Int J Appl Earth Obs Geoinformation 92 (2020) 102178

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Species dominance and above ground biomass in the Białowieża Forest, Poland, described by airborne hyperspectral and lidar data

Gaia Vaglio Laurin[a], Nicola Puletti[b], Mirko Grotti[b], Krzysztof Stereńczak[c,*], Aneta Modzelewska[c], Maciej Lisiewicz[c], Rafał Sadkowski[c], Łukasz Kuberski[c], Gherardo Chirici[d], Dario Papale[a]

[a] University of Tuscia, Department for Innovation in Biological, Agro-Food and Forest Systems, Viterbo, Italy
[b] Council for Agricultural Research and Economics, Research Centre for Forestry and Wood (CREA-FL), Arezzo, Trento, Italy
[c] Department of Geomatics, Forest Research Institute, Sękocin Stary, 05-090, Raszyn, Poland
[d] Università degli Studi di Firenze, Department of Agricultural, Food and Forestry Systems, Via San Bonaventura, 13-50145, Firenze, Italy

ABSTRACT

The objective of this research is to test and evaluate hyperspectral and lidar data to derive information on tree species dominance and above ground biomass in the Białowieża Forest in Poland. This forest is threatened by climate change, fire, bark beetles attacks, and logging, with changes in species composition and dominance. In this conservation valuable area, the monitoring of forest resources is thus critical.

Results indicate that vegetation indices from hyperspectral data can support species dominance detection: using a Classification and Regression Trees algorithm the three main plot types (dominated by Deciduous, Spruce, and Pines species) were classified with an Overall Accuracy > 0.9. The accuracy decreased when a ‘Mixed’ group was added to account for very heterogeneous plots, and plots dominated by Spruce were not correctly detected. Hyperspectral vegetation indices were also used to estimate the level of species dominance in the forest plots, using a Multivariate Multiple Linear Regression model; the obtained accuracy varied according to groups, being higher for Deciduous (R^2 = 0.87), compared to Pines (R^2 = 0.61), and to Spruce-dominated plots (R^2 = 0.37).

Lidar data were employed to estimate above ground biomass, using an exponential regression model; overall the R^2 resulted equal to 0.66 but ranged from 0.57 to 0.78 when considering subgroups according to species dominance; the addition of hyperspectral vegetation indices improved the result only for Pines.

The illustrated methods provide a reliable description of important forest characteristics and simplify resource monitoring, supporting local authorities to address the challenges imposed by climate change and other forest threats.

1. Introduction

Active conservation and management is a requirement for maintaining forest resources, and linking knowledge to action helps to face environmental change, providing conservation solutions. To this end, resource monitoring and robust ecological data are needed (Larson et al., 2013).

Stand delineation and species composition estimation are considered key data to support forest inventory and mapping, as well as forest management decision making (Leckie et al., 2003). Changes in forest species composition may results in changes in the whole ecosystem and its biodiversity, as evidenced in eastern United States after that bark beetles altered the forest soil chemistry (Arthur et al., 2017); or in China where changes in species composition and community structure caused impacts on above ground biomass and soil carbon density (Hu et al., 2015); and also in the Białowieża Forest in Poland, where different intensity of stand management practices impacted birds (Czeszczewik et al., 2015) and beetles (Jaworski et al., 2019) species composition and abundance. Species-level and forest type information is also critical for sustainable forest management (SFM), the main approach in forest policy in Europe, that is assessed by various indicators including forest type and species/dominance information (Barbati et al., 2014).

Above ground biomass (AGB) is a fundamental parameter for carbon accounting and reporting, and for timber production; links to biodiversity conservation are also well-known (Barlow et al., 2016; Cheng et al., 2018; Verkerk et al., 2014). Climate change can negatively impact AGB, as reported in China forests (Zhang and Liang, 2014); or in United States forest, according to climate change scenarios (Wang et al., 2017). In European forests increased extreme weather events, such as prolonged drought or storms and floods, have clear impacts on carbon sequestration capacity (Lindner et al., 2010).

Remote sensing plays a relevant role in providing forest information thanks to its ability to extrapolate point data to larger extents, representing a fundamental tool to monitor vast areas (Chirici et al., 2020). Hyperspectral sensors are well suited to collect species level information: with hundreds of radiometric fine bands, each representing a tiny portion of the electromagnetic spectrum, this sensor

* Corresponding author.
E-mail address: K.Sterenczak@ibles.waw.pl (K. Stereńczak).

https://doi.org/10.1016/j.jag.2020.102178
Received 16 January 2020; Received in revised form 30 April 2020; Accepted 12 June 2020
0303-2434/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

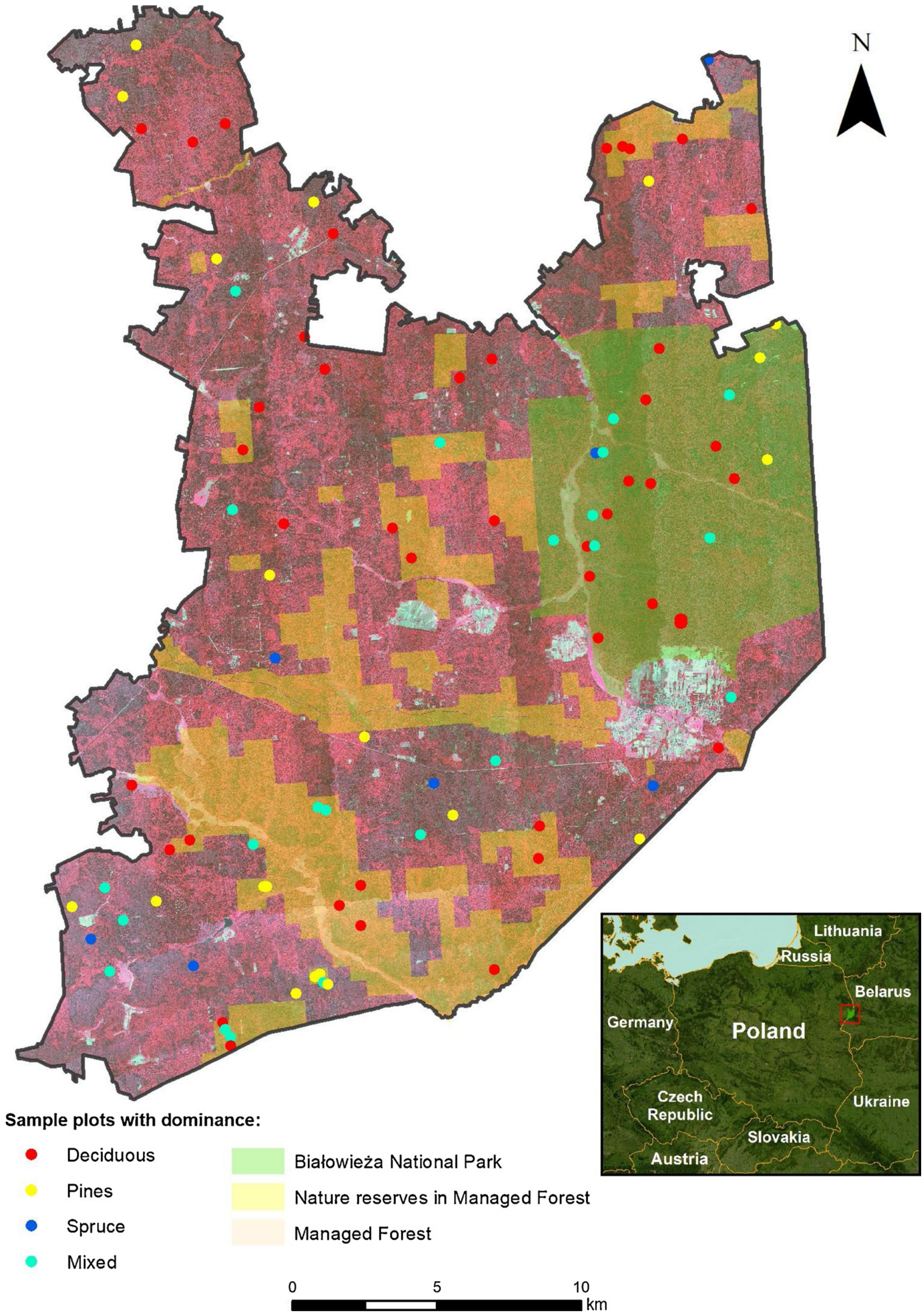

Fig. 1. The Białowieża Forest: areas with different protection and management status and distribution of the field plots used in the research.

allows the extraction of reflection characteristics of different species (Lucas et al., 2008). Hyperspectral data were successfully employed to discriminate among different species groups or ecological units and guilds in temperate forests (Schull et al., 2011; Treitz and Howarth, 2000), and in tropical forest (Vaglio Laurin et al., 2016). However, species classification can still remain a challenge, especially in dense forest stands with high species intermixture (Heinzel and Koch, 2012; Modzelewska et al., 2020).

Forest type classification could also be achieved using satellite data with improved spectral capability such as Sentinel 2, without recurring to costly hyperspectral data (Puletti et al., 2018; Vaglio Laurin et al., 2016), as also demonstrated in the Polish Carpathian mountains (Grabska et al., 2019). However, the prediction at fine spatial scale of canopy biodiversity, of vegetation associations, or of level of dominance of species or guilds, requires the use of hyperspectral data (Asner and Martin, 2009; Leutner et al., 2012; Vaglio Laurin et al., 2014a, b; Modzelewska et al., 2020).

Lidar (light detection and ranging) usually generates highly accurate biomass estimates, thanks to its ability to provide detailed vertical forest structure information and based on the strong links occurring between forest height and density, and tree biomass (Chirici et al., 2016; Vaglio Laurin et al., 2017). Several studies were developed to estimate AGB in temperate forests, including in mixed stands in Germany (Latifi et al., 2010, 2015) and in Czech Republic (Brovkina et al., 2015). Joined lidar and hyperspectral data use was also previously investigated, e.g. in northern Italy (Vaglio Laurin et al., 2017), and in the western Carpathians forests (Brovkina et al., 2015). Combining data from different sensors can be an advantage for forest characterization, because of the complementarity of the information content (Koch 2010). This is the case of tree species mapping, when the spectral information is merged with the structural information from lidar (Dalponte et al., 2008; Jones et al., 2010); or AGB estimation, when hyperspectral data added to lidar metrics result in an increase of predictions accuracy (Anderson et al., 2008; Vaglio Laurin et al., 2014a, b).

Field data on AGB are usually collected in plots and are fundamental to validate the models used to create biomass estimations. The impact of the plot size in the accuracy of the estimates was investigated also in temperate European forests (Maleki and Kiviste, 2015; Stereńczak et al., 2018a). Small plots are commonly affected by the edge effects caused by trees standing near the plot border and having part of them (e.g. the crown) inside and part outside the plot. Sometimes field data are collected in small plots for other purposes: when small plots represent the only available information, procedures to artificially enlarge it applying a buffer around the edge might be of help, if the area can be assumed as sufficiently homogeneous (Maleki and Kiviste, 2015; Stereńczak et al., 2018b).

The objective of this research is to test and evaluate the use of airborne hyperspectral and lidar data, separately and jointly, to derive critical information on stand species dominance and AGB in the Białowieża Forest in Poland. This information is urgently requested as Białowieża Forests (BF) resources are threatened by different factors including: fire (Szczygieł et al., 2016; Szczygiel et al., 2018); recurrent bark beetles attacks (Stereńczak et al., 2019) promoted by climate warming (Boczoń et al., 2018); and recent logging activities (Żmihorski et al., 2018). BF is also facing accelerated changes in species composition and dominance, caused by human-induced and natural disturbance, such as bark beetle invasions and consequent dieback of selected species, with expansion of their competitors. In recent years the BF area registered important spruce (*Picea abies*) and ash (*Fraxinus excelsior*) dieback, with hornbeam (*Carpinus betulus*) expansion (Miścicki, 2016; Cholewińska et al., 2018).

2. Methods

2.1. Study area

The Białowieża Forest UNESCO World Heritage site is one of the oldest European forest. It is located at the Poland - Belarus border and, with an extent of 141,885 ha of forest –partly primeval- and a buffer zone of 166,708 ha, represents an invaluable conservation area and the remaining part of the vast primeval forest that once covered central Europe. The BF includes lowland conifers and deciduous species, with presence of very old trees and extensive undisturbed areas rich in deadwood; it hosts 59 mammal, over 250 bird, 13 amphibian, 7 reptile and over 12,000 invertebrate species, and it is home to the largest free-roaming population of European Bison. The forest benefits from legal protection in both countries, and includes an area in which stands are actively managed, a forest reserve, and a national park. This, and additional information on BF, is provided by Kujawa et al. (2016). This study was carried out on Polish part of Białowieża Forest (Fig. 1) which covers approximately 62 000 ha.

2.2. Remote sensing and field data

Remote sensing and field data were collected through the LIFE + ForBioSensing project over the entire Polish part of the BF UNESCO World Heritage site. The LIFE + project started in 2014 with the purpose to develop and apply an innovative monitoring approach, providing comprehensive illustration of forest stands dynamics, and moving from point or plot level to large scale monitoring to improve the efficiency of management strategies.

In 2015, 685 circular plots of 500 m^2 (12.62 m radius), randomly distributed throughout the Polish part of BF according to a stratified random sampling design, were set up. Height, diameter at breast height (DBH), and species information were recorded, and AGB was computed at tree level according to regional allometric models (Cienciala et al., 2008; Socha and Wężyk, 2004; Tabacchi et al., 2011). The original size of the field plots was designed for field-based monitoring purposes, but such size is of limited usefulness when linking to remote sensing data, as large tree crowns (r > = 8) cause important edge effects. To overcome this issue, the size of field plots was artificially increased using lidar data and following the method proposed by Stereńczak et al. (2018b), that defines homogeneous patches as areas with similar forest structure according to ALS data. Thus, ALS-based height and density metrics were computed for the 685 plots, and for neighbouring areas of 500 m^2 pixel size. The difference in ALS-based metrics between plots and neighboring areas was computed, retaining only those plots for which this difference resulted below a 10 % threshold. The radius of these plots was increased up to 25 m, obtaining 100 plots (14.6 % of the total) of 1962.5 m^2, characterized by similar height and density with respect to the original ones and located in small forest patches internally homogeneous. The AGB values computed for the original small plots were increased proportionally to the new area. Species dominance information from original plots was also increased proportionally to the area, after checking the homogeneity in composition through the map developed by Kamińska et al. (2018), based on ALS and orthophotos data analysis. The occurring species were partitioned in three groups: Deciduous (main deciduous species); Pines (mainly *Pinus sylvestris*); and Spruce (*Picea abies* only). A dominance group was assigned when the individuals of a group represented more than the 60 % of the individuals recorded in the plot; when this threshold was not reached the plot was labeled as Mixed.

In July 2015 an airborne survey collected hyperspectral images and Airborne Laser Scanning (ALS) data in strips covering all the BF plots, allowing the extraction of the corresponding information over the 100 plots.

Hyperspectral data were acquired with HySpex VNIR-1800 (Ground Sampling Distance (GSD) - 2.5 m) and SWIR-384 (GSD - 5 m) sensors, and were radiometrically corrected. Even if the topography of the area is gentle, data were geometrically corrected using a Digital Terrain Model derived by ALS data. Atmospheric correction was performed by MODTRAN5 model in ATCOR4 (Richter and Schläpfer, 2019) and applying a spectral smoothing filter (Persson and Strang, 2003). Shadows were manually masked, and VNIR and SWIR bands were stacked and resampled at 2.5 m spatial resolution. Hyperspectral pixels having > 60 % of area included in the plot were extracted. Vegetation indices were computed using the vegetation analysis tool available in ENVI software (Exelis Visual Information Solutions. 2010. Whitepaper: Vegetation Analysis: Using Vegetation Indices in ENVI. Boulder, Colorado). Within the plots, for each band and vegetation index, minimum, maximum, mean, median, coefficient of variation, and standard deviation were computed.

The ALS point cloud was collected using a full-waveform Riegl LMS-6800i sensor at 500 m altitude, with average density equal to 7 points/m^2, ± 30° maximum scan angle, and footprint size equal to 0.25 m. 135 individual flight lines were collected with 40 % overlap. Data classification was conducted in Terrascan software and data processing in R statistical package, to obtain Canopy Height Model and a Digital Terrain Model rasters, resampled at 0,5 m spatial resolution. The forest metrics included in Table 1 were extracted at plot level using the rLIDAR package (R Development Core Team, 2019). The metrics were extracted from lidar pulses filtered above different heights from the ground, with the purpose to exclude the influence of the dense forest understory.

2.3. Data analysis and techniques

A Pearson correlation analysis was used to select the hyperspectral vegetation indices and bands mostly correlated with the percentage of dominance of Deciduous, Pines and Spruce species occurring in the

Table 1
Plot level height metrics derived from the LiDAR dataset.

Metric	Metrics
All returns	Maximum and average tree height
All returns above 1 m height	Average tree height
All returns above 2 m height	Average tree height
All returns above 3 m height	Average tree height
First returns above 2 m height	Maximum and average tree height

plots.

A Classification And Regression Trees (CART) approach, with hyperspectral data as input, was used to classify the plots in the following 'dominance' groups: Deciduous, Pines, Spruce, and Mixed. Then, a Multivariate Multiple Linear Regression (MMLR) technique was used to predict at plot level the percentage of dominance of each group. Finally, an exponential stepwise regression model was applied to predict above ground biomass using lidar metrics, and also testing the addition of hyperspectral inputs. All the different models were validated with leave one out approach (LOO), a special case of k-fold cross validation, that was selected considering the limited number of samples in certain vegetation groups.

CART decision tree is a non-parametric machine learning model for regression and classification problems (Breiman et al., 1984). To find solutions, a decision tree makes sequential, hierarchical decision about the outcome variable based on the predictor data. CART extracts subgroups of observations within which the explanatory variables are relatively homogeneous, and between which the response variable is relatively distinct. Advantages in CART use include: the ability to handle both numerical and categorical data; the use of a white box model in which explanations for the observed condition are provided by Boolean logic; the possibility to validate models by means of statistical tests; no assumption made on training data or prediction residuals; the efficiency in large dataset analysis; the robustness against collinearity; and the inbuilt feature selection that removes irrelevant predictor features.

MMLR is used to model the linear relationship between more than one independent variable and more than one dependent variable. It regresses each dependent variable separately on the predictors, initially producing different and separate models with related responses. However, the covariance among predictors needs to be taken into account when determining their contribution. This is done with multivariate analysis of variance, meaning that modified hypothesis tests are used to determine whether a predictor contributes to a model (Fox and Weisberg, 2011; Johnson and Wichern, 2007). Computationally, MMLR gives the same coefficients, standard errors, t-and p-values and confidence intervals as one would estimate with individual multiple linear regression for each of the dependent variables. When a correlation structure among the dependent variables is present, a single multivariate regression is more efficient than regressions analyses for each dependent variable separately (Breiman and Friedman, 1997).

In above ground biomass regression, a stepwise procedure was used. In regression modeling, the stepwise procedure is a method of fitting in which the choice of predictive variables is carried out by an automatic procedure; in each step, a variable is considered for addition to or subtraction from the set of explanatory variables (Hocking, 1976). Different criteria can be used to include or not the variable including *F*-tests or *t*-tests, adjusted R^2, Akaike information criterion, Bayesian information criterion. The following flowchart summarizes the methodological approach (Fig. 2).

3. Results

3.1. Dominance groups and hyperspectral data

Applying the > 60 % threshold to field records, the 100 plots resulted dominated by the following species/groups: Spruce (*Picea abies*) 7 plots; Pines (all *Pinus* species) 21 plots; Deciduous (all deciduous species) 50 plots; and Mixed (no species/groups over the 60 % threshold) 22 plots.

The percentage (or level) of dominance in the plots was correlated with hyperspectral bands reflectance and vegetation indices values. Only the three indices mostly correlated with each of the vegetation groups were retained for further analysis, and are the indices presented in Table 2. In fact, even though the correlations are influenced by plot number per group, the results are similar for indices and bands. This suggested the use of vegetation indices for testing, as they facilitate data analysis by reducing the number of variables without losing information, and allowing future comparisons with data collected from other sensors.

The retained vegetation indices provide information on vegetation senescence (CAI), photosynthetic activity (MRENDVI, MRESR, TVI, MTVI2, TCARI), and water content (NDWI and NDMI).

For classification and regression purposes, four indices were employed to reduce multicollinearity: CAI, MRESR, TCARI and NDMI. The Cellulose Absorption Index (CAI, Daughtry, 2001) is based on the shortwave infrared range and informs on dry carbon that is present in large amounts in woody materials and senescent/dead vegetation. The Modified Red Edge Simple Ratio (MRESR, Sims and Gamon, 2002) and Transformed Chlorophyll Absorption Reflectance Index (TCARI, Haboudane et al., 2004) report on photosynthetic activity, exploiting narrowband red and near-infrared reflectance (including the red edge transition from chlorophyll absorption to near-infrared leaf scattering); these indices are sensitive to the combined effects of foliage chlorophyll concentration, canopy leaf area, foliage clumping, and canopy architecture. The Normalized Difference Moisture Index (NDMI, Gao, 1995) is based on near-infrared and shortwave infrared reflectance and exploits the known water absorption features and the light penetration depth to make integrated measurements of total column water content.

3.2. Classification according to dominance groups

CART, with vegetation indices as input, was used to perform the classification of plots into three (Spruce, Pines, and Deciduous) and four (Spruce, Pines, Deciduous, and Mixed) classes. Prior to classification, to avoid introducing strongly correlated inputs, Pearson correlation values were computed among the indices in Table 2, to retain only those with correlation values < $|0.5|$, and namely: Cai_max, Mresr_max, Ndmi_sd, and Tcari_median (Fig. 3).

The classifications were validated with Leave-One-Out (LOO) procedure. Tables 3 and 4 report the confusion matrices, the overall accuracies, and the average variable importance of the input features.

Both classifications produced high overall accuracies, but the addition of the Mixed group increased the task difficulty. Specifically, the Spruce group was not predicted at all, due to confusion with Pines and Mixed groups, that might be the result of the spectral similarity and the limited number of samples in the Spruce group. The order of importance of the input variables remained the same with 3 or 4 groups, with Tcari_median being the most important input.

3.3. Estimate of dominance levels

To estimate the level of dominance in the plots, a Multivariate Multiple Linear Regression model was set up for the Deciduous, Pines and Spruce groups, after checking the linearity of the relationships between the groups and the considered vegetation indices.

Information on the level of dominance can be relevant, as a plot dominated for 95 % by a certain species/group can be ecologically quite different from a plot dominated at 65 % level. The four vegetation indices used for CART classification were used as inputs (Cai_max, Mresr_max, Ndmi_sd, and Tcari_median), allowing second order polynomials in the model. According to MANOVA test results, all the terms resulted significant except 'CAI_max^2', that was excluded. The model

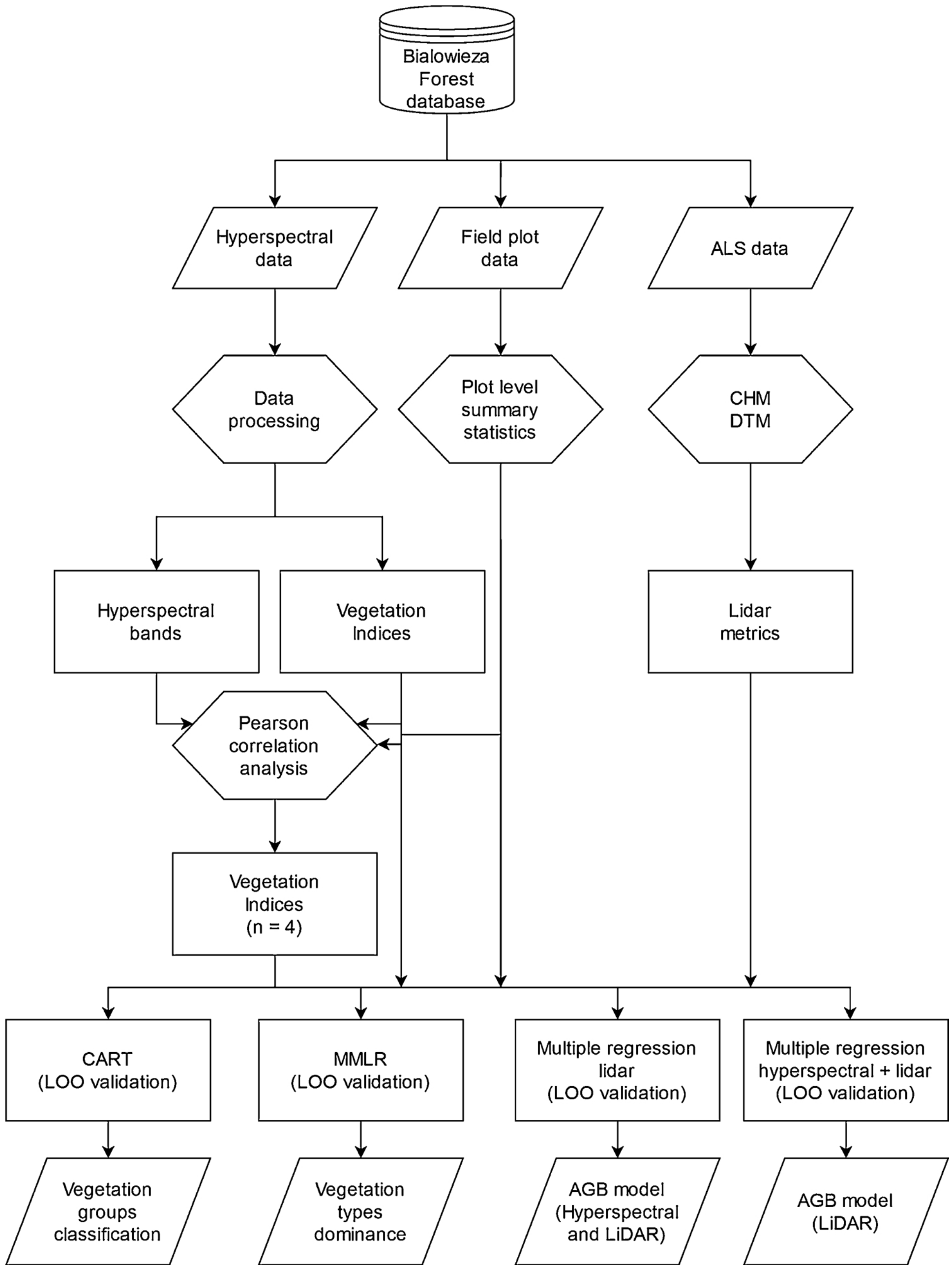

Fig. 2. Flowchart of the methodological approach.

Table 2
Hyperspectral vegetation indices considered in this study. The Pearson correlation values are reported for the three indices and the three bands mostly correlated with the percentage of dominance in the Spruce, Pines, and Deciduous groups.

Dominance Class	Veg. Index	Pearson *r*	Band	Pearson *r*
Spruce	CAI_max (Daughtry, 2001)	0.89	80max (666.7 – 670.5 nm)	0.90
(7 plots)	MRESR_max (Sims and Gamon, 2002)	0.88	81max (669.9 – 673.7 nm)	0.89
	MRENDVI_max (Sims and Gamon, 2002)	0.87	79max (663.6 – 667.4 nm)	0.88
Pines	NDMI_sd (Gao, 1995)	−0.63	382sd (2134.2 – 2138 nm)	−0.67
(21 plots)	NDMI_cv (Gao, 1995)	−0.63	365sd (2042 – 2045.8 nm)	−0.66
	NDWI_min (Gao, 1995)	−0.63	242sd (1374.9 – 1378.7 nm)	−0.66
Deciduous	MTVI2_median (Haboudane et al., 2004)	−0.79	92median (705.1 – 708.9 nm)	0.75
(50 plots)	TVI_median (Broge and Leblanc, 2001)	0.79	91median (701.9 – 705.7 nm)	0.75
	TCARI_median (Haboudane et al., 2004)	−0.78	90median (698.7 – 702.5 nm)	0.75

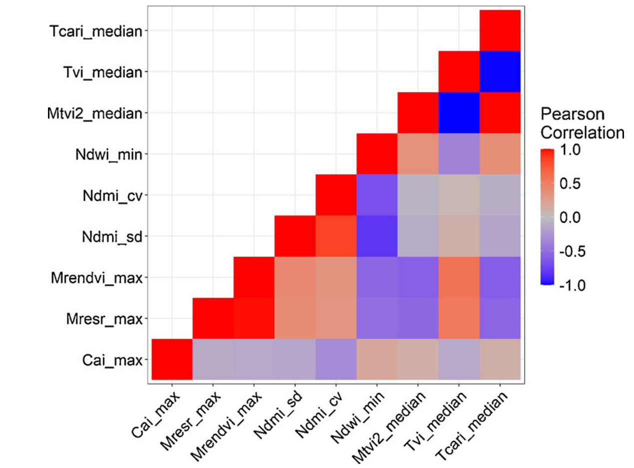

Fig. 3. Pearson correlation among the 9 indices presented in Table 2.

Table 3
CART results for the 3 groups (Spruce, Pines, and Deciduous) classification.

		Predicted		
		Deciduous	Pines	Spruce
Obs.	**Deciduous**	49	0	1
	Pines	0	19	2
	Spruce	0	3	4
	Overall accuracy	0.92		
	Variable importance Tcari_median	53.07		
	Variable importance Mresr_max	21.67		
	Variable importance Ndmi_sd	17.91		
	Variable importance Cai_max	7.35		

Table 4
CART results for the 4 groups (Spruce, Pines, Deciduous, and Mixed) classification.

		Predicted			
		Deciduous	Pines	Spruce	Mixed
Obs.	Deciduous	45	0	0	5
	Pines	0	19	0	2
	Spruce	0	3	0	4
	Mixed	0	5	0	17
	Overall accuracy	0.81			
	Variable importance Tcari_median	48.80			
	Variable importance Mresr_max	24.16			
	Variable importance Ndmi_sd	20.58			
	Variable importance Cai_max	6.45			

Table 5
MMLR results (LOO validated) for the prediction of species dominance in the BF plots.

Species	Adjusted R^2	RMSE%
Deciduous	0.87	13.1
Pines	0.61	19.5
Spruce	0.37	16.6

was validated using LOO, and the accuracy per each vegetation group is reported in Table 5, and scatterplots are presented in Fig. 4.

The MMLR results and the scatterplots indicate an accurate prediction of the level of dominance only for the Deciduous group, that resulted the dominant one in 47 plots (those with predicted percentage value > 60 %). This result is consistent with those obtained with CART classification, but here the added value is that MMLR allows to quantify the exact amount of Deciduous trees in a given plot.

Only in10 plots the predicted Pines percentages result above the 60 % threshold, while from field data the plots assigned to the Pines group were 21. Finally, none of the plots has a Spruce percentage > 60 %, against the 7 plots assigned to Spruce group from field records. In summary, the test shows with MMLR is possible to evaluate the amount of Deciduous and non-Deciduous trees in a given plot, but not to distinguish the amount of other groups.

3.4. Above ground biomass

An exponential multiple regression, validated with LOO approach, was used to predict the above ground biomass of the plots, using the inputs selected by the stepwise procedure, and checking the distribution of errors to exclude the presence of biases. The stepwise selection is based on Akaike Information Criteria computed among different models, and the inputs selected in the final models have an associated significance p-value.

Six groups of plots were tested:

- the four groups according to dominant species/group (Deciduous, n = 50; Pines, n = 21; Spruce, n = 7; and Mixed, n = 22)
- all the plots (n = 100)
- all the plots without the Mixed group (n = 78).

With lidar data, the 'mean plot height' metric based on all returns was always included, excepting for the Mixed group, for which the 'mean plot height' computed using all the returns above 1 m of height (thus excluding the returns from the lower understory) was included. For the Spruce group the regression is not robust, with no input being significant, due to low number of plots in modelling.

With lidar and hyperspectral joined data, the previously included lidar metrics remained unchanged. Hyperspectral included indices were: Pines (Tcari_median and Cai_max, both significant); all the plots and the Deciduous group (Tcari _median, not significant); all the plots without Mixed group (Tcari_median, significant). Fig. 5 illustrates the linearity of the relationship for predicted vs. observed AGB.

The accuracy of LOO validated results are presented in Table 6, together with the included inputs for each model. The results evidence that the addition of hyperspectral data in certain cases can improve the model accuracy.

4. Discussion and conclusions

Vegetation indices obtained from remote sensing are effective variables for quantitative and qualitative evaluations of vegetation cover, vigor, and growth dynamics, among other applications (Xue and Su, 2017).

The present study confirms the usefulness of VIs in this study area. All the vegetation indices that obtained higher variable importance in the classification models bring information on the chlorophyll content, showing that for the detection of Deciduous, Pines, and Spruce groups, the presence of narrowband indices from the red-edge portion of the spectra is critical.

The classification accuracies resulted high for both CART models, with and without the Mixed class. In the model that includes the four classes (Deciduous, Pines, Spruce, and Mixed), the Spruce class is completely confused with Pines and Mixed classes. Instead, in the model based on three classes only (without Mixed) the Spruce class is only partly confused with Pines. The major confusion that occurs when using four classes indicates that too much spectral similarity exists between Spruce and Mixed groups, in addition to unbalanced samples number. Therefore, even in presence of a satisfactory overall accuracy, it is advisable to exclude the use of a Mixed class in the classification efforts of this study area.

There are limitations in the classification approach that have to be

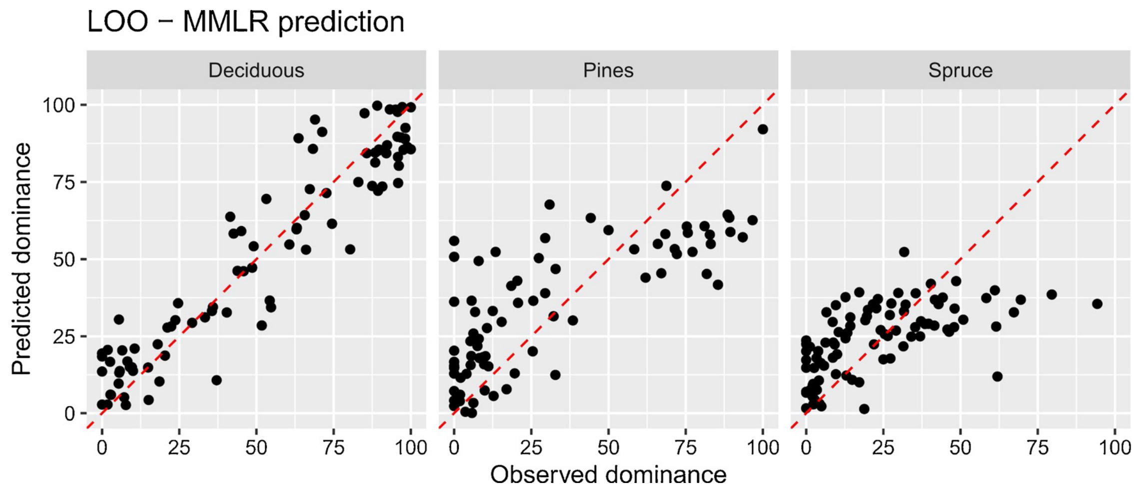

Fig. 4. Scatterplots for MMLR regressions predicted vs. observed values for level (%) of dominance in Deciduous, Pines and Spruce groups.

reminded. First, the accuracy results are referred to a reference database, characterized by its own inaccuracy (Kamińska et al., 2018); then, the procedure of increasing the field samples area (Stereńczak et al., 2018b) could introduce additional errors with respect to species composition. Even when keeping these limitations in mind, the present research illustrates an approach to cope with small forest plots, that are common reference data in forest analysis; it also highlights the value of hyperspectral data to characterize at fine scale the vegetation types of the study area.

The MMLR-based predictions of dominance levels are characterized by high variability in results accuracy; overall, the method provides satisfactory results only to estimate the amount of dominance for the Deciduous group (R^2 = 0.87). This is possibly due to spectral similarity between the other two groups and the unbalanced sampling among groups. Even if with this method is only possible to evaluate the percentage of Deciduous and non-Deciduous tree species in a given area, and not to detect the percentages of other groups, for BF conservation management this is already valuable information for management purposes, especially with respect to pests that often affect conifer species in BF forst. The results of this grouping approach could be improved by more balanced sampling; the approach is valuable for simplifying data collection and analysis, and thus facilitating repeated monitoring. This method could be applied to predict the composition of certain groups of species and monitor their dynamics, especially where

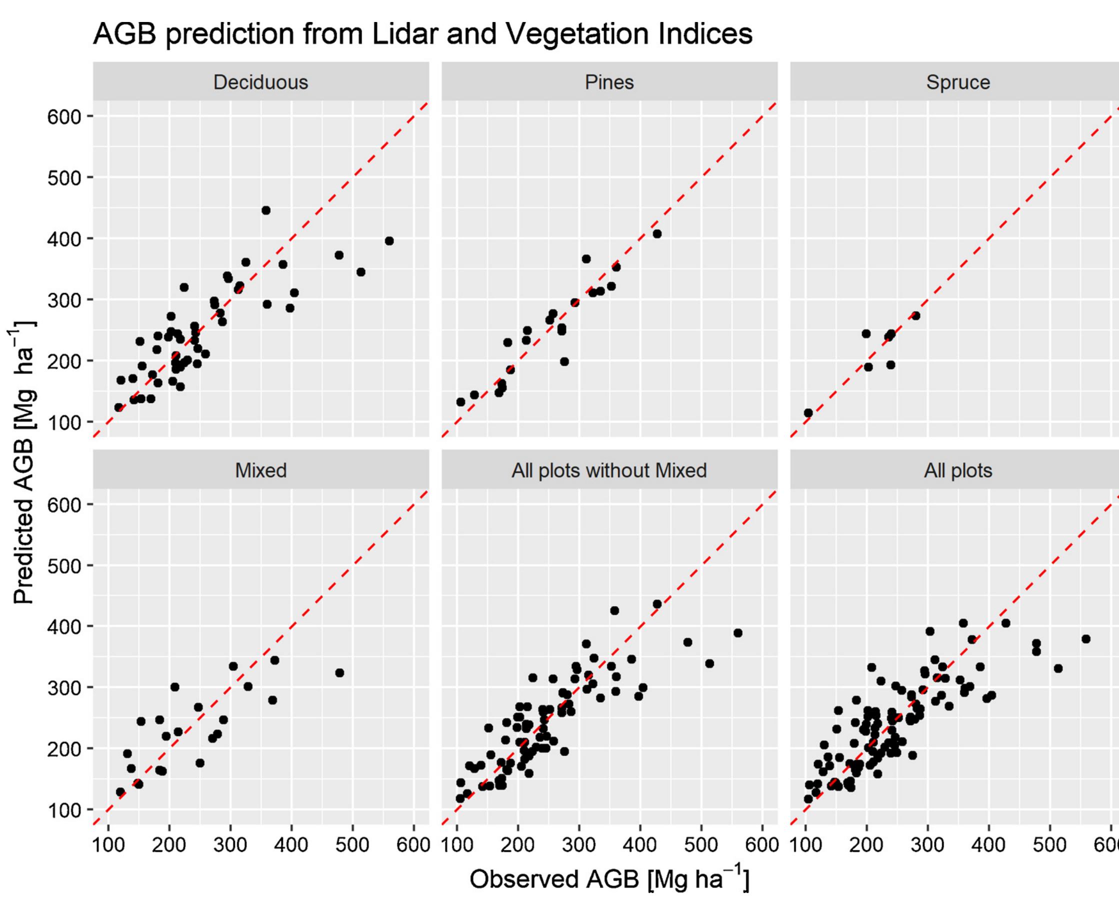

Fig. 5. Scatterplots of predicted vs. observed AGB values, according to stepwise regression and selected lidar metrics plus hyperspectral vegetation indices.

Table 6
Accuracy and inputs in exponential multiple regression tests, validated with LOO procedure. The * indicates not significant inputs (p value > 0.05).

Groups	Lidar adj. R^2	Lidar RMSE (%)	Inputs lidar	Lidar + Hyper_VI adj. R^2	Lidar + Hyper_VI RMSE (%)	Inputs lidar + hyperspectral VIs
Deciduous (n = 50)	0.67	23.56	Mean plot height (all returns)	0.69	24.06	Mean plot height (all returns) + *Tcari_median**
Pines (n = 21)	0.78	15.17	Mean plot height (all returns)	0.84	16.39	Mean plot height (all returns) + *Tcari_median, Cai_max*
Spruce (n = 7)	0.68	26.90	Mean plot height (all returns)* + Mean plot height (all returns above 1 m height)*	0.66	28.22	Mean plot height (all returns)* + Mean plot height (all returns above 1 m height)*
Mixed (n = 21)	0.57	28.17	Mean plot height (all returns above 1 m height)	0.57	28.17	Mean plot height (all returns above 1 m height)
All plots (n = 100)	0.66	22.36	Mean plot height (all returns)	0.67	21.03	Mean plot height (all returns) + *Tcari_median**
All plots without Mixed (n = 78)	0.69	20.95	Mean plot height (all returns)	0.72	22.60	Mean plot height (all returns) + *Tcari_median*

pest problems or accelerated climate impacts cause relevant changes in the arboreal assemblages.

The results of the AGB estimation using ALS data alone, and adding to ALS data the hyperspectral vegetation indices, were similar ($R^2 = 0.69$ and $R^2 = 0.70$, respectively). These results are comparable to those obtained for a similar ecosystem in a Czech forests, where Brovkina et al. (2015) obtained with lidar inputs in AGB regression a slightly higher accuracy ($R^2 = 0.77$, without cross-validation), with limited improvement from hyperspectral data addition ($R^2 = 0.79$). In a Finnish mixed forest Kankare et al. (2013) lidar based AGB estimate resulted in R^2 of 0.71. In a German temperate forest Latifi et al. (2012) achieved a relative error of 32–45 % for plot level biomass based on lidar metrics and of 35–45 % with joined hyperspectral inputs. However, other authors reported a significant improvement when using fused data from the two sensors (Anderson et al., 2008; Vaglio Laurin et al., 2014a, b).

It is worth to note that the procedure adopted to increase the plot area assumes that to similar structure corresponds similar amount of biomass. Above ground biomass also depends on species wood density, and errors in species composition could have been introduced, influencing the results. The procedure, although not perfect, allowed the objective detection of similar fragment of stands, and even considering the potential introduced errors, the obtained results are in line with those presented by other authors in similar ecosystems.

Furthermore, the AGB estimation results suggest that the improvement brought by the addition of hyperspectral data to ALS could occur for specific vegetation types: here a relevant increase in accuracy was observed for the Pines group (R^2 from 0.78 to 0.84) when adding chlorophyll/LAI (Tcari_median index) and senescence (CAI_max) information. Also Popescu et al. (2004) observed quite different accuracies in the estimates of AGB for Pines and Deciduous species, based on combined lidar and hyperspectral data. It is recognized that differences in the 3D structure of the species assemblages, or the density of the understory layer, influence the penetration of the lidar pulses (Wassihun et al., 2019). The predictive power of hyperspectral data can be higher when lidar relationships with biomass are weaker, favoring the joint data use. Hyperspectral data can also support the stratification of vegetation types prior to biomass estimate, to improve the estimates in selected vegetation types.

Overall, this research confirms the value of hyperspectral and ALS data to derive valuable forest characteristics in the BF area. ALS data proved to be suited to characterize forest above ground biomass. Vegetation indices based on red-edge bands were useful for detecting groups of species. Sentinel 2 data is equipped with red-edge bands, but it lacks the spatial resolution needed to fine-detail the forest composition, for which hyperspectral airborne data appear to be necessary. New opportunities could arise by the use of the hyperspectral satellite PRISMA, which data will be soon available; PRISMA is equipped with a panchromatic camera (5 m spatial resolution), that joined with the fine radiometric hyperspectral bands could improve the detection of groups of species or guilds without recurring to an airborne survey. However, for AGB estimation at fine scale, the use of airborne lidar data remains fundamental.

With the presented methods, a reliable description of the most important Bielowieza forest characteristics was produced: this kind of approach can simplify the monitoring operations in broad areas and could support local authorities in deriving useful information to conserve and sustainably manage the valuable BF resources.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Gaia Vaglio Laurin: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Nicola Puletti:** Software, Formal analysis, Methodology, Writing - review & editing, Validation. **Mirko Grotti:** Software, Formal analysis, Validation. **Krzysztof Stereńczak:** Data curation, Investigation, Methodology, Funding acquisition. **Aneta Modzelewska:** Data curation, Software. **Maciej Lisiewicz:** Data curation, Investigation. **Rafał Sadkowski:** Data curation, Software. **Łukasz Kuberski:** Data curation, Investigation. **Gherardo Chirici:** Writing - review & editing, Formal analysis. **Dario Papale:** Supervision, Resources.

Acknowledgments

This work is a result of networking between 3 UE LIFE projects: Samfix (Saving Mediterranean Forests from Invasions of Xylosandrus beetles and associated pathogenic fungi, LIFE17 NAT/IT/000609); ForBioSensing (Comprehensive monitoring of stand dynamics in Białowieża Forest supported with remote sensing techniques, LIFE13 ENV/PL/000048); and FRESh LIFE (Demonstrating Remote Sensing integration in sustainable forest management, LIFE14/IT000414). The research was supported by the Polish National Fund for Environmental Protection and Water Management (485/2014/WN10/OP-NM-LF/D).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jag.2020.102178.

References

Anderson, J.E., Plourde, L.C., Martin, M.E., Braswell, B.H., Smith, M.L., Dubayah, R.O., Blair, J.B., 2008. Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. Remote Sens. Environ. 112 (4), 1856–1870.
Arthur, M.A., Weathers, K.C., Lovett, G.M., Weand, M.P., Eddy, W.C., 2017. A beech bark disease induced change in tree species composition influences forest floor acid–base chemistry. Can. J. For. Res. 47 (7), 875–882.
Asner, G.P., Martin, R.E., 2009. Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. Front. Ecol. Environ. 7 (5), 269–276.
Barbati, A., Marchetti, M., Chirici, G., Corona, P., 2014. European forest types and forest Europe SFM indicators: tools for monitoring progress on forest biodiversity conservation. For. Ecol. Manage. 321, 145–157.
Barlow, J., Lennox, G.D., Ferreira, J., Berenguer, E., Lees, A.C., Mac Nally, R., Parry, L., 2016. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. Nature 535 (7610), 144.
Boczoń, A., Kowalska, A., Ksepko, M., Sokołowski, K., 2018. Climate warming and drought in the bialowieza forest from 1950–2015 and their impact on the dieback of Norway spruce stands. Water 10 (11), 1502.
Breiman, L., Friedman, J.H., 1997. Predicting multivariate responses in multiple linear regression. J. R. Stat. Soc. Series B Stat. Methodol. 59 (1), 3–54.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman & Hall/CRC.
Broge, N.H., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sens. Environ. 76 (2), 156–172.
Brovkina, O., Zemek, F., Fabiánek, T., 2015. Aboveground biomass estimation with airborne hyperspectral and LiDAR data in Tesinske Beskydy Mountains. Beskydy 8 (1), 35–46.
Cheng, Y., Zhang, C., Zhao, X., von Gadow, K., 2018. Biomass-dominant species shape the productivity-diversity relationship in two temperate forests. Ann. For. Sci. 75 (4), 97.
Chirici, G., McRoberts, R.E., Fattorini, L., Mura, M., Marchetti, M., 2016. Comparing echo-based and canopy height model-based metrics for enhancing estimation of forest aboveground biomass in a model-assisted framework. Remote Sens. Environ. 174, 1–9.
Chirici, G., Giannetti, F., McRoberts, R.E., Travaglini, D., Pecchi, M., Maselli, F., Chiesi, M., Corona, P., 2020. Wall-to-wall spatial prediction of growing stock volume based on Italian National Forest Inventory plots and remotely sensed data. Int. J. Appl. Earth Obs. Geoinf. 84, 101959.
Cholewińska, O., Keczyński, A., Smerczyński, I., Jaroszewicz, B., 2018. European ash (Fraxinus excelsior L.) dieback in a core area of Białowieża National Park. Natl. Parks Nat. Reserves 37 (2), 3–18 (in Polish).
Cienciala, E., Apltauer, J., Exnerová, Z., Tatarinov, F., 2008. Biomass functions applicable to oak trees grown in Central-European forestry. J. For. Sci. 54 (3), 109–120.
Czeszczewik, D., Zub, K., Stanski, T., Sahel, M., Kapusta, A., Walankiewicz, W., 2015. Effects of forest management on bird assemblages in the Bialowieza Forest, Poland. iForest-Biogeosci. For. 8 (3), 377.
Dalponte, M., Bruzzone, L., Gianelle, D., 2008. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. IEEE Trans. Geosci. Remote. Sens. 46 (5), 1416–1427.
Daughtry, C., 2001. Discriminating crop residues from soil by short-wave infrared reflectance. Agron. J. 93, 125–131.
Fox, J., Weisberg, S., 2011. An {R} Companion to Applied Regression, second edition. Sage, Thousand Oaks CA. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.
Gao, B., 1995. Normalized difference water index for remote sensing of vegetation liquid water from space. Proc. SPIE. Int. Soc. Opt. Eng. 2480, 225–236.
Grabska, E., Hostert, P., Pflugmacher, D., Ostapowicz, K., 2019. Forest stand species mapping using the sentinel-2 time series. Remote Sens. 11 (10), 1197.
Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. Remote Sens. Environ. 90 (3), 337–352.
Heinzel, J., Koch, B., 2012. Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. Int. J. Appl. Earth Obs. Geoinf. 18, 101–110.
Hocking, R.R., 1976. A Biometrics invited paper. The analysis and selection of variables in linear regression. Biometrics 32 (1), 1–49.
Hu, Y., Su, Z., Li, W., Li, J., Ke, X., 2015. Influence of tree species composition and community structure on carbon density in a subtropical forest. PLoS One 10 (8), e0136984.
Jaworski, T., Plewa, R., Tarwacki, G., Sućko, K., Hiszczański, J., Horák, J., 2019. Ecologically similar saproxylic beetles depend on diversified deadwood resources: from habitat requirements to management implications. For. Ecol. Manage. 449 (117462).
Johnson, R., Wichern, D., 2007. Applied Multivariate Statistical Analysis, sixth edition. Prentice-Hall.
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., Sadkowski, R., 2018. Species-related single dead tree detection using multi-temporal ALS data and CIR imagery. Remote Sens. Environ. 219, 31–43.
Kankare, V., Vastaranta, M., Holopainen, M., Räty, M., Yu, X., Hyyppä, J., Viitala, R., 2013. Retrieval of forest aboveground biomass and stem volume with airborne scanning LiDAR. Remote Sens. 5 (5), 2257–2274.
Kujawa, A., Orczewska, A., Falkowski, M., Blicharska, M., Bohdan, A., Buchholz, L., Nowak, S., 2016. The Białowieża forest–a UNESCO natural heritage site–protection priorities. For. Res. Pap. 77 (4), 302–323.
Larson, A.J., Belote, R.T., Williamson, M.A., Aplet, G.H., 2013. Making monitoring count: project design for active adaptive management. J. For. 111 (5), 348–356.
Latifi, H., Nothdurft, A., Koch, B., 2010. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. Forestry 83 (4), 395–407.
Latifi, H., Fassnacht, F., Koch, B., 2012. Forest structure modeling with combined airborne hyperspectral and LiDAR data. Remote Sens. Environ. 121, 10–25.
Latifi, H., Fassnacht, F.E., Müller, J., Tharani, A., Dech, S., Heurich, M., 2015. Forest inventories by LiDAR data: a comparison of single tree segmentation and metric-based methods for inventories of a heterogeneous temperate forest. Int. J. Appl. Earth Obs. Geoinf. 42, 162–174.
Leckie, D.G., Gougeon, F.A., Walsworth, N., Paradine, D., 2003. Stand delineation and composition estimation using semi-automated individual tree crown analysis. Remote Sens. Environ. 85 (3), 355–369.
Leutner, B.F., Reineking, B., Müller, J., Bachmann, M., Beierkuhnlein, C., Dech, S., Wegmann, M., 2012. Modelling forest α-diversity and floristic composition—on the added value of LiDAR plus hyperspectral remote sensing. Remote Sens. 4 (9), 2818–2845.
Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Lexer, M.J., 2010. Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. For. Ecol. Manage. 259 (4), 698–709.
Lucas, R., Bunting, P., Paterson, M., Chisholm, L., 2008. Classification of Australian forest communities using aerial photography, CASI and HyMap data. Remote Sens. Environ. 112 (5), 2088–2103.
Maleki, K., Kiviste, A., 2015. Effect of sample plot size and shape on estimates of structural indices: a case study in mature silver birch (Betula pendula Roth) dominating stand in Järvselja. For. Stud. 63 (1), 130–150.
Miścicki, 2016. Changes in the stands of the Białowieża National Park from 2000 to 2015. For. Res. Pap. 77 (4), 371–379.
Modzelewska, A., Fassnacht, F.E., Stereńczak, K., 2020. Tree Species Identification within an extensive forest area using airborne hyperspectral data. Int. J. Appl. Earth Observ. Geoinf. 84, 101960.
Persson, P.O., Strang, G., 2003.). Smoothing by savitzky-golay and legendre filters. Mathematical Systems Theory in Biology, Communications, Computation, and Finance. Springer, New York, NY, pp. 301–315.
Popescu, S.C., Wynne, R.H., Scrivani, J.A., 2004. Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. For. Sci. 50 (4), 551–565.
Puletti, N., Chianucci, F., Castaldi, C., 2018. Use of Sentinel-2 for forest classification in Mediterranean environments. Ann. Silvic. Res 42, 32–38.
R Development Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
Richter, R., Schläpfer, D., 2019. ATCOR-4 User Guide, Version 7.3. 0, April 2019. Atmospheric/Topographic Correction for Airborne Imagery. ReSe Applications LLC, Wil, Switzerland.
Schull, M.A., Knyazikhin, Y., Xu, L., Samanta, A., Carmona, P.L., Lepine, L., Myneni, R.B., 2011. Canopy spectral invariants, part 2: application to classification of forest types from hyperspectral data. J. Quant. Spectrosc. Radiat. Transf. 112 (4), 736–750.
Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens. Environ. 81 (2-3), 337–354.
Socha, J., Wężyk, P., 2004. Empirical formulae to assess the biomass of the above-ground part of pine trees. Forestry 7 (2), 04.
Stereńczak, K., Lisańczuk, M., Parkitna, K., Mitelsztedt, K., Mroczek, P., Miścicki, S., 2018a. Influence of number and size of sample plots on modelling growing stock volume based on airborne laser scanning. Wood 61, 201.
Stereńczak, K., Lisańczuk, M., Erfanifard, Y., 2018b. Delineation of homogeneous forest patches using combination of field measurements and LiDAR point clouds as a reliable reference for evaluation of low resolution global satellite data. For. Ecosyst. 5, 1.
Stereńczak, K., Mielcarek, M., Modzelewska, A., Kraszewski, B., Fassnacht, F.E., Hilszczański, J., 2019. Intra-annual Ips typographus outbreak monitoring using multi-temporal GIS analysis based on hyperspectral and ALS data in the Białowieża Forests. For. Ecol. Manage. 442, 105–116.
Szczygieł, R., Kwiatkowski, M., Kołakowski, B., 2016. Forest fire risk at Bialowieza primeval forest. Bezpieczeństwo i Technika Pożarnicza 43.
Szczygiel, R., Kwiatkowski, M., Kolakowski, B., 2018. Influence of bark beetle infestation on the forest fire risk in the Bialowieza Forest. SYLWAN 162 (11), 955–964.
Tabacchi, G., Di Cosmo, L., Gasparini, P., 2011. Aboveground tree volume and phytomass prediction equations for forest species in Italy. Eur. J. For. Res. 130 (6), 911–934.

Treitz, P., Howarth, P., 2000. High spatial resolution remote sensing data for forest ecosystem classification: an examination of spatial scale. Remote Sens. Environ. 72 (3), 268–289.
Vaglio Laurin, G., Chen, Q., Lindsell, J.A., Coomes, D.A., Del Frate, F., Guerriero, L., Valentini, R., 2014a. Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. Isprs J. Photogramm. Remote. Sens. 89, 49–58.
Vaglio Laurin, G.V., Chan, J.C.W., Chen, Q., Lindsell, J.A., Coomes, D.A., Guerriero, L., Valentini, R., 2014b. Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. PLoS One 9 (6), e97910.
Vaglio Laurin, G., Puletti, N., Hawthorne, W., Liesenberg, V., Corona, P., Papale, D., Valentini, R., 2016. Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. Remote Sens. Environ. 176, 163–176.
Vaglio Laurin, G., Pirotti, F., Callegari, M., Chen, Q., Cuozzo, G., Lingua, E., Papale, D., 2017. Potential of ALOS2 and NDVI to estimate forest above-ground biomass, and comparison with lidar-derived estimates. Remote Sens. 9 (1), 18.
Verkerk, P.J., Mavsar, R., Giergiczny, M., Lindner, M., Edwards, D., Schelhaas, M.J., 2014. Assessing impacts of intensified biomass production and biodiversity protection on ecosystem services provided by European forests. Ecosyst. Serv. 9, 155–165.
Wang, W.J., He, H.S., Thompson, F.R., Fraser, J.S., Dijak, W.D., 2017. Changes in forest biomass and tree species distribution under climate change in the northeastern United States. Landsc. Ecol. 32 (7), 1399–1413.
Wassihun, A.N., Hussin, Y.A., Van Leeuwen, L.M., Latif, Z.A., 2019. Effect of forest stand density on the estimation of above ground biomass/carbon stock using airborne and terrestrial LIDAR derived tree parameters in tropical rain forest, Malaysia. Environ. Syst. Res. 8 (1), 27.
Xue, J., Su, B., 2017. Significant remote sensing vegetation indices: a review of developments and applications. J. Sens. 2017.
Zhang, Y., Liang, S., 2014. Changes in forest biomass and linkage to climate and forest disturbances over Northeastern China. Glob. Chang. Biol. 20 (8), 2596–2606.
Żmihorski, M., Chylarecki, P., Orczewska, A., Wesołowski, T., 2018. Białowieża Forest: a new threat. Science 361 (6399), 238.