

## ASSESSING THE IMPORTANCE OF TOPOGRAPHIC VARIABLES FOR THE SPATIAL DISTRIBUTION OF TREE SPECIES IN A TROPICAL MOUNTAIN FOREST

DANIEL KÜBLER, PATRICK HILDEBRANDT, SVEN GÜNTER, BERND STIMM, MICHAEL WEBER, REINHARD MOSANDL, JOHANA MUÑOZ, OMAR CABRERA, NIKOLAY AGUIRRE, JÖRG ZEILINGER and BRENNER SILVA

With 4 figures and 2 tables

Received 15 July 2015 · Accepted 17 December 2015

**Summary:** Availability and improved access to high-resolution digital terrain models (DTM) enables new approaches for the analysis of spatially explicit biological data. In this study, the spatial distribution of 16 tree species in a tropical mountain rain forest in South Ecuador and its relationship with topographic variables was evaluated at a fine-scale ecological level using two presence-only species distribution modelling techniques: The maximum entropy model (Maxent) and the ecological niche factor analysis (ENFA). Spatially explicit tree data stem from long-term forest monitoring plots in three microcatchments with a total area of 11.1 ha. Topographic variables were derived from a high-resolution DTM. Model performance was assessed by the true skill statistic (TSS) and area under curve (AUC) of the receiver operator characteristic (ROC), using both a k-fold approach and null-models. Performance varied among species and techniques, but generally Maxent models showed better performance than ENFA models. Furthermore, the ecological plausibility of the models was confirmed by comparing them with a previously established forest type classification. Among the explanatory topographic variables, elevation and a Topographic Position Index (TPI) appear as the main determinants for the distribution of most of the tree species. This study demonstrates that even on a small scale, the use of presence-only species distribution modelling techniques is a viable option for modelling suitable habitat for tree species in tropical mountain rain forests, indicating suitability for supporting stand-level planning and site-species matching techniques for natural forest management.

**Zusammenfassung:** Die zunehmende Verfügbarkeit von hochauflösenden digitalen Geländemodellen ermöglicht neue Ansätze zur Analyse von räumlich expliziten biologischen Daten. In dieser Studie haben wir die räumliche Verteilung von 16 Baumarten und deren Zusammenhang mit topographischen Variablen auf einer feinskaligen ökologischen Ebene in einem tropischen Bergregenwald in Südecuador evaluiert. Dafür haben wir zwei auf Präsenzdaten basierende Techniken zur Habitatmodellierung angewandt: Die Maximum-Entropie Methode (Maxent) und „Ecological Niche Factor Analysis“ (ENFA). Räumlich explizite Baumdaten stammen aus Langzeitbeobachtungsflächen von drei kleinen Wassereinzugsgebieten mit einer Fläche von 11.1 ha. Topographische Variablen wurden aus einem hochauflösenden digitalen Geländemodell abgeleitet. Zur Bewertung der Modelle wurden die True Skill Statistic (TSS) und die Fläche unterhalb der Kurve der Receiver Operating Characteristic (ROC) für sowohl k-fache Kreuzvalidierungen als auch Null-Modelle berechnet. Die Güte der Modelle hat zwischen den beiden Techniken und verschiedenen Arten variiert, aber generell schnitten Maxent-Modelle besser als ENFA-Modelle ab. Des Weiteren haben wir die ökologische Plausibilität der Modelle überprüft, indem wir sie mit einer bereits bestehenden Waldtypenklassifizierung verglichen haben. Unter den erklärenden topographischen Variablen scheinen die Höhe über dem Meeresspiegel und ein Topographischer Positions Index (TPI) die bestimmenden Faktoren für die Verteilung der meisten Baumarten zu sein. Diese Studie belegt, dass selbst auf kleinen Skalen auf Präsenzdaten basierende Techniken zur Habitatmodellierung angewandt werden können, um geeignete Habitate für Baumarten in tropischen Bergregenwäldern abzubilden. Dies deutet darauf hin, dass diese Techniken zur Unterstützung von Planungen auf Bestandesebene und zur Auswahl von geeigneten Standorten für Baumarten im Rahmen der Bewirtschaftung von Naturwäldern geeignet sind.

**Keywords:** Montane forest, species distribution modelling, Maxent, ENFA, biogeography, Ecuador

### 1 Introduction

For land use and conservation planning and sustainable forest management, accurate information about tree species distribution and its ecological drivers on different spatial scales is essential. While there is a long research history for temperate forests about the relationship between the distribution of tree spe-

cies and environmental variables (ELLENBERG 2009; BURNS and HONKALA 1990a, 1990b), the knowledge about this subject is scarce for tropical forests. This holds especially true for the Andean tropical montane forest, which is an epicenter of global biodiversity and endemism (MYERS et al. 2000), yet still one of the least-studied tropical regions on the planet (PITMAN et al. 2011).

Only few studies have previously analyzed the relationship between environmental variables and species distribution in steep tropical mountain forests at local scales (SVENNING et al. 2009; JONES et al. 2011; AIBA et al. 2004; HOMEIER et al. 2010; LEDO et al. 2013; FAGUA et al. 2013; BALDECK et al. 2013). In these studies, topographic variables (e.g. elevation, aspect) were used as proxies for environmental factors such as light availability, micro-climate or edaphic conditions, which have a direct effect on plant growth. The studies generally used a spatial resolution of at least 10–20 m for both topographic and species data (but see LEDO et al. 2013), and a discrete habitat type classification to represent topography (but see BALDECK et al. 2013), which means that gradual micro-topographic habitat changes were not taken into account.

In this study, we use species distribution modelling (SDM) (GUISAN and THUILLER 2005) to assess the importance of topographic variables for the distribution of the 16 most abundant tree species in a 11.1-ha sample plot in a tropical mountain forest in southern Ecuador. SDM, a growing field in ecological biogeographical research, biodiversity conservation, and ecosystem management (ELITH et al. 2006), is rarely used at a very fine scale and in a small study area, which is probably linked to the scarcity of high-resolution environmental variables and datasets as well as species information at those fine scales (KHATCHIKIAN et al. 2011). Although SDM is generally used for predictive modelling of species environmental requirements, numerous examples exist where those methods were used to infer the importance of environmental factors for the distribution of species (e.g. WOLLAN et al. 2008; LAUSCH et al. 2011; HERMOSILLA et al. 2011; VEDEL-SØRENSEN et al. 2013; CAUWER et al. 2014). SDM allows for the consideration of gradual microtopographic habitat changes because species occurrences and continuous topographic variables can be used without any prior spatial aggregation or habitat type classification.

The main objectives of the study were to investigate the performance of two commonly used SDM algorithms (Ecological Niche Factor Analysis (ENFA) and Maxent) to model the distribution of tree species at a very fine scale, and subsequently, to assess the importance of different topographic variables for the spatial distribution of tree species.

## 2 Methods

### 2.1 Study area

Our study site is located within the privately owned Reserva Biológica San Francisco (RBSF, 3°58'S, 79°04'W), on the eastern slope of the Cordillera El Consuelo in the Andes of southern Ecuador. The RBSF has an altitude of 1800–3150 m asl and lies along the road between the towns Loja and Zamora in the Zamora-Chinchi Province, bordering on the North of the Podocarpus National Park. The altitude of our study site is 1850–2150 m asl. Mean annual precipitation at 1952 m amounts to 2100 mm and annual temperature averages 15 °C. The site is characterized by steep environmental gradients and a strong topographical heterogeneity, which in turn strongly defines the floristic composition and forest structure (HOMEIER et al. 2010). For a comprehensive description of the geology, climate, flora, and fauna of the RBSF, we refer to BECK et al. (2008a).

### 2.2 Tree dataset

Our tree dataset stems from permanent study plots, which were installed for a silvicultural forest management experiment and first measured in 2003 (GÜNTER et al. 2008). The size of the experimental plots on the ground is 13 ha, which corresponds to 11.1 ha in projected map view. The plots are subdivided between three microcatchments, have areas of 5, 4 and 4 ha on the ground, and a maximal distance of 870 m between farthest points (east–west direction, projected map view).

All trees on these sample plots with a diameter at breast height (DBH)  $\geq$  20 cm were measured for DBH, mapped, individually labeled, and identified to species level. Trees with DBH  $<$  20 cm were not assessed on the whole sample plot area, but on smaller subplots, and are therefore not included in this study. We included the 16 most abundant species, with total occurrences in the sampling plots ranging from 55–196 trees for each species, corresponding to 4.95–17.66 trees/ha. The included tree species represent 49% of abundance and 52% of basal area of all trees in the sampling plots (Tab. 1). Figure A-1 shows the distribution of all trees with DBH  $\geq$  20 cm of those species.

For the elevation range of the permanent study plots, HOMEIER (2008) distinguished three main forest types in the RBSF based on the combination

**Tab. 1:** Characteristics of the 16 species included in the analysis. The assignment of each tree species to a forest types is based on HOMEIER (2008). The altitudinal ranges of species are based on all available occurrence data in Ecuador in the Global Biodiversity Information Facility (GBIF)

Species	Species code	N	Abundance [N / ha]	Basal area [m <sup>2</sup> / ha]	Forest type	Altitudinal range in Ecuador [m asl]
<i>Alchornea grandiflora</i> Müll. Arg.	Alchgran	123	11.13	3.22	II, III	900–2300
<i>Alzatea verticillata</i> Ruiz & Pav.	Alzavert	93	8.42	3.43	II, III	1200–2250
<i>Cecropia andina</i> Cuatrec.	Cecrandi	81	7.33	2.12	I	1540–2400
<i>Cecropia angustifolia</i> Trécul	Cecrangu	194	17.55	5.32	I	0–2300
<i>Cedrela montana</i> Moritz ex Turcz.	Cedrmont	55	4.98	1.39		1400–3200
<i>Clusia ducunoides</i> Engl.	Clusducu	96	8.69	2.27	II, III	250–3030
<i>Graffenrieda emarginata</i> (Ruiz & Pav.) Triana	Grafemar	86	7.78	1.50	I, II, III	1120–2900
<i>Guarea kunthiana</i> A. Juss.	Guarkunt	109	9.86	4.95	I	80–3100
<i>Heliocarpus americanus</i> L.	Heliamer	73	6.61	2.26	I	50–2615
<i>Hyeronima asperifolia</i> Pax & K. Hoffm.	Hyeraspe	82	7.42	3.28	I	1.8–3000
<i>Meriania franciscana</i> C. Ulloa & Homeier	Merifran	68	6.15	1.83	I	1890–2500
<i>Nectandra lineatifolia</i> (Ruiz & Pav.) Mez	Nectline	60	5.43	1.64	I, II	200–3000
<i>Nectandra membranacea</i> (Sw.) Griseb.	Nectmemb	75	6.79	1.67	I	80–2900
<i>Sapium glandulosum</i> (L.) Morong	Sapiglan	63	5.70	1.80	I	30–2150
<i>Tabebuia chrysantha</i> G. Nicholson	Tabechry	97	8.78	5.36	I	0–1800
<i>Tapirira guianensis</i> Aubl.	Tapiguia	63	5.70	2.84	I	2–1800

of different vegetation classification approaches. Forest type I, the tallest and most speciose forest, grows in major ravines and gentle lower slopes on altitudes below 2200 m asl. The canopy in this forest type reaches 25–30 m, with some emergents reaching up to 35 m. Common trees of this forest type include *Piptocoma discolor* (Asteraceae), *Tabebuia chrysantha* (Bignoniaceae), *Hyeronima asperifolia* and *Sapium glandulosum* (Euphorbiaceae), *Nectandra lineatifolia* and *N. membranacea* (Lauraceae), *Meriania* sp., *Miconia punctata* and other *Miconia* spp (Melastomataceae), *Inga* spp (Mimosaceae), *Morus insignis*, *Naucleopsis glabra*, and *Ficus* spp (Moraceae), *Prumnopitys montana* (Podocarpaceae), *Micropholis guyanensis* (Sapotaceae) and *Heliocarpus americanus* (Tiliaceae). After the formation of gaps, distinct succession phases can be observed in forest type I, with fast-growing pioneer species (e.g. *Piptocoma discolor*, *Cecropia andina*, *Heliocarpus americanus*) being subsequently replaced by late-successional species and long-lived pioneers such as *Tabebuia chrysantha* and *Cedrela* sp. (HOMEIER and BRECKLE 2008).

