

## Article

# Linking Land Surface Phenology and Vegetation-Plot Databases to Model Terrestrial Plant $\alpha$ -Diversity of the Okavango Basin

Rasmus Revermann <sup>1,\*</sup>, Manfred Finckh <sup>1</sup>, Marion Stellmes <sup>2</sup>, Ben J. Strohbach <sup>3</sup>, David Frantz <sup>2</sup> and Jens Oldeland <sup>1</sup>

<sup>1</sup> Department of Biodiversity, Ecology and Evolution of Plants, University of Hamburg, Biocentre Klein Flottbek, Ohnhorststr. 18, 22609 Hamburg, Germany; manfred.finckh@uni-hamburg.de (M.F.); jens.oldeland@uni-hamburg.de (J.O.)

<sup>2</sup> Department of Environmental Remote Sensing and Geoinformatics, Faculty of Regional and Environmental Sciences, Trier University, Behringstraße 21, 54296 Trier, Germany; stellmes@uni-trier.de (M.S.); frantz@uni-trier.de (D.F.)

<sup>3</sup> School of Natural Resources and Spatial Sciences, Namibia University of Science and Technology, P/Bag 13388 Windhoek, Namibia; bstrohbach@nust.na

\* Correspondence: rasmus.revermann@uni-hamburg.de; Tel.: +49-40-42816-408; Fax: +49-40-42816-539

Academic Editors: Susan L. Ustin, Parth Sarathi Roy and Prasad S. Thenkabail

Received: 19 November 2015; Accepted: 18 April 2016; Published: 29 April 2016

**Abstract:** In many parts of Africa, spatially-explicit information on plant  $\alpha$ -diversity, *i.e.*, the number of species in a given area, is missing as baseline information for spatial planning. We present an approach on how to combine vegetation-plot databases and remotely-sensed land surface phenology (LSP) metrics to predict plant  $\alpha$ -diversity on a regional scale. We gathered data on plant  $\alpha$ -diversity, measured as species density, from 999 vegetation plots sized 20 m  $\times$  50 m covering all major vegetation units of the Okavango basin in the countries of Angola, Namibia and Botswana. As predictor variables, we used MODIS LSP metrics averaged over 12 years (250-m spatial resolution) and three topographic attributes calculated from the SRTM digital elevation model. Furthermore, we tested whether additional climatic data could improve predictions. We tested three predictor subsets: (1) remote sensing variables; (2) climatic variables; and (3) all variables combined. We used two statistical modeling approaches, random forests and boosted regression trees, to predict vascular plant  $\alpha$ -diversity. The resulting maps showed that the Miombo woodlands of the Angolan Central Plateau featured the highest diversity, and the lowest values were predicted for the thornbush savanna in the Okavango Delta area. Models built on the entire dataset exhibited the best performance followed by climate-only models and remote sensing-only models. However, models including climate data showed artifacts. In spite of lower model performance, models based only on LSP metrics produced the most realistic maps. Furthermore, they revealed local differences in plant diversity of the landscape mosaic that were blurred by homogenous belts as predicted by climate-based models. This study pinpoints the high potential of LSP metrics used in conjunction with biodiversity data derived from vegetation-plot databases to produce spatial information on a regional scale that is urgently needed for basic natural resource management applications.

**Keywords:** Angola; Botswana; dry tropical forests; EVI; Miombo; MODIS; Namibia; phenological metrics; predictive modeling; species density

## 1. Introduction

Globally, biodiversity is declining at a high rate [1], and international treaties, such as the Convention on Biological Diversity, pledged to halt biodiversity loss. Paramount for safeguarding

biodiversity is a better understanding of biodiversity patterns and spatially-explicit information. The recent discussion on ‘essential biodiversity variables’ has shown that remote sensing applications are indispensable in the process and are needed to monitor changes in biodiversity over large areas with a consistent methodology [2,3]. In this context, field-based ecological data also play a prominent role as baseline data for biodiversity models and as ground truth information for remote sensing applications. In a recent effort, the initiative on a global index of vegetation-plot databases (GIVD) created a meta-database containing over 200 existing vegetation-plot databases worldwide with over three million vegetation plots [4,5]. These databases harbor an enormous potential as ground truth information for future remote sensing studies and spatial modeling approaches from local to global scales; yet, so far this potential remains unexploited. However, there are a few studies combining MODIS data with vegetation databases, e.g., to predict tree species richness in the USA [6] or to analyze vegetation responses to drought in Dutch dune ecosystems [7]. The main reasons for the missing integration of remote sensing, spatial modeling and ecological research are not only the differing traditions of the disciplines, but are rooted in the mismatch of the spatial scales of satellite imagery and the size of ecological field sites. However, in recent years remote sensing products have diversified, and even more importantly, many have become readily available at no cost with appropriate spatial and temporal resolution; for a review from the remote sensing perspective, see Wang *et al.* [8]. Likewise, the availability of field-based data has increased and has become more accessible through the formation of global (meta-) databases.

Vegetation plots are samples of a specific area of the landscape and vary in size depending on vegetation type and the purpose of the study: the size of vegetation plots in woodlands and forests commonly ranges from 400 m<sup>2</sup> to 25,000 m<sup>2</sup> [9]. Typically, they hold information of the co-occurring plant species, their cover or abundance, vegetation structure and are often connected to abiotic parameters, such as soil properties. Generally, vegetation plots are stored in a vegetation database compiling information of several vegetation plots within a region. From these databases, one can extract information on plant diversity. Diversity has many dimensions, and its measurement strongly depends on the observed spatial scale [10]. Commonly, diversity is treated in three different components: (1)  $\alpha$ -diversity defined as the diversity of a vegetation plot; (2)  $\beta$ -diversity is the difference in species composition between vegetation plots; and (3)  $\gamma$ -diversity reflects the diversity at the landscape level, *i.e.*, the species pool of all sampled vegetation plots [10]. The  $\alpha$ -diversity measure “species density” is often regarded as the “common currency” in diversity research [11] and is defined as the number of species present in a given area, e.g., in a vegetation plot.

Turner *et al.* [12] list two main approaches to assess biodiversity using remote sensing: (i) direct measurements where species are recognized based on their spectral properties; or (ii) indirect ones where no direct link is established, but instead, relies on the spatially-explicit localization of distinct vegetation units. Closely linked to these vegetation units are properties, such as  $\alpha$ -diversity, *i.e.*, the average species number of a defined site or habitat, that we seek to extrapolate using statistical models [13]. In ecology, the establishment of species distribution models (SDMs) as a standard tool to make predictions for unsurveyed areas based on field surveys has boosted the integration of robust statistical methods for predictive modeling [14]. In predictive modeling, statistical algorithms are used to relate the attributes in question, *i.e.*, the response variable, to a set of environmental predictor variables, such as climate data or remotely-sensed information.

Spectral properties of vegetation change throughout the seasons due to changes of biophysical and biochemical properties, *i.e.*, pigment, sugar and water content of leaves in the canopy, above ground biomass or vegetation structure. As such, land surface phenology (LSP) metrics can be derived from dense spectral observations reflected in remotely-sensed vegetation indices across large areas [15,16]. Software, such as TIMESAT [17], is frequently used to derive various LSP metrics, *i.e.*:(1) temporal metrics defining phenological stages of the vegetation (e.g., start and end of the green season); (2) biomass-related metrics (e.g., integrals or amplitudes); and (3) seasonality-related metrics (e.g., the green-up rate).

As every vegetation unit has a more or less unique combination of phenological metrics, LSPs are highly suited to distinguish different vegetation types, as demonstrated by Fan *et al.* [18], who used LSP to identify rubber plantations in fragmented tropical forests. Moreover, LSP metrics served for mapping above-ground woody biomass [19] and have been successfully applied in species distribution modeling [20,21], change detection [22,23] and for vegetation mapping [24,25]. So far, no study tested the suitability of LSP metrics for modeling plant  $\alpha$ -diversity. However, especially biomass-related LSP metrics, *i.e.*, integrals or base values, could be promising predictors for plant  $\alpha$ -diversity due to strong empirical linkages of above ground biomass and species richness [26,27].

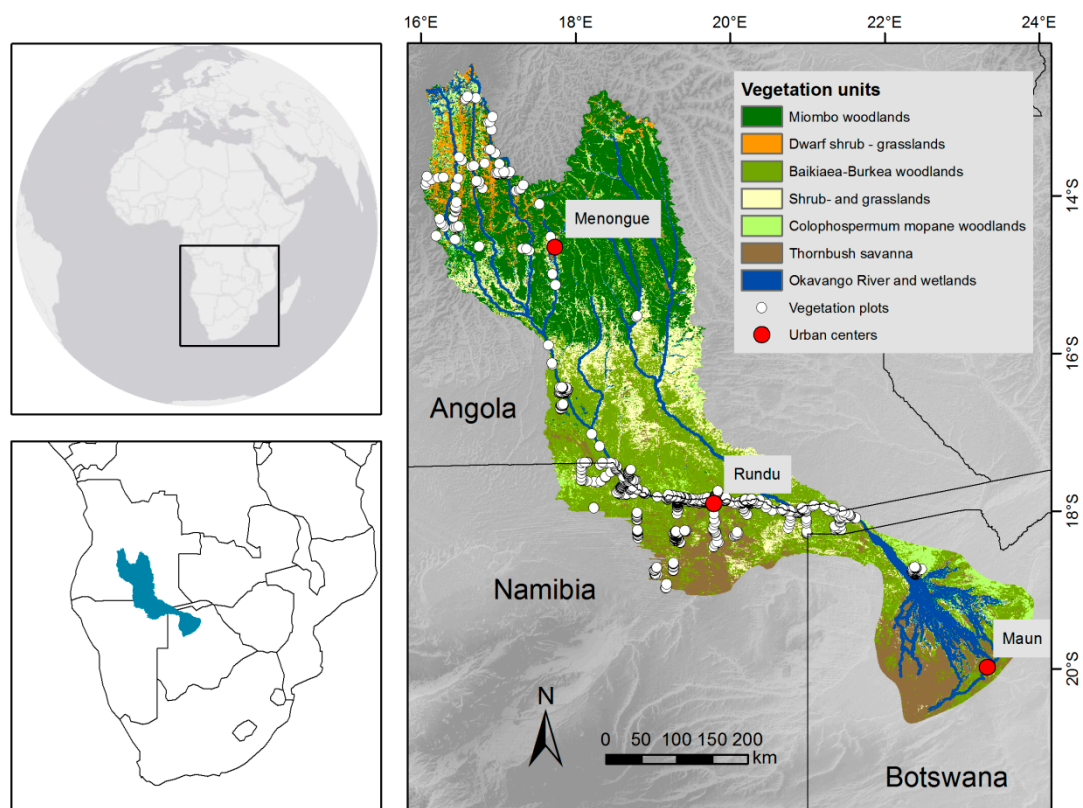
Generally, climate is regarded as the main driver of biogeographic patterns at large spatial scales [28]. However, with increasing spatial resolution, factors, such as topography, soil properties or disturbance patterns, gain importance. It has been shown that climatic predictors serve as large-scale determinants, while land cover data increase the predictive power of species distribution models on finer resolutions of 1 km to 20 km [28,29]. Most studies on plant  $\alpha$ -diversity using remote sensing data focus either on global or continental scales with very coarse resolution (100 km) [30] or have a small extent, but operate on fine grain sizes (1 m to 30 m) [31–33]. The study of Saatchi *et al.* [34] is an exception in this regard and covers the entire Amazon basin at 1- to 5-km spatial resolution.

