree traits differentiated according to the examined tree species.

The Mann-Whitney U test was used to determine significant differences between 'correctly classified' and 'incorrectly classified' trees. It was conducted for the examined tree traits (height and crown area) for each season and species.

Results

Classification results

The resulting tree species maps are shown in Figure 3. All classifications produced high accuracies (OA \geq 78 per cent and $\kappa > 0.74$) (Figure 4).

The highest levels of accuracies were obtained for the early summer acquisition (83–94 per cent OA and 0.80–0.92 κ).

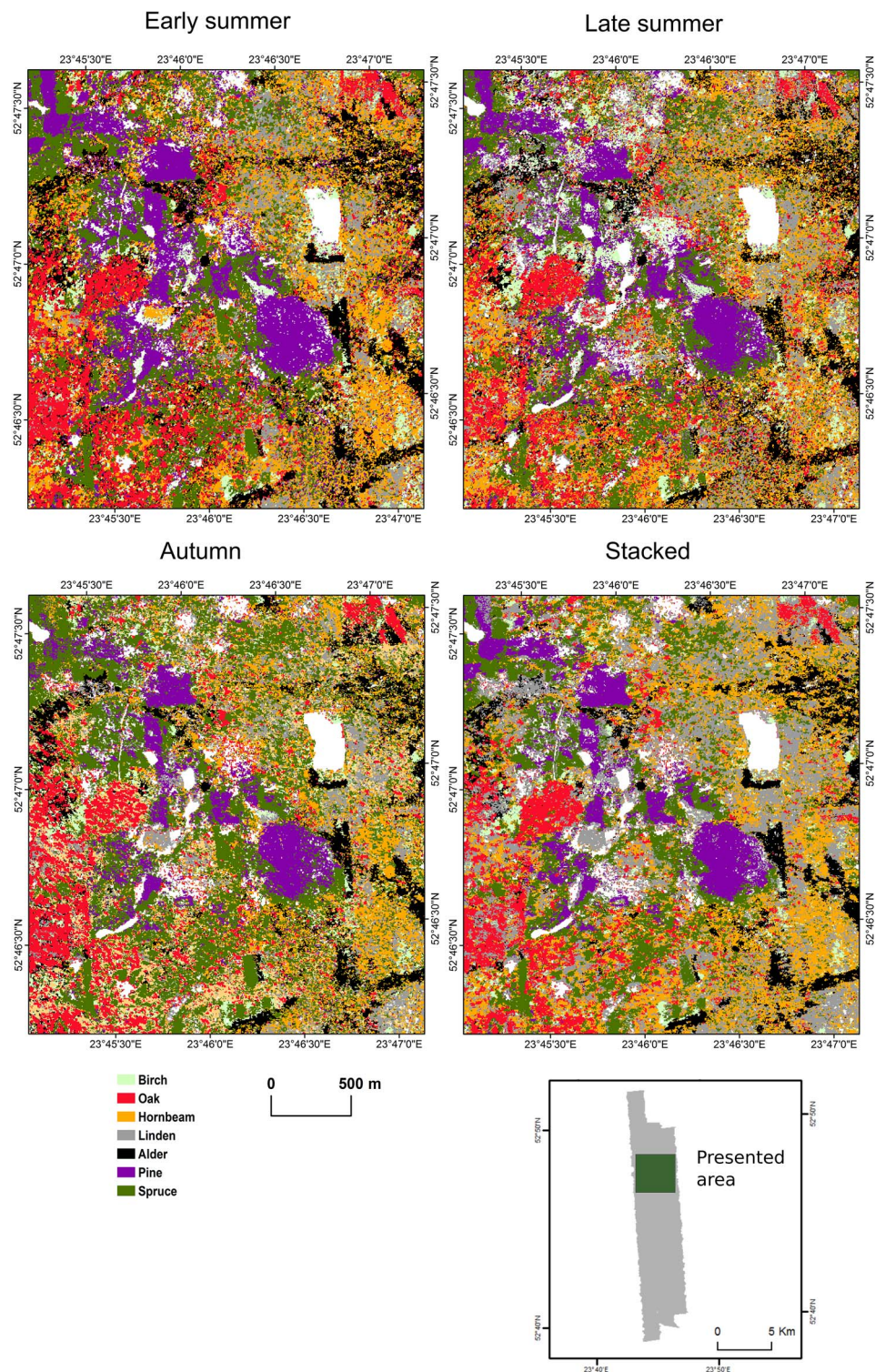


Figure 3 Classification results—tree species maps obtained for particular seasons (subset presented to enable comparison among seasons).

Autumn produced the lowest accuracies (78–89 per cent OA and 0.74–0.87 κ). The multitemporal dataset classification resulted in slightly higher accuracies than the early summer ones (84–94 per cent OA and 0.81–0.92 κ). The McNemar test results indicate

significant differences between early summer and autumn classification results (P -value=0.01). However, there were no significant differences between the summer and the other two acquisitions.