Forest type II can be found on upper slopes and ridges between 1900–2100 m asl, with a canopy height of 15 m. This forest type, characterized by a thick humus layer, is dominated by the frequent tree species *Alzatea verticillata* (Alzateaceae), *Dictyocaryum lamarckianum* and *Wettinia aequatorialis* (Arecaceae), *Weinmannia pinnata*, *W. sorbifolia* and *W. spruceana* (Cunoniaceae), *Abarema keillipii* (Fabaceae), *Hyeronima moritzinana* (Euphorbiaceae), *Ocotea aciphylla* (Lauraceae), *Graffenrieda emarginata* and *Miconia calophylla* (Melastomataceae), *Podocarpus oleifolius* (Podocarpaceae) and *Matayba inelegans* (Sapindaceae). In contrast to forest type I, early and late successional phases show a very similar species composition in forest type II (HOMEIER and BRECKLE 2008).

In forest type III, found between 2100 and 2250 m asl, the trees usually do not surpass 12 m. Characteristic tree species include *Hedyosmum translucidum* (Chloranthaceae), *Clusia* cf. *ducunoides*, *Clusia* spp and *Tovomita meddeliana* (Clusiaceae), *Weinmannia haenkeana* and *W. ovata* (Cunoniaceae), *Purdiaea nutans* (Clethraceae), *Alchornea grandiflora*

(Euphorbiaceae), *Endlicheria oreocola*, *Licaria subsessilis*, *Ocotea benthamiana* and *Persea subcordata* (Lauraceae), *Eschweilera sessilis* (Lecythidaceae), *Graffenrieda emarginata* (Melastomataceae), *Calypttranthes pulchella* and *Myrcia* sp. (both Myrtaceae) and *Podocarpus oleifolius* (Podocarpaceae).

### 2.3 Topographic variables

The modelled area in this study corresponds to the area of the permanent sample plots. In order to define the habitat characteristics of those plots, we derived topographic variables based on a digital terrain model (DTM) with a spatial resolution of 1 m (covering 132267 raster data cells), which was created with data from an airborne LIDAR sensor. Derived topographic variables were calculated using the open source software SAGA GIS (SAGA Development Team 2008) and the R-package “RSAGA” (BRENNING 2008). In order to avoid edge effects, variables were first calculated over a larger extent and subsequently, a mask of the permanent sample plots was applied.

Both SDM algorithms used in this study (see Section 2.4) are considered to be relatively robust against multicollinearity between predictors (HIRZEL et al. 2002; ELITH et al. 2011). However, while model performance might not be affected by collinearity, model interpretation can be hindered and Maxent tends to overfit models (ELITH et al. 2011). In order to prevent this, we examined cross-correlations between potential topographic variables in a preliminary analysis. Only variables with pairwise Pearson correlations coefficients of  $r \leq 0.45$  were retained for the modelling process (Tab. A-1). Based on assumed ecological relevance for tree species distribution, we included the following variables in the final models: Elevation, slope, aspect, SAGA Wetness Index (WI) and Topographic Position Index (TPI) (Fig. A-2).

For the calculation of slope and aspect, we used a 2nd degree polynomial fit. Slope is an indicator of the intensity of gravitational and disturbance processes acting upon vegetation, which play an important role in our study area (MUENCHOW et al. 2012; VORPAHL et al. 2012). Mean slope on the permanent sample plots was  $31.2^\circ$ , with the first and third quartiles being  $25.8^\circ$  and  $36.6^\circ$ , respectively.

Aspect, an inherent circular variable, was converted into two separate continuous variables denominated northness and eastness by using the sine and cosine transformations, respectively. They quantify the degree to which the aspect is north and east. As our study plots are located on a north-

ern flank, values for northness are mostly positive towards one, with little variation between plots. Northness was therefore not included in the analysis. Eastness, on the other hand, was included, because it varies between plots and has an ecological importance as prevalent wind direction and climate influence is usually in an east-west gradient in the RBSF (ROLLENBECK 2006).

Soil moisture is an important factor structuring local pattern of species distributions (PÉLISSIER et al. 2002). To account for this, we calculated the SAGA Wetness Index, which represents relative local soil moisture availability. It is calculated similarly to the more commonly used Topographic Wetness Index, but is based on a modified catchment area calculation (BÖHNER et al. 2002).

Former studies showed a strong influence of topography on both biotic and abiotic factors in our study area (WILCKE et al. 2011; HOMEIER et al. 2010; WERNER et al. 2012; VORPAHL et al. 2012; SVENNING et al. 2009). Therefore, we calculated a TPI (GUISAN et al. 1999), ranging from positive values expressing ridges and upper slopes to negative values describing valleys and lower slopes. The TPI is scale-dependent; for its calculation, a moving circular window is used, whose size can be varied. In a pre-analysis, we assessed a range of different sizes for plausibility and retained a TPI with a moving window of 100 m radius.

Available spatial explicit data about environmental conditions, such as precipitation (e.g. FRIES et al. 2014), temperature, soil properties and landslide risk, proved to be too coarse to be included in the present study. In Section 4.3, we therefore discuss how individual topographic variables relate to environmental conditions in our study area.

### 2.4 Species distribution modelling

SDM is based on the niche concept defined by HUTCHINSON (1957), which assumes that species can persist in only a limited range of environmental conditions. According to this concept, the distribution of species is linked to an n-dimensional structure that is made up of different environmental factors. Models estimate the niche space occupied by species by linking species occurrences with environmental layers in a given study extent (FRANKLIN and MILLER 2009). However, how exactly SDM relates to the niche concept is currently discussed in an ongoing debate (ELITH and LEATHWICK 2009; JIMÉNEZ-VALVERDE et al. 2008; SOBERÓN 2010; ARAÚJO and

GUIBAN 2006). In our study, we followed JIMÉNEZ-VALVERDE et al. (2008) and considered that different modelling methods predict a gradient of potential–realized distributions, where potential distribution refers to the places where a species could live and realized distribution to the places where a species actually lives.

Generally, SDM algorithms can be classified as either presence-absence or presence-only models (ELITH et al. 2006). Having a complete tree census, absence data for a given species could be simulated for the modeled area based on the occurrence data, e.g. by randomly assigning an arbitrary number of pixels with a minimal distance to all occurrence points of that species as absence points. However, it would be difficult to distinguish between absences where species are absent because environmental conditions at a site are not suitable (“true absences” sensu HIRZEL and LE LAY (2008), “environmental absences” sensu LOBO et al. (2010)) and absences where a species might not be present on an environmentally suitable site due to dispersal limitations, biotic interactions, historic events and demographic stochasticity (“fallacious absences” sensu HIRZEL and LE LAY (2008), “contingent absences” sensu LOBO et al. (2010)). We assume that this is especially true at local scales and in species-rich ecosystems (WILLIS and WHITTAKER 2002) and for large organisms such as trees (JONES et al. 2008). We therefore used presence-only techniques for our analyses (HIRZEL and LE LAY 2008; JIMÉNEZ-VALVERDE et al. 2008).

In recent years, a number of factors relating to data quality has been identified that might reduce the performance of presence-only techniques. In this section we will mention how these factors relate to data from a complete tree census at a local scale, which has rarely been analyzed with SDM (but see HENGL et al. 2009). Such data has certain particularities in comparison with other datasets commonly used for SDM.

Spatial autocorrelation, a pattern in which observations are related to one another by the geographic distance between the observation (LEGENDRE and FORTIN 1989), is present in virtually all spatially explicit ecological datasets (LENNON 2000). Although it can artificially inflate validation measures for models (VELOZ 2009; MERCKX et al. 2011; SEGURADO et al. 2006), most studies using SDM do not directly account for the effects of spatial dependence (DORMANN et al. 2007; ELITH and LEATHWICK 2009). Contrary to SDM algorithms for presence/absence data (DORMANN et al. 2007; CRASE et al. 2012), little work has focused on methods for taking into

account the spatial autocorrelation of occurrences for presence-only algorithms (VELOZ 2009). We addressed this issue in our model evaluation technique (see Section 3.1.3) and in the discussion (see Section 4.2.3).

Geographical sampling bias can occur when some (usually easily accessible) areas are sampled more than others and can severely diminish model quality (PHILLIPS et al. 2009; SYFERT et al. 2013). We acknowledge that our data only represents a small part of the true range of environmental conditions in which the tree species occur. However, we believe that no geographical sampling bias exists within our modeled area, because sampling intensity is equal for each site during a full tree census. Therefore, we applied no method for dealing with geographical sampling bias during the model building process.

#### 2.4.1 Maxent

Maxent is a machine-learning technique that fits a probability distribution of species occurrence with environmental layers over the entire study area (PETERSON 2006). We selected Maxent for our analysis because it showed not only a good model performance in comparison to other presence-only SDM algorithms (ELITH et al. 2006; HERNANDEZ et al. 2008; WISZ et al. 2008; PEARSON et al. 2007), but also allows to infer the importance of environmental variables for species distribution (ELITH et al. 2011).

Maxent default configuration is to allow selecting from a range of functional forms (“feature types”) to describe the relationship between the probability of occurrence and an environmental variable. Those functional forms are linear, product, quadratic, hinge, threshold and categorical (ELITH et al. 2011). According to SYFERT et al. (2013), we fitted models with the default auto feature option and compared them with models fitted with simple functional forms, excluding hinge, threshold and categorical features (results not published). The models fitted with simple functional forms performed similarly to those with auto features, but seemed ecologically more sensible and much less prone to overfitting, without unrealistically steep slopes and abrupt changes. Therefore, we report models fitted without hinge, threshold and categorical features in this paper.

Maxent uses random background samples for the model fitting process. The intention of those samples is not to pretend that a species is absent at the selected sites, but to provide a sample of the set

of conditions available in the modeled area (PHILLIPS et al. 2009). We created 10,000 random samples within the modeled area with the R-package “dismo” (HIJMANS et al. 2013), which were also used during the evaluation of model performance (see Section 2.5). As pointed out beforehand, we assumed that no geographical sample bias exists in the model area and therefore, an unweighted and completely random creation of background samples was justified (ELITH et al. 2011).

Other settings of Maxent were left on default; convergence threshold (0.00001), maximum iterations (1,000), and regularization multiplier ( $r = 1$ ). Maxent analysis was carried out using Maxent software (Version 3.3.3k, PHILLIPS et al. 2006) and R-package “dismo” (HIJMANS et al. 2013).

