



## Article

# A Random Forest Modelling Procedure for a Multi-Sensor Assessment of Tree Species Diversity

Giorgos Mallinis <sup>1,\*</sup>, Irene Chrysafis <sup>1</sup>, Georgios Korakis <sup>2</sup>, Eleanna Pana <sup>1</sup> and Apostolos P. Kyriazopoulos <sup>3</sup>

<sup>1</sup> Forest Remote Sensing and Geospatial Analysis Laboratory, Democritus University of Thrace, GR 68200 Orestiada, Greece; echrysaf@fmenr.duth.gr (I.C.); eleannapana@gmail.com (E.P.)

<sup>2</sup> Laboratory of Forest Botany; Democritus University of Thrace, GR 68200 Orestiada, Greece; gkorakis@fmenr.duth.gr

<sup>3</sup> Laboratory of Range Science, Department of Forestry and Management of the Environment and Natural Resources, Democritus University of Thrace, GR 68200 Orestiada, Greece; apkyriaz@fmenr.duth.gr

\* Correspondence: gmallin@fmenr.duth.gr

Received: 20 March 2020; Accepted: 7 April 2020; Published: 9 April 2020



**Abstract:** Earth observation data can provide important information for tree species diversity mapping and monitoring. The relatively recent advances in remote sensing data characteristics and processing systems elevate the potential of satellite imagery for providing accurate, timely, consistent, and robust spatially explicit estimates of tree species diversity over forest ecosystems. This study was conducted in Northern Pindos National Park, the largest terrestrial park in Greece and aimed to assess the potential of four satellite sensors with different instrumental characteristics, for the estimation of tree diversity. Through field measurements, we originally quantified two diversity indices, namely the Shannon diversity index ( $H'$ ) and Simpson's diversity ( $D1$ ). Random forest regression models were developed for associating remotely sensed spectral signal with tree species diversity within the area. The models generated from the use of the WorldView-2 image were the most accurate with a coefficient of determination of up to 0.44 for  $H'$  and 0.37 for  $D1$ . The Sentinel-2 -based models of tree species diversity performed slightly worse, but were better than the Landsat-8 and RapidEye models. The coefficient of variation quantifying internal variability of spectral values within each plot provided little or no usage for improving the modelling accuracy. Our results suggest that very-high-spatial-resolution imagery provides the most important information for the assessment of tree species diversity in heterogeneous Mediterranean ecosystems.

**Keywords:** biodiversity indices; Sentinel-2; Landsat-8; RapidEye; machine learning; Mediterranean forest habitats; WorldView-2

## 1. Introduction

Forest ecosystems cover nearly one third of the Earth's land surface and contain over 80% of the terrestrial biodiversity [1]. Forest biodiversity expressing the variability among living organisms in forest ecosystems is an important factor for the terrestrial ecosystems' functions and process [2] and several studies have demonstrated the positive linkages between forest biodiversity and ecosystem stability and health [3]. Other studies have also demonstrated the role of forest biodiversity in forest ecosystem services provision, such as biomass production, carbon storage, wild fruits production, nutrient use and retention, and water provision [4,5]. Yet, direct and indirect factors, such as habitat loss, landscape degradation, fire, soil erosion, anthropogenic activities, demographics, and climate change threaten and lead to a decline in the forest biodiversity levels [6]. In particular, Mediterranean forest habitats, characterized as one of the richest biodiversity hotspots worldwide are among the most threatened on Earth [7,8], identified as priority area for conservation [9].

In such areas, biodiversity monitoring systems are needed for identifying the location, the magnitude, and the rate of changes in order to prevent or mitigate losses and enhance the resilience capacity of forest areas through appropriate management plans and conservation strategies.

Terrestrial ecological field surveys are the most accurate procedure for data collection on forest tree diversity, providing detailed and precise information for specific sites. However, the terrestrial sampling approach has many deficiencies related to personnel costs, time, and logistic support requirements [10], as well as site accessibility constraints over remote areas or regions with specific geographic, climatic and socioeconomic conditions [11]. In addition, the bias associated with the variability of the experts employed in the field measurement procedures can also lead to a lack of standardization among field measurements [12].

In contrast, remote sensing (RS), especially following improvements over the last two decades in the spatial, spectral, temporal and radiometric characteristics of earth observation sensors, as well as in the costs associated with RS data acquisition and analysis, can complement in-situ measurements for providing consistent, spatially explicit measurement of forest tree diversity [3,11,13,14]. In this respect, several studies have evaluated the use of a variety of sensors and methodological approaches over different forest environments, primarily focusing on  $\alpha$ -diversity estimation [15]. Remote sensing can be used for tree species diversity estimation both in an indirect or in a direct approach [16], with the latter involving individual targets mapping (communities or species) [17]. On the other hand, remote sensing can be used indirectly for species distribution mapping [14] or for the establishment of direct relationships between pixel values and field measured diversity [11,17,18]. This modelling procedure between field measured diversity values and remote sensing data, is the most widely used approach and it is closely related to the Spectral Variation Hypothesis (SVH) originally formulated by Palmer [18]. SVH relates the spectral diversity or the heterogeneity of the RS data to environmental diversity that corresponds to a higher number of species [2]. According to the SVH, species diversity is analogous to habitat heterogeneity [18], which determines the spectral response recorded by a remote sensing sensor, without considering the taxonomic group under study [19].

Previous studies had identified that the relationship between spectral diversity and biological diversity relationship is highly variable, related to parameters such as forest composition, sensor characteristics and modelling approach [3]. The influence of these parameters on the strength of the relationship can be identified either through simulation-based studies or empirical studies with different methodological settings.

For example, Foody and Carter [20] predicted species richness and abundance in tropical rain forests using artificial neural networks and a Landsat TM image. Landsat ETM+ imagery along with linear regression has also been widely used in related studies as by Mohammadi et al. [21] for tree diversity mapping in Iran, and Hernández-Stefanoni et al. [10] in a tropical forest in Mexico. More recently, Landsat-8 Operational Land Imager (OLI) imagery has also been evaluated for savanna woodland species diversity estimation [22]. Using hyperspectral data from the Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) sensor, Carlson et al. [23] assessed forest tree diversity in a tropical forest ecosystem using linear regression analysis. Hyperspectral data recorded by EO-1 Hyperion has also been used by Kalacska et al. [24] in a tropical dry forest. Very-high-spatial-resolution multispectral data from an Unmanned Aerial Vehicle (UAV) platform have also been explored for richness and diversity monitoring [25]. Some other studies, albeit few, have also used active sensors extensively for forest tree diversity measurement as in the study of Bouvier et al. [26], where airborne laser scanner (ALS) data has been assessed in a mixed forest area in France, for richness and diversity monitoring.

However, only in a few cases has a multiple-scale, multi-source framework been adopted for exploring the influence of sensor's different characteristics and resolution in the strength of the spectral diversity—tree species  $\alpha$ -diversity relationship. In such a comparative study, Nagedra et al. [27] developed regression models using images of different spatial and spectral resolution (i.e., Landsat-7 and IKONOS) in tropical dry forests. The performance of Sentinel-2 Multispectral Instrument (MSI)

and Landsat-8 OLI satellite sensors, has been recently evaluated by Toressani et al. [2] for tree species diversity estimation in an alpine conifer forest. Forest beta-diversity was evaluated by Khare et al. [28] in a subtropical deciduous forest using images acquired from Pléiades 1A, RapidEye, and Landsat-8 OLI satellite sensors. Ozdemir et al. [29] evaluated diversity correlation with texture measures calculated from RapidEye, SPOT-5, and Aster images. Finally, Chrysafis et al. [30], using a Lasso regression approach, evaluated Sentinel-2 MSI and RapidEye sensors for tree diversity estimation in a heterogeneous forest in northwest Greece.

Since only a limited number of studies have evaluated the effect of image spectral and spatial resolutions in the estimation of tree species diversity, the aim of this study was to assess, at multiple scales, tree species  $\alpha$ -diversity, using Landsat-8 OLI, Sentinel-2 MSI, RapidEye, and WorldView-2 optical satellite imagery, in forest habitats of Greece's largest terrestrial National Park.

Since tree species diversity detection from satellite imagery is expected to vary with image resolution [3], we selected a range of satellite sensors with different instrumental characteristics. In addition, Landsat-8 OLI and Sentinel-2 MSI images are accessible cost-free, offering the opportunity to minimize costs for forest ecosystems mapping and monitoring as for any project based on remote sensing [31].