The aim of this study was to test the suitability of MODIS EVI land surface phenology metrics at 250-m spatial resolution to predict vascular plant  $\alpha$ -diversity derived from the vegetation-plot database of the Okavango Basin. We used two statistical model algorithms, boosted regression trees (BRT) and random forests (RF), and compared the performances of the models on three predictor subsets: (1) only LSP metrics and topography; (2) only climate data; and (3) the entire set of predictor variables including both LSP metrics and climate data. Finally, we analyzed the  $\alpha$ -diversity maps generated for the Okavango Basin using the different models and datasets. In doing this, we aimed to provide recommendations for generating spatially-explicit maps on plant  $\alpha$ -diversity on a regional level with comparatively high spatial resolution to support natural resource management and conservation applications.

## 2. Data and Methods

### 2.1. Study Site

The Okavango Basin is situated in southern Africa and is shared by the countries of Angola, Namibia and Botswana (Figure 1). The Okavango River and its tributaries originate on the Angolan Central Plateau, where the large majority of the runoff is generated [35]. The middle reaches of the river form the border between Angola and Namibia before entering Botswana, where it terminates in the Okavango Delta, one of the largest inland deltas of the world. The course of the river follows a strong environmental gradient from its source on the Angolan Central Plateau at altitudes of 1850 m a.s.l. to the Okavango Delta in Botswana at around 940 m a.s.l. Mean annual temperature increases from northwest to southeast from 18 °C to 24 °C. Precipitation shows an inverse trend: the Angolan Central Plateau features a sub-humid climate with mean annual precipitation of over 1400 mm, and the delta receives less than 500 mm per year [36]. Accordingly, vegetation changes along the course of the river. Miombo forests are the dominant vegetation type of the Angolan Central Plateau with its gently rolling landscape. However, topography has a strong impact on local vegetation patterns; mid- and bottom slopes of the valleys feature geoxylic grasslands, and the valley bottoms of many tributaries harbor wetlands [37]. As climate becomes drier, the closed Miombo woodlands give way to the more open *Baikiaea-Burkea* woodlands of the middle reaches. The area surrounding the delta to the east is dominated by *Colophospermum mopane* woodlands, while the driest areas to the west and south of the delta are covered by thornbush savannas formed by various *Acacia* communities [38].



**Figure 1.** Location of the Okavango Basin in southern Africa. The map of the Okavango Basin shows major vegetation units modified after Stellmes *et al.* [39] and the location of vegetation plots used in this study. The three major urban centers of the basin, Menongue, Rundu and Maun, are indicated by a red dot. The map datum is WGS84, and the background shows the SRTM digital elevation model. The extent of the study area, the Okavango Basin, follows the definition of The Future Okavango (TFO) project [40]. For a map on observed species density, see Figure S1.

## 2.2. Data

### 2.2.1. Vegetation Data

Quantitative information on the vegetation, especially on the large Angolan share of the Okavango Basin, is scarce and limited to descriptive studies from the pre-independence era, *i.e.*, before 1975 [41–44]. During The Future Okavango (TFO) project, we initiated an extensive plot-based vegetation survey based on a random stratified sampling design to ensure coverage of all major vegetation types of the basin (GIVD database ID: AF-00-009) [37,38,45–47]. However, the remoteness, limited access and the danger of land mines posed restrictions on the sampling. On vegetation plots sized 20 m × 50 m, all vascular plant species were recorded. Vegetation surveys were carried out during the growing seasons (November to May) of the years 2011 to 2014. Additionally, data from the National Phytosociological Database of Namibia were used (GIVD database ID: AF-NA-001) [48]. For the present study, we only considered plots from terrestrial vegetation, *i.e.*, forests, woodlands and grasslands, as plots from semi-terrestrial and aquatic vegetation units were too small to properly relate to MODIS data. To avoid mixed pixel problems, we only selected plots that were not located at the edge of vegetation units and had a minimum distance of 500 m between plots, *i.e.*, there was only one vegetation plot within one MODIS pixel. In total, 999 vegetation plots were selected for modeling. This dataset comprises the best available data for the region. However, some vegetation units were underrepresented, such as the thornbush savanna in the southwest of the Okavango Delta and the transition zone between

Miombo woodlands and *Baikiaea-Burkea* woodlands. As a plant  $\alpha$ -diversity measure, we derived species density per  $10^3 \text{ m}^2$ , i.e., the number of vascular plant species per vegetation plot [10].

### 2.2.2. MODIS

We compiled a MODIS-enhanced vegetation index (EVI) time series with a spatial resolution of  $250 \text{ m} \times 250 \text{ m}$  (MOD13Q1 product). The main requisition of an appropriate vegetation index is its capability of differentiating biomass at a certain point in time, as well as tracing phenological changes reliably [49]. We used the standard MODIS vegetation 16-day EVI product, because this vegetation index overcomes some limitations of the NDVI that are of relevance in our study area. Thus, the EVI is less sensitive to the background signal, such as soil brightness, and it does not saturate as fast with high biomass values. Moreover, still, inherent atmospheric effects should be lessened [50,51]. Using TIMESAT [17], we derived eleven land surface phenology (LSP) metrics based on the 16-day EVI composite time series ranging from July 2000 to June 2012 (Table 1). As a consequence of the Southern Hemisphere, the start of the year was set to the middle of the year, 1 July, when most deciduous species have shed their leaves and annual plants have died back. In order to reduce the effect of the inter-annual variability of LSP, we used the long-term mean of the annual metrics. Additionally, we used the mean and the minimum of the near infrared (NIR) channel of the surface reflectance product (MOD13Q1) to differentiate between vegetation-scarce surfaces with different brightness, such as water and soil.

### 2.2.3. Topography

We selected three predictor variables describing topography (Table 1), as it has been shown that the local topography of the Angolan Central Plateau creates micro-climatic conditions strongly influencing vegetation patterns [52,53]. Moreover, water availability plays a primary role in the semi-arid parts of the Okavango Basin. Based on the global digital elevation model SRTM (Shuttle Radar Topography Mission) with a horizontal resolution of  $90 \text{ m} \times 90 \text{ m}$ , we calculated the topographic position index (TPI) [54], the topographic wetness index (TWI) [55] and the topographic ruggedness index (TRI) [56] in the open source GIS SAGA [57]. Subsequently, the topographic attributes were resampled using bilinear interpolation to match the MODIS resolution.

### 2.2.4. Climate

Weinzierl *et al.* [58] provided a regionalization of climate data from 1950 to 2000 for the Okavango Basin based on orographic parameters and a geographically-weighted regression algorithm with a resolution of  $1 \text{ km} \times 1 \text{ km}$ . The original climate data stem from the regional climate model REMO for the domain of south central Africa forced with the global circulation model ECHAM [59]. We resampled the regionalized data using bilinear interpolation to match the resolution of the MODIS data. Based on monthly values of minimum temperature, maximum temperature and monthly precipitation, we derived 15 bioclimatic predictors using the “dismo” package in R [60].

To test whether the quality of predictions depended on climate, we additionally tested a second climate dataset compiled from two sources: (1) precipitation data were obtained from the gridded African Rainfall Climatologies Version 2 with a spatial resolution of  $0.1^\circ$  (ARC2, [61]); input data of the ARC2 data are 3-hourly satellite-based infrared measurements and daily gauge measurements; (2) temperature data were derived from the Climate Research Unit (CRU) TS3.10 dataset [62]. CRU is based on meteorological stations across the global land area and has a spatial resolution of  $0.5^\circ$ . These climate data were subject to the same treatment as the REMO climate data, and the same bioclimatic predictor variables were calculated. For results based on the modeling using the second climate dataset, refer to the Supplementary Material.



**Table 1.** Description of predictor variables and data sources. All variables excluded from modelling after screening for collinearity among predictor variables are denoted with an asterisk. SRTM: digital elevation model of shuttle radar topography mission; REMO: regional climate model for the domain of south central Africa forced with the global circulation model ECHAM. RS TOPO, remote sensing and topography.