The importance of each topographic variable was assessed for each species with a jackknife test that builds several models using the same occurrence data, but a different set of predictor variables. At first, models for each species were created excluding one variable each run and then, models were run with only one variable at a time. Finally, the regularized training gain of those different models and models created with all predictor variables were compared for each species.

We further examined response curves for each species showing how logistic output (probability of presence) varies over the complete range of each topographic variable when creating a model using only the corresponding variable (PHILLIPS et al. 2006).

#### 2.4.2 Ecological Niche Factor Analysis (ENFA)

The ENFA compares the ecological conditions of sites where a species occurs with conditions of the entire study area (HIRZEL et al. 2002). We chose the ENFA for our analysis because it is well suited for determining the importance of environmental variables for the characterization of the ecological niche of a given species (BASILLE et al. 2008).

The ENFA, conceptually similar to a principal component analysis, reduces the environmental variables introduced as predictors to a small number of orthogonal factors. The first factor, termed “marginality”, contains the most information and describes the preference of a given species for specific environmental conditions among the whole set of possible conditions. Negative and positive coefficients for each environmental variable indicate that the species prefers lower and higher values than the mean of the study area respectively (BASILLE et al. 2008). The ab-

solute value of the marginality expresses the ecological importance of the habitat factor for the species (HIRZEL et al. 2002). The ENFA was carried out using the R-package “adehabitat” (CALENGE 2006).

#### 2.5 Evaluation of model performance

We assessed the performance of the produced models using two evaluation criteria: the threshold-independent area under the curve (AUC) of the receiver operating characteristic (ROC) (FIELDING and BELL 1997) and the threshold-dependent true skill statistics (TSS) (ALLOUCHE et al. 2006). These metrics were calculated for (i) final models (created with the complete available dataset), (ii) different permutations of null-models to test if the final models differ significantly from what would be expected by chance alone (RAES and STEEGE 2007), (iii) randomly split k-fold cross validation data partitions (FIELDING and BELL 1997), and (iv) spatially separated data partitions to examine the effect of spatial autocorrelation on model performance (VELOZ 2009). The methods applied for data partitioning and randomization (ii–iv) are described in the Appendix. Additionally, our results were compared with those of previous work about tree species characteristics in our study area (HOMEIER et al. 2010; HOMEIER et al. 2008).

The use of multiple metrics is important, because each metric evaluates a different aspect of predictive performance (ELITH and GRAHAM 2009). Despite using a k-fold data partitioning technique for some of those metrics, the complete available data were used to create the final models (FIELDING and BELL 1997).

##### 2.5.1 AUC

The AUC is a common metric in presence-only SDM (MERCKX et al. 2011). It was initially used for presence/absence techniques, but can be employed for presence-only techniques as well, because plotting sensitivity against a random sample of background locations is sufficient to define an ROC curve (WILEY et al. 2003; PHILLIPS et al. 2006). However, rather than expressing the ability of the model to discriminate between suitable and unsuitable habitat, as it is the case with presence/absence methods, the AUC is now the probability that a randomly chosen presence site is ranked above a random background site (PHILLIPS et al. 2006).

Whereas values of the AUC range from 0.5 (random prediction) to 1 (perfect accuracy) when used

with presence/absence data, the maximum achievable AUC is less than 1 for presence-only data (WILEY et al. 2003). The threshold of the AUC for more informative than random or good models varies according to different sources and ranges from 0.6–0.85 (MERCKX et al. 2011). In this study, we interpreted values of the AUC of 0.6–0.7 as poor, 0.7–0.8 as average, 0.8–0.9 as good, and 0.9–1 as excellent (MERCKX et al. 2011).

The same 10,000 random background points previously created for the Maxent models were used as pseudo-absences for the calculation of the AUC.

### 2.5.2 TSS

Although widely (and often exclusively) used for the evaluation of presence-only models (MERCKX et al. 2011), the reliability of the AUC as the only quality metric for model evaluation has been challenged in recent years (LOBO et al. 2008; JIMÉNEZ-VALVERDE 2012). Therefore, we additionally calculated the threshold-dependent true skill statistic (TSS), which is similar to the widespread Kappa statistic, but less dependent on the prevalence of the modelled species (ALLOUCHE et al. 2006). It is defined as

$$TSS = sensitivity + specificity - 1$$

where sensitivity is the proportion of observed presences that are correctly predicted as presences, and specificity is the proportion of observed absences that are correctly predicted as absences (ALLOUCHE et al. 2006). Values of TSS < 0.2 can be considered as poor, 0.2–0.6 as fair to moderate and >0.6 as good (LANDIS and KOCH 1977).

In order to calculate the TSS, the continuous habitat predictions have to be transformed into a binary format. As the intensity of mapped predictions vary between different model algorithms, threshold selection should be objective and not rely on arbitrary values (ELITH and GRAHAM 2009). For our analysis, we calculated a threshold score for each permutation according to the method based on maximizing the sum of sensitivity and specificity (Max SSS), which is equivalent to maximizing the TSS (LIU et al. 2013).

### 2.5.3 Comparison with existing knowledge

While model evaluation in SDM mainly focuses on predictive performance, it is also important to evaluate the ecological realism of models (FRANKLIN

and MILLER 2009). To accomplish this, we compared our models with results previously obtained by HOMEIER and BRECKLE (2008) and HOMEIER et al. (2010). Those authors described three different forest types in the elevation range of our permanent sample plots (see Section 2.2 for a short description of these types) and assigned each species to one or multiple forest types (see Tab. S1 in the Supporting Information of HOMEIER et al. 2010). We performed a hierarchical cluster analysis based on the marginality factor of the ENFA for each species and each topographic variable in order to define different groups of tree species. We then assessed if these cluster groups agreed with the previously defined forest types (HOMEIER et al. 2008). Subsequently, we explored for each species to what extent the grouping based on ENFA models concurred with the classification of HOMEIER et al. (2010).

## 3 Results

### 3.1 Evaluation of model performances

Figure 1 summarizes the results of the different quantitative evaluation metrics for the 16 species and for the average across all species.

#### 3.1.1 Significance of final models

For null-models, the average AUC score of all one-sided 95 % confidence intervals across all quantity ranges of randomly sampled locations was 0.62 for Maxent and 0.58 for ENFA models (red squares in panel aa and ab, Fig. 1). AUC scores across all species of the final models averaged to 0.76 for Maxent (ranging from 0.61 to 0.90) and 0.74 (0.65–0.80) for ENFA (blue triangles in panel aa and ab, Fig. 1). For 15 of the 16 species, AUC scores of the final models for Maxent were higher than the 95 % confidence interval of the respective null-models. Only *Nectandra lineatifolia* performed significantly worse than the respective null-models (panel ay, Fig. 1). The final ENFA models of all 16 species showed significantly higher AUC values than the respective null-models.

The one-sided 95 % confidence intervals of the TSS scores of the null-models averaged to 0.20 for Maxent and 0.16 for ENFA across all quantity ranges. In comparison, final Maxent models scored an average TSS value of 0.46 (0.23–0.67), while the average value of ENFA models was 0.41 (0.29–0.51). Results

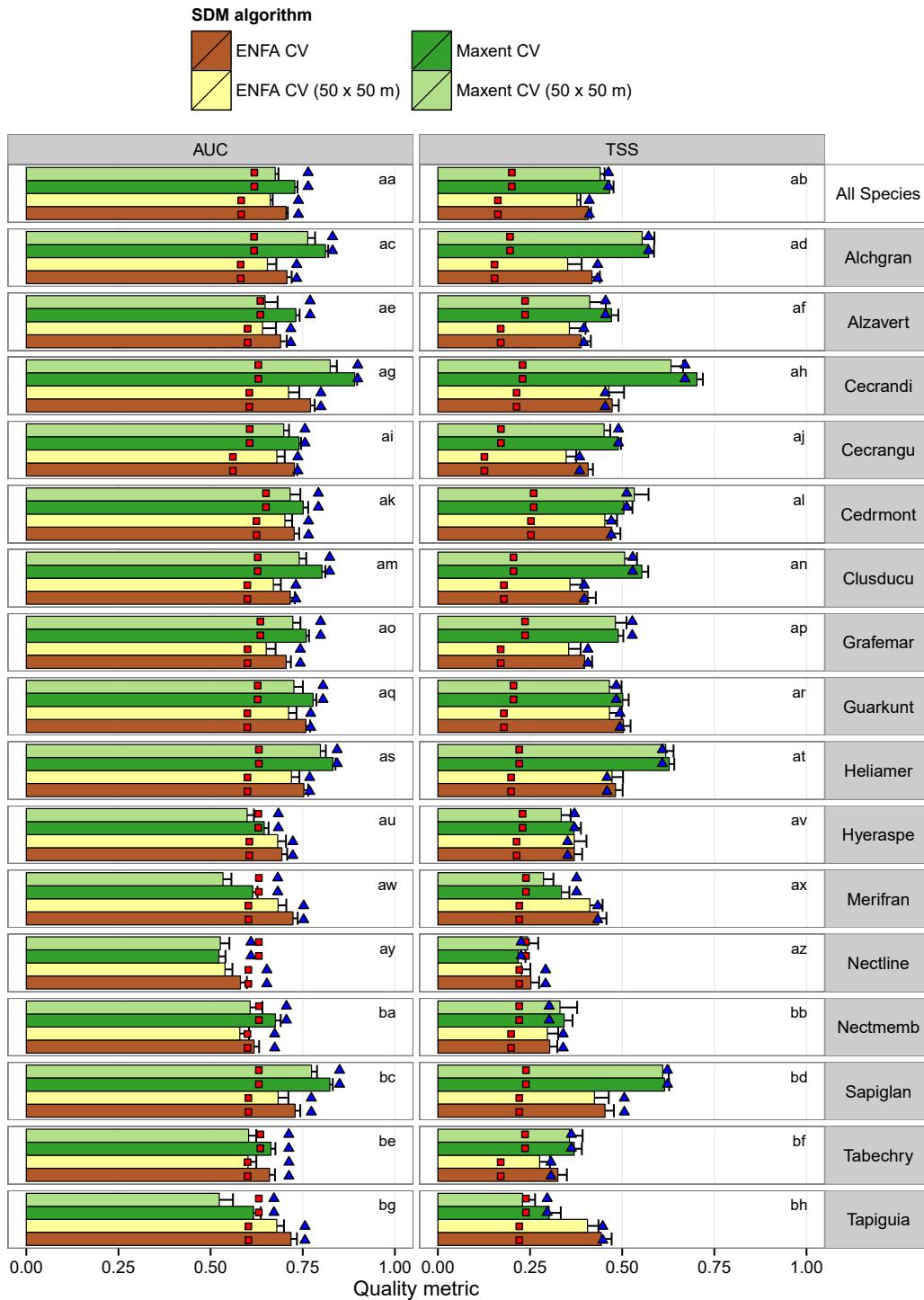


Fig. 1: Model evaluation for each species (ac–bh) and for the mean across all species (aa, ab). Results for AUC and TSS are on the left and right side respectively. ENFA models are indicated by brown bars and Maxent models by green bars. Dark green/brown colors show random data partitioning for cross-validation (CV), while light green/brown colors show cross-validation with spatially separated data partitions (CV 50 x 50 m). Error bars indicate the 0.95 confidence interval. Red squares symbolize the upper 0.95 quantile of the quality metrics for the null-models of the respective quantity range. Blue triangles show the quality metrics of the final models for each species



for individual species showed the same trend as results for the AUC: all final ENFA models and 15 out of 16 Maxent models performed significantly better than the respective null-models. The only species performing worse than the null-model was *Nectandra lineatifolia* (panel az, Fig. 1).