Contrary to the majority of previous studies employing multiple linear regression, we adopted a random forest (RF) regression modelling procedure [32]. RF includes integrated cross-validation for assessing the accuracy of the model and is capable of ranking important variables and addressing data that interacts in a non-linear manner [32–36]. Furthermore, a drawback in the use of multiple linear regression relates to the assumptions of linearity and the absence of collinearity amongst variables. However, these assumptions are rarely observed in forest and remote sensing data, underlining the need for the use of more robust statistical methods [37].

Specifically, this study aims to answer the following questions: (i) Which are the most relevant spectral bands for tree diversity estimation? (ii) What is the strength of the relationship between remotely sensed data and field-measured diversity indices? (iii) Which sensor provides the best remote-sensing-based approach for tree species diversity mapping? And finally, (iv) is random forest regression an efficient approach for modelling tree diversity?

## 2. Study Area

The Northern Pindos National Park in northwest Greece has a total area 1969 sq. km and it is among the largest protected areas of the country, including eleven sites belonging to the Natura 2000 European Network of Protected Areas. The mountainous landscape comprises an intense relief with high topographical diversity. Due to the montane character of the area, the prevailing climate type is transitional from Mediterranean to continental and locally varies depending on elevation and aspect; the amount of precipitation is among the highest in Greece, a fact that is imprinted to the extended plant coverage of the area. Annual rainfall ranges between 1000 and 2250 mm, whereas the annual maximum temperature reaches 34 °C.

The vegetation at lower altitudes (400–1000 m) is composed of thermophilous, deciduous oak species, such as Hungarian oak (*Quercus frainetto*), Turkish oak (*Q. cerris*), Macedonian oak (*Q. trojana*), Downy oak (*Q. pubescens*). Pure or mixed stands of deciduous tree species, such as the hop hornbeam (*Ostrya carpinifolia*), oriental hornbeam (*Carpinus orientalis*), common hornbeam (*Carpinus betulus*), and manna ash (*Fraxinus ornus*), are also found in this zone. Most important forest ecosystems are found at higher altitudes (900–1600 m) in the National Park. They consist of pure and mixed stands of black pine (*Pinus nigra*), beech (*Fagus sylvatica*), and King's Boris fir (*Abies borisii-regis*). Bosnian pine (*Pinus leucodermis*) formation occupy the upper mountain slopes and form the tree limit, while sub-alpine grasslands extend above 1800 m to peaks. Riparian vegetation along the rivers and streams within the Park is comprised mainly of oriental plane (*Platanus orientalis*) galleries.

### 3. Materials and Methods

#### 3.1. Remote Sensing Data and Preprocessing

Since the main aim of the study was a comparative multi-resolution assessment of tree species diversity, we used sensor data with different spectral and spatial resolutions. The medium spatial resolution imagery evaluated in this study included Sentinel-2 MSI (S-2) and Landsat-8 OLI (L-8) images acquired on 25th August 2017 and 12th July 2017 respectively. Both sensors have spectral bands covering visible to shortwave infrared (SWIR) spectrum, and a 12-bit radiometric resolution. The spatial resolution of the Landsat-8 OLI bands used is 30m, while Sentinel-2 MSI bands are available at 10 m, 20 m, and 60 m spatial resolution (Table 1). The images provided by the European Space Agency (Sentinel-2) and United States Geological Survey (Landsat-8) were preprocessed at surface reflectance and orthorectified to correct for relief displacement.

**Table 1.** Instrumental characteristics of the multispectral images used in the present study.

Spectral Band	Landsat 8 OLI		Sentinel-2 MSI		RapidEye		WorldView-2	
	Central Wavelength (μm)/Spatial (m) Resolution		Central Wavelength (μm)/Spatial (m) Resolution		Central Wavelength (μm)/Spatial (m) Resolution		Central Wavelength (μm)/Spatial (m) Resolution	
Coastal (C)	B1	0.443/30	B1	0.443/60			B1	0.427/2
Blue (B)	B2	0.482/30	B2	0.490/10	B1	0.475/5	B2	0.478/2
Green (G)	B3	0.561/30	B3	0.560/10	B2	0.555/5	B3	0.546/2
Yellow (Y)							B4	0.608/2
Red (R)	B4	0.655/30	B4	0.665/10	B3	0.657/5	B5	0.659/2
Red edge 1 (RE1)			B5	0.705/20	B4	0.715/5	B6	0.724/2
Red edge 2 (RE2)			B6	0.740/20				
NIR narrow1 (NIR <sub>n1</sub> )			B7	0.783/20				
Near infrared (NIR)	B5	0.865/30	B8	0.842/10	B5	0.805/5	B7	0.831/2
NIR narrow 2 (NIR <sub>n2</sub> )			B8a	0.865/20				
Near infrared 2 (NIR 2)							B8	0.908/2
Shortwave infrared (SWIR1)	B6	1.609/30	B11	1.610/20				
Shortwave infrared (SWIR2)	B7	2.201/30	B12	2.190/20				

Higher spatial, but lower spectral, resolution RapidEye 3A level imagery acquired on 25th August 2017 was also evaluated for tree diversity mapping. RapidEye (RE) is a commercial optical Earth observation mission with a pixel-size of 6.5 m and five optical bands in the visible near-infrared part of the electromagnetic spectrum (VNIR), with a 12-bit dynamic range. The level 3A provides images radiometrically corrected at-sensor, orthorectified to a map projection, and resampled to a 5-m pixel size.

Commercial very-high-spatial-resolution (2 m) WorldView-2 images, available at Ortho Ready Standard Imagery (OR2A) level, acquired on 15th of July 2015, were also used. The OR2A are radiometrically and sensor corrected, mapped to the average base elevation of the terrain covered by each individual scene. For the orthorectification process of the WorldView-2 (WV-2) images, ground control points identified on very large scale orthoimages and a 5-m digital elevation model were used.

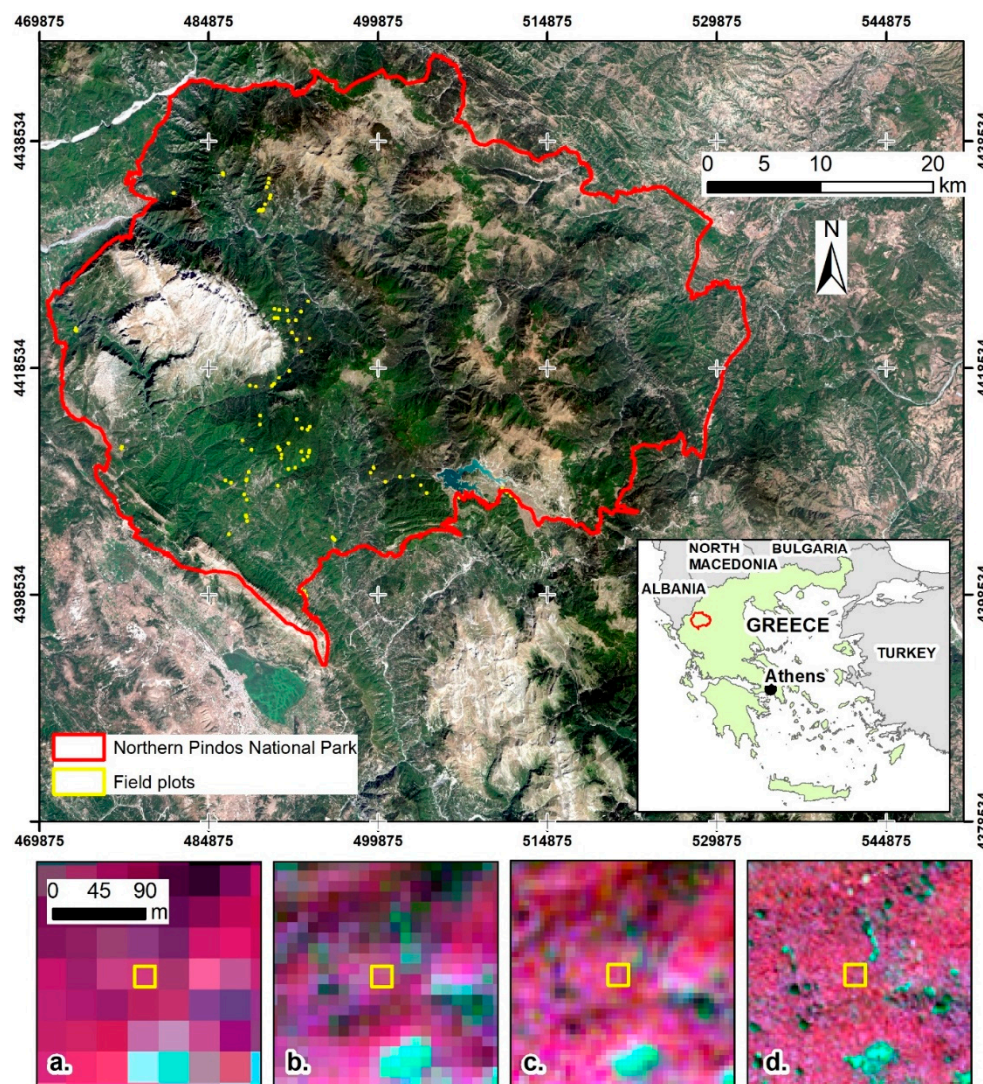
Finally, for the conversion of the RE and WV-2 DN values into surface reflectance, the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) algorithm/code was employed [38].