Dataset	Variable	Variable Description	Dataset
RS TOPO	Amplitude	maximum of EVI–minimum of EVI	MODIS EVI time series
	BaseValue	base value of EVI in the course of year	MODIS EVI time series
	LargeIntegral	total integral of EVI in the course of year	MODIS EVI time series
	SmallIntegral *	integral of EVI above BaseValue	MODIS EVI time series
	NIR	near infrared band	MODIS EVI time series
	NIR_min *	minimum of the near infrared band	MODIS EVI time series
	MaxFit *	maximum fitted value of EVI	MODIS EVI time series
	RateDecrease *	rate of senescence (slope of the line connecting the annual peak and the point at the end of greenness)	MODIS EVI time series
	RateIncrease *	rate of green up (slope of the line connecting the point of the onset of greenness and the annual peak)	MODIS EVI time series
	SeasonEnd	day of year, end of greening	MODIS EVI time series
	SeasonLength	number of days, duration of greening	MODIS EVI time series
	SeasonMid	day of year, peak of greening	MODIS EVI time series
	SeasonStart	day of year, start of the greening	MODIS EVI time series
	TPI	topographic position index	SRTM 90 m
	TRI	topographic ruggedness index	SRTM 90 m
	TWI	topographic wetness index	SRTM 90 m
CLIMATE	bio1	annual mean temperature (°C)	REMO
	bio2 *	mean diurnal range (°C) (mean of monthly (max temp–min temp))	REMO
	bio3	isothermality ((BIO2/BIO7) × 100)	REMO
	bio4	temperature seasonality (standard deviation × 100)	REMO
	bio5 *	max temperature of warmest month (°C)	REMO
	bio6 *	min temperature of coldest month (°C)	REMO
	bio7	temperature annual range (BIO5 to BIO6) (°C)	REMO
	bio8 *	mean temperature of wettest quarter (°C)	REMO
	bio9 *	mean temperature of driest quarter (°C)	REMO
	bio10 *	mean temperature of warmest quarter (°C)	REMO
	bio11 *	mean temperature of coldest quarter (°C)	REMO
	bio12	annual precipitation (mm)	REMO
	bio15	precipitation seasonality (coefficient of variation)	REMO

### 2.3. Statistical Modeling

To test whether remote sensing data or climate data are better suited to predict plant  $\alpha$ -diversity, we tested three subsets of the predictor variables: (1) remote sensing data and topographic data denoted as remote sensing and topography (“RS TOPO”); (2) only climatic data “CLIMATE”; (3) all predictor variables “ALL” (Table 1).

Collinearity among predictor variables can lead to erroneous estimation of the parameters of a statistical model and, hence, cause misleading interpretations [63]. Therefore, all predictors were screened prior to modelling and tested for collinearity using a Spearman rank-correlation test ( $r_s$ ). For visualizing the strength of the correlation, we used the R package “corrplot” [64]. For all pairs of variables with  $|r_s| > 0.7$ , the variable better reflecting ecological processes determining vegetation patterns was selected [63].

The choice of the statistical model type is a common source of the algorithmic prediction error [65]. Thus, we tested and compared two modeling techniques that have been used in various areas of ecological modeling and have been demonstrated to have a high predictive power [66]: boosted regression trees (BRT) [67] and random forests (RF) [68,69]. BRT combines the strength of traditional regression trees and boosting, the adaptive, stage-wise combination of a multitude of individual models. High predictive performance is enabled through accommodating non-linear relationships and fitting interactions among predictor variables. We used the R packages “gbm” [70] and ‘caret’ [71]

to compute BRT assuming a Poisson distribution of the response variable. There are three important parameters that need to be set by the user: interaction depth, number of trees and the learning rate. We systematically varied the three parameters using a 10-fold cross-validation to find the optimal settings for each data subset [72].

RF builds multiple regression trees based on bootstrap samples with each tree being grown on a randomized subsample of the predictor variables. A large number of trees is grown without pruning, and final results are averaged. In RF, the specification of model parameters has less influence on model output. We operated RF with default settings for the number of variables used at each split (number of candidate variables divided by 3); the number of trees to grow was set to 1000. RF was calculated using the R package ‘randomForest’ [73].

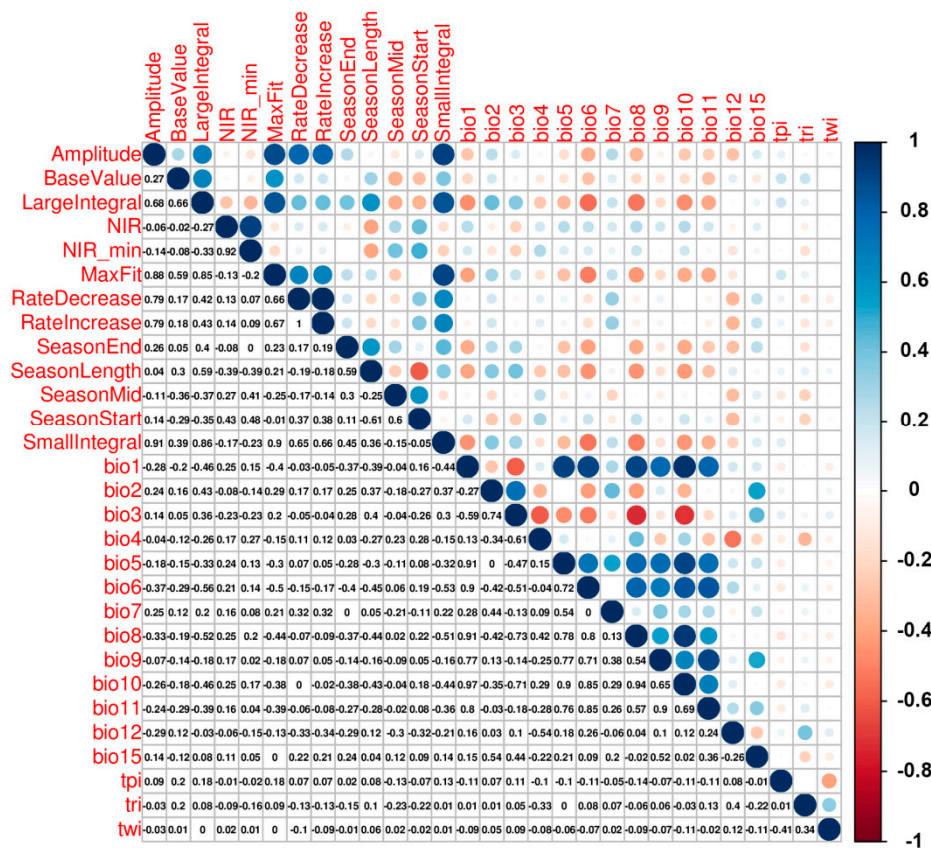
BRT and RF offer slightly different measures of variable importance, and the measures cannot be compared directly among model types. In BRT, variable importance is measured as the relative influence of each variable averaged over all trees. For RF, we display the increase in the mean square error of the prediction [73].

For validation, the dataset was split into training and test data samples with a ratio of 80:20 using random stratified sampling. The following criteria of model performance were calculated: explained variance, the Pearson’s coefficient of correlation ( $r_p$ ) between the predicted and observed values of species density, the coefficient of determination  $R^2$ , the root mean square error (RMSE), and the relative root mean square error (rRMSE in percent) [71]. The models were calibrated on training data and then used to predict plant  $\alpha$ -diversity of the entire Okavango Basin at 250-m spatial resolution. All analyses were carried out in R [74].

### 3. Results

#### 3.1. Model Building and Validation

After screening for collinearity, seven LSP metrics, three topography predictors and six climate variables were selected for modeling out of the 29 potential predictor variables (Figure 2, Table 1). Models on all subsets (“RS TOPO”, “CLIMATE” and “ALL”) showed a clear correlation of predicted and observed values of plant  $\alpha$ -diversity, and values of the Pearson correlation ( $r_p$ ) ranged from 0.69 to 0.80 on test data. In order to compare  $r_p$  values based on confidence intervals, we computed the z-scores based on the Fisher transformation. Comparisons revealed no significant differences in the correlation for “ALL” and “CLIMATE” models. However, “RS TOPO” models showed consistently significantly lower correlation in comparison to “ALL” (BRT: z-value  $-2.475$ ,  $p$ -value 0.007; RF: z-value  $-1.758$ ,  $p$ -value 0.039) and “CLIMATE” (BRT: z-value  $-2.475$ ,  $p$ -value 0.007; RF: z-value  $-1.758$ ,  $p$ -value 0.039). The RMSE was moderately high with values of 9.3 to 11.0 species per  $10^3 \text{ m}^2$ , and the relative RMSE ranged from 26.8% to 31.4%. The  $R^2$  indicated a moderate goodness-of-fit ranging from 0.48 to 0.64. The explained variance ranged from 43% to 67% (Table 2). The two statistical model algorithms BRT and RF performed almost equally well regarding all performance criteria. Only the explained deviance was consistently higher in BRT for all datasets than in RF. The difference of the performance criteria between training and test data was much higher in RF than in BRT. Regarding the different input data, the “ALL” models performed best, closely followed by “CLIMATE”, while “RS TOPO” models exhibited the poorest performance.



**Figure 2.** Correlation matrix of predictor variables measured by Spearman's rank-correlation coefficient ( $r_s$ ) ranging from  $-1$  to  $1$ . The lower half of the diagonal gives the numeric value of  $r_s$ ; the upper diagonal visualizes the correlation coefficient: the size of the circles corresponds to the strength of the correlation; red denotes negative and blue positive correlation coefficients. For details on predictor variables, see Table 1.

**Table 2.** Validation results for the two model types boosted regression trees (BRT) and random forests (RF) on the three subsets of the predictor variables: remote sensing and topography ("RS TOPO"), only climate data ("CLIMATE") and all data ("ALL"). The following performance measures were calculated: explained variance (expl. var. (%)), Pearson's correlation coefficient ( $r_p$ ) between observed and predicted values, the coefficient of determination ( $R^2$ ), the root mean square error (RMSE, in species per  $10^3 \text{ m}^2$ ), and the relative root mean square error (rRMSE in percent). The results for training and test data are displayed (training 80% of the data and testing 20%).