### 3.1.2 Random data partitioning

For cross validation with random data partitioning, average scores across all species for Maxent models were higher than for ENFA models for both AUC (0.75 vs. 0.71, panel aa, Fig. 1) and TSS (0.47 vs. 0.41, panel ab, Fig. 1).

The mean of AUC scores for models with random data partitioning (CV, dark green/brown) ranged from 0.54 to 0.87 for Maxent and 0.59 to 0.78 for ENFA. Mean scores for the TSS ranged from 0.18 to 0.68 for Maxent and 0.25 to 0.53 for ENFA.

Applying the classification of MERCKX et al. (2011) to the AUC scores, 5 Maxent models can be considered as good, 5 as average, 5 as poor, and one as random. For ENFA, 11 models were average, 4 poor, and one random. Conversely, the classification of TSS scores according to LANDIS and KOCH (1977) indicated that all models performed fairly to moderately for ENFA, and for Maxent, 3 models performed good and 13 fairly to moderately (Fig. 2).

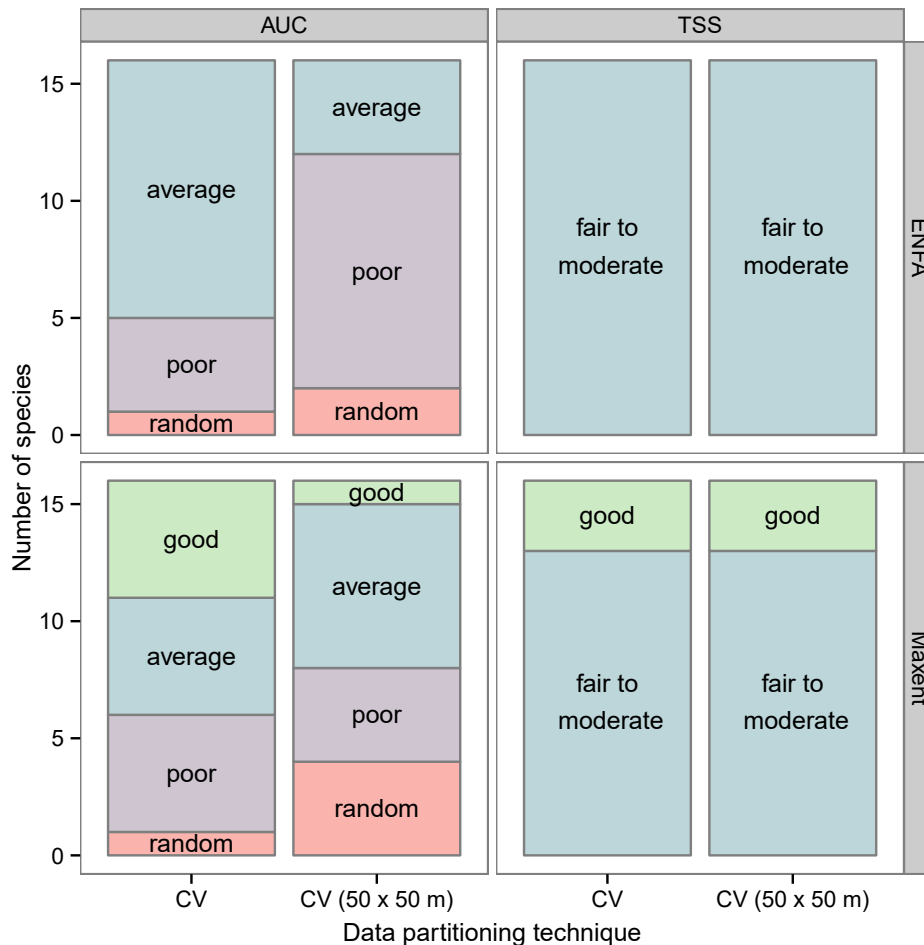


Fig. 2: Summary of the model performance for all 16 species. AUC scores (left) were classified into quality classes (<0.6 random, 0.6–0.7 poor, 0.7–0.8 average, 0.8–0.9 good, 0.9–1 excellent, MERCKX et al. 2011). The classification scale of LANDIS and KOCH (1977) was used for TSS scores (right, <0.2 poor, 0.2–0.6 fair to moderate, >0.6 good). Results are shown for SDM algorithms (Maxent and ENFA) and data partitioning techniques (cross-validation with random partitions (CV) and cross-validation with spatially separated partitions (CV 50 x 50 m))

### 3.1.3 Spatially separated data partitioning

For models with spatially separated data partitions (CV (50 x 50 m), light green/brown, Fig. 1), mean AUC scores of Maxent and ENFA ranged from 0.48 to 0.82 and 0.53 to 0.72 respectively. Mean TSS scores ranged from 0.13 to 0.58 for Maxent and from 0.19 to 0.47 for ENFA. These AUC scores indicated that one Maxent model was good, 7 average, 4 poor, and 4 random, whereas 4 ENFA models were average, 10 poor, and 2 random. Mean TSS scores, on the other hand, indicated 16 fair to moderate models for ENFA, and 3 good and 13 fair to moderate models for Maxent (Fig. 2).

Compared to cross validation with random data partitioning, evaluation metrics for nearly all species were lower for cross validation with spatially separated data partitioning (exception: TSS for *Cedrela montana* and AUC and TSS for *Nectandra lineatifolia*, panel al, ay, and az, Fig. 1). For AUC/ENFA, AUC/Maxent, TSS/ENFA and TSS/Maxent, scores for randomly separated data partitioning were 0.04, 0.05, 0.03, and 0.03 higher than for spatially separated data partitioning, respectively. Furthermore, variation between the 50 model runs for each species, evaluation metric, and data partitioning technique was higher for spatially separated data partitioning compared to random data partitioning (average 0.95 confidence intervals across all species and evaluation metrics was 0.015 and 0.027, respectively).

### 3.2 Most important variables for characterization of suitable habitat for tree species

The contribution of ecological variables for the distribution of each species according to Maxent and ENFA is compared in figure A-3. These results

are based on final models created with the complete dataset. Considering that *Nectandra lineatifolia* performed worse than the null-model in 4 of the 8 presented evaluation metrics, we acknowledged that this species cannot be satisfactorily modeled with our data and excluded it for the final models.

The importance of variables was calculated with two conceptually different methods (see Section 2.4). Therefore, we assumed that figure A-3 should only be assessed visually in order to examine to what extent results of the two SDM algorithms have a similar trend. For most species, the relative importance according to Maxent and ENFA showed a similar pattern (e.g. *Cecropia andina*), whereas few other species, such as *Cedrela montana* or *Meriania franciscana* showed more pronounced differences.

When considering the average across all 16 species, ENFA and Maxent models generally agreed on the importance of each topographic variable for species distribution (Tab. 2). The absolute coefficient of the marginality factor of the ENFA models indicated that the most important variables were elevation (0.70) and TPI (0.41), whereas Wetness Index (0.31), eastness (0.26) and slope (0.17) only had minor importance. Likewise, according to the Maxent jackknife analysis, the most influential variables when used individually (“Gain with only variable”) were elevation (0.34) and TPI (0.17), while eastness (0.09), Wetness Index (0.08) and slope (0.03) were less influential.

The jackknife test for Maxent, where each variable was withdrawn in turn (“Gain without variable”), identified elevation (0.34) as the variable that mostly decreased training gain when excluded. Consequently, this variable contained the most information not present in the four remaining variables, which were of lesser importance (0.45–0.48).

Tab. 2: The mean importance of topographic variables across all species according to ENFA and Maxent. The absolute value of marginality of the ENFA describes how far species depart from the mean available conditions in the study area. For Maxent, “Gain with only variable” shows the regularized training gain of models created individually with only one ecological variable compared with models created with all variables, whereas “Gain without variable” compared the regularized training gain of models with one variable excluded compared with models created with all variables. Values in parenthesis indicate the 0.95 confidence interval

Variable	ENFA: Absolute marginality	Maxent: Gain with only variable	Maxent: Gain without variable
Eastness	0.26 (0.15–0.37)	0.09 (0.05–0.13)	0.45 (0.31–0.59)
Elevation	0.70 (0.59–0.81)	0.34 (0.22–0.46)	0.34 (0.22–0.46)
Slope	0.17 (0.11–0.23)	0.03 (0.02–0.04)	0.48 (0.34–0.62)
TPI 100m	0.41 (0.33–0.50)	0.17 (0.10–0.25)	0.46 (0.32–0.59)
Wetness Index	0.31 (0.26–0.37)	0.08 (0.03–0.13)	0.48 (0.34–0.62)

For the two most important topographic variables, elevation and TPI, Maxent response curves were compared to the representations of the niche-environment system in our study area (Fig. 3). Response curves for elevation show a continuous descending or ascending trend for most species, whereas response curves for TPI are characterized by a local maximum for the majority of species.

### 3.3 Comparison with previous classifications

Three clusters (clustering height of 1.49) were identified with a hierarchical cluster analysis based on the marginality factors of each species and topographic variable (Tab. A-2, Fig. A-4). *Nectandra lineatifolia*, not being significantly different from the null-models, was not included in this cluster analysis.

Group A, consisting only of the species *Tapirira guianensis*, showed a distinct preference for lower values of eastness and higher values of slope. Group B consisted of 10 species, showing a clear preference for lower elevations, valleys (lower coefficients of TPI and higher coefficients of Wetness Index), and (less clearly) for higher coefficients of eastness and slope over the mean of the available conditions in the study area. Group C, composed of 4 species, on the other hand, preferred higher elevation and ridges (higher coefficients for TPI and lower coefficients of Wetness Index), whereas preferences for eastness and slope varied between species of this group. Hence, group B seemed to be equivalent to the definition of forest type I, while group C reflected forest types II and III as defined by HOMEIER et al. (2008).

The previous assignment of individual species to forest types by HOMEIER et al. (2010) is well reflected in the groups resulting from the cluster analysis (Tab. A-2). Only three species are not congruent with this assignment; for *Cedrela montana*, no information was provided, *Graffenrieda emarginata* was assigned to all 3 forest types and *Tapirira guianensis* was assigned to an individual cluster group, although HOMEIER et al. (2010) classified this species as belonging to forest type I.

### 3.4 Prediction maps of habitat suitability

Maps of habitat suitability for each species were created based on the results for ENFA and Maxent models. For the prediction, the models created with the complete data set for each species were used. The maps (Fig. 4) visually emphasize the cluster groups identified in the previous section. Within the

three experimental plots, species of group B generally had higher scores for habitat suitability on lower elevations and in valleys, whereas species of group C preferred higher elevations and ridges.

Further, it can be observed that the predicted habitat suitability score is generally higher for ENFA models than for Maxent models. This is a result of the fact that different model algorithms generate mapped predictions with differing intensities (ELITH and GRAHAM 2009). It can be compensated for by using objective methods for threshold selection instead of arbitrary values during the transformation of continuous predictions into discrete classes.