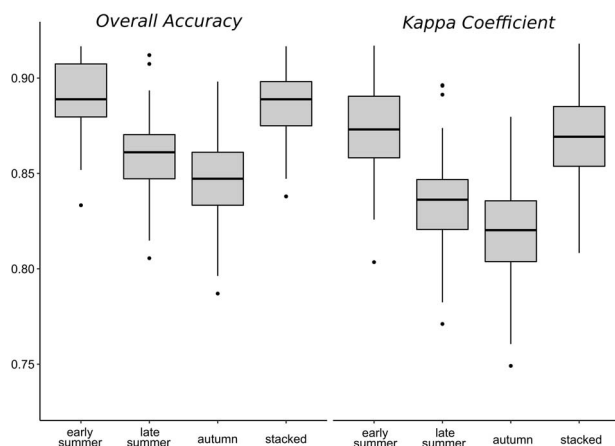


Figure 4 Overall and Kappa accuracies compared among different terms applied. Box and whisker plots present the range of accuracies obtained from the iterative classification with 100 iterations.

Classification accuracy varied between species. Alder is classified with the highest accuracy (≥ 83 per cent for both PA and UA among all seasons), followed by birch (≥ 78 per cent PA and ≥ 86 per cent UA), spruce (≥ 86 per cent PA and ≥ 77 per cent UA) and pine (≥ 83 per cent PA and ≥ 86 per cent UA). Oak, linden and hornbeam are classified with lower accuracies (PA ≥ 65 per cent, ≥ 68 per cent, ≥ 45 per cent and UA ≥ 68 per cent, ≥ 68 per cent, ≥ 41 per cent, respectively). Differences between late summer and early summer or late summer and autumn accuracies are both rather small (2–3 per cent). A notable difference is the one between early summer and autumn (~ 5 per cent). Particularly for linden, hornbeam and pine (5–10 per cent of difference for PA/UA), we observed stronger differences. Coniferous species, alder and birch are generally classified with higher accuracy. The results obtained for the stacked multitemporal dataset are higher than the late summer or autumn ones but do not exceed accuracies obtained with early summer data (Figure 5). However, comparing classification results for particular species, the McNemar test did not indicate significant differences between each pair of acquisition results (P -value > 0.05 for each compared pair).

Misclassified tree individuals—results

The investigated tree traits show differences between correctly and incorrectly classified trees for mostly three species: spruce (in early summer), oak (in late summer and autumn) and in all seasons for birch (Figures 6 and 7).

The crown areas of correctly classified trees generally span a large value range, especially for birch, linden and oak (Figure 6). Misclassified birches, oaks and to a lesser degree spruce show a tendency towards smaller crown areas as compared to the correctly classified trees. These differences are significant for birch during all seasons, for oaks during late summer and autumn and for spruce in early summer (Figure 6). In all cases, the median crown area of the misclassified trees was at least 10 m^2 below the median of correctly classified trees for these three species.

Focusing on tree height, in late summer and autumn, the heights of misclassified trees vary more (for example, in oaks,

the interquartile range is at least 5 m wider). Shorter trees are misclassified frequently. This can be observed for birches, oaks, pines and spruces (the difference of median values is up to 10 m). In alders and lindens, we do not observe differences in height between the groups of correctly and incorrectly classified ones. Interestingly, hornbeams are rather misclassified when they are higher. Statistically significant differences were found for birch in all seasons, where trees lower than 20 m are classified incorrectly and the difference of medians of correctly and incorrectly classified trees is about 10 m. Similarly, in oaks, the height difference between correctly and incorrectly classified trees is about 6 m and was found to be significant in late summer and autumn. Compared to the results for birch, the variation in height of incorrectly classified oaks is wider (Figure 7).

Discussion

Here, we used three hyperspectral images acquired in different stages of the growing season (July, August, October) across 1 year (2015) to classify tree species in a heterogeneous temperate forest. We tested for differences in classification accuracy between the individual images and whether stacking the three images improves the classification result. Besides this, we examined whether tree size, expressed as height and crown area influences the ability of the classification workflow to classify a tree correctly.

Our results identified the beginning of July as the best period to acquire hyperspectral data in order to classify tree species (highest OA of the three examined time points). One reason for the higher accuracy obtained in July might be that all the species have already leafed-out and all but linden have passed the flowering stage (March–May), but have not started ageing yet (Delpierre *et al.*, 2016; Tomanek and Witkowska-Żuk, 2008). The second best results were obtained for August while the worst results were found for the autumn image. This is somewhat surprising as we expected that leaf tinting may have a positive effect on the separability of species. However, former studies comparing the usefulness of hyperspectral data from different seasons for classifying tree species, also reported data from June (Tagliabue *et al.*, 2016), July (Voss and Sugumaran, 2008) or August (Richter *et al.*, 2016) as better than datasets acquired in autumn. One reason for the lower accuracies in autumn may relate to the sun-sensor geometry. Lower sun elevation in autumn causes longer shadows and reduces the number of sunlit pixels, which results in an overall lower quality of the images acquired in October. The negative influence of shadows for classifying woody species has recently been reported for very high-resolution UAV data (Lopatin *et al.*, 2019) and may also affect the data examined here. Furthermore, some of the trees might have already leafed-off in our particular dataset. The year 2015 was a very dry year in Poland (Boczoń *et al.*, 2018; Łabędzki and Bąk, 2015) which may have led stressed trees to shed their leaves earlier than their typical leaf-shedding time in late October (Kim *et al.*, 2016). Another reason for the reduced accuracy of the autumn image may lay in the variability of the leaf-tinting processes which can vary notable even within a single species and across small spatial scales due to differences in site conditions or genetic variations of the tree species. The influence of environmental variables on