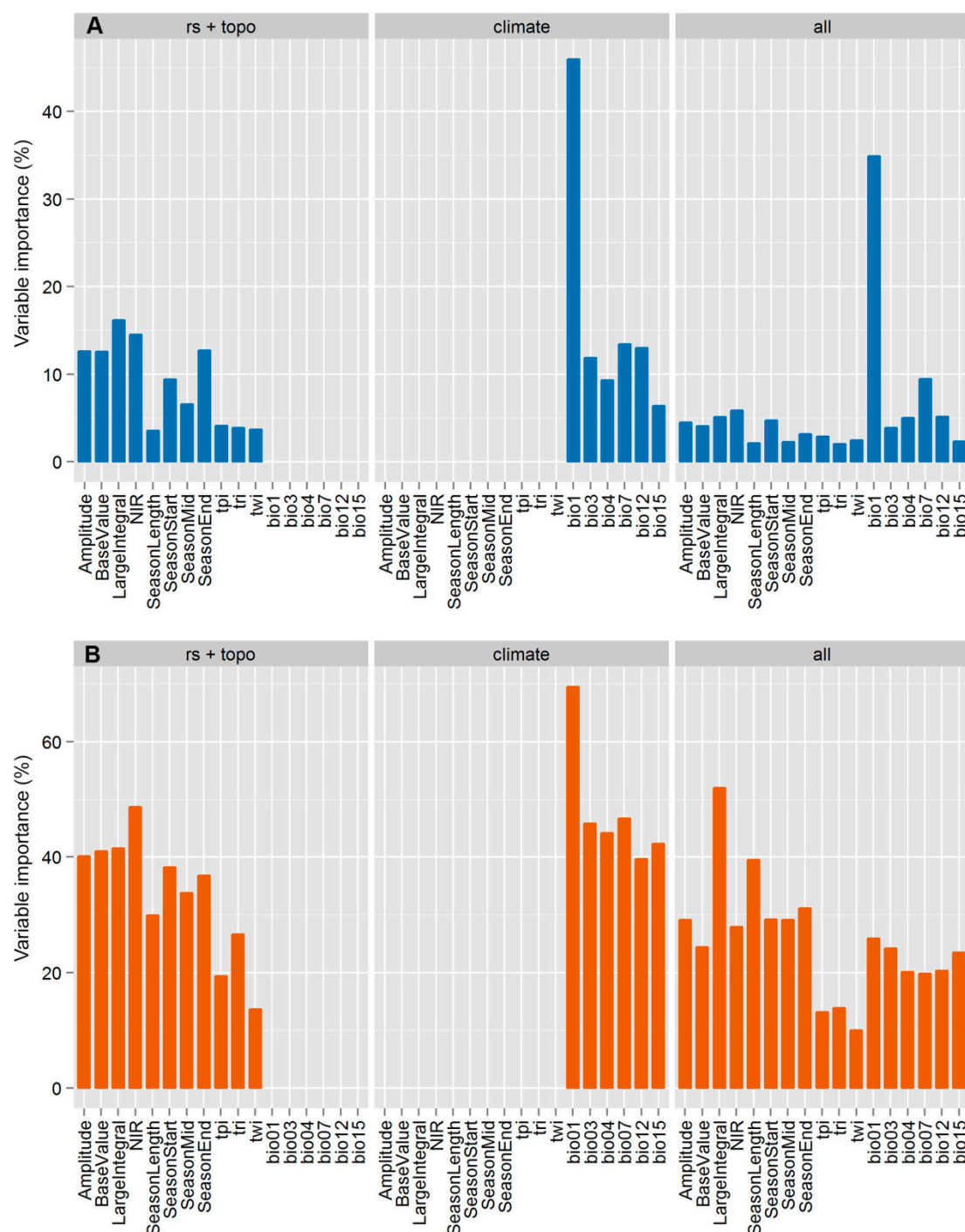
Model	Dataset	Expl. var.	Correlation ( $r_p$ )		$R^2$		RMSE		rRMSE	
		Train (%)	Train	Test	Train	Test	Train	Test	Train	Test
BRT	RS TOPO	54	0.80	0.69	0.60	0.48	10.1	11	28.8	31.8
	CLIMATE	61	0.82	0.80	0.68	0.63	9.1	9.3	25.8	26.8
	ALL	67	0.86	0.80	0.74	0.64	8.3	9.3	23.5	26.8
RF	RS TOPO	43	0.94	0.70	0.89	0.49	5.9	10.9	16.7	31.4
	CLIMATE	50	0.94	0.78	0.87	0.61	5.8	9.6	16.6	27.6
	ALL	54	0.95	0.79	0.90	0.63	5.3	9.4	15.2	27.0

### 3.2. Variable Importance

Variable importance varied among the three subsets of predictor variables and between the model types. However, a few general trends were evident (Figure 3). The topographic variables had only limited influence in all model runs. In the "RS TOPO" dataset, the "NIR" and "LargeIntegral" were



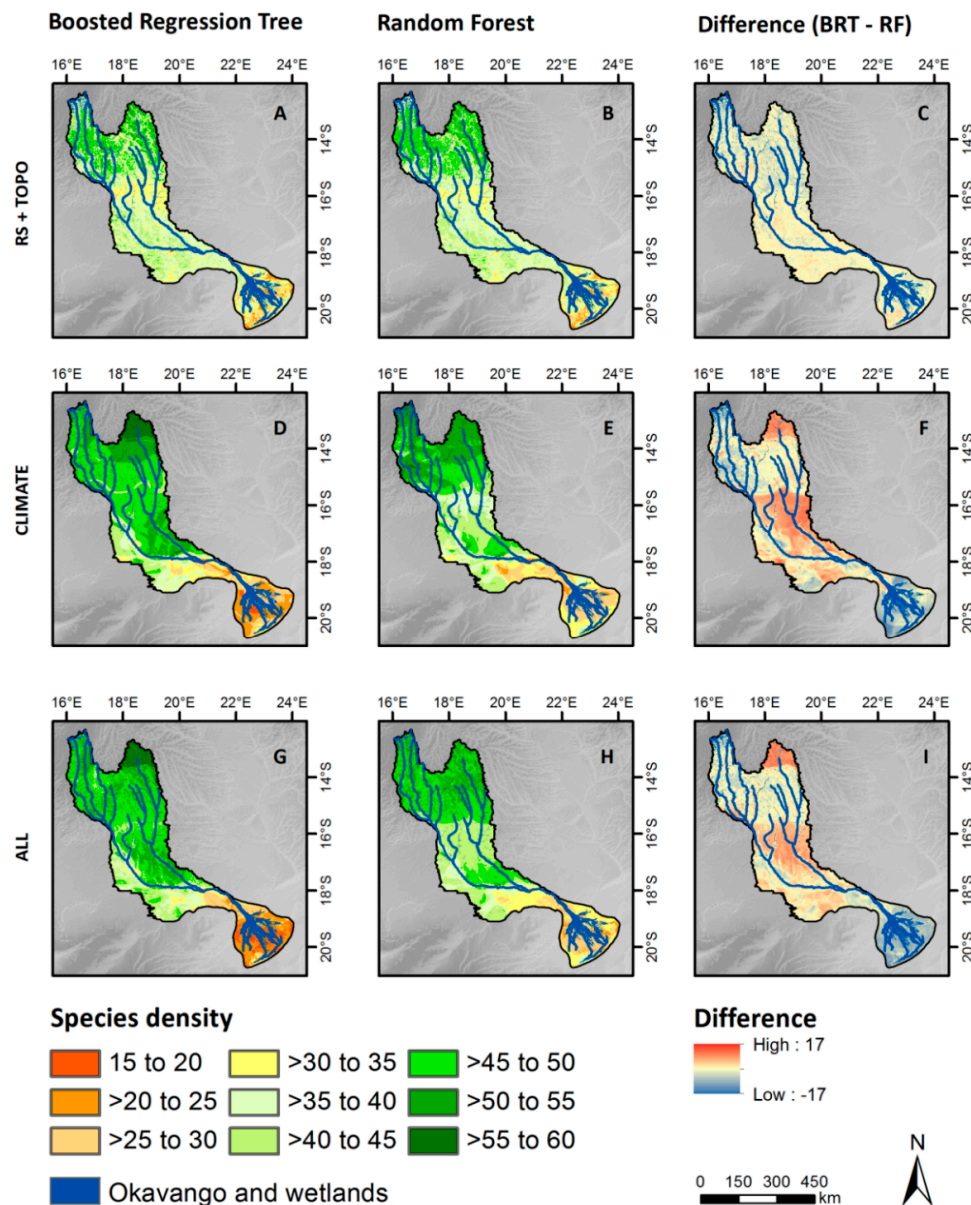
the two most important variables. Most of the biomass-related metrics were superior to the temporal metrics, *i.e.*, length, start or end of the season. Among the climatic variables, annual mean temperature ('bio1') was the most important predictor throughout all model runs. Precipitation-related variables did not have much predictive power. In BRT, in the dataset "ALL", climatic predictors had the highest importance, while the opposite was observed in RF, where LSP metrics yielded higher predictive power than climatic predictors (Figure 3).



**Figure 3.** Variable importance for the two model types: (A) BRT, calculated as the relative influence (%); (B) RF, calculated as the increase in MSE (%). As the calculation of variable importance differs among BRT and RF, only the ranking of the variables can be compared, but not the absolute values.

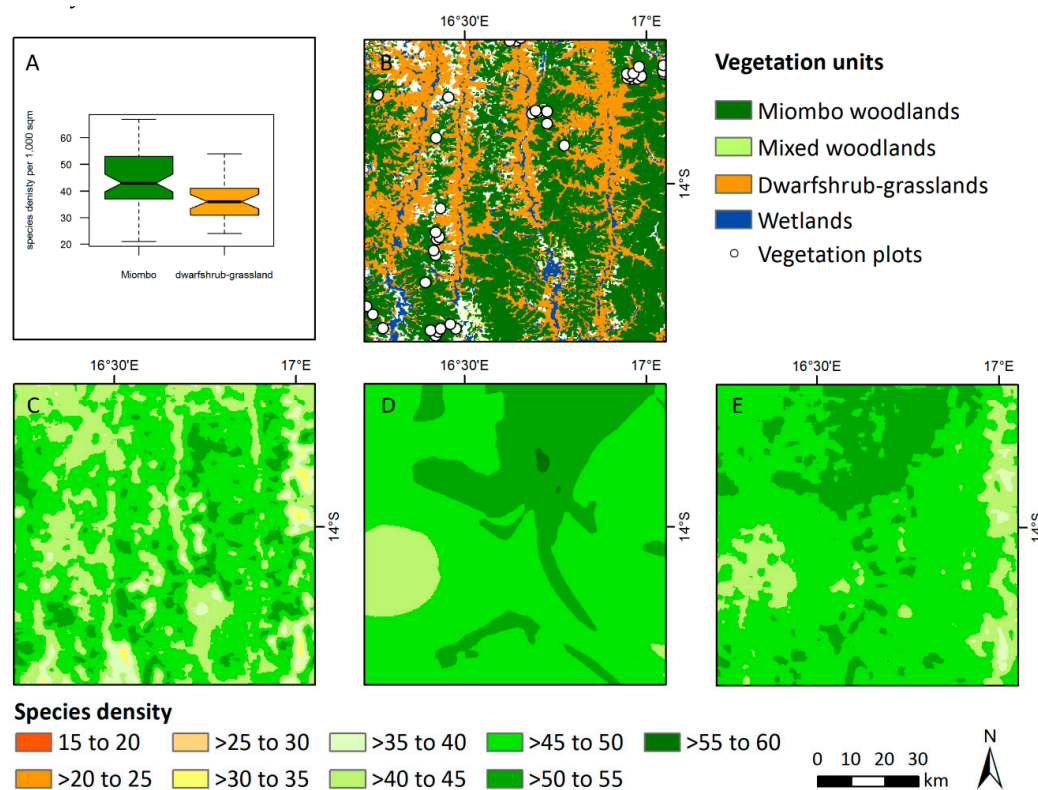
### 3.3. Patterns of Plant Alpha Diversity