## 4 Discussion

### 4.1 Limitations

#### 4.1.1 Subset of the full range of environmental conditions

The tree presence data used for model calibration only represent a very small subset of the full range of environmental conditions of the species, corresponding to “an artificially constrained geographic space” (RAES 2012). Several implications of this restriction have been identified by previous studies about SDM, which may also apply to our results.

First, response curves are possibly influenced by the restricted environmental ranges, with strongest effects towards the upper and lower ends (THUILLER et al. 2004). As mentioned in the methods section, we first fitted Maxent models with hinge and threshold features types, which resulted in unrealistic habitat suitability maps and response curves. Overfitting in the tails of the response curves could be one possible explanation for this. However, by limiting feature types to linear, product and quadratic features, overfitting seemed to have been reduced considerably (Fig. 3). In the case of the ENFA, only linear dependencies within the species niche are fitted (HIRZEL et al. 2002). Therefore, we assumed that restricted data only produced a minor effect on the upper and lower environmental ranges for ENFA results.

Second, prediction of habitat suitability should not be realized to areas beyond the range of environmental conditions the models were calibrated with (SÁNCHEZ-FERNÁNDEZ et al. 2011; THUILLER et al. 2004). This is especially the case for the region of our study area, which is characterized by consider-

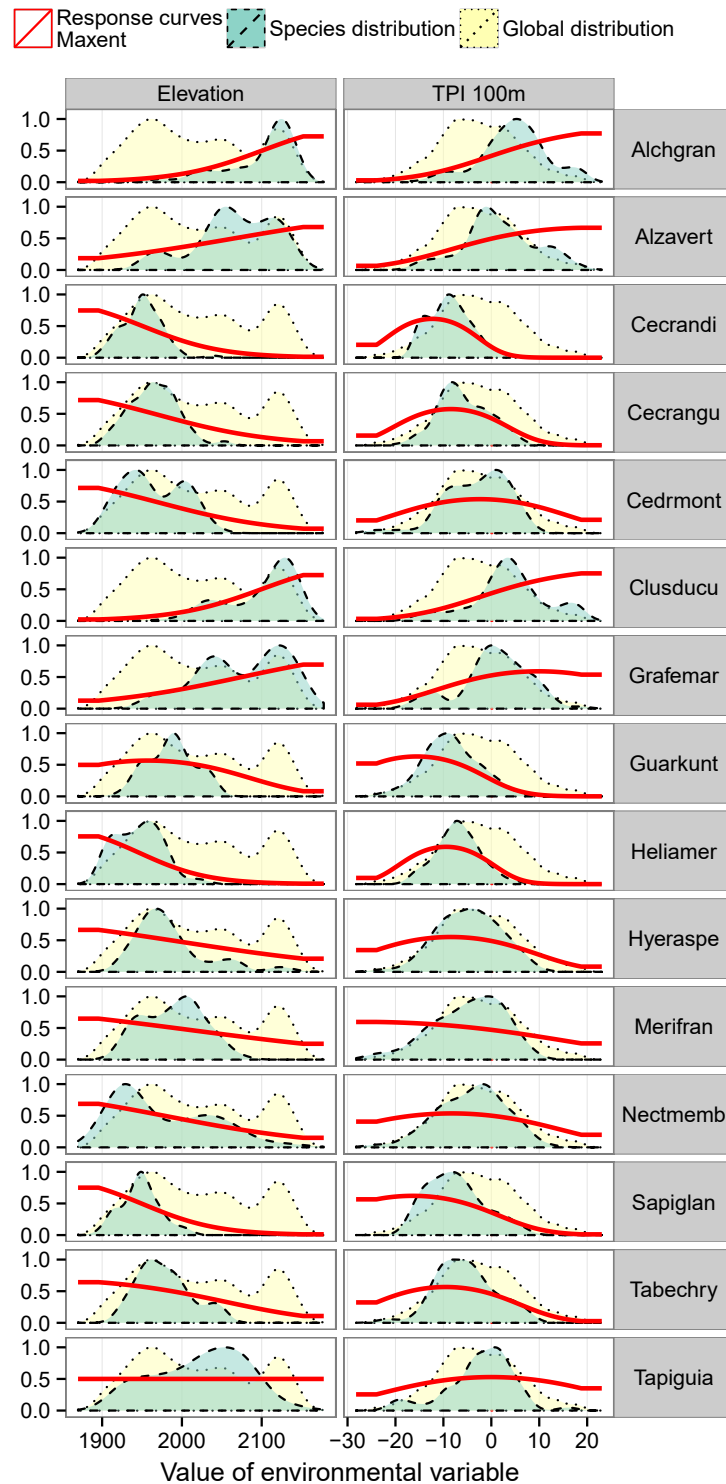


Fig. 3: Response curves for the Maxent models and representation of the niche-environment system in the study area for the two most important variables (elevation and TPI) and 15 species. The distribution of the available topographic variables in the study area ("Global distribution") is compared to the distribution of topographic variables on sites where the species occurs ("Species distribution")

able climatic variability (ROLLENBECK 2006) – over a distance of 30 km in east-west direction, BECK et al. (2008b) reported a gradient of mean annual precipitation ranging from 3103 to 1068 mm on equal elevations (1950 m asl). If habitat suitability were to be predicted to a wider area in future studies, it would therefore be highly advisable to include precipitation and other relevant climatic factors as predictor variables (see e.g. ENGELBRECHT et al. 2007). Furthermore, the resulting habitat suitability maps should be verified with additionally sampled independent test data.

Third, like habitat suitability, the importance of topographic variables cannot be extrapolated beyond the range of environmental conditions either. This is especially true for species with a wide distribution range, such as *Tabebuia chrysantha*, occurring from Northern Mexico to Bolivia and in both dry and humid forests (VOLLAND-VOIGT et al. 2011). In our study area, this species shows a clear preference for negative values of TPI (Tab. A-2), which correspond to valleys. However, this does not allow inferring preferences for topographic positions of this species in other regions of its distribution range.

The above-mentioned points are important to acknowledge, but they do not conflict with the main objectives of this study, which are focused on the local scale.

#### 4.1.2 Niche or neutral processes

The focus on the local scale however might result in a different potential problem, which is linked to the ongoing debate in recent decades whether niche or neutral processes are more important for species distribution in multi-species communities. Niche differentiation with respect to resources remains one of the most prominent theories for tropical tree species (HUTCHINSON 1957). It postulates that different trees species specialize on different habitats, where they are competitively dominant and relatively more abundant. The neutralist theories, on the other hand, propose that dispersal limitations and stochastic processes drive species composition, and that niche differentiation only plays a limited role for species coexistence (HUBBELL 2001). In the context of SDM, spatial patterns in species distributions caused by neutral processes might be erroneously interpreted as being caused by niche processes, if those patterns happen to be spatially correlated with environmental variables.

Previous studies about the influence of both processes at local scales in tropical forests reached

diverging results. HARMS et al. (2001) concluded that niche processes only played a limited role in the maintenance of species diversity, which was partially supported by some studies (QUEENBOROUGH et al. 2007; VALENCIA et al. 2004; GETZIN et al. 2014). Conversely, numerous other studies only found little evidence for this hypothesis and concluded that niche processes played a predominant role for the distribution of tree species at local scales in tropical forests (GUNATILLEKE et al. 2006; CHUYONG et al. 2011; YAMADA et al. 2010; ITOH et al. 2003; JOHN et al. 2007; JONES et al. 2008). BALDECK et al. (2013) and CHANG et al. (2013) recently pointed out that most of the studies cited in this paragraph only used environmental variables related to topography, and that by including spatially explicit soil resource variables in addition to topographic variables alone, the variation in the tree community composition in tropical forests explained by the environment greatly increased. They concluded that the role of niche processes may have been underestimated in previous studies.

This indicates that especially in regions characterized by very heterogeneous topographic conditions, such as our study area, niche processes are most important for the distribution of species. This is congruent with previous studies in our study area (BUSSMANN 2003; HOMEIER et al. 2008; HOMEIER et al. 2010), which showed a strong influence of topography on the occurrence of different forest types. Therefore, it seems acceptable to assume that species distribution patterns in our study area are mainly caused by niche processes, and that the use of SDM is justified for our study area.

## 4.2 Aspects of SDM

### 4.2.1 Model evaluation

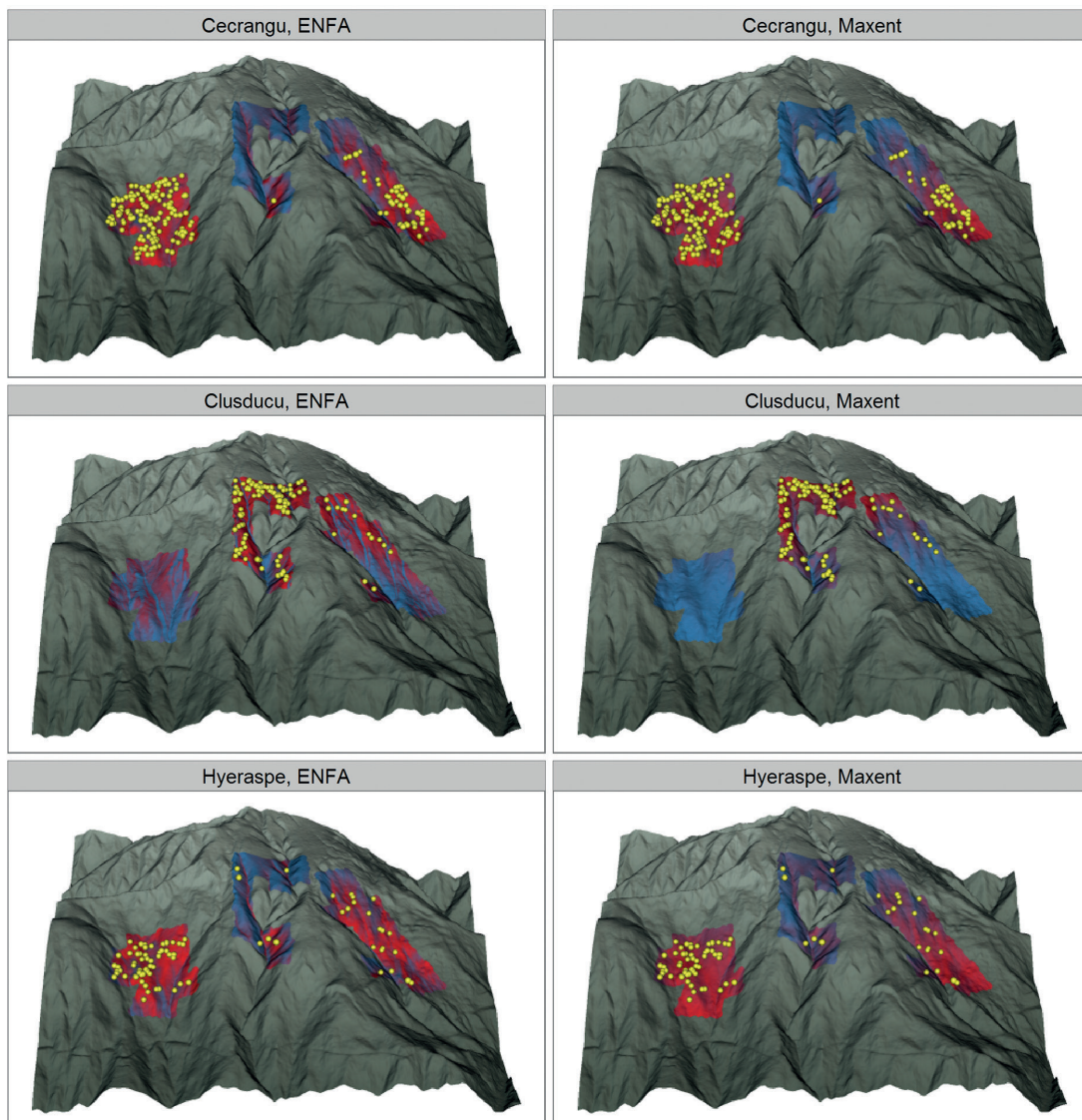
AUC values of our models were low for many species. For example, for cross validation with random data partitioning, AUC values indicated that 6 Maxent models and 5 ENFA models could be considered as poor or random. In contrast, according to TSS values, all models performed at least fairly to moderately. Likewise, all final ENFA and 15 out of 16 final Maxent models performed significantly better than the null-models. Additionally, the ecological interpretation of our models was very similar to pre-existent knowledge about forest types and the membership of different tree species to these forest types (HOMEIER et al. 2008; HOMEIER

et al. 2010). Those previous results were obtained in the same study area (RBSF), but using independent sample plots and different methods.