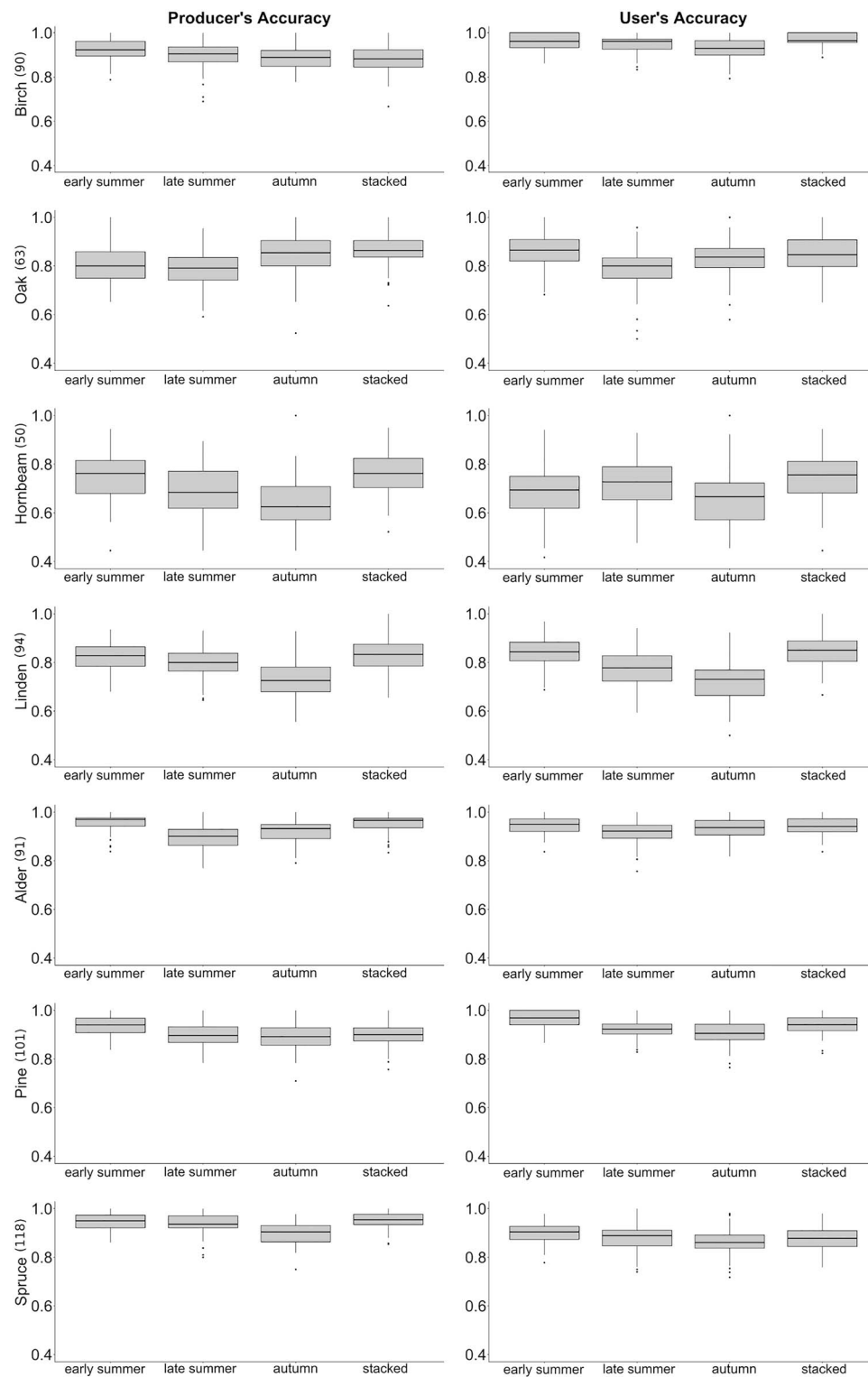


Figure 5 Producer's and user's accuracies for particular species (number of individuals per class in brackets).

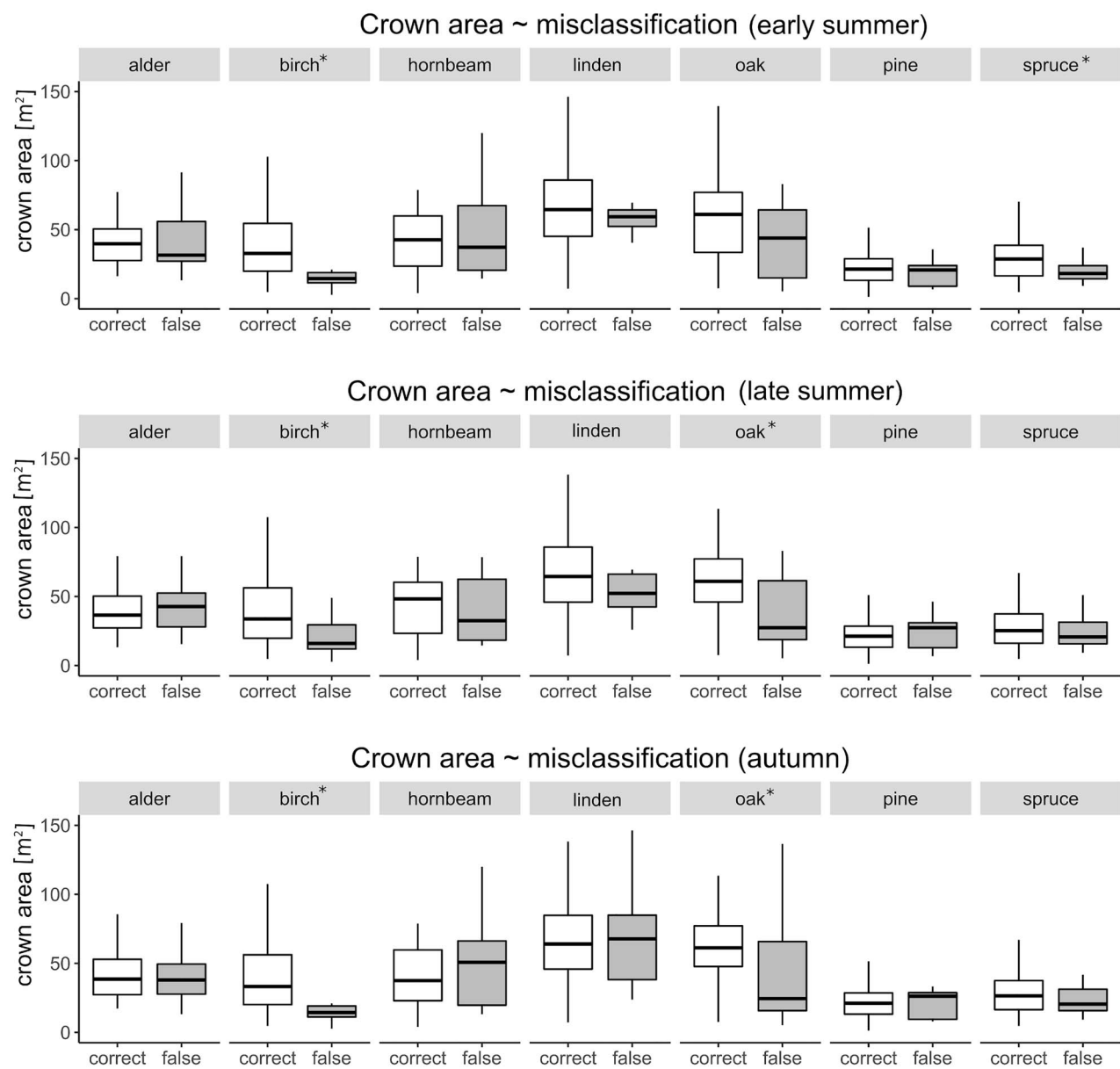


Figure 6 Crown area of correctly and incorrectly classified trees (*—statistically significant differences based on Mann–Whitney U test).

the timing of leaf senescence processes as observed with remote sensing data has been reported before (e.g. Wolter *et al.* 1995). In comparison to the individual classification results, the classification using all three images together resulted in higher accuracies than the autumn or late summer ones and a similar, slightly higher accuracy than obtained with the best single dataset—early summer. Key *et al.* (2001) made similar observations in their study based on airborne multispectral data (RGB + NIR). In their case, the multitemporal dataset also did not outperform the best single dataset. Contrarily, studies based on multispectral Landsat satellite data reported notably improved results for multitemporal datasets (e.g. Mickelson *et al.* 1998, Wolter *et al.*, 1995). These deviating results may partly be

explained with the differences in spatial grain and spectral coverage of the applied datasets. Changes in illumination conditions are likely to increase the (intra-class) spectral variability in high-resolution multitemporal airborne datasets like the ones used in Key *et al.* (2001). These may cancel out the potential benefit of increased spectral information provided by the multitemporal imagery which clearly benefited the Landsat-based studies. When having a closer look at the traits of the misclassified trees, we observed clear trends for some of the species. For birch, the classification produced different accuracies for trees of different heights and crown areas (Figures 6 and 7). Shorter trees with smaller crowns were more frequently misclassified than larger trees. As birch is a species with a generally narrow crown

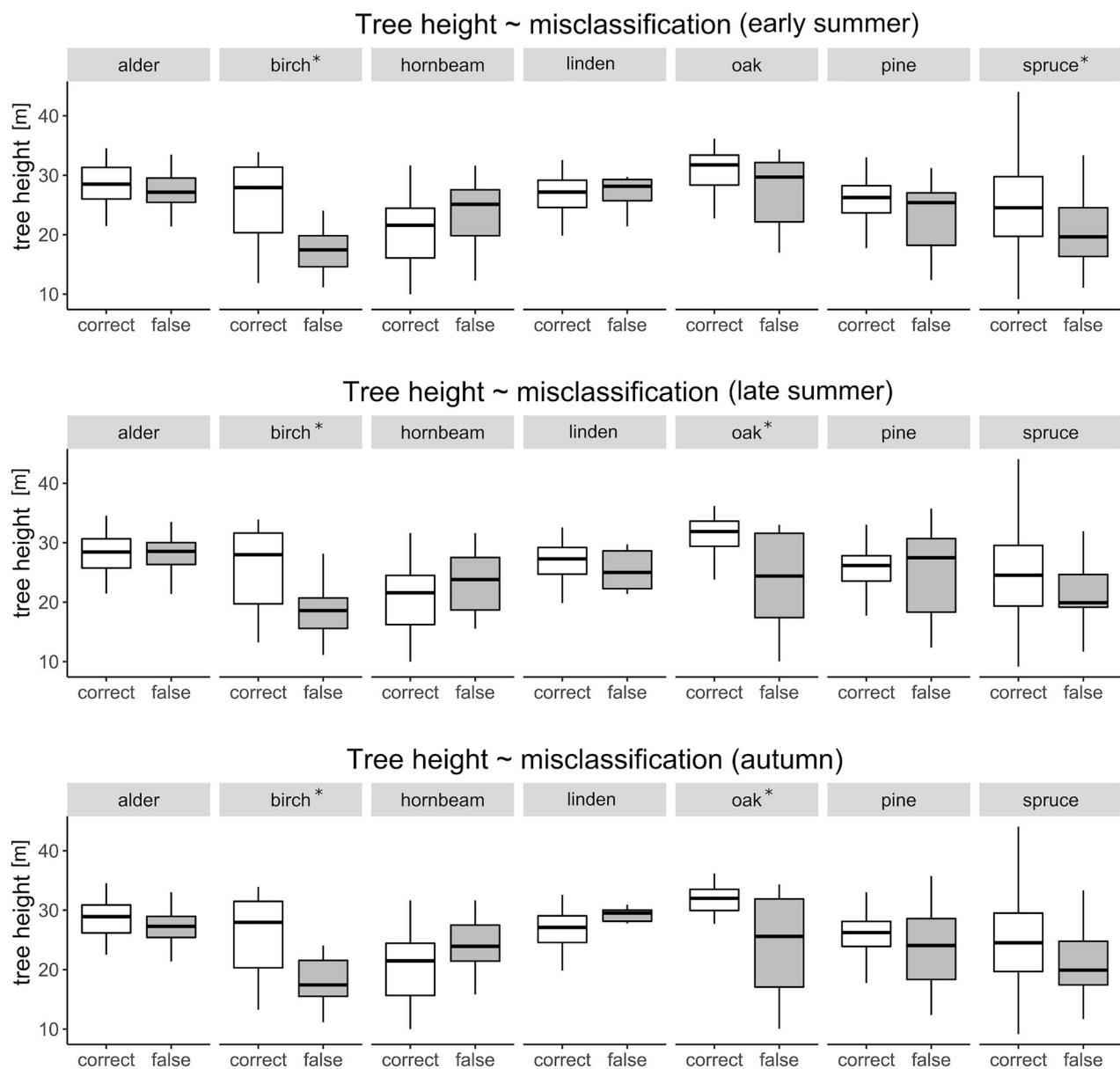


Figure 7 Trees height of correctly and incorrectly classified trees (*—statistically significant differences based on Mann-Whitney U test).

when compared to the other broad-leaved species, birches with smaller crowns probably ‘disappear’ among other trees’ crowns and may also suffer more than other species from problems related to mixed pixels. The median crown area of misclassified birch trees was found to be below 25 m², which is less than a single hyperspectral pixel which supports the assumption that mixed pixels might have been a problem. Interestingly, birch has also been a challenging species to classify in boreal forests where birch is typically the only commonly occurring broadleaved species (e.g. [Dalponte et al. 2013](#)). The authors explained the low accuracies for birch with the sparse number of samples available for birch in their study, but the problem of mixed pixels may have also contributed to the problem.

In early summer, the number of misclassified oak trees is notably lower than in the other seasons. Particularly oaks with smaller crowns are misclassified more frequently in late summer and autumn. One reason for that might be the time of linden’s flowering, which is in July ([Tomanek and Witkowska-Żuk, 2008](#)). In our study area, oaks grow in stands with lindens and hornbeams and are most frequently misclassified with those species. Flowering lindens in July might have increased the spectral contrast between the species and thereby helped to correctly classify even small-crowned or lower oak trees.

Spruce trees’ height and crown metrics significantly influence the classification results, particularly in early summer. Similarly, as discussed for birch, spruces also tend to have very narrow

crowns that do not occupy a lot of space in a pixel. Hence, mixed pixels might have also been a major problem here. The reason why this tendency was more pronounced in early summer than in the other months remains unclear at this point and would require a more detailed investigation.

Alder, pine, linden and hornbeam trees do not show significant differences when we compare crown area and tree heights of correctly and incorrectly classified trees. One reason for that might be that alder and pine trees are generally classified with higher accuracies due to their tendency to grow in rather pure stands in the study area (frequently 90–100 per cent of alder or pine). In such situations, mixed pixel problems due to other co-occurring species do not exist. Focusing on pure stands generally improves classification accuracies as shown by some studies presenting very high accuracies for woody species classifications using multi- and hyperspectral datasets (e.g. [Adelabu et al. 2013](#), [Fassnacht et al. 2014](#)). We suppose that for linden and hornbeam the exactly opposite situation applies. The intermixture in stands with linden, hornbeam and oak is generally high, which leads to generally increased misclassification rates in these stands ([Modzelewska et al., 2020](#)). If the rate of misclassification is generally high even trees with bigger crowns are being misclassified frequently and hence the differences between the two classes are less pronounced. In summary—crown metrics explain the misclassification of individual trees for species that have smaller crowns (e.g. spruce and birch) and species that tend to grow in stands with higher intermixture rates (e.g. oaks). Misclassified trees of species that tend to grow in rather uniform, mono-specific stands like alder or pine ([Modzelewska et al., 2020](#)) show no clear relation to tree size. Linden and hornbeam do not fit this categorization, but both species generally occur with lower frequencies in the study area and suffer from generally lower classification rates which may be one reason for not finding clear differences in the crown and height values of the correctly and incorrectly classified individuals. Hornbeam trees are shade-tolerant and typically grow in mixed stands dominated by oaks. They are lower than co-occurring species ([Sikkema et al., 2016](#)), so parts of their crowns can be hidden under other crowns. Linden is a typical shade-tolerant species that grow in close proximity to other trees. It usually occurs in dense forests with other deciduous species ([Eaton et al., 2016](#)). In consequence, their crowns are often mixed with other species' crowns, leading to a mixed spectral signal.

Hence, the factors influencing tree species classification results are related not only to phenology but also to the particular species' traits. The latter can affect the classification even more than phenology. Our results show that it is worth investigating the locally optimal season (that appears to be early summer for central Europe, but certainly may differ for other parts of the Earth). Implementing multitemporal dataset is also a good solution, but the result did not outperform the best single dataset in our study, so we would suggest using the best single one, when possible. Furthermore, taking into account tree traits can be the key to avoid some of the misclassification; however, this topic needs to be investigated more deeply in the future.

Conclusions

Concerning the three examined image acquisition dates, we found the image from July and the multitemporal stacked

dataset as performing better than images from August or October. We assume that because each individual hyperspectral dataset encompasses a huge amount of spectral information, the added value of multitemporal datasets of hyperspectral images is not as large as in the case of multispectral images.

Concerning the examined tree traits, we found significant differences in tree height and the crown area between correctly and incorrectly classified trees for birch (in all seasons), oak (in late summer and autumn) and spruce (in early summer). When one of the tree traits significantly influenced the species classification, the other investigated trait was also significant.

Future studies may benefit from actively considering such findings by for example integrating the crown area actively in the classification workflow, either as an additional predictor variable in an object-based approach or by conducting separate classifications for smaller and larger tree individuals. Another aspect worthwhile to be investigated in future studies is the neighbourhood conditions of misclassified trees. As discussed above, it is likely that the species composition of the directly neighbouring trees have a direct effect on the likelihood of a given tree to be classified correctly. However, corresponding analyses are still missing. One reason for this may be that such investigations would require a full inventory of the location and species across larger spatially continuous areas which are not available in many existing datasets. Auxiliary information collected by unmanned aerial vehicles or the application of synthetic datasets may be alternatives to approach this research question in future studies.

Conflict of interest statement

None declared.

Funding

This work was financially supported by the Project LIFE+ For-BioSensing PL 'Comprehensive monitoring of stand dynamics in Białowieża Forest supported with remote sensing techniques', co-funded by Life Plus (contract number LIFE13 ENV/PL/000048), and Poland's National Fund for Environmental Protection and Water Management (contract number 485/2014/WN10/OP-NM-LF/D).

Data availability statement

The data underlying this article will be shared on reasonable request to the corresponding author.

References

- Adelabu, S., Mutanga, O., Adam, E. and Cho, M.A. 2013. Exploiting machine learning algorithms for tree species classification in a semiarid woodland using rapid eye image. *J. Appl. Remote. Sens.* **7** (1).
- Boczoń, A., Kowalska, A., Ksepko, M. and Sokółowski, K. 2018. Climate warming and drought in the Białowieża Forest from 1950–2015 and their impact on the dieback of Norway spruce stands. *Water (Switzerland)* **10**, 1–18. doi: [10.3390/w10111502](https://doi.org/10.3390/w10111502).
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46. doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104).
- Dalponte, M., Ørka, H. O., Gobakken, T., Gianelle, D. and Næsset, E. 2013. Tree species classification in boreal forests with hyperspectral

- data. *IEEE Trans Geosci Remote Sens* **51**(5), 2632–2645. doi: [10.1109/TGRS.2012.2216272](https://doi.org/10.1109/TGRS.2012.2216272)
- Delpierre, N., Vitasse, Y., Chuine, I., Guillemot, J., Bazot, S., Rutishauser, T., et al. 2016 Temperate and boreal forest tree phenology: from organ-scale processes to terrestrial ecosystem models. *Ann. For. Sci.* **73**, 5–25. doi: [10.1007/s13595-015-0477-6](https://doi.org/10.1007/s13595-015-0477-6).
- Eaton, E., Caudullo, G. and de Rigo, D. 2016 *Tilia cordata*, *Tilia platyphyllos* and other limes in Europe: distribution, habitat, usage and threats. In *European Atlas of Forest Tree Species*. J., San-Miguel-Ayán, D., de Rigo, G., Caudullo, T., Houston Durrant, A., Mauri (eds.). Publ. Off. EU, Luxembourg City, pp. 184–185.
- Faliński, J.B. 1986 *Geobotany. Vegetation Dynamics in Temperate Lowland Primeval Forests*. Springer, Dordrecht, pp. 39–111.
- Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., et al. 2016 Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **186**, 64–87. doi: [10.1016/j.rse.2016.08.013](https://doi.org/10.1016/j.rse.2016.08.013).
- Fassnacht, F.E., Neumann, C., Forster, M., Buddenbaum, H., Ghosh, A., Clasen, A., et al. 2014 Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 2547–2561. doi: [10.1109/JSTARS.2014.2329390](https://doi.org/10.1109/JSTARS.2014.2329390).
- Forzieri, G., Moser, G. and Catani, F. 2012 Assessment of hyperspectral MIVIS sensor capability for heterogeneous landscape classification. *ISPRS J. Photogramm. Remote Sens.* doi: [10.1016/j.isprsjprs.2012.09.011](https://doi.org/10.1016/j.isprsjprs.2012.09.011).
- Ghosh, A., Fassnacht, F.E., Joshi, P.K. and Kochb, B. 2014 A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales. *Int. J. Appl. Earth Obs. Geoinf.* **26**, 49–63. doi: [10.1016/j.jag.2013.05.017](https://doi.org/10.1016/j.jag.2013.05.017).
- Green, A.A., Berman, M., Switzer, P., Craig, M.D. 1988 A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sens.* **26**, 65–74. doi: doi.org/10.1109/36.3001.
- Hill, R.A., Wilson, A.K., George, M. and Hinsley, S.A. 2010 Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Appl. Veg. Sci.* **13**, 86–99. doi: [10.1111/j.1654-109X.2009.01053.x](https://doi.org/10.1111/j.1654-109X.2009.01053.x).
- Jędrzejewska, B. and Jędrzejewski, W. 1998 Białowieża primeval Forest—abiotic conditions and habitats. 4–27. doi: [10.1007/978-3-662-35364-6_2](https://doi.org/10.1007/978-3-662-35364-6_2).
- Jones, T.G., Coops, N.C., Sharma, T. 2010. Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific northwest, Canada. *Remote Sens. Environ.* **114** (12), 2841–2852.
- Kamińska, A., Lisiewicz, M., Stereńczak, K., Kraszewski, B. and Sadkowski, R. 2018 Species-related single dead tree detection using multi-temporal ALS data and CIR imagery. *Remote Sens. Environ.* **219**, 31–43. doi: [10.1016/j.rse.2018.10.005](https://doi.org/10.1016/j.rse.2018.10.005).
- Kern, A., Marjanović, H. and Barcza, Z. 2020 Spring vegetation green-up dynamics in Central Europe based on 20-year long MODIS NDVI data. *Agric. For. Meteorol.* **287**. doi: [10.1016/j.agrformet.2020.107969](https://doi.org/10.1016/j.agrformet.2020.107969).
- Keskitalo, J., Bergquist, G. and Jansson, S. 2005 A cellular timetable of autumn senescence. *Plant Physiol.* doi: [10.1104/pp.105.066845](https://doi.org/10.1104/pp.105.066845).
- Key, T., Warner, T.A., McGraw, J.B. and Fajvan, M.A. 2001 A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest. *Remote Sens. Environ.* **75**, 100–112. doi: [10.1016/S0034-4257\(00\)00159-0](https://doi.org/10.1016/S0034-4257(00)00159-0).
- Kim, J., Woo, H.R.R. and Nam, H.G.G. 2016 Toward systems understanding of leaf senescence: an integrated multi-omics perspective on leaf senescence research. *Mol. Plant* **9**, 813–825. doi: [10.1016/j.molp.2016.04.017](https://doi.org/10.1016/j.molp.2016.04.017).
- Korpela, I., Mehtätalo, L., Markelin, L., Seppänen, A. and Kangas, A. 2014 Tree species identification in aerial image data using directional reflectance signatures. *Silva Fenn.* **48**. doi: <https://doi.org/10.14214/sf.1087>.
- Lopatin, J., Dolos, K., Kattenborn, T., Fassnacht, F. E. 2019. How canopy shadow affects invasive plant species classification in high spatial resolution remote sensing, *Remote Sens. Ecol. Conserv.*, **5** (4), 302–317.
- Łabędzki, L. and Bąk, B. 2015 Drought in Poland in 2015 and an assessment of impacts in permanent grassland. *Wiadomości Melior. i Łąkarskie* **3**, 102–106.
- Leckie, D.G., Tinis, S., Nelson, T., Burnett, C., Gougeon, F.A., Cloney, E., et al. 2005 Issues in species classification of trees in old growth conifer stands. *Can. J. Remote Sens.* **31**, 175–190. doi: [10.5589/m05-004](https://doi.org/10.5589/m05-004).
- Melgani, F. and Bruzzone, L. 2004 Estimation of biophysical parameters from optical remote-sensing images with high-order residues. *IEEE Int. Geosci. Remote Sens. Symp.* doi: [10.1109/IGARSS.2004.1368700](https://doi.org/10.1109/IGARSS.2004.1368700).
- Meyer, D. 2012 Support vector machines. the Interface to LIBSVM in package. Vienna University of Technology, Vienna, p. e1071.
- Mickelson, J.G., Civco, D.L. and Silander, J.a. 1998 Delineating forest canopy species in the Northeastern United States using multi-temporal TM imagery. *Photogramm. Eng. Remote Sens.* **64**, 891–904.
- Miścicki, S. 2016 Changes in the stands of the Białowieża National Park from 2000 to 2015. *For. Res. Pap.* **77**, 371–379. doi: [10.1515/frp-2016-0038](https://doi.org/10.1515/frp-2016-0038).
- Modzelewska, A., Fassnacht, F.E. and Stereńczak, K. 2020 Tree species identification within an extensive forest area with diverse management regimes using airborne hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* **84**, 101960. doi: [10.1016/j.jag.2019.101960](https://doi.org/10.1016/j.jag.2019.101960).
- Mountrakis, G., Im, J. and Ogole, C. 2011 Support vector machines in remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* **66**, 247–259. doi: [10.1016/j.isprsjprs.2010.11.001](https://doi.org/10.1016/j.isprsjprs.2010.11.001).
- Norsk Elektro Optikk AS, HySpex hyperspectral cameras. <https://www.hyspex.no/products/> (accessed on 20 September, 2019).
- Ørka, H.O., Næsset E., Bollandsås O.M. 2009. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sens. Environ.* **113** (6), 1163–1174.
- Pal, M. and Mather, P.M. 2005 Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* **26**, 1007–1011. doi: [10.1080/01431160512331314083](https://doi.org/10.1080/01431160512331314083).
- Panchen, Z.A., Primack, R.B., Gallinat, A.S., Nordt, B., Stevens, A.D., Du, Y., et al. 2015 Substantial variation in leaf senescence times among 1360 temperate woody plant species: implications for phenology and ecosystem processes. *Ann. Bot.* **116**, 865–873. doi: [10.1093/aob/mcv015](https://doi.org/10.1093/aob/mcv015).
- R Development Core Team, 2015. The R Project for Statistical Computing. <http://www.R-project.org/> (accessed on 12 July 2016).
- Richter, R., Reu, B., Wirth, C., Doktor, D. and Vohland, M. 2016 The use of airborne hyperspectral data for tree species classification in a species-rich central European forest area. *Int. J. Appl. Earth Obs. Geoinf.* **52**, 464–474. doi: [10.1016/j.jag.2016.07.018](https://doi.org/10.1016/j.jag.2016.07.018).
- Sikkema, R., Caudullo, G. and de Rigo, D. 2016 *Carpinus betulus* in Europe: distribution, habitat, usage and threats. In *European Atlas of Forest Tree Species*. J., San-Miguel-Ayán, D., de Rigo, G., Caudullo, T., Houston Durrant, A., Mauri (eds.). Publ. Off. EU, Luxembourg City, pp. 74–75.
- Stereńczak, K. 2013 Determination of stand density using data from airborne laser scanning|Określenie zagęszczenia drzewostanów z wykorzystaniem danych z lotniczego skanowania laserowego. *Sylwan* **157**, 607–617.
- Stereńczak, K., Kraszewski, B., Mielcarek, M. and Piasecka, Ż. 2017 Inventory of standing dead trees in the surroundings of communication routes – The contribution of remote sensing to potential risk assessments. *For. Ecol. Manage.* **402**, 76–91. doi: [10.1016/j.foreco.2017.07.018](https://doi.org/10.1016/j.foreco.2017.07.018).

- Stereńczak, K., Kraszewski, B., Mielcarek, M., Piasecka, Ż., Lisiewicz, M. and Heurich, M. 2020 Mapping individual trees with airborne laser scanning dAata in an European lowland forest using a self-calibration algorithm. *Int. J. Appl. Earth Obs. Geoinf.* **93**, 102191.
- Story, M. and Congalton, R.G. 1986 Accuracy assessment: a user's perspective. **52**, 397–399.
- Tagliabue, G., Panigada, C., Colombo, R., Fava, F., Cilia, C., Baret, F., et al. 2016 Forest species mapping using airborne hyperspectral APEX data. *Misc. Geogr.* **20**, 28–33. doi: [10.1515/mgrsd-2016-0002](https://doi.org/10.1515/mgrsd-2016-0002).
- Tomanek, J. and Witkowska-Żuk, L. 2008 *Botanika leśna*. Powszechne Wydawnictwo Rolnicze i Leśne, Warsaw.
- Trier, Ø.D., Salberg, A.B., Kermit, M., Rudjord, Ø., Gobakken, T., Næsset, E., et al. 2018 Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data. *Eur. J. Remote Sens.* **51**, 336–351. doi: [10.1080/22797254.2018.1434424](https://doi.org/10.1080/22797254.2018.1434424).
- Vapnik, V.N. 1998 Statistical learning theory. *Dermatol. Int.* doi: [10.2307/1271368](https://doi.org/10.2307/1271368).
- Vincheh, Z.H. and Arfania, R. 2017 Lithological mapping from OLI and ASTER multispectral data using matched filtering and spectral analogues techniques in the Pasab-e-Bala area, Central Iran. *Open J. Geol.* **7**, 1494–1508. doi: [10.4236/ojg.2017](https://doi.org/10.4236/ojg.2017).
- Voss, M. and Sugumaran, R. 2008 Seasonal effect on tree species classification in an urban environment using hyperspectral data, LiDAR, and an object-oriented approach. *Sensors* **8**, 3020–3036. doi: [10.3390/s8053020](https://doi.org/10.3390/s8053020).
- Wolter, P.T., Mladenoff, D.J., Host, G.E., Crow and T.R. 1995. Improved forest classification in the northern Lake states using multi-temporal landsat imagery. *Photogramm. Eng. Remote Sensing* **61**(9), 1129–1143.