The predicted plant  $\alpha$ -diversity in the basin ranged from 15 to 65 species per  $10^3 \text{ m}^2$ . All derived maps showed that the Miombo woodlands of the upper reaches of the Okavango River featured the highest plant  $\alpha$ -diversity, reaching values of over 60 species per  $10^3 \text{ m}^2$  (Figure 4). Generally, plant  $\alpha$ -diversity followed a decreasing trend southwards. The *Baikiaea-Burkea* woodlands of the middle reaches took an intermediate position, while the area around the Okavango Delta in Botswana showed the lowest values of 20 to 25 species per  $10^3 \text{ m}^2$ . Furthermore, the surroundings of the larger urban centers Rundu and Menongue depicted the absolute lowest plant  $\alpha$ -diversity.



**Figure 4.** Plant alpha diversity (species density per  $10^3 \text{ m}^2$ ) in the Okavango Basin predicted by boosted regression trees (BRT) and random forests (RF) and the difference between the two model algorithms displayed for the three datasets 'RS TOPO' (A,B,C), 'CLIMATE' (D,E,F), and 'ALL' (G,H,I). BRT (A,D,G), RF (B,E,H), difference (C,F,I). For a map on observed species density, see Figure S2.

The predictions of BRT and RF were similar for “RS TOPO”, but showed regional differences for the models built on the datasets “CLIMATE” and “ALL” (Figure 4C,F,I). On these datasets, BRT predicted higher plant  $\alpha$ -diversity for the Miombo woodlands of the far northeast of the basin and for the *Baikiaea-Burkea* woodlands of the middle reaches of the Okavango River. In contrast, RF predicted higher values than BRT in the thornbush savanna of the delta region. The maps based on the models on “CLIMATE” and “ALL” datasets showed belts of undifferentiated plant  $\alpha$ -diversity (Figure 4D,E,G,H). In contrast, “RS TOPO” showed fine-scale patterns of the landscape of the Miombo region (Figure 4A,B, Figure 5).



**Figure 5.** Plant alpha diversity (species density per  $10^3 \text{ m}^2$ ) in the Miombo region. (A) Observed species density of the vegetation units “Miombo woodlands” and “dwarf shrub/grassland” in the upper reaches of the Okavango Basin derived from vegetation-plot database. “Miombo woodlands” (mean = 44.0, SD = 10.8) exhibit significantly higher species density than “dwarf shrub-grasslands” (mean = 35.9, SD = 7.3) according to a two-group *t*-test ( $p < 0.001$ ); (B) Major vegetation units of the area according to Stellmes *et al.* [39] and location of vegetation plots; (C) Species density as predicted by BRT on the “RS TOPO” dataset; (D) Species density predicted by BRT on the “CLIMATE” dataset; (E) Species density predicted by BRT on the “ALL” dataset.

## 4. Discussion

### 4.1. Model Evaluation and Quality of Predictions

Pearson *et al.* [65] divided the prediction error of species distribution models in two components: (1) the algorithmic prediction error emanating from the choice of the statistical model and other parameters set during the modeling exercise; and (2) quality of the input data. In our study, the performance of BRT and RF models was very similar for all tested performance criteria showing good to moderate performance (Table 2). While there was little difference between the plant  $\alpha$ -diversity maps of BRT and RF on the “RS TOPO” dataset (Figure 4A,B), the maps based on models including climate data showed discrepancies between the two model algorithms (Figure 4C–F; for a detailed

discussion, see the discussion on climate data below). Thus, depending on the dataset, the algorithmic prediction error varies in magnitude, although BRT and RF are both machine learning techniques based on regression trees and exhibited comparable model performance.

#### 4.2. Data Quality

The response variable in this study was derived from two vegetation-plot databases. As shown by García Márquez *et al.* [75], spatial bias is an inherent problem of many vegetation-plot databases and can lead to the wrong model predictions. In Angola, very limited accessibility and the danger of land mines put strong restrictions on a purely random stratified sampling design. Consequently, some areas of the basin and some vegetation units are under-sampled, e.g., the vegetation belt in the transition from the Miombo woodlands to the *Baikiaea-Burkea* woodlands at the base of the Angolan Central Plateau. Furthermore, the data of the thornbush savanna surrounding the Okavango Delta are scarce. The spatial bias of the response variable may also explain considerably high RMSE values. Beyond that, regions with lower sampling intensity showed the highest discrepancies between the two model algorithms on the datasets “CLIMATE” and “ALL”. However, generally, the vegetation database contains a sufficient number of plots and represents the best available vegetation dataset for the region.

The relatively coarse resolution of MODIS data might also be a potential error source, where especially small vegetation units are acquired in mixed pixels and, thus, are negatively affecting the proper linkage to the smaller vegetation plots. Hence, substituting MODIS imagery with spatially appropriate remote sensing data could improve predictions. However, at the current state, deriving LSP based on Landsat for tropical Africa remains problematic, as one image at least every 16 days is required [76]. This is not the case in tropical Africa due to the reduced data availability for this region in the Landsat archive [77] and missing clear sky images from the wet season. Nevertheless, the increasing revisit frequency of the medium-resolution satellite-platforms Landsat and Sentinel-2 might account for this problem, alleviating the direct derivation of LSP at the required spatial scale.

#### 4.3. Patterns of Plant Alpha Diversity

In general, the derived maps based on MODIS LSP (“RS TOPO”) showed a realistic pattern of plant  $\alpha$ -diversity when compared to the vegetation map of the region [39]. The highest plant  $\alpha$ -diversity was predicted for the more mesic regions of the upper reaches of the Okavango Basin and steadily decreased southwards. Hence, plant  $\alpha$ -diversity followed the environmental gradient of decreasing precipitation and increasing temperatures in a north-south direction. This pattern is in line with the globally-observed phenomenon of a latitudinal gradient of species richness [78]. However, the gradient is a rough abstraction with many exceptions, and the underlying process are still debated [79]. Apart from global or continental maps, there are no previous studies depicting plant  $\alpha$ -diversity of the Okavango Basin. The global map on vascular plant diversity of Barthlott *et al.* [30] operates on a spatial scale of 10,000 km<sup>2</sup> and features only three diversity zones for the Okavango Basin. According to this map, plant  $\alpha$ -diversity ranges from 500 to 2000 species per 10,000 km<sup>2</sup>. Naturally, the number of species increases with increasing plot size or reference area of a map. However, the increase in species number with increasing area is system dependent and, thus, results in species area curves that vary according to vegetation type [80]. Therefore, given this scale dependency of plant  $\alpha$ -diversity, our results cannot be easily scaled up to the larger map units of Barthlott *et al.* [30] to directly compare the data. Nevertheless, it becomes evident that due to the high spatial resolution, our maps reveal diversity patterns with unprecedented detail showing more than a purely latitudinal gradient. On the Angolan Central Plateau in the north of the Okavango Basin, two major vegetation units occur in close proximity following the pattern of the gently rolling topography of the landscape: Miombo woodlands dominate on elevated areas, while geoxylic grasslands inhabit the slopes [37]. The measured plant  $\alpha$ -diversity of the Miombo woodlands was significantly higher than plant  $\alpha$ -diversity of the geoxylic grasslands (Figure 5A). The models on the ‘RS TOPO’ dataset were capable of capturing this difference, but not the ‘CLIMATE’ models (Figure 5). The major urban centers of the basin showed low plant  $\alpha$ -diversity, which



can be explained by their spectral similarity to open vegetation types or shrub-dominated thornbush savanna also featuring low diversity.

Incorporating climate data into modelling species densities did improve model performance when compared to remote sensing-only models “RS TOPO” (Table 2). However, a visual evaluation of the resulting maps revealed artefacts in the presented patterns, *i.e.*, the predicted patterns of plant  $\alpha$ -diversity did not match existing patterns in the vegetation of the Okavango Basin (Figure 1). Maps produced by BRT and RF on the full set of predictor variables (including climate, but also remote sensing information, dataset “ALL”) showed less obvious artifacts. Nevertheless, the maps exhibited sharp borders with abruptly changing values of plant  $\alpha$ -diversity (Figure 4). Only in some cases did these changes coincide with climatic borders resulting in actual alteration of land cover, *i.e.*, at the southern foothills of the Angolan Central Plateau. Moreover, the differences between the predictions of plant  $\alpha$ -diversity by BRT and RF were much larger when climate data were included in the modeling. The differences did not follow a systematic pattern, but showed a spatial pattern (Figure 4C,F,I). Thus, the error can be related to the fact that the model algorithms give different weight to the climatic predictor variables (Figure 3).

Models including climate (“CLIMATE” and “ALL”) reproduced large-scale climatic gradients resulting in belts of undifferentiated plant  $\alpha$ -diversity. In contrast, models based on LSP metrics and topography (“RS TOPO”) produced by far better maps as judged by experts. The maps depict local differences in plant  $\alpha$ -diversity reflecting the mosaic of the landscape that is blurred in the climate models, as evident from the Miombo region (Figure 5). Climatic predictors are known to be large-scale determinants, while land cover predictors gain importance on smaller spatial scales [29,51]. Therefore, extra- and azonal vegetation types pose challenges in predictive modeling if climatic predictors are included and make careful checks or even post-processing required [81].

#### 4.4. Biophysical Meaning of LSP Metrics

The productivity-diversity hypothesis [82] links the variation in species diversity to productivity measured as plant biomass and proposes a hump-backed relationship. However, four decades after it was first hypothesized, the exact form of the relationship between plant  $\alpha$ -diversity and biomass, as well as its generality across biomes is still hotly debated ([26,27], and citations therein). While empirical studies usually measure biomass in  $\text{kg} \cdot \text{ha}^{-1}$ , we had to rely on LSP metrics derived from EVI time series as a proxy. The “LargeIntegral” can be considered an indicator for total biomass [83,84], the “Amplitude” for the build-up of life biomass during the vegetation period [84] and ‘BaseValue’ as the share of biomass that remains after senescence of the vegetation during the dry season [83].

In this study, we showed that biomass-related LSP metrics (e.g., “LargeIntegral”, “Amplitude” and “BaseValue”) are good predictors for plant  $\alpha$ -diversity (Figure 3). The models for the Okavango Basin showed that areas with low productivity, such as the dry thornbush savanna, featured low species numbers, while the mesic Miombo woodlands exhibited the highest productivity and also the highest number of species. However, we did not find the originally-proposed unimodal relationship, but diverse response functions of plant  $\alpha$ -diversity to above ground biomass-related LSP metrics in all BRT and RF models (sigmoidal, linear and bimodal relationships; Figure S5).

Although vegetation indices, such as the EVI, are well-known proxies for above ground biomass, they tend to saturate at high biomass values [85]. Consequently, the forecasted effect of low plant  $\alpha$ -diversity at sites with high biomass [27,82] could be blurred and, hence, might restrict our observed response to an apparently linear relationship. Furthermore, the generality of the productivity-diversity relationship is still debated and could be biome or even formation specific. Sampling across multiple biomes and plant formations ranging from grasslands, savannas to forests as in the case of this study could lead to superimposition of multiple relationships, resulting in an overall weak linear response. Nonetheless, the productivity-diversity relationship could provide an important theoretical background for spatially-modelled  $\alpha$ -diversity based on remotely-sensed LSP metrics.

#### 4.5. Do Additional Climate Data Improve Models and Maps?

In spite of higher model performance of the models incorporating climate data, the resulting diversity maps are unrealistic and not meaningful when compared to actual observations (Figures 4 and 5). The low quality of the spatial representation of plant  $\alpha$ -diversity of the climate models could be related to the coarse spatial resolution of the climate data. The study region of the Okavango Basin encompasses an area with very limited historic climate data available to calibrate regional climate models. One reason could therefore be that the regionalized climatologies of the regional climate model REMO do not capture the climatic patterns in the Okavango Basin well enough. We thus tested a second climate dataset from independent sources: for temperature, we used data from the Climate Research Unit (CRU) [62]; for precipitation, we used the remotely-sensed information from the African Rainfall Climatologies (ARC2) [61]. The resulting models had comparable model performance and contained different, but similar artifacts (Table S1, Figure S6). In conclusion, artifacts in modeled diversity maps were not related to the source of climate data, but the problem is inherent to using climate data as predictors for modeling plant  $\alpha$ -diversity of the ecosystems of southern Africa on a medium spatial resolution. Modeling tree diversity of the Amazon basin, Saatchi *et al.* [34] came to a similar conclusion that gridded climate data cannot fully capture landscape-scale variation in plant  $\alpha$ -diversity, as the patterns are, apart from climate, controlled by local phenomena, such as soil properties, geology, nutrient availability and past history of the area. Land cover, in turn, is a result of large-scale (climatic) gradients, but also mirrors site conditions and the history of disturbance events. For the Okavango Basin, predicted patterns of plant  $\alpha$ -diversity are similar to patterns of the land cover classification of Stellmes *et al.* [39] (Figure 5B,C), hence supporting the assumption of Turner *et al.* [12] that land cover is a good proxy for estimating diversity.

The fact that the chosen performance criteria did not identify the models delivering the most realistic maps as the best ones is highly problematic and poses fundamental questions on how to judge the validity of models. At the same time, it highlights the importance of cross-checking model results with experts and revising the resulting maps within an ecological context. Not to treat statistical significance synonymously with ecological relevance is paramount if communicating scientific results to stakeholders and policy makers [86].

One explanation may lie in the ecology of the studied ecosystems. To a large extent, savanna ecosystems are disturbance driven; especially fire has played a major role in their evolution and maintenance [87,88]. Midgley and Bond [89] therefore argued that climatic predictors are not an ideal choice to model these ecosystems. Therefore, remote sensing predictors depicting the current land cover irrespective of the potential natural vegetation serve as better predictors. Nevertheless, this is not reflected by the higher model performance of the “RS TOPO” models. However, in RF, the LSP predictor variables had higher predictive power than climatic predictors, while in BRT, the opposite was the case. This also explains the different patterns of the resulting maps. Including further remote sensing-based predictors depicting fire will be promising. In savanna ecosystems, the fire frequency and the timing of fire in the vegetation period are of paramount importance [90]. On the one hand, short fire return periods may impede tree generation, capturing trees permanently in the sapling stage, the so-called “demographic-bottleneck” [88]. On the other hand, fires early in the dry season mainly affect the herbaceous layer, while hot, late dry season fires are more likely to also impact canopy species. The corresponding parameter can be derived from the MODIS active fire product (MOD14A1 and MYD14A1) and MODIS burned area (MCD45, 500-m resolution) [90] and included in the modeling.

Another important ecological feature shaping the spatial pattern of the dwarf shrub-grasslands of the Angolan Central Plateau is the frequent occurrence of nocturnal frost in the low-lying valleys during the dry season [52,53]. Generally, adaptations to frost are limited in the flora of tropical Africa. Therefore, the regular frost events reduce the species pool of the dwarf shrub-grasslands to a large extent to frost avoidance specialists protecting their buds underground, e.g., dwarf shrubs or so-called “geoxyles” [91], or under dry leaf matter, e.g., many tufted C4 grasses. To develop topographically-corrected climate

datasets showing spatial and temporal extents of cold air during night frost events will thus be a promising way forward to improve vegetation modeling in tropical highlands.

## 5. Conclusions

Vegetation-plot databases harbor a great potential to provide response variables for modeling ecosystem properties using remote sensing data. In this study, we showed that plant  $\alpha$ -diversity derived from such databases can be used for predicting plant  $\alpha$ -diversity of unsurveyed areas using land surface phenology derived from MODIS EVI time series. The models for the Okavango Basin showed that the Miombo woodlands of the Angolan Central Plateau feature the highest plant  $\alpha$ -diversity and that plant  $\alpha$ -diversity decreases southwards, reaching the lowest values in the thornbush savanna surrounding the Okavango Delta. In spite of higher model performance, models incorporating resampled climate data did not produce realistic maps on plant  $\alpha$ -diversity. The suitability of climate predictors for modeling plant  $\alpha$ -diversity on a medium spatial resolution has therefore to be questioned. Using MODIS LSP metrics as predictor variables has several advantages for modeling plant diversity. First, the global coverage ensures transferability of modeling frameworks to other regions. Second, the medium spatial resolution is fine enough to display local patterns of the landscape mosaic. Third, using land cover-related predictor variables instead of climatic predictors improves the representation of extra- and azonal vegetation types. The presented modeling approach combines plot-based ecological field data with continuous remote sensing data and, hence, enables predictions of ecosystem properties for vast, unsurveyed areas as they exist in many parts of the world. In this way, the approach may contribute to systematic conservation planning, as it provides the much needed spatial information for, e.g., identifying biodiversity hot spots or the delimitation of protected areas.

**Supplementary Materials:** The supplementary materials of this paper are available online at [www.mdpi.com/2072-4292/8/5/370/s1](http://www.mdpi.com/2072-4292/8/5/370/s1). Table S1: Validation results for the two model types boosted regression trees (BRT) and random forest (RF) on the three subsets of the predictor variables (a) remote sensing and topography 'RS TOPO' (b) only climate data derived from CRU and ARC2 'CLIMATE CRU/ARC2', (c) all data 'ALL' ('RS TOPO' and 'CLIMATE CRU/ARC2'). The following performance measures were calculated: explained variance (expl. var. [%]), Pearson's correlation coefficient (rp) between observed and predicted values, coefficient of determination ( $R^2$ ), the root mean square error (RMSE, in species per  $10^3 \text{ m}^2$ ) and the RMSE normalized by the mean, the relative root mean square error (rRMSE in per cent). The results for training and testing data are displayed (training 80% of the data and testing 20%); Figure S1: Observed values of alpha diversity plotted against predicted values on training data for (A) BRT on data set 'RS TOPO'; (B) RF on data set 'RS TOPO'; (C) BRT on data set 'CLIMATE'; (D) RF on data set 'CLIMATE'; (E) BRT on data set 'ALL'; (F) RF on data set 'ALL'; Figure S2: Observed plant alpha diversity (species density per  $10^3 \text{ m}^2$ ). Data is based on 999 vegetation plots sized  $20 \times 50 \text{ m}$ ; Figure S3: Model residual for the two model types: boosted regression trees (A,C,E) and random forest (B,D,E) on the three datasets: 'RS TOPO' (A,B); 'CLIMATE' (C,D); 'ALL'; (G,H). Furthermore, we calculated variograms to check for spatial autocorrelation but no severe spatial autocorrelation was detected; Figure S4: Plant alpha diversity (species density per  $10^3 \text{ m}^2$ ) predicted by the two model types: BRT (A,C,E) and random forest (B,D,F) on the three data (sub-)sets: 'RS TOPO' (A,B); 'CLIMATE' (D,E); 'ALL' (E,F); Figure S5: Partial dependence plots of the LSP metrics 'Amplitude' (A,D), 'BaseValue' (B,E) 'LargeIntegral' (C,F) for the two model types BRT (A–C) and RF (D–F); Figure S6: Plant alpha diversity (species density per  $10^3 \text{ m}^2$ ) predicted by the two model types: BRT (A,D) and random forest (B,E) on the second climate data set CRU/ARC2 (A,B); and on the entire data set comprising the second climate data set CRU/ARC2 and remote sensing data (D,E) and the difference between the two model algorithms (C and F). For a map on observed species density see Figure S1.

**Acknowledgments:** Research was funded by the German Federal Ministry of Education and Research (BMBF) in the context of The Future Okavango (TFO) project, Grant Number 01LL0912A. The MODIS and SRTM data were retrieved from the online Data Pool, courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, [https://lpdaac.usgs.gov/data\\_access/data\\_pool](https://lpdaac.usgs.gov/data_access/data_pool). Furthermore, we thank all people involved in the field work and the local communities of the study area for their support during field work. We thank Thomas Weinzierl and Torsten Weber for providing climate data. Finally, we acknowledge the thoughtful comments of the anonymous reviewers that helped to significantly improve the manuscript.

**Author Contributions:** All authors contributed to the final version of the manuscript. R.R. designed the study, compiled environmental data, carried out statistical modeling and wrote the first draft of the manuscript. J.O. continuously contributed to the modeling part and study design. B.S., M.F. and R.R. carried out field work. M.S. and D.F. processed satellite data and produced the LSP metrics.

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

## References

1. Butchart, S.H.M.; Walpole, M.; Collen, B.; van Strien, A.; Scharlemann, J.P.W.; Almond, R.E.A.; Baillie, J.E.M.; Bomhard, B.; Brown, C.; Bruno, J.; *et al.* Foster Global biodiversity: Indicators of recent declines. *Science* **2010**, *328*, 1164–1168. [[CrossRef](#)] [[PubMed](#)]
2. Pettorelli, N.; Skidmore, A.K. Agree on biodiversity metrics to track from space. *Nature* **2015**, *523*, 403–405.
3. Pereira, H.; Ferrier, S.; Walters, M. Essential biodiversity variables. *Science* **2013**, *339*, 277–278. [[CrossRef](#)] [[PubMed](#)]
4. Dengler, J.; Jansen, F.; Glöckler, F.; Peet, R.K.; De Cáceres, M.; Chytrý, M.; Ewald, J.; Oldeland, J.; Lopez-Gonzalez, G.; Finckh, M.; *et al.* The Global Index of Vegetation-Plot Databases (GIVD): A new resource for vegetation science. *J. Veg. Sci.* **2011**, *22*, 582–597. [[CrossRef](#)]
5. Jansen, F.; Glöckler, F.; Chytrý, M.; De Cáceres, M.; Ewald, J.; Lopez-gonzalez, G.; Oldeland, J.; Peet, R.K.; Schaminée, J.H.J.; Dengler, J. News from the Global Index of Vegetation-Plot Databases (GIVD): The metadata platform, available data, and their properties. *Biodivers. Ecol.* **2012**, *4*, 77–82. [[CrossRef](#)]
6. Nightingale, J.M.; Fan, W.; Coops, N.C.; Waring, R.H. Predicting tree diversity across the United States as a function of modeled gross primary production. *Ecol. Appl.* **2008**, *18*, 93–103. [[CrossRef](#)] [[PubMed](#)]
7. Van Rooijen, N.M.; de Keersmaecker, W.; Ozinga, W.A.; Coppin, P.; Hennekens, S.M.; Schaminée, J.H.J.; Somers, B.; Honnay, O. Plant Species Diversity Mediates Ecosystem Stability of Natural Dune Grasslands in Response to Drought. *Ecosystems* **2015**, *18*, 1383–1394. [[CrossRef](#)]
8. Wang, K.; Franklin, S.E.; Guo, X.; Cattet, M. Remote sensing of ecology, biodiversity and conservation: A review from the perspective of remote sensing specialists. *Sensors* **2010**, *10*, 9647–9667. [[CrossRef](#)] [[PubMed](#)]
9. Sutherland, W.J. *Ecological Census Techniques: A handbook*; University Press: Cambridge, UK, 1997; Volume 12.
10. Magurran, A. *Measuring Biological Diversity*; Blackwell Science: Malden, MA, USA, 2004.
11. Gaston, K.J.; Spicer, J.I. *Biodiversity: An Introduction*, 2nd ed.; Blackwell Publishing: Oxford, UK, 2004.
12. 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](#)]
13. Gillespie, T.W.; Foody, G.M.; Giorgi, A.P.; Saatchi, S.; Ambientali, S.; Mattioli, V. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* **2008**, *32*, 203–221. [[CrossRef](#)]
14. Elith, J.; Leathwick, J.R. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time—Appendix. *Annu. Rev. Ecol. Evol. Syst.* **2009**, *40*, 677–697. [[CrossRef](#)]
15. Justice, C.O.; Townshend, J.R.G.; Holben, B.N.; Tucker, C.J. Analysis of the phenology of global vegetation using meteorological satellite data. *Int. J. Remote Sens.* **1985**, *6*, 1271–1318. [[CrossRef](#)]
16. Jönsson, P.; Eklundh, L. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1824–1832. [[CrossRef](#)]
17. Jönsson, P.; Eklundh, L. TIMESAT—A program for analyzing time-series of satellite sensor data. *Comput. Geosci.* **2004**, *30*, 833–845. [[CrossRef](#)]
18. Fan, H.; Fu, X.; Zhang, Z.; Wu, Q. Phenology-Based Vegetation Index Differencing for Mapping of Rubber Plantations Using Landsat OLI Data. *Remote Sens.* **2015**, *7*, 6041–6058. [[CrossRef](#)]
19. 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](#)]
20. Cord, A.F.; Klein, D.; Gernandt, D.S.; de la Rosa, J.A.P.; Dech, S. Remote sensing data can improve predictions of species richness by stacked species distribution models: A case study for Mexican pines. *J. Biogeogr.* **2014**, *41*, 736–748. [[CrossRef](#)]
21. Tuanmu, M.-N.; Viña, A.; Bearer, S.; Xu, W.; Ouyang, Z.; Zhang, H.; Liu, J. Mapping understory vegetation using phenological characteristics derived from remotely sensed data. *Remote Sens. Environ.* **2010**, *114*, 1833–1844. [[CrossRef](#)]



22. Fensholt, R.; Horion, S.; Tagesson, T.; Ehammer, A.; Ivits, E.; Rasmussen, K. Global-scale mapping of changes in ecosystem functioning from earth observation-based trends in total and recurrent vegetation. *Glob. Ecol. Biogeogr.* **2015**, 1003–1017. [[CrossRef](#)]
23. Stellmes, M.; Röder, A.; Udelhoven, T.; Hill, J. Mapping syndromes of land change in Spain with remote sensing time series, demographic and climatic data. *Land Use Policy* **2013**, 30, 685–702. [[CrossRef](#)]
24. Senf, C.; Pflugmacher, D.; van der Linden, S.; Hostert, P. Mapping Rubber Plantations and Natural Forests in Xishuangbanna (Southwest China) Using Multi-Spectral Phenological Metrics from MODIS Time Series. *Remote Sens.* **2013**, 5, 2795–2812. [[CrossRef](#)]
25. Hüttich, C.; Gessner, U.; Herold, M.; Strohbach, B.J.; Schmidt, M.; Keil, M.; Dech, S. On the Suitability of MODIS Time Series Metrics to Map Vegetation Types in Dry Savanna Ecosystems: A Case Study in the Kalahari of NE Namibia. *Remote Sens.* **2009**, 1, 620–643. [[CrossRef](#)]
26. Tredennick, A.T.; Adler, P.B.; Grace, J.B.; Harpole, W.S.; Borer, E.T.; Seabloom, E.W.; Anderson, T.M.; Bakker, J.D.; Biederman, L.A.; Brown, C.S.; *et al.* Comment on “Worldwide evidence of a unimodal relationship between productivity and plant species richness”. *Science* **2015**, 351, 457. [[CrossRef](#)] [[PubMed](#)]
27. Fraser, L.H.; Pither, J.; Jentsch, A.; Sternberg, M.; Zobel, M.; Askarizadeh, D.; Bartha, S.; Beierkuhnlein, C.; Bennett, J.A. Worldwide evidence of a unimodal relationship between productivity and plant species richness. *Science* **2015**, 349, 302–306. [[CrossRef](#)] [[PubMed](#)]
28. Pearson, R.G.; Dawson, T.P.; Liu, C. Modelling species distributions in Britain: A hierarchical integration of climate and land-cover data. *Ecography* **2004**, 27, 285–298. [[CrossRef](#)]
29. Luoto, M.; Virkkala, R.; Heikkinen, R.K. The role of land cover in bioclimatic models depends on spatial resolution. *Glob. Ecol. Biogeogr.* **2007**, 16, 34–42. [[CrossRef](#)]
30. Barthlott, W.; Mutke, J.; Rafiqpoor, D.; Kier, G.; Kreft, H. Global Centers of Vascular Plant Diversity. *Nov. Acta Leopoldina* **2005**, 92, 61–83.
31. Viedma, O.; Torres, I.; Pérez, B.; Moreno, J.M. Modeling plant species richness using reflectance and texture data derived from QuickBird in a recently burned area of Central Spain. *Remote Sens. Environ.* **2012**, 119, 208–221. [[CrossRef](#)]
32. Feilhauer, H.; Schmidtlein, S. Mapping continuous fields of forest alpha and beta diversity. *Appl. Veg. Sci.* **2009**, 12, 429–439. [[CrossRef](#)]
33. 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](#)]
34. Saatchi, S.; Buermann, W.; ter Steege, H.; Mori, S.; Smith, T.B. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sens. Environ.* **2008**, 112, 2000–2017. [[CrossRef](#)]
35. Steudel, T.; Göhmann, H.; Flügel, W.-A.; Helmschrot, J. Assessment of hydrological dynamics in the upper Okavango River Basins. *Biodivers. Ecol.* **2013**, 5, 247–261. [[CrossRef](#)]
36. Weber, T. Okavango Basin—Climate. *Biodivers. Ecol.* **2013**, 5, 15–17. [[CrossRef](#)]
37. Revermann, R.; Gomes, A.; Goncalves, F.M.; Lages, F.; Finckh, M. Cusseque—Vegetation. *Biodivers. Ecol.* **2013**, 5, 59–63. [[CrossRef](#)]
38. Revermann, R.; Finckh, M. Okavango Basin—Vegetation. *Biodivers. Ecol.* **2013**, 5, 29–35. [[CrossRef](#)]
39. Stellmes, M.; Frantz, D.; Finckh, M.; Revermann, R. Okavango Basin—Earth Observation. *Biodivers. Ecol.* **2013**, 5, 23–27. [[CrossRef](#)]
40. Wehberg, J.; Weinzierl, T. Okavango Basin—Physicogeographical setting. *Biodivers. Ecol.* **2013**, 5, 11–13. [[CrossRef](#)]
41. Gossweiler, J.; Mendonça, F.A. *Carta Fitogeográfica de Angola*; República Portuguesa Ministério das Colónias: Lisbon, Portugal, 1939.
42. Barbosa, L.A.G. *Carta Fitogeográfica de Angola*; Instituto de Investigação Científica de Angola: Luanda, Angola, 1970.
43. Monteiro, R.F.R. *Estudo da Flora e da Vegetação das Florestas Abertas do Plantalto do Bié*; Instituto de Investigação Científica de Angola: Luanda, Angola, 1970.
44. Dos Santos, R.M. *Itinários Florísticos e carta da Vegetação do Cuando Cubango*; Instituto de Investigação Científica Tropical: Lisbon, Portugal, 1982.
45. Wallenfang, J.; Finckh, M.; Oldeland, J.; Revermann, R. Impact of shifting cultivation on dense tropical woodlands in southeast Angola. *Trop. Conserv. Sci.* **2015**, 8, 863–892.
46. Revermann, R.; Finckh, M. Caiundo—Vegetation. *Biodivers. Ecol.* **2013**, 5, 91–96. [[CrossRef](#)]

47. Revermann, R.; Gomes, A.L.; Gonçalves, F.M.; Wallenfäng, J.; Hoche, T.; Jürgens, N.; Finckh, M. Vegetation Database of the Okavango Basin. *Phytocoenologia* **2016**. [[CrossRef](#)]
48. Strohbach, B.; Kangombe, F. National Phytosociological Database of Namibia. *Biodivers. Ecol.* **2012**, *4*, 298. [[CrossRef](#)]
49. Sonnenschein, R.; Kuemmerle, T.; Udelhoven, T.; Stellmes, M.; Hostert, P. Differences in Landsat-based trend analyses in drylands due to the choice of vegetation estimate. *Remote Sens. Environ.* **2011**, *115*, 1408–1420. [[CrossRef](#)]
50. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
51. Waring, R.H.; Coops, N.C.; Fan, W.; Nightingale, J.M. MODIS enhanced vegetation index predicts tree species richness across forested ecoregions in the contiguous U.S.A. *Remote Sens. Environ.* **2006**, *103*, 218–226. [[CrossRef](#)]
52. Revermann, R.; Finckh, M. Cusseque—Microclimate. *Biodivers. Ecol.* **2013**, *5*, 47–50. [[CrossRef](#)]
53. Finckh, M.; Revermann, R.; Aidar, M.P.M. Climate refugees going underground—A response to Maurin *et al.* (2014). *New Phytol.* **2016**, 904–909. [[CrossRef](#)] [[PubMed](#)]
54. Wilson, J.P.; Gallant, J.C. *Terrain Analysis—Principles and Applications*; Wiley: New York, NY, USA, 2000.
55. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [[CrossRef](#)]
56. Riley, S.J.; DeGloria, S.D.; Elliot, R. A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. *Intermt. J. Sci.* **1999**, *5*, 23–27.
57. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [[CrossRef](#)]
58. Weinzierl, T.; Conrad, O.; Böhner, J.; Wehberg, J. Regionalization of Baseline Climatologies and Time Series for the Okavango Catchment. *Biodivers. Ecol.* **2013**, *5*, 235–245. [[CrossRef](#)]
59. Jacob, D. A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. *Meteorol. Atmos. Phys.* **2001**, *77*, 61–73. [[CrossRef](#)]
60. Hijmans, R.J.; Phillips, S.; Leathwick, J.; Elith, J. Dismo: Species Distribution Modeling. Available online: <https://CRAN.R-project.org/package=dismo> (accessed on 10 June 2015).
61. Novella, N.S.; Thiaw, W.M. African Rainfall Climatology Version 2 for Famine Early Warning Systems. *J. Appl. Meteorol. Climatol.* **2013**, *52*, 588–606. [[CrossRef](#)]
62. Harris, I.; Jones, P.D.; Osborn, T.J.; Lister, D.H. Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 Dataset. *Int. J. Climatol.* **2014**, *34*, 623–642. [[CrossRef](#)]
63. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; *et al.* Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 027–046. [[CrossRef](#)]
64. Wei, T. Corrplot: Visualization of a correlation matrix 2013. Available online: <https://CRAN.R-project.org/package=corrplot> (accessed on 10 July 2015).
65. Pearson, R.G.; Thuiller, W.; Araújo, M.B.; Martinez-Meyer, E.; Brotons, L.; McClean, C.; Miles, L.; Segurado, P.; Dawson, T.P.; Lees, D.C. Model-based uncertainty in species range prediction. *J. Biogeogr.* **2006**, *33*, 1704–1711. [[CrossRef](#)]
66. Elith, J.; Graham, C.H.; Anderson, R.P.; Dudík, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; *et al.* Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **2006**, *29*, 129–151. [[CrossRef](#)]
67. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [[CrossRef](#)]
68. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **2006**, *9*, 181–199. [[CrossRef](#)]
69. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
70. Ridgeway, G. Gbm: Generalized Boosted Regression Models. Available online: <https://CRAN.R-project.org/package=gbm> (accessed on 15 July 2015).
71. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; the R Core Team; *et al.* caret: Classification and Regression Training. Available online: <https://CRAN.R-project.org/package=caret> (accessed on 15 July 2015).

72. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)] [[PubMed](#)]
73. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
74. R: A language and environment for statistical computing, R Core Team (R Foundation for Statistical Computing): Vienna, Austria, 2015.
75. García Márquez, J.; Dormann, C.; Sommer, J.H.; Schmidt, M.; Thiombiano, A.; Da Sylvestre, S.; Chatelain, C.; Dressler, S.; Barthlott, W. A methodological framework to quantify the spatial quality of biological databases. *Biodivers. Ecol.* **2012**, *4*, 25–39. [[CrossRef](#)]
76. Archibald, S.; Scholes, R. Leaf green-up in a semi-arid African savanna-separating tree and grass responses to environmental cues. *J. Veg. Sci.* **2007**, *18*, 583–594. [[CrossRef](#)]
77. Kovalskyy, V.; Roy, D.P. The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30m Landsat data product generation. *Remote Sens. Environ.* **2013**, *130*, 280–293. [[CrossRef](#)]
78. Willig, M.R.; Kaufman, D.M.; Stevens, R.D. Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. *Annu. Rev. Ecol. Evol. Syst.* **2003**, *34*, 273–309. [[CrossRef](#)]
79. Gaston, K.J. Global patterns in biodiversity. *Nature* **2000**, *405*, 220–227. [[CrossRef](#)] [[PubMed](#)]
80. Dengler, J. Which function describes the species-area relationship best? A review and empirical evaluation. *J. Biogeogr.* **2009**, *36*, 728–744. [[CrossRef](#)]
81. Jürgens, N.; Oldeland, J.; Hachfeld, B.; Erb, E.; Schultz, C. Ecology and spatial patterns of large-scale vegetation units within the central Namib Desert. *J. Arid Environ.* **2013**, *93*, 59–79. [[CrossRef](#)]
82. Grime, J.P. Control of species density in herbaceous vegetation. *J. Environ. Manag.* **1973**, *1*, 151–167.
83. Helman, D.; Lensky, I.; Tessler, N.; Osem, Y. A Phenology-Based Method for Monitoring Woody and Herbaceous Vegetation in Mediterranean Forests from NDVI Time Series. *Remote Sens.* **2015**, *7*, 12314–12335. [[CrossRef](#)]
84. DeFries, R.; Hansen, M.; Townshend, J. Global discrimination of land cover types from metrics derived from AVHRR pathfinder data. *Remote Sens. Environ.* **1995**, *54*, 209–222. [[CrossRef](#)]
85. Huete, R.; Liu, H.L.H.; Van Leeuwen, W.J.D. The use of vegetation indices in forested regions: Issues of linearity and saturation. In Proceedings of the 1997 IEEE International Geoscience and Remote Sensing, 1997. IGARSS '97. Remote Sensing—A Scientific Vision for Sustainable Development, Singapore, 3–8 August 1997.
86. Martínez-Abraín, A. Statistical significance and biological relevance: A call for a more cautious interpretation of results in ecology. *Acta Oecol.* **2008**, *34*, 9–11. [[CrossRef](#)]
87. Bond, W.J.; Keeley, J.E. Fire as a global “herbivore”: The ecology and evolution of flammable ecosystems. *Trends Ecol. Evol.* **2005**, *20*, 387–394. [[CrossRef](#)] [[PubMed](#)]
88. Sankaran, M.; Ratnam, J.; Hanan, N.P. Tree-grass coexistence in savannas revisited—Insights from an examination of assumptions and mechanisms invoked in existing models. *Ecol. Lett.* **2004**, *7*, 480–490. [[CrossRef](#)]
89. Midgley, G.F.; Bond, W.J. Future of African terrestrial biodiversity and ecosystems under anthropogenic climate change. *Nat. Clim. Chang.* **2015**, *5*, 823–829. [[CrossRef](#)]
90. Stellmes, M.; Frantz, D.; Finckh, M.; Revmann, R.; Röder, A.; Hill, J. Fire frequency, fire seasonality and fire intensity within the Okavango region derived from MODIS fire products. *Biodivers. Ecol.* **2013**, *5*, 351–362. [[CrossRef](#)]
91. White, F. The underground forests of Africa: A preliminary review. *Gard. Bull. Singapore* **1976**, *11*, 57–71.