The relatively low values of the AUC compared to the other quality metrics might be caused by two implications of its use with presence-only models. First, the maximum achievable AUC is no longer 1, but  $1 - a/2$ , where  $a$  is the fraction of the area covered by the species true distribution, which is normally unknown (WILEY et al. 2003; PHILLIPS et al. 2006). Considering that we only used the 16 most common species, values of  $a$  may be high, and consequently, the maximum achievable AUC might be low. Second, the AUC depends partly on the total extent of the

study area. It increases with larger extents, where environmental conditions differ much from those of the area where the species occurs (LOBO et al. 2008). Our relatively small study area therefore might have contributed to the low values of AUC.

The fact that our models performed well for the majority of species according to the TSS and the comparison with null-models, combined with the fact that the ecological interpretation agreed with previous studies, suggest that AUC underestimated the performance of the models. Our results therefore underline the importance of using more than one performance criterion for the evaluation of SDM (LOBO et al. 2008).



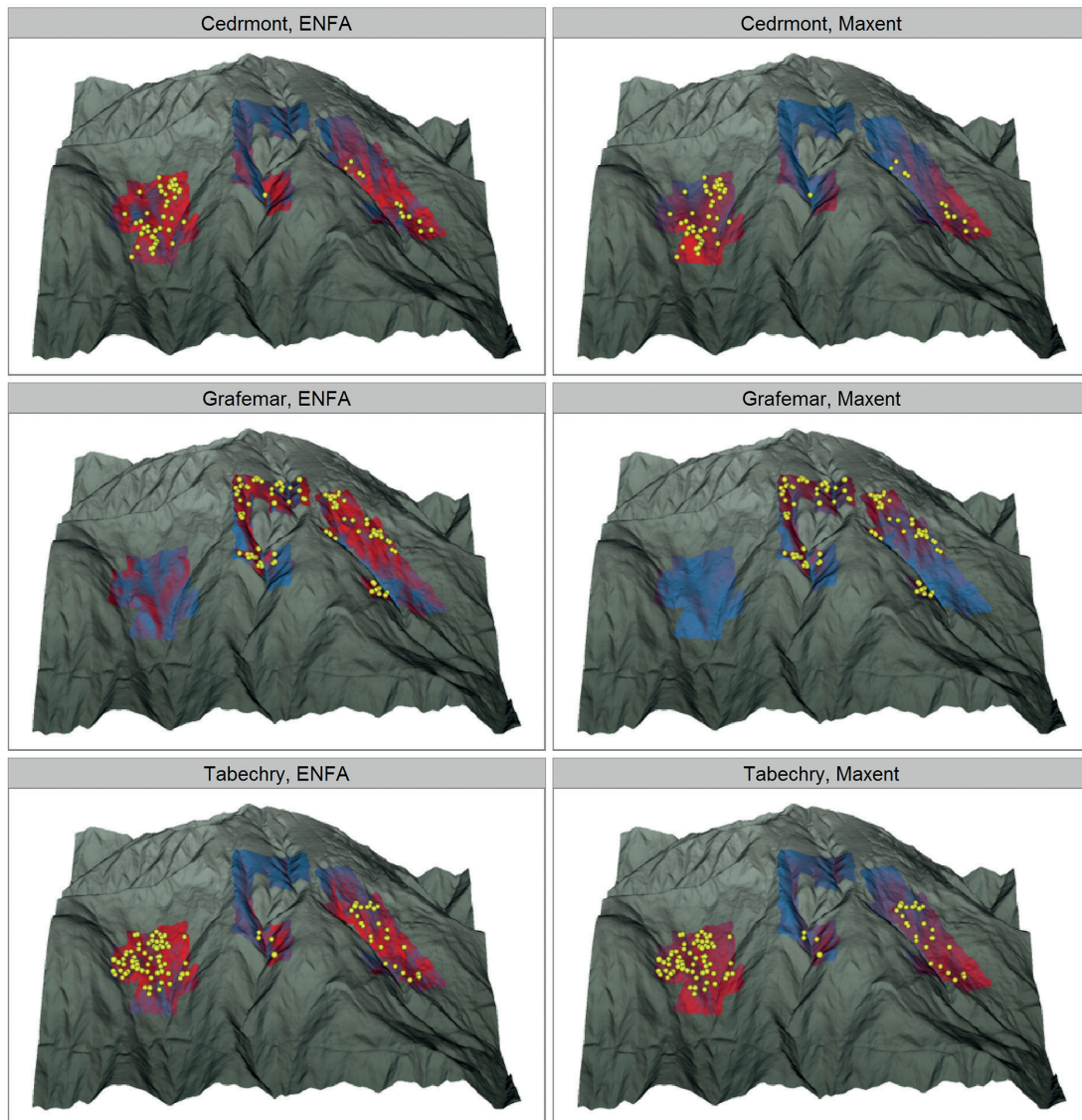


Fig. 4: Habitat suitability maps of 6 trees species from final ENFA and Maxent models. Blue colors indicate unsuitable conditions and red colors suitable conditions. Yellow points show the registered occurrence of each species, which were used for model creation. Species of cluster group B (*Cecropia angustifolia*, *Cedrela montana*, *Hyeronima asperifolia* and *Tabebuia chrysantha*) tend to have areas predicted as suitable in valleys and lower altitudes, whereas areas with high predicted suitability for species of cluster group C (*Clusia ducuoides* and *Graffenrieda emarginata*) are located on ridges and higher altitudes

#### 4.2.2 Comparison between results of ENFA and Maxent

For most species, Maxent models were more robust than ENFA models (exceptions are, e.g., *Meriania franciscana* and *Tapirira guianensis*, Fig. 1). This is congruent with findings from previous studies, where Maxent showed better performances than other SDM algorithms (HERMOSILLA et al. 2011; ELITH et al. 2006). The fact that Maxent fits models with different feature classes (linear, prod-

uct, quadratic, hinge, threshold and categorical, PHILLIPS and DUDÍK 2008), whereas ENFA only fits linear relationships (HIRZEL et al. 2002), might be an explanation for this. Even though we only allowed linear, product and quadratic features for Maxent (see Section 2.4.1), this algorithm likely fits complex species-environment relationships still better than the ENFA.

In terms of variable importance across all species, the results of both SDM algorithms were very similar. This confirms findings from previous stud-

ies that used Maxent and ENFA to analyze the importance of environmental variables (MCKINNEY et al. 2012; HERMOSILLA et al. 2011).

#### 4.2.3 Effect of spatial autocorrelation on model performance

Model quality metrics were lower for cross validation with spatially separated data partitions compared to randomly separated data partitions (see Section 3.1.3). This could be interpreted as a correction for quality metrics, which may have been falsely inflated due to the presence of spatial autocorrelation in the tree occurrence data (VELOZ 2009; MERCKX et al. 2011). However, this approach might be overly simplistic in our case for a number of reasons.

First, spatial autocorrelation can be divided into “spurious” autocorrelation, an artefact of experimental design caused for example by biased sampling, and “real” spatial autocorrelation, which corresponds to the actual spatial structure of species distribution. “Spurious” spatial autocorrelation should be removed from the data, whereas no data should be discarded for “real” autocorrelation (KAMINO et al. 2012). Having a complete tree census with a sampling intensity of 100 %, we assumed that no spatial sampling bias occurred for the modelled area, and that existing spatial autocorrelation in our occurrence data is therefore “real”.

Second, “real” spatial autocorrelation can further be separated into two types of distinct processes: endogenous and exogenous. Endogenous processes are associated directly with the biological occurrence data, while exogenous processes are independent from the occurrence data. For example, tree clustering could be caused by limited ability to disperse away from a parent tree (endogenous process), or by fine-scale variation in soil conditions (exogenous process) (FRANKLIN and MILLER 2009). In the context of SDM, spatial autocorrelation in occurrence data could therefore be considered either as noise hindering the analysis of species-habitat association, or as a direct result of species-habitat association.

Third, subsampling presence points in order to ensure that clusters of training data are not excessively clustered around test data reduces spatial autocorrelation (VELOZ 2009), but also jettisons information and therefore reduces the predictive performance of models (LENNON 2000; SEGURADO et al. 2006).

These reasons show that, at least in our study, a reduction of model performance when using spatially separated data partitions for cross vali-

ation must not necessarily mean that models without a correction of spatial autocorrelation are falsely inflated.

#### 4.3 Interpretation of the importance of individual topographic variables

##### 4.3.1 Elevation

Elevation, whose predominant influence on species distribution at intermediate and large scales is well established in phytogeography (HUMBOLDT and BONPLAND 1805; GENTRY 1988), was identified by our study as the most important topographic variable for the distribution of tree species at a local scale. Similar results on small scales in tropical forests were reported previously (LIEBERMAN et al. 1985; WANG et al. 2009; LAN et al. 2009; LEDO et al. 2013). The altitudinal gradient is an indirect factor (no direct effect on species distribution, also known as proxy factors), which correlates with direct factors (physiological effect on plant growth). In terms of climatic direct factors, FRIES et al. (2009) reported an annual average lapse rate of  $-0.52 \text{ K}/100 \text{ m}$  in the study area for the altitudinal range of the permanent sample plots (1850–2150 m asl), which translates to a difference of annual average temperatures of 1.56 K between the lowest and highest sites in the sample plots. Similarly, average annual precipitation increases by 250 mm/100 m (corresponding to a difference of 750 mm between lowest and highest sites) and fog input augments by 40 mm/100 m (corresponding to a difference of 120 mm between lowest and highest sites) (ROLLENBECK 2006). While these climatic elevation gradients play an important role for the distribution of tree species, additional ecological gradients related to elevation have to be taken into account. With regards to soil properties, the general trend in tropical mountains that increasing elevation results in decreasing soil fertility and decreased N availability (TANNER et al. 1998) was confirmed for the RBSF (WILCKE et al. 2008; WOLF et al. 2011). Further, in our study area the mineral soil depth decreases, the thickness of the organic layer and pH increase and soil texture becomes coarser in the A horizon with increasing elevation (WILCKE et al. 2008; LIESS et al. 2011; WOLF et al. 2011).