#### 3.2. Field Data Collection and Tree Diversity Indices

The field data were collected during the 2018 summer period (July - August). A total number of 100 sampling plots, 20×20 m each, were established using the gradsect method of survey [39] (Figure 1). To implement this method, field plots were located randomly along gradsects that were oriented along the strongest gradients of topographic variables. To establish the gradsects, the elevation, slope, and aspect of the area were used along with the road network. Within each plot, the coordinates from



both the center and one corner were recorded with high accuracy Global Navigation Satellite System (GNSS) in order to minimize positional inaccuracies. Then, the tree species were identified, and all the individual trees with a diameter at breast height (dbh) larger than 8 cm were recorded. As a result, the recorded trees included all the canopy overstory individuals as well as a number of the midstory and understory individuals whose diameter exceeded 8 cm. Records of the tree species in each plot were used to calculate diversity indices on the basis of the number of individuals (presence–absence).



**Figure 1.** Extent of the study area and 100 field plots' location. In the lower row, subsets of the Landsat-8 Operational Land Imager (OLI) (a), Sentinel-2 Multispectral Instrument (MSI) (b), RapidEye (c), and WorldView-2 (d) satellite sensors used in illustrating the spatial resolution differences (R: NearInfrared, G: Red, Blue: Green) are shown.

In order to compare tree diversity levels among regions, temporal periods, etc., the use of mathematical functions known as diversity indices is well established [40]. We used two indices (Table 2) to account for richness and evenness components of diversity. Species richness refers to the absolute number of species present in the ecosystem, while evenness considers the relevant abundance of the species. The Shannon Index ( $H'$ ) of diversity [41] takes both species abundance and species richness into account [42]. In addition, we used one other diversity index related to the original Simpson index  $D$  giving more emphasis to the evenness component of diversity [43]. The Simpson's

diversity ( $D1$ ) is the complement of Simpson's original formulation and represents the probability that the two individuals represent different species [40].

**Table 2.** Diversity indices used in the study.

Species Diversity Index	Formula	Reference
Shannon Index ( $H'$ )	$H' = -\sum_{i=1}^S p_i \times \ln(p_i)$	[41]
Simpson's diversity ( $D1$ )	$D_1 = 1 - \sum_{i=1}^S p_i^2$	[44]
where $p_i$ is the proportion of the $i$ th species in the sampling plot		

### 3.3. Spectral Information Extraction

Spectral information inherent in the original image bands of each sensor was considered for tree-diversity modelling. For the Landsat-8 OLI bands and Sentinel-2 MSI 20 and 60 m bands, pixel values corresponding to each plot center were extracted and used for the analysis. For all other images, the mean spectral values of the pixels falling within each plot were computed. Furthermore, for these images, having a finer pixel size than plot size, the Coefficient of Variation (CV) was computed as a measure for the quantification of the amount of the variability in forest vegetation within each plot [2,22].

### 3.4. Statistical Modelling

According to previous studies, linear regression models have been widely used for tree diversity patterns prediction [14]. Therefore, we originally examined the data in terms of the assumptions required for implementing a linear regression approach. In a preliminary step using the Shapiro–Wilk test and QQ normal probability plots, we identified that the requirements of linear regression for normality and homoscedasticity could not be met for our data. Therefore, in order to identify the correlation between spectral information and diversity indices, we used the Spearman rank coefficient, a non-parametric measure of correlation [45]. Subsequently, image-based predictive models for the tree-diversity indices were developed using a Random Forest (RF) regression algorithm [32], as implemented in the randomForest package [46].

Originally, using a Spearman's rank analysis, we eliminated highly correlated variables in order to minimize instability during variable importance computation. Then, we applied a variable selection procedure based on backward elimination process for determining the variables of the highest predictive value [47]. We selected this approach since several studies have demonstrated that the identification and inclusion of the most important variables in a reduced parsimonious model improves the efficiency of prediction, minimizes the influence of noisy predictors and redundant data, and increases the interpretability of the final model [48,49].

More specifically, the variables of each image were compared one by one through the Spearman rank matrices. We then grouped these variables according to the area of the electromagnetic spectrum (i.e., visible near infrared and shortwave infrared) recorded by the respective sensor band, and we removed from further analysis the highly correlated variables ( $r$  higher or equal to 0.7 with a  $p$ -value smaller than 5%) with the highest ranking predictor in the respective part of the spectrum [47,50].

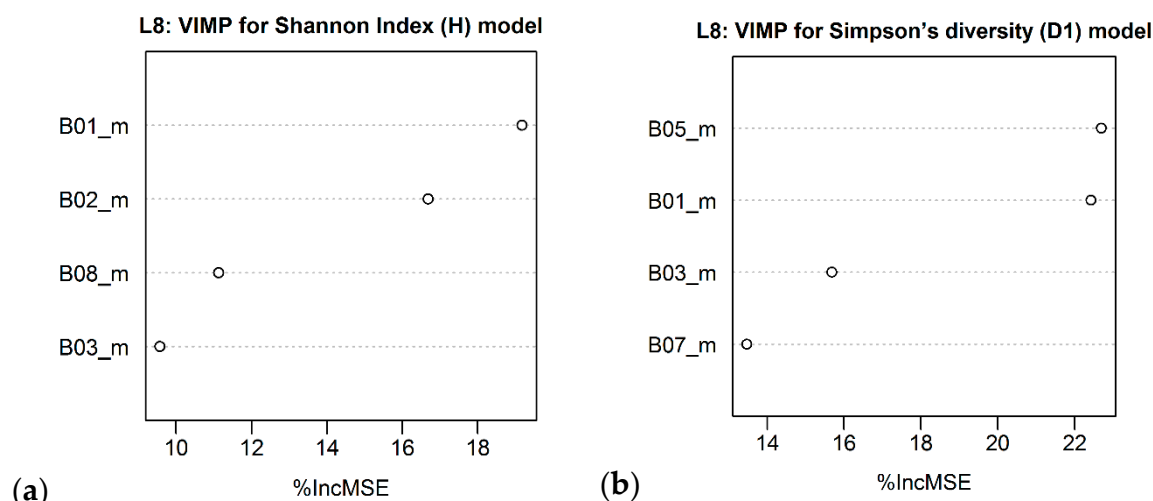
The remaining variables were used for the initial training of the RF models for each image, and the variable importance rankings were generated considering the percentage increase in the mean squared prediction error (%IncMSE) [32,51]. Finally, a backward elimination selection procedure was followed so as to produce the most accurate model with the smallest subset of variables retained for each sensor [52].

For all models, the number of regression trees grown (ntree) was optimized based on an out-of-bag estimate error, the number of random variables used in each tree (mtry) and the minimal size of the

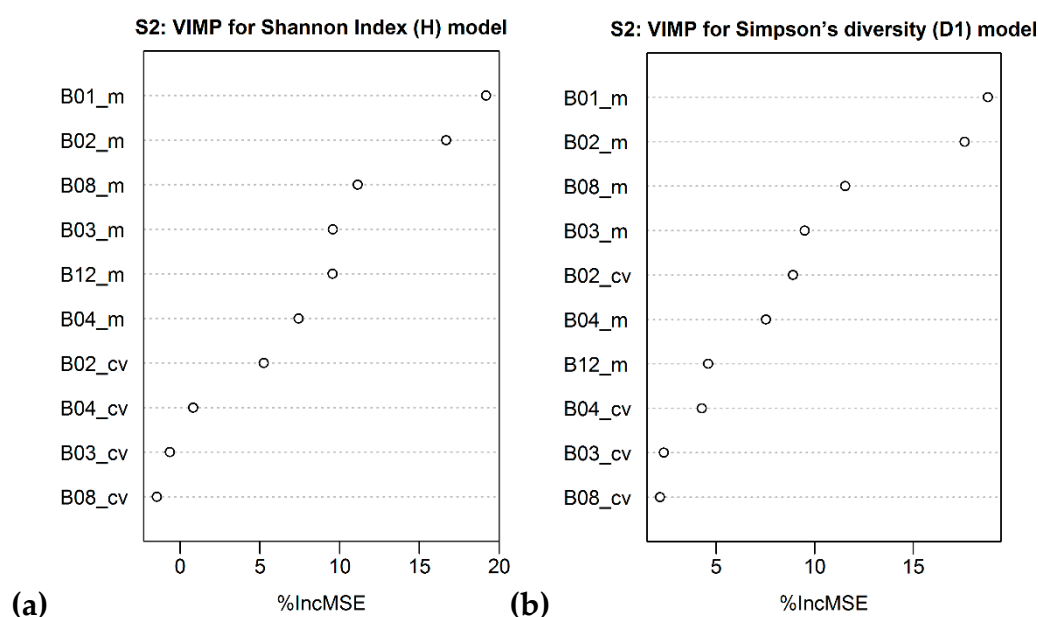
terminal nodes of the trees (nodesize) was set to default values [53]. All models were evaluated based on their performance to predict the field measured indices. Thus, the coefficient of determination ( $R^2$ ), the Root Mean Square Error (RMSE), and the Relative Squared Error (RSE) have been calculated following a 10-fold cross-validation.

#### 4. Results

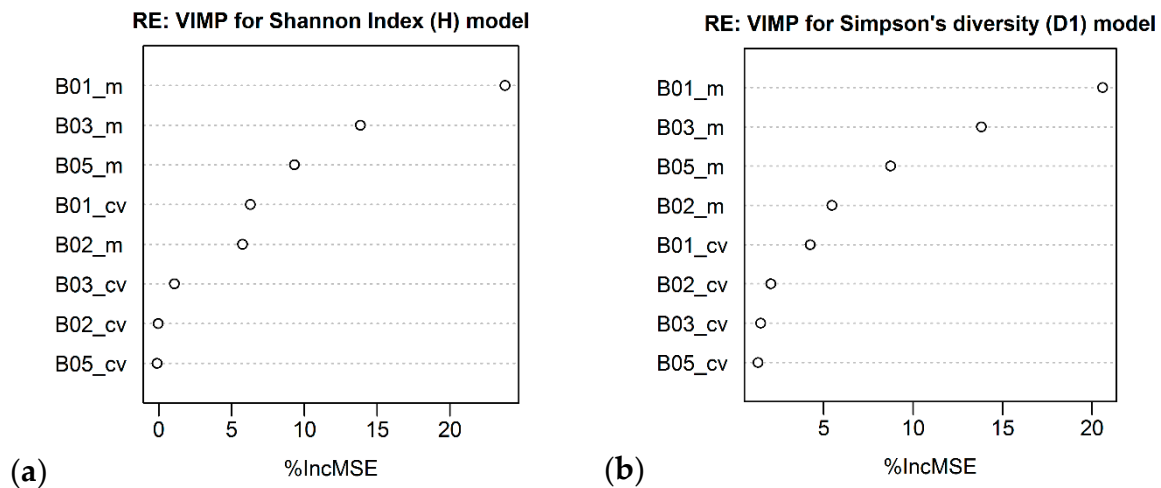
The identification of the variables with the most predictive power was carried out for each image using the percentage increase in the mean squared prediction error (%IncMSE) of RF models (Figures 2–5).



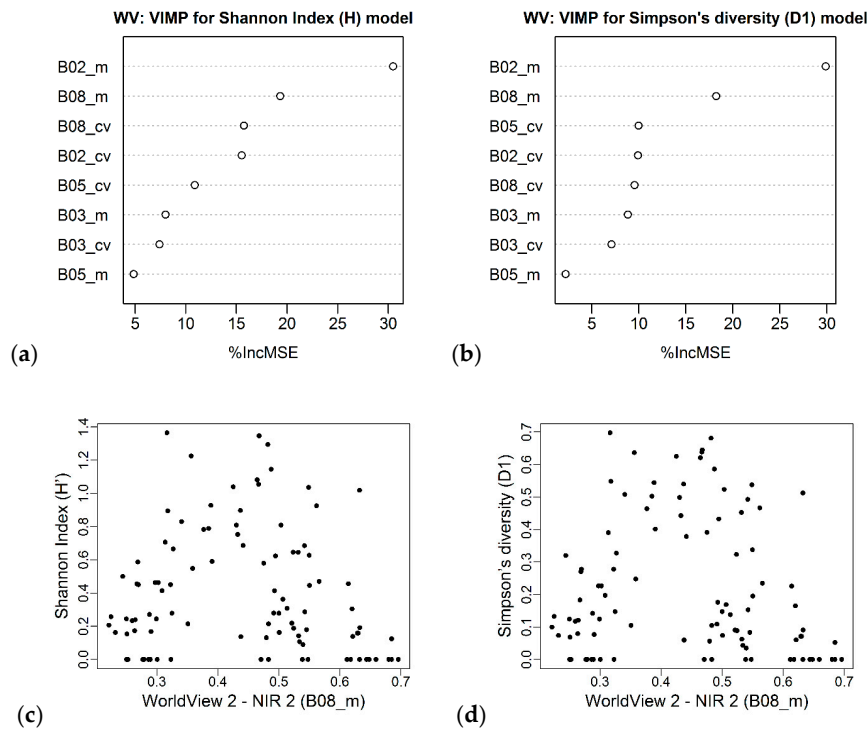
**Figure 2.** Importance of Landsat-8 OLI bands using the percentage increase in mean squared prediction error (%IncMSE) for (a) Shannon's ( $H'$ ) and (b) Simpson's diversity ( $D1$ ) indices. Higher %IncMSE indicates greater variable importance.



**Figure 3.** Importance of Sentinel-2 MSI bands using the percentage increase in mean squared prediction error (%IncMSE) for (a) Shannon's ( $H'$ ) and (b) Simpson's diversity ( $D1$ ) indices. Higher %IncMSE indicates greater variable importance.



**Figure 4.** Importance of RapidEye bands using the percentage increase in mean squared prediction error (%IncMSE) for (a) Shannon's ( $H'$ ) and (b) Simpson's diversity ( $D1$ ) indices prediction. Higher %IncMSE indicates greater variable importance.



**Figure 5.** Importance of WorldView-2 bands using the percentage increase in mean squared prediction error (%IncMSE) for (a) Shannon's ( $H'$ ) and (b) Simpson's diversity ( $D1$ ) indices. Scatterplots of (c) Shannon's ( $H'$ ) and (d) Simpson's diversity ( $D1$ ) indices with the second NIR band (B08\_m) of WorldView-2 are also presented.

Figure 2 presents the hierarchical ranking of the relative importance for the predictor bands of Landsat-8 dataset using the percentage increase in mean squared prediction error (%IncMSE). The near-infrared band (B05) was the most important variable for both diversity indices ( $H'$ ,  $D1$ ). The coastal band (B01) band was ranked second for the prediction of  $H'$  and  $D1$  indices.

In the case of the Sentinel-2 MSI dataset (Figure 3), the coastal band (B01) band was the most important variable for both diversity indices, followed by the green band (B03). The results also indicate that mean spectral values are in general more important than the coefficient of variation

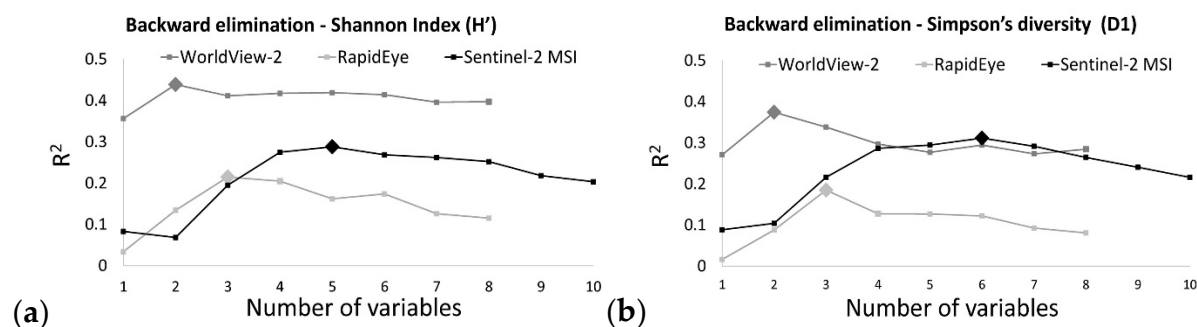


quantifying spectral variation within each plot. The coefficient of variation in the blue band (B02) was the most important among these spectral variation measures. The mean blue band (B02) was also ranked higher than most of the variables in both diversity indices.

Results on the importance of RapidEye bands (Figure 4) indicate that the blue (B01) and red (B03) bands were the most significant variables for both diversity indices. Moreover, in RapidEye models, spectral variation within each plot was less important than the average spectral values recorded.

Finally, in the case of WorldView-2 dataset, the blue (B02) and the second NIR band (B08) were ranked higher from all other WorldView-2 bands. Even though none of the CV features are ranked first, the plots indicate an increased importance of spectral variation compared to lower spatial resolution Sentinel-2 MSI and RapidEye sensors.

Figure 5 also illustrates the non-linear nature of the relationship between the second NIR band (B08\_m) of WorldView-2 and (c) Shannon's ( $H'$ ) and (d) Simpson's diversity ( $D1$ ) indices, confirming the use of random forest approach for developing remote-sensing-based predictive models for tree diversity. Based on the results of variable selection using the backward elimination method for the different images (Figure 6), the higher relationship between field measured diversity indices and the remotely sensed data is identified for Shannon's  $H'$  index (Table 3).



**Figure 6.** The optimal number of variables based on the backward elimination process for estimating (a) Shannon index ( $H'$ ) and (b) Simpson's diversity ( $D1$ ), using Random Forest models.

The most accurate model for  $H'$  is derived from the WorldView-2 data (Table 3) considering two variables ( $R^2 = 0.44$ ,  $RMSE = 0.29$ ). The sentinel-2 MSI model included five variables ( $R^2 = 0.29$ ,  $RMSE = 0.35$ ), slightly higher than the models of Landsat-8 ( $R^2 = 0.27$ ,  $RMSE = 0.34$ ) and RapidEye models ( $R^2 = 0.21$ ,  $RMSE = 0.35$ ).

WorldView-2 image was also more accurate for Simpson's diversity  $D1$  prediction ( $R^2 = 0.37$ ,  $RMSE = 0.19$ ). In this case, the Sentinel-2 MSI images were also highly correlated with  $D1$  ( $R^2 = 0.31$ ,  $RMSE = 0.20$ ), including six variables in the respective model. Landsat-8 ( $R^2 = 0.24$ ,  $RMSE = 0.20$ ) and RapidEye ( $R^2 = 0.18$ ,  $RMSE = 0.21$ ) models presented slightly lower accuracies (Table 3).

**Table 3.** Goodness-of-fit measures between predicted tree species diversity from Landsat 8 Operational Land Imager (OLI), Sentinel-2 Sentinel-2 Multispectral Instrument (MSI), RapidEye and WorldView 2 imagery and observed Shannon Index ( $H'$ ) and Simpson's diversity ( $D1$ ), as well as variables included in the optimal models. The values in the parenthesis correspond to the accuracy ( $R^2$ ) obtained when the full variable set for an image was used in the random forest (RF) models.

Sensor	Shannon Index ( $H'$ )		Simpson's diversity ( $D1$ )	
	Variables	Accuracy	Variables	Accuracy
Landsat – 8 OLI	B05_m	$R^2 = 0.27$ (0.25) RMSE = 0.34 RSE = 0.73	B05_m	$R^2 = 0.24$ (0.22) RMSE = 0.20 RSE = 0.76
	B01_m		B01_m	
	B03_m		B03_m	
	B07_m		B07_m	
Sentinel-2 MSI	B01_m	$R^2 = 0.29$ (0.27) RMSE = 0.33 RSE = 0.71	B01_m	$R^2 = 0.31$ (0.28) RMSE = 0.20 RSE = 0.69
	B02_m		B02_m	
	B08_m		B08_m	
	B03_m		B03_m	
	B12_m		B02_cv B04_m	
RapidEye	B01_m	$R^2 = 0.21$ (0.18) RMSE = 0.35 RSE = 0.79	B01_m	$R^2 = 0.18$ (0.14) RMSE = 0.21 RSE = 0.82
	B03_m		B03_m	
	B05_m		B05_m	
WorldView 2	B02_m	$R^2 = 0.44$ (0.43) RMSE = 0.29 RSE = 0.56	B02_m	$R^2 = 0.37$ (0.38) RMSE = 0.19 RSE = 0.63
	B08_m		B08_m	

## 5. Discussion

Remote sensing has become increasingly popular for mapping and monitoring the status of tree diversity [3,54]. Nowadays, with the increased number of remote sensing systems available, their improved instrumental characteristics, easier access, and increased affordability, the challenge relates to the efficient exploitation of the information content and better integration of the remotely sensed data in the diversity monitoring and conservation process [16]. Research is needed for assessing the scientific applicability of available sensors and the appropriate scale of measurement over different forest ecosystems [13] for complementing and sustaining, or even replacing [55] the field based measurements of tree diversity.

The comparative analysis of the models developed from WorldView-2, RapidEye, Sentinel-2 MSI and Landsat-8 OLI imagery, indicated that the WorldView-2 imagery was able to better approximate the tree species diversity indices in this study site in Northwest Greece. Sentinel-2 MSI models outperformed Landsat-8 OLI image models, while RapidEye models were ranked fourth in terms of productive accuracy. The advantage of the WorldView-2 dataset can be attributed to the very high spatial resolution of the images, which allows individual tree crowns to be visually identified [56]. Previous studies have also indicated that tree diversity prediction through spectral responses varies with spatial scale [3]. Khare et al. [28] calculated various diversity indices for different spatial resolutions and found that tree diversity was better explained at the finest (i.e., 0.5 m) pixel size evaluated. However, in another study in a dry tropical forest environment, it was found that Landsat imagery presented better correlation with tree diversity compared to high-spatial-resolution IKONOS data [27]. The authors related this finding to the very fine scale of IKONOS data for the specific landscape, decomposing tree crowns to sunlit and shadow-covered pixels, suggesting that larger pixel sizes than a single crown reduce the spectral variability as the reflectance signals is the average of larger ranges of surface conditions [27].

Large-scale simulation studies through field spectroscopy can be found in other ecosystem settings (i.e., grasslands), where studies suggested that the detection of  $\alpha$ -diversity declined with lower spatial

resolution and the optimal pixel size for  $\alpha$ -diversity prediction approximates the size of an individual plant leaf or crown [57]. Yet again, in such ecosystem types, other experimental studies present findings indicating a controversy regarding the optimal pixel size [58]. In forest areas, the challenge is aggravated by the fact that optimal spatial resolution is not constant, being primarily affected by the spatial and structural parameters of the forest stands [59]. Therefore, there is a need for experimental studies across biomes, ecosystems, and sites for developing, if possible, operational, generic, applicable approaches for the remote sensing of forest biodiversity [3].

Sentinel-2 MSI and Landsat-8 OLI images, having a spatial resolution not allowing for the direct delineation of individual tree crowns, presented relatively lower levels of accuracy than the WorldView-2 images. The Landsat-8 OLI models developed in our study, presented similar levels of accuracy with the models developed by Madonsela et al. [22] in a savanna woodland area and by Nagendra et al., [27] in a dry tropical forest. Sentinel-2 MSI models, despite the lower accuracy in comparison to WorldView-2 models, presented satisfactory results ( $R^2 = 0.31$  for Simpson's diversity). This could be explained by the "surrogacy" concept [3], which suggest that spectral diversity is linked with various types of field tree diversity through physical and ecological rules. This results in a satisfactory correlation between spectral measurements and field-measured diversity, even when pixel size is larger than the size of individual trees.

Medium spatial resolution remotely sensed data, such as the Sentinel-2 MSI imagery available free to the research community, can have practical application as a screening tool to identify critical tree diversity areas [22]. The same data can be fused with information on habitat types, soil, climate, etc., for developing complex predictive models of tree diversity [27]. The RapidEye image was found less suited from all images for assessing tree species diversity despite having a finer spatial scale than Landsat-8 OLI and Sentinel-2 MSI images. A possible reason for this might be the lower spectral resolution of the RapidEye image. Previous studies suggested that increased spectral resolution resulted in an increase in accuracy of diversity estimation [60] and that a large number of spectral bands is preferable to few bands [61]. But this might be a weak hypothesis since the higher spatial resolution models of the WorldView-2 images relied on a limited number of variables (two to four). Potential radiometric distortions caused by resampling the original nominal 6.5-m spatial resolution to a 5-m pixel size (Level 3A) might be another reason driving the lower accuracy of RapidEye diversity models.

The lower accuracy of the RapidEye image might also be related to the absence of information on the SWIR part of the spectrum. This band has been incorporated in almost all Sentinel-2 MSI and Landsat-8 OLI diversity models. SWIR has essential spectral information useful for characterization of vegetation [22] and has been strongly correlated with species richness in a tropical forest ecosystem [62]. SWIR spectral bands have been also reported to provide critical information for enhancing the predictive accuracies regarding forest attributes along with the lower the visible part of the spectrum (notably blue band) and forest structure [63,64]. The blue band was also a very important variable for local diversity estimation in all sensors. The linkages the blue part of the electromagnetic spectrum and forest structure were also noted in Mediterranean [63] and boreal forest environments [65]. In general spectral features in the visible region (400–700 nm) are reflective of leaf chemistry and pigment content, which results in strong absorption features [66]. Specifically, the relative high importance of the CV of the blue band for the Sentinel-2 MSI and WorldView-2 data, might be related to the absorption in the blue by carotenoids [67].

Blue band, adjacent to the coastal one, was also a very important variable for the final Sentinel-2 MSI, RapidEye and Worldview-2 models, similar to the findings of Nagedra et al. [27]. Interestingly, the blue band presented very high correlation with the coastal one, only in the case of the Landsat-8 OLI image (Spearman rank correlation = 0.966) and therefore it was removed from the subsequent RF modelling procedure.

The NIR band, which is widely considered as a key band for diversity estimation [60], has been also included in all diversity models. Information from the coastal part of the electromagnetic spectrum was also a very important for the Sentinel-2 MSI and Landsat-8 OLI models. The reflectance measured

in the coastal band is related to the chlorophyll content of plants and often used for forest biomass and carbon estimations [68] and its' strength for species richness estimation has been demonstrated in a savanna woodland [22].

In a previous study, it was identified that the inclusion of the red-edge spectral information was an essential parameter for tree diversity estimation, equal or even higher than the spatial resolution [28]. In our study, the red-edge bands of the Sentinel-2 MSI, RapidEye, and WorldView-2 images presented a weak relationship with all diversity indices and high correlation with the NIR bands, as indicated by the Spearman measure, and therefore were not included in the RF modelling procedure.

In order to quantify the spectral heterogeneity within each plot, the coefficient of variation was calculated for the 10-m Sentinel-2 MSI, RapidEye, and Worldview-2 bands. However, both the variable importance plots and the final variable selected within each model confirmed the weak relationship of this measure of spectral variation with tree diversity. Madonsela et al. [22] also noted that the relationship between spectral indices and measures of tree species diversity declined significantly when surrogate measures of variability were used as predictors of tree species diversity in a tropical savanna environment. Nagendra et al. [27] also reported that spectral/spatial variation of reflectance values within field sampling plots provides either non-significant or weakly significant correlations with species richness and diversity. Meng et al. [37], using multispectral (10-m) and panchromatic (2.5-m) Spot-6 data for forest diversity estimation, also noted that no measures related to the internal variability of forest stands (i.e., texture) were included in the final models. They suggested that this is related to the fact that diversity indices measure only a quantitative aspect of diversity, not being able to correlate with the qualitative spectral variability resulting from different types of mixture, i.e., the pronounced spectral differentiation between an evergreen conifer and a deciduous species in contrast with the similar spectral values between two conifer species. Nevertheless, as Lu and Weng underline [69], the importance of adding information related to the variability of spectral values, increases as the spatial resolution increases.

Unlike to the majority of the previous studies, we did not use a linear regression approach for assessing the relationship between remote sensing data and field the measurements of tree species diversity. RF regression does not require any data distribution assumptions and can detect interactions and higher order relationships between independent variables without a priori specification of these terms [50]. Furthermore, RF are particularly appealing due to their ability to generalize even under a limited training samples regime, as often is the case in remote sensing applications [48]. In our study, the relatively modest or low accuracy achieved by the RF models, might also be related to the field data collection strategy for the tree diversity indices calculation. Since we considered all trees with dbh larger than 8 cm, a small number of these individuals included in the sample may not be directly observed from optical sensor systems. Nevertheless, increased presence of understory species is closely related to canopy gaps, resulting also in an increased spectral contribution of lower canopy strata and understory to forest reflectance [70].

We adopted a two-step process, removing highly intercorrelated variables for reducing the computational complexity of the algorithm and the dimension of the input data. Several related publications have shown that the predictive power of RFs may benefit from variable selection [50,71]. Finally, in terms of the most appropriate diversity index to be used with remote sensing data, the results indicate that, in our site, a remotely sensed spectral signal correlates better with Shannon's Index than Simpson's diversity, since, across all sensors, the Sentinel-2 MSI model provided the highest coefficient of determination. However, similar to the insights from previous in-depth comparative studies [44], the identification of one ideal diversity index might not be feasible.

## 6. Conclusions

In this study, we compared remotely sensed data from four different sensors, namely Landsat-8 OLI, Sentinel-2 MSI, RapidEye, and WorldView-2, for tree species local diversity estimation over



the forest habitats of the Northern Pindos National Park in northwest Greece using a random forest modelling procedure.

The models relying on the very-high-spatial-resolution WorldView-2 image provided the highest accuracy for the prediction of the richness and evenness components of diversity. The use of the Sentinel-2 MSI imagery also provides an attractive option for tree species diversity estimation considering the accuracy results, when compared to the relative higher spatial resolution RapidEye imagery.

The variable selection procedure improved the predictive power and robustness of RF models compared to the original full dataset models and allowed for the identification of the important spectral bands for tree diversity estimation. The Coefficient of Variation (CV), quantifying internal variability within each plot, provided little or no usage for improving the modelling accuracy. This research confirmed that the  $\alpha$ -diversity is scale and sensor dependent. Freely available Sentinel-2 MSI images characterized by a good compromise between spatial and spectral resolutions, seems promising for providing accurate  $\alpha$ -diversity index estimates through remote sensing techniques. This study, as well as previous ones, emphasized the tradeoffs existing between spatial and spectral resolutions for local tree diversity estimation. Further research is required to determine the optimal combination of spectral and spatial resolutions over different biomes and forest stand characteristics.

**Author Contributions:** Conceptualization, G.M.; Methodology, G.M., G.K. and I.C.; Investigation, G.M., I.C. and G.K.; Data Analysis G.M., I.C. and G.K.; Writing—Original Draft Preparation, G.M. and I.C.; Writing—Review & Editing, G.M., I.C., G.K., E.P., A.P.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the project “Conservation and sustainable capitalization of biodiversity in forested areas-BIOPROSPECT” (Reg. No: BMP1/2.1/2336/2017), implemented within the framework of INTERREG V-A COOPERATION PROGRAMME BALKAN MEDITERRANEAN 2014 – 2020 programme, co-funded by the European Union and national funds of the participating countries.

**Acknowledgments:** The Landsat-8 OLI and Sentinel-2 MSI data were available from the US Geological Survey and European Space Agency at no-cost. The authors are grateful to Athanasios Stampoulidis and the staff of the Northern Pindos National Park Management Body for their help during field sampling activities.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Aerts, R.; Honnay, O. Forest restoration, biodiversity and ecosystem functioning. *BMC Ecol.* **2011**, *11*, 29. [[CrossRef](#)] [[PubMed](#)]
2. Torresani, M.; Rocchini, D.; Sonnenschein, R.; Zebisch, M.; Marcantonio, M.; Ricotta, C.; Tonon, G. Estimating tree species diversity from space in an alpine conifer forest: The Rao's Q diversity index meets the spectral variation hypothesis. *Ecol. Inform.* **2019**, *52*, 26–34. [[CrossRef](#)]
3. Wang, R.; Gamon, J.A. Remote sensing of terrestrial plant biodiversity. *Remote Sens. Environ.* **2019**, *231*, 111218. [[CrossRef](#)]
4. Gamfeldt, L.; Snäll, T.; Bagchi, R.; Jonsson, M.; Gustafsson, L.; Kjellander, P.; Ruiz-Jaen, M.C.; Fröberg, M.; Stendahl, J.; Philipson, C.D.; et al. Higher levels of multiple ecosystem services are found in forests with more tree species. *Nat. Commun.* **2013**, *4*, 1340. [[CrossRef](#)]
5. Sugden, A.M. Tree diversity improves forest productivity. *Science* **2018**, *362*, 41–43.
6. Gauquelin, T.; Michon, G.; Joffre, R.; Duponnois, R.; Génin, D.; Fady, B.; Bou Dagher-Kharat, M.; Derridj, A.; Slimani, S.; Badri, W.; et al. Mediterranean forests, land use and climate change: A social-ecological perspective. *Reg. Environ. Chang.* **2018**, *18*, 623–636. [[CrossRef](#)]
7. Médail, F.; Monnet, A.-C.; Pavon, D.; Nikolic, T.; Dimopoulos, P.; Bacchetta, G.; Arroyo, J.; Barina, Z.; Albassatneh, M.C.; Domina, G.; et al. What is a tree in the Mediterranean Basin hotspot? A critical analysis. *For. Ecosyst.* **2019**, *6*, 17. [[CrossRef](#)]
8. Food and Agriculture Organization of United Nations (FAO). *State of Mediterranean Forests 2013 (SoMF 2013)*; Besacier, C., Garavaglia, V.D.S., Eds.; FAO: Rome, Italy, 2013; ISBN 9789251075388.
9. Myers, N.; Mittermeier, R.A.; Fonseca, G.A.B.; Fonseca, G.A.B.; Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **2000**, *403*, 853. [[CrossRef](#)]

10. Hernández-Stefanoni, J.L.; Gallardo-Cruz, J.A.; Meave, J.A.; Rocchini, D.; Bello-Pineda, J.; López-Martínez, J.O. Modeling  $\alpha$ - and  $\beta$ -diversity in a tropical forest from remotely sensed and spatial data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *19*, 359–368. [\[CrossRef\]](#)
11. Rocchini, D.; Hernández-Stefanoni, J.L.; He, K.S. Advancing species diversity estimate by remotely sensed proxies: A conceptual review. *Ecol. Inform.* **2015**, *25*, 22–28. [\[CrossRef\]](#)
12. Löhmus, A.; Löhmus, P.; Runnel, K. A simple survey protocol for assessing terrestrial biodiversity in a broad range of ecosystems. *PLoS ONE* **2018**, *13*, e0208535. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Anderson, C.B. Biodiversity monitoring, earth observations and the ecology of scale. *Ecol. Lett.* **2018**, *21*, 1572–1585. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Saatchi, S.; Giorgi, A.P.; Gillespie, T.W.; Rocchini, D.; Foody, G.M. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* **2008**, *32*, 203–221.
15. Whittaker, R.H. Evolution and measurement of species diversity. *Taxon* **1972**, *21*, 213–251. [\[CrossRef\]](#)
16. Turner, W.; Spector, S.; Gardiner, N.; Fladeland, M.; Sterling, E.; Steininger, M. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* **2003**, *18*, 306–314. [\[CrossRef\]](#)
17. White, J.C.; Gómez, C.; Wulder, M.A.; Coops, N.C. Characterizing temperate forest structural and spectral diversity with Hyperion EO-1 data. *Remote Sens. Environ.* **2010**, *114*, 1576–1589. [\[CrossRef\]](#)
18. Palmer, M.W.; Earls, P.G.; Hoagland, B.W.; White, P.S.; Wohlgemuth, T. Quantitative tools for perfecting species lists. *Environmetrics* **2002**, *13*, 121–137. [\[CrossRef\]](#)
19. Rocchini, D.; Boyd, D.S.; Féret, J.-B.; Foody, G.M.; He, K.S.; Lausch, A.; Nagendra, H.; Wegmann, M.; Pettorelli, N. Satellite remote sensing to monitor species diversity: Potential and pitfalls. *Remote Sens. Ecol. Conserv.* **2016**, *2*, 25–36. [\[CrossRef\]](#)
20. Foody, G.M.; Cutler, M.E.J. Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing. *J. Biogeogr.* **2003**, *30*, 1053–1066. [\[CrossRef\]](#)
21. Mohammadi, J.; Shataee, S. Possibility investigation of tree diversity mapping using Landsat ETM+ data in the Hyrcanian forests of Iran. *Remote Sens. Environ.* **2010**, *114*, 1504–1512. [\[CrossRef\]](#)
22. Madonsela, S.; Cho, M.A.; Ramoelo, A.; Mutanga, O. Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS J. Photogramm. Remote Sens.* **2017**, *133*, 116–127. [\[CrossRef\]](#)
23. Carlson, K.M.; Asner, G.P.; Hughes, R.F.; Ostertag, R.; Martin, R.E. Hyperspectral Remote Sensing of Canopy Biodiversity in Hawaiian Lowland Rainforests. *Ecosystems* **2007**, *10*, 536–549. [\[CrossRef\]](#)
24. Kalacska, M.; Sanchez-Azofeifa, G.A.; Rivard, B.; Caelli, T.; White, H.P.; Calvo-Alvarado, J.C. Ecological fingerprinting of ecosystem succession: Estimating secondary tropical dry forest structure and diversity using imaging spectroscopy. *Remote Sens. Environ.* **2007**, *108*, 82–96. [\[CrossRef\]](#)
25. Getzin, S.; Wiegand, K.; Schöning, I. Assessing biodiversity in forests using very high-resolution images and unmanned aerial vehicles. *Methods Ecol. Evol.* **2012**, *3*, 397–404. [\[CrossRef\]](#)
26. Bouvier, M.; Durrieu, S.; Gosselin, F.; Herpigny, B. Use of airborne lidar data to improve plant species richness and diversity monitoring in lowland and mountain forests. *PLoS ONE* **2017**, *12*, e0184524. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Nagendra, H.; Rocchini, D.; Ghate, R.; Sharma, B.; Pareeth, S. Assessing plant diversity in a dry tropical forest: Comparing the utility of landsat and ikonos satellite images. *Remote Sens.* **2010**, *2*, 478–496. [\[CrossRef\]](#)
28. Khare, S.; Latifi, H.; Rossi, S. Forest beta-diversity analysis by remote sensing: How scale and sensors affect the Rao's Q index. *Ecol. Indic.* **2019**, *106*, 105520. [\[CrossRef\]](#)
29. Ozdemir, I.; Mert, A.; Ozkan, U.Y.; Aksan, S.; Unal, Y. Predicting bird species richness and micro-habitat diversity using satellite data. *For. Ecol. Manage.* **2018**, *424*, 483–493. [\[CrossRef\]](#)
30. Chrysafis, I.; Mallinis, G.; Korakis, G.; Dragozi, E. Forest diversity estimation using Sentinel-2 and RapidEye imagery: A case study of the Northern Pindos National Park. In Proceedings of the Seventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2019), Cyprus, 18–21 March 2019; Papadavid, G., Themistocleous, K., Michaelides, S., Ambrosia, V., Hadjimitsis, D.G., Eds.; SPIE: Washington, DC, USA, 2019; p. 50.
31. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10. [\[CrossRef\]](#)
32. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)

33. Ehrlinger, J.; Rajeswaran, J.; Blackstone, E.H. ggRandomForests: Exploring random forest survival. *R Vignette* **2014**.
34. Genuer, R.; Poggi, J.; Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [\[CrossRef\]](#)
35. Evans, J.S.; Murphy, M.A.; Holden, Z.A.; Cushman, S.A. *Modeling Species Distribution and Change Using Random Forest BT—Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*; Drew, C.A., Wiersma, Y.F., Huettmann, F., Eds.; Springer: New York, NY, USA, 2011; pp. 139–159, ISBN 978-1-4419-7390-0.
36. Smith, P.F.; Ganesh, S.; Liu, P. A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *J. Neurosci. Methods* **2013**, *220*, 85–91. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Meng, J.; Li, S.; Wang, W.; Liu, Q.; Xie, S.; Ma, W. Estimation of forest structural diversity using the spectral and textural information derived from SPOT-5 satellite images. *Remote Sens.* **2016**, *8*, 125. [\[CrossRef\]](#)
38. ENVI. Atmospheric Correction Module: QUAC and FLAASH User's Guide. In *Atmospheric Correction Module*; Version 4.7; ITT Visual Information Solutions: White Plains, NY, USA, 2009; p. 44.
39. Gillison, A.N.; Brewer, K.R.W. The use of gradient directed transects or gradsects in natural resource surveys. *J. Environ. Manage.* **1985**, *20*, 103–127.
40. Daly, J.A.; Baetens, M.J.; De Baets, B. Ecological Diversity: Measuring the Unmeasurable. *Mathematics* **2018**, *6*, 119. [\[CrossRef\]](#)
41. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [\[CrossRef\]](#)
42. Ifo, S.A.; Moutsambote, J.-M.; Koubouana, F.; Yoka, J.; Ndzai, S.F.; Bouetou-Kadilamio, L.N.O.; Mampouya, H.; Jourdain, C.; Bocko, Y.; Mantota, A.B.; et al. Tree Species Diversity, Richness, and Similarity in Intact and Degraded Forest in the Tropical Rainforest of the Congo Basin: Case of the Forest of Likouala in the Republic of Congo. *Int. J. For. Res.* **2016**, *2016*, 7593681. [\[CrossRef\]](#)
43. Simpson, E.H. Measurement of Diversity. *Nature* **1949**, *163*, 688. [\[CrossRef\]](#)
44. Morris, E.K.; Caruso, T.; Fischer, M.; Hancock, C.; Obermaier, E.; Prati, D.; Maier, T.S.; Meiners, T.; Caroline, M.; Wubet, T.; et al. Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* **2014**, 3514–3524. [\[CrossRef\]](#)
45. Wood, E.M.; Pidgeon, A.M.; Radeloff, V.C.; Keuler, N.S. Image texture as a remotely sensed measure of vegetation structure. *Remote Sens. Environ.* **2012**, *121*, 516–526. [\[CrossRef\]](#)
46. Liaw, A.; Wiener, M. Breiman and Cutler's Random Forests for Classification and Regression. *Package RandomForest*. 2015. Available online: <http://CRAN.R-project.org/package=randomForest> (accessed on 19 March 2020). R package version 4.6-12.
47. Desbordes, P.; Ruan, S.; Modzelewski, R.; Pineau, P.; Vauclin, S.; Gouel, P.; Michel, P.; Di Fiore, F.; Vera, P.; Gardin, I. Predictive value of initial FDG-PET features for treatment response and survival in esophageal cancer patients treated with chemo-radiation therapy using a random forest classifier. *PLoS ONE* **2017**, *12*, e0173208. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Chrysafis, I.; Mallinis, G.; Gitas, I.; Tsakiri-Strati, M. Estimating Mediterranean forest parameters using multi seasonal Landsat 8 OLI imagery and an ensemble learning method. *Remote Sens. Environ.* **2017**, *199*, 154–166. [\[CrossRef\]](#)
49. Speiser, J.L.; Miller, M.E.; Tooze, J.; Ip, E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **2019**, *134*, 93–101. [\[CrossRef\]](#)
50. Yuchi, W.; Gombojav, E.; Boldbaatar, B.; Galsuren, J.; Enkhmaa, S.; Beejin, B.; Naidan, G.; Ochir, C.; Legtseg, B.; Byambaa, T.; et al. Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. *Environ. Pollut.* **2019**, *245*, 746–753. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Ismail, R.; Mutanga, O. A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 45–51. [\[CrossRef\]](#)
52. Dube, T.; Mutanga, O.; Elhadi, A.; Ismail, R. Intra-and-inter species biomass prediction in a plantation forest: Testing the utility of high spatial resolution spaceborne multispectral Rapideye sensor and advanced machine learning algorithms. *Sensors* **2014**, *14*, 15348–15370. [\[CrossRef\]](#)
53. *Advances in Neural Networks—ISNN 2006*; Wang, J.; Yi, Z.; Zurada, J.M.; Lu, B.-L.; Yin, H. (Eds.) Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2006; Volume 3972, ISBN 978-3-540-34437-7.

54. Turner, W.; Rondinini, C.; Pettorelli, N.; Mora, B.; Leidner, A.K.; Szantoi, Z.; Buchanan, G.; Dech, S.; Dwyer, J.; Herold, M.; et al. Free and open-access satellite data are key to biodiversity conservation. *Biol. Conserv.* **2015**, *182*, 173–176. [[CrossRef](#)]
55. Khare, S.; Latifi, H.; Ghosh, S.K. Multi-scale assessment of invasive plant species diversity using Pléiades 1A, RapidEye and Landsat-8 data. *Geocarto Int.* **2018**, *33*, 681–698. [[CrossRef](#)]
56. Strahler, A.H.; Woodcock, C.E.; Smith, J.A. On the nature of models in remote sensing. *Remote Sens. Environ.* **1986**, *20*, 121–139. [[CrossRef](#)]
57. Wang, R.; Gamon, J.A.; Cavender-Bares, J.; Townsend, P.A.; Zygielbaum, A.I. The spatial sensitivity of the spectral diversity-biodiversity relationship: An experimental test in a prairie grassland. *Ecol. Appl.* **2018**, *28*, 541–556. [[CrossRef](#)] [[PubMed](#)]
58. Gholizadeh, H.; Gamon, J.A.; Townsend, P.A.; Zygielbaum, A.I.; Helzer, C.J.; Hmimina, G.Y.; Yu, R.; Moore, R.M.; Schweiger, A.K.; Cavender-Bares, J. Detecting prairie biodiversity with airborne remote sensing. *Remote Sens. Environ.* **2019**, *221*, 38–49. [[CrossRef](#)]
59. Marceau, D.J.; Gratton, D.J.; Fournier, R.A.; Fortin, J.P. Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. *Remote Sens. Environ.* **1994**, *49*, 105–117. [[CrossRef](#)]
60. Rocchini, D.; Ricotta, C.; Chiarucci, A. Using satellite imagery to assess plant species richness: The role of multispectral systems. *Appl. Veg. Sci.* **2007**, *10*, 325–331. [[CrossRef](#)]
61. Wang, R.; Gamon, J.A.; Schweiger, A.K.; Cavender-Bares, J.; Townsend, P.A.; Zygielbaum, A.I.; Kothari, S. Influence of species richness, evenness, and composition on optical diversity: A simulation study. *Remote Sens. Environ.* **2018**, *211*, 218–228. [[CrossRef](#)]
62. Hernández-Stefanoni, J.L.; Dupuy, J.M.; Castillo-Santiago, M.A. Assessing species density and abundance of tropical trees from remotely sensed data and geostatistics. *Appl. Veg. Sci.* **2009**, *12*, 398–414. [[CrossRef](#)]
63. De La Cueva, A. V Structural attributes of three forest types in central Spain and Landsat ETM plus information evaluated with redundancy analysis. *Int. J. Remote Sens.* **2008**, *29*, 5657–5676. [[CrossRef](#)]
64. Mallinis, G.; Koutsias, N.; Makras, A.; Karteris, M. Forest parameters estimation in a European Mediterranean landscape using remotely sensed data. *For. Sci.* **2004**, *50*, 450–460.
65. Ardö, J. Volume quantification of coniferous forest compartments using spectral radiance recorded by Landsat Thematic Mapper. *Int. J. Remote Sens.* **1992**, *13*, 1779–1786. [[CrossRef](#)]
66. Hesketh, M.; Sánchez-Azofeifa, G.A. The effect of seasonal spectral variation on species classification in the Panamanian tropical forest. *Remote Sens. Environ.* **2012**, *118*, 73–82. [[CrossRef](#)]
67. Ollinger, S.V. Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytol.* **2011**, *189*, 375–394. [[CrossRef](#)] [[PubMed](#)]
68. Eckert, S. Improved forest biomass and carbon estimations using texture measures from worldView-2 satellite data. *Remote Sens.* **2012**, *4*, 810–829. [[CrossRef](#)]
69. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [[CrossRef](#)]
70. Rautiainen, M.; Lukeš, P. Spectral contribution of understory to forest reflectance in a boreal site: An analysis of EO-1 Hyperion data. *Remote Sens. Environ.* **2015**, *171*, 98–104. [[CrossRef](#)]
71. Karlson, M.; Ostwald, M.; Reese, H.; Sanou, J.; Tankoano, B.; Mattsson, E. Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest. *Remote Sens.* **2015**, *7*, 10017–10041. [[CrossRef](#)]