These strong gradients of direct factors help to understand why elevation is the most important variable for species distribution, even at small scales. However, disentangling the relative contribution of each of the direct factors linked to elevation is beyond the scope of our study.



### 4.3.2 TPI

Maxent and ENFA models identified TPI as the second most important variable for the distribution of tree species. This result was expected, because a different forest type can be found on ridges compared to valleys (HOMEIER et al. 2008), and the continuous values of the TPI express the position of each site on a topographical gradient ranging from valleys to mid-slopes to ridges (GUISAN et al. 1999). This is congruent with findings from previous studies examining the influence of topography on forest composition in tropical forests (TAKYU et al. 2002; WEBB et al. 1999; QUEENBOROUGH et al. 2007; VALENCIA et al. 2004; LAN et al. 2009; WANG et al. 2009; CHUYONG et al. 2011). Like elevation, the topographic position is an indirect factor having strong effects on direct factors by creating a variety of microhabitats along a topographic gradient. WEBB et al. (1999) concluded that differences in forest structure and composition between topographic positions were mainly caused by frequent disturbances on ridges, which were more exposed to chronic and catastrophic wind events than valleys. In our study area, this hypothesis is supported by the fact that trees have lower canopy heights as well as lower height to DBH ratios on ridges compared to valleys (HOMEIER et al. 2010), which might be an adaptation to exposure to wind. However, tree turnover rates are actually higher in valleys than on ridges in our study area (HOMEIER et al. 2010), which suggests that factors related to topography other than wind exposure play a more important role for the distribution of tree species. Periodic waterlogging processes in valleys during high rain events may explain higher turnover rates in valleys, because they force trees to have shallow rooting systems, which may result in higher mortality through tree falls (ASHTON et al. 2011). Furthermore, ridge-top soils generally have higher humus concentrations and are more acid and poorer in nutrients (especially N) than valley-bottom soils (WILCKE et al. 2011; WOLF et al. 2011). This can be partly explained by lateral transport of water and matter and the removal of mineral soil on ridges caused by landslides, which are mainly controlled by topography in the RBSF (VORPAHL et al. 2012). Additionally, throughfall in valleys is more abundant and nutrient-richer than on ridges (OESKER et al. 2008) and herbivory and leaf litter decomposition is considerably lower on ridges than on valleys (WERNER and HOMEIER 2015). Moreover, ridges are more prone to drought (BUSSMANN 2003) due to higher solar radiation com-

pared to valleys, although this might be mitigated by the fact that ridges receive more scavenged cloud water than valleys (HOMEIER et al. 2010).

WERNER and HOMEIER (2015) argue that habitat heterogeneity along the topographical gradient from valleys to ridges is caused by a combination of the above-mentioned factors and maintained through a plant-mediated, nutrient-driven feedback cycle. While fast-growing tree species adapted to aboveground competition (i.e. light) are more competitive in valleys, slow-growing tree species adapted to belowground competition in nutrient-poor soils are more competitive on ridges. These slow-growing species are generally characterized by large investments in long-living foliage with antiherbivore defenses (COLEY et al. 1985), which, in combination with mechanically resistant leaves due to increased wind and drought exposure, result in poor litter decomposability. This causes an accumulation of litter and mor humus, which in turn immobilizes basic cations, contributes to soil acidification and further enhances soil nutrient limitations on upper slopes and ridges.

### 4.3.3 Other topographic variables

Compared to elevation and TPI, the remaining variables played a less important role across all tree species. With regards to slope, results of previous studies (WANG et al. 2009; BALDECK et al. 2013), where this variable showed a higher influence, are not supported by our findings. We explain this by the fact that slope showed a relatively small variation in our sample plots (first and third quartiles 25.8° and 36.6° respectively), and therefore contributed only little to the total variation in environmental conditions. Similarly, aspect (eastness) varied little in our sample plots (first and third quartiles of eastness -0.72 and -0.23 respectively), which may explain its low influence compared to other studies (WANG et al. 2009). The preference of *Tapirira guianensis* for negative values of eastness and positive values of slope might be explained by its ability to sprout from uprooted trees (NEGRELLE 1995) and the resulting competitive advantage on sites with frequent disturbances.

One possible explanation for the low importance of the Wetness Index for species distribution might be related to its higher spatial variability within our sample plots compared to the other 4 topographic variables, which are smoother and more continuous (Fig. A-2). This seems to be a consequence of the high-resolution DTM, and therefore a Wetness Index calculated based on a coarser DTM might result in a higher importance for species distribution.

## 5 Conclusion

In this study, we showed that SDM is an adequate tool to analyze the ecological relationships between species and the environment at local scales in tropical forests, as our results about the importance of topographic variables were congruent with previous studies. All topographic predictors used in this study were derived from a high-resolution DTM, which will be increasingly available in the future. This will allow applying SDM-based methods in a wider range of situations and regions. Apart from inference, predictive modelling of species habitat suitability can be used for several applications. Examples include the identification of priority areas for biodiversity conservation, a more efficient planning of field inventories, identification of suitable species and sites for reforestation projects (e.g., restoration with natural species), and pre-selection of possible sites for seed collection. Additionally, when combined with forest inventory data, topographic variables could be used to create high-resolution spatial predictions of forest stand parameters or carbon stocks. Those predictions could then serve as a planning instrument for sustainable forest management and conservation by facilitating the creation of adequate small-scale management units.

## Acknowledgements

We thank Nature and Culture International (NCI) for providing the research station and the Ministerio del Ambiente del Ecuador for granting the research permits to our projects. We also express our gratitude to the numerous student assistants involved in the fieldwork, Benjamin Rothfuss Dair for language editing and two anonymous reviewers for their helpful comments on an earlier draft.

This work was funded by the German Research Foundation (DFG) in the framework of Research Unit FOR402, FOR816 and PAK 824.

## References

- AIBA, S.-I.; KITAYAMA, K. and TAKYU, M. (2004): Habitat associations with topography and canopy structure of tree species in a tropical montane forest on Mount Kinabalu, Borneo. In: *Plant Ecology* 174, 147–161. DOI: [10.1023/B%3AVEGE.0000046059.92806.49](https://doi.org/10.1023/B%3AVEGE.0000046059.92806.49)
- ALLOUCHE, O.; TSOAR, A. and KADMON, R. (2006): Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). In: *Journal of Applied Ecology* 43, 1223–1232. DOI: [10.1111/j.1365-2664.2006.01214.x](https://doi.org/10.1111/j.1365-2664.2006.01214.x)
- ARAÚJO, M. B. and GUIBAN, A. (2006): Five (or so) challenges for species distribution modelling. In: *Journal of Biogeography* 33, 1677–1688. DOI: [10.1111/j.1365-2699.2006.01584.x](https://doi.org/10.1111/j.1365-2699.2006.01584.x)
- ASHTON, M. S.; SINGHAKUMARA, B.; GUNATILLEKE, N. and GUNATILLEKE, S. (2011): Sustainable forest management for mixed-dipterocarp forests: a case study in Southwest Sri Lanka. In: GÜNTER, S.; WEBER, M.; STIMM, B. and MOSANDL, R. (eds.): *Silviculture in the tropics*. Berlin, Heidelberg, 193–213.
- BALDECK, C. A.; HARMS, K. E.; YAVITT, J. B.; JOHN, R.; TURNER, B. L.; VALENCIA, R.; NAVARRETE, H.; DAVIES, S. J.; CHUYONG, G. B.; KENFACK, D.; THOMAS, D. W.; MADAWALA, S.; GUNATILLEKE, N.; GUNATILLEKE, S.; BUNYAVEJCHEWIN, S.; KIRATIPRAYOON, S.; YAACOB, A.; SUPARDI, M. N. N. and DALLING, J. W. (2013): Soil resources and topography shape local tree community structure in tropical forests. In: *Proceedings of the Royal Society B: Biological Sciences* 280. DOI: [10.1098/rspb.2012.2532](https://doi.org/10.1098/rspb.2012.2532)
- BASILLE, M.; CALENGE, C.; MARBOUTIN, É.; ANDERSEN, R. and GAILLARD, J.-M. (2008): Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. In: *Ecological Modelling* 211, 233–240. DOI: [10.1016/j.ecolmodel.2007.09.006](https://doi.org/10.1016/j.ecolmodel.2007.09.006)
- BECK, E.; BENDIX, J.; KOITKE, I.; MAKESCHIN, F. and MOSANDL, R. (eds.) (2008a): *Gradients in a tropical mountain ecosystem of Ecuador*. Berlin, Heidelberg.
- BECK, E.; MOSANDL, R.; RICHTER, M. and KOITKE, I. (2008b): The investigated gradients. In: BECK, E.; BENDIX, J.; KOITKE, I.; MAKESCHIN, F. and MOSANDL, R. (eds.): *Gradients in a tropical mountain ecosystem of Ecuador*. Berlin, Heidelberg, 55–62. DOI: [10.1007/978-3-540-73526-7\\_7](https://doi.org/10.1007/978-3-540-73526-7_7)
- BÖHNER, J.; KÖTHE, R.; CONRAD, O.; GROSS, J.; RINGELER, A. and SELIGE, T. (2002): Soil regionalisation by means of terrain analysis and process parameterisation. In: MICHELL, E.; NACHTERGAELE, F. and MONTANARELLA, L. (eds.): *Soil classification 2001*. Ispra, 213–222.
- BRENNING, A. (2008): Statistical geocomputing combining R and SAGA: the example of landslide susceptibility analysis with generalized additive models: SAGA – seconds out. In: *Hamburger Beiträge zur Physischen Geographie und Landschaftsoekologie* 19, 23–32.
- BURNS, R. M. and HONKALA, B. H. (eds.) (1990a): *Silvics of North America: Conifers*. Washington, DC.
- (eds.) (1990b): *Silvics of North America: Hardwoods*. Washington, DC.
- BUSSMANN, R. (2003): The vegetation of Reserva Biológica San Francisco, Zamora-Chinchipe, southern Ecuador – a phytosociological synthesis. In: *Lyonia* 3, 145–254.

- CALENGE, C. (2006): The package “adehabitat” for the R software: a tool for the analysis of space and habitat use by animals. In: *Ecological Modelling* 197, 516–519. DOI: [10.1016/j.ecolmodel.2006.03.017](https://doi.org/10.1016/j.ecolmodel.2006.03.017)
- CAUWER, V. D.; MUYS, B.; REVERMANN, R. and TRABUCCO, A. (2014): Potential, realised, future distribution and environmental suitability for *Pterocarpus angolensis* DC in southern Africa. In: *Forest Ecology and Management* 315, 211–226. DOI: [10.1016/j.foreco.2013.12.032](https://doi.org/10.1016/j.foreco.2013.12.032)
- CHANG, L.-W.; ZELENY, D.; LI, C.-F.; CHIU, S.-T. and HSIEH, C.-F. (2013): Better environmental data may reverse conclusions about niche- and dispersal-based processes in community assembly. In: *Ecology* 94, 2145–2151. DOI: [10.1890/12-2053.1](https://doi.org/10.1890/12-2053.1)
- CHUYONG, G. B.; KENFACK, D.; HARMS, K. E.; THOMAS, D. W.; CONDIT, R. and COMITA, L. S. (2011): Habitat specificity and diversity of tree species in an African wet tropical forest. In: *Plant Ecology* 212, 1363–1374. DOI: [10.1007/s11258-011-9912-4](https://doi.org/10.1007/s11258-011-9912-4)
- COLEY, P. D.; BRYANT, J. P. and CHAPIN, F. S. (1985): Resource availability and plant antiherbivore defense. In: *Science* 230, 895–899. DOI: [10.1126/science.230.4728.895](https://doi.org/10.1126/science.230.4728.895)
- CONDIT, R.; ASHTON, P. S.; BAKER, P.; BUNYAVEJCHEWIN, S.; GUNATILLEKE, S.; GUNATILLEKE, N.; HUBBELL, S. P.; FOSTER, R. B.; ITOH, A.; LAFRANKIE, J. V.; LEE, H. S.; LOSOS, E.; MANOKARAN, N.; SUKUMAR, R. and YAMAKURA, T. (2000): Spatial patterns in the distribution of tropical tree species. In: *Science* 288, 1414–1418. DOI: [10.1126/science.288.5470.1414](https://doi.org/10.1126/science.288.5470.1414)
- CRASE, B.; LIEDLOFF, A. C. and WINTLE, B. A. (2012): A new method for dealing with residual spatial autocorrelation in species distribution models. In: *Ecography* 35, 879–888. DOI: [10.1111/j.1600-0587.2011.07138.x](https://doi.org/10.1111/j.1600-0587.2011.07138.x)
- DORMANN, C.; MCPHERSON, J. M.; ARAÚJO, M. B.; BIVAND, R.; BOLLIGER, J.; CARL, G.; G. DAVIES, R.; HIRZEL, A.; JETZ, W.; KISSLING, D. W.; KÜHN, I.; OHLEMÜLLER, R.; PERES-NETO, P. R.; REINEKING, B.; SCHRÖDER, B.; SCHURR, F. M. and WILSON, R. (2007): Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. In: *Ecography* 30, 609–628. DOI: [10.1111/j.2007.0906-7590.05171.x](https://doi.org/10.1111/j.2007.0906-7590.05171.x)
- ELITH, J. and GRAHAM, C. H. (2009): Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. In: *Ecography* 32, 66–77. DOI: [10.1111/j.1600-0587.2008.05505.x](https://doi.org/10.1111/j.1600-0587.2008.05505.x)
- ELITH, J. and LEATHWICK, J. R. (2009): Species distribution models: ecological explanation and prediction across space and time. In: *Annual Review of Ecology Evolution and Systematics* 40, 677–697. DOI: [10.1146/annurev.ecolsys.110308.120159](https://doi.org/10.1146/annurev.ecolsys.110308.120159)
- ELITH, J.; H. GRAHAM, C.; P. ANDERSON, R.; DUDÍK, M.; FERRIER, S.; GUISAN, A.; J. HIJMANS, R.; HUETTSMANN, F.; R. LEATHWICK, J.; LEHMANN, A.; LI, J.; G. LOHMANN, L.; A. LOISELLE, B.; MANION, G.; MORITZ, C.; NAKAMURA, M.; NAKAZAWA, Y.; MCC. M. OVERTON, J.; TOWNSEND PETERSON, A.; J. PHILLIPS, S.; RICHARDSON, K.; SCACHETTI-PEREIRA, R.; E. SCHAPIRE, R.; SOBERÓN, J.; WILLIAMS, S.; S. WISZ, M. and E. ZIMMERMANN, N. (2006): Novel methods improve prediction of species’ distributions from occurrence data. In: *Ecography* 29, 129–151. DOI: [10.1111/j.2006.0906-7590.04596.x](https://doi.org/10.1111/j.2006.0906-7590.04596.x)
- ELITH, J.; PHILLIPS, S. J.; HASTIE, T.; DUDÍK, M.; CHEE, Y. E. and YATES, C. J. (2011): A statistical explanation of MaxEnt for ecologists. In: *Diversity and Distributions* 17, 43–57. DOI: [10.1111/j.1472-4642.2010.00725.x](https://doi.org/10.1111/j.1472-4642.2010.00725.x)
- ELLENBERG, H. H. (2009): *Vegetation ecology of Central Europe*. Cambridge.
- ENGBRECHT, B. M. J.; COMITA, L. S.; CONDIT, R.; KURSAR, T. A.; TYREE, M. T.; TURNER, B. L. and HUBBELL, S. P. (2007): Drought sensitivity shapes species distribution patterns in tropical forests. In: *Nature* 447, 80–82. DOI: [10.1038/nature05747](https://doi.org/10.1038/nature05747)
- FAGUA, J. C.; CABRERA, E. and GONZALEZ, V. H. (2013): The effect of highly variable topography on the spatial distribution of *Aniba perutilis* (Lauraceae) in the Colombian Andes. In: *Revista de Biología Tropical* 61, 301–309. DOI: [10.15517/rbt.v61i1.11129](https://doi.org/10.15517/rbt.v61i1.11129)
- FIELDING, A. H. and BELL, J. F. (1997): A review of methods for the assessment of prediction errors in conservation presence/absence models. In: *Environmental Conservation* 24, 38–49.
- FRANKLIN, J. and MILLER, J. A. (2009): *Mapping species distributions: spatial inference and prediction*. Cambridge.
- FRIES, A.; ROLLENBECK, R.; GÖTTLICHER, D.; NAUSS, T.; HOMMEIER, J.; PETERS, T. and BENDIX, J. (2009): Thermal structure of a megadiverse Andean mountain ecosystem in southern Ecuador and its regionalization. In: *Erdkunde* 63, 321–335. DOI: [10.3112/erdkunde.2009.04.03](https://doi.org/10.3112/erdkunde.2009.04.03)
- FRIES, A.; ROLLENBECK, R.; BAYER, F.; GONZALEZ, V.; OÑATE-VALVIESO, F.; PETERS, T. and BENDIX, J. (2014): Catchment precipitation processes in the San Francisco valley in southern Ecuador: combined approach using high-resolution radar images and in situ observations. In: *Meteorology and Atmospheric Physics* 126, 13–29. DOI: [10.1007/s00703-014-0335-3](https://doi.org/10.1007/s00703-014-0335-3)
- GBIF (Global Biodiversity Information Facility): Occurrence data of 16 species, 1141 records. <http://www.gbif.org/> (accessed on 4 November 2015).
- GENTRY, A. H. (1988): Changes in plant community diversity and floristic composition on environmental and geographical gradients. In: *Annals of the Missouri Botanical Garden* 75, 1–34. DOI: [10.2307/2399464](https://doi.org/10.2307/2399464)
- GETZIN, S.; WIEGAND, T. and HUBBELL, S. P. (2014): Stochastically driven adult–recruit associations of tree species on Barro Colorado Island. In: *Proceedings of the Royal Society B: Biological Sciences* 281. DOI: [10.1098/rspb.2014.0922](https://doi.org/10.1098/rspb.2014.0922)

- GUISAN, A. and THUILLER, W. (2005): Predicting species distribution: offering more than simple habitat models. In: *Ecology Letters* 8, 993–1009. DOI: [10.1111/j.1461-0248.2005.00792.x](https://doi.org/10.1111/j.1461-0248.2005.00792.x)
- GUISAN, A.; WEISS, S. and WEISS, A. (1999): GLM versus CCA spatial modeling of plant species distribution. In: *Plant Ecology* 143, 107–122. DOI: [10.1023/A:1009841519580](https://doi.org/10.1023/A:1009841519580)
- GUNATILLEKE, C. V. S.; GUNATILLEKE, I. A. U. N.; ESUFALI, S.; HARMS, K. E.; ASHTON, P. M. S.; BURSLEM, D. F. R. P. and ASHTON, P. S. (2006): Species-habitat associations in a Sri Lankan dipterocarp forest. In: *Journal of Tropical Ecology* 22, 371. DOI: [10.1017/S0266467406003282](https://doi.org/10.1017/S0266467406003282)
- GÜNTER, S.; CABRERA, O.; WEBER, M.; STIMM, B.; ZIMMERMANN, M.; FIEDLER, K.; KNUTH, J.; BOY, J.; WILCKE, W.; IOST, S.; MAKESCHIN, F.; WERNER, F.; GRADSTEIN, R. and MOSANDL, R. (2008): Natural forest management in neotropical mountain rain forests – an ecological experiment. In: BECK, E.; BENDIX, J.; KOTTKE, I.; MAKESCHIN, F. and MOSANDL, R. (eds.): *Gradients in a tropical mountain ecosystem of Ecuador*. Berlin, Heidelberg, 347–359. DOI: [10.1007/978-3-540-73526-7\\_33](https://doi.org/10.1007/978-3-540-73526-7_33)
- HARMS, K. E.; CONDIT, R.; HUBBELL, S. P. and FOSTER, R. B. (2001): Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. In: *Journal of Ecology* 89, 947–959. DOI: [10.1111/j.1365-2745.2001.00615.x](https://doi.org/10.1111/j.1365-2745.2001.00615.x)
- HENGL, T.; SIERDSEMA, H.; RADOVIC, A. and DILO, A. (2009): Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. In: *Ecological Modelling* 220, 3499–3511. DOI: [10.1016/j.ecolmodel.2009.06.038](https://doi.org/10.1016/j.ecolmodel.2009.06.038)
- HERMOSILLA, C.; ROCHA, F. and VALAVANIS, V. (2011): Assessing *Octopus vulgaris* distribution using presence-only model methods. In: *Hydrobiologia* 670, 35–47. DOI: [10.1007/s10750-011-0671-y](https://doi.org/10.1007/s10750-011-0671-y)
- HERNANDEZ, P. A.; FRANKE, I.; HERZOG, S. K.; PACHECO, V.; PANIAGUA, L.; QUINTANA, H. L.; SOTO, A.; SWENSON, J. J.; TOVAR, C.; VALQUI, T. H.; VARGAS, J. and YOUNG, B. E. (2008): Predicting species distributions in poorly-studied landscapes. In: *Biodiversity & Conservation* 17, 1353–1366. DOI: [10.1007/s10531-007-9314-z](https://doi.org/10.1007/s10531-007-9314-z)
- HIJMANS, R. J.; PHILLIPS, S.; LEATHWICK, J. and ELITH, J. (2013): *dismo: Species distribution modeling*. <http://CRAN.R-project.org/package=dismo>
- HIRZEL, A. H. and LE LAY, G. (2008): Habitat suitability modelling and niche theory. In: *Journal of Applied Ecology* 45, 1372–1381. DOI: [10.1111/j.1365-2664.2008.01524.x](https://doi.org/10.1111/j.1365-2664.2008.01524.x)
- HIRZEL, A. H.; HAUSSER, J.; CHESSEL, D. and PERRIN, N. (2002): Ecological-Niche Factor Analysis: how to compute habitat-suitability maps without absence data? In: *Ecology* 83, 2027–2036. DOI: [10.1890/0012-9658\(2002\)083\[2027:ENFAHT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2)
- HOMEIER, J. (2008): The influence of topography on forest structure and regeneration dynamics in an Ecuadorian montane forest. In: GRADSTEIN, S. R.; HOMEIER, J. and GANSERT, D. (eds.): *The tropical mountain forest. Patterns and processes in a biodiversity hotspot*. Göttingen, 97–108.
- HOMEIER, J. and BRECKLE, S.-W. (2008): Gap dynamics in a tropical lower montane forest in South Ecuador. In: BECK, E.; BENDIX, J.; KOTTKE, I.; MAKESCHIN, F. and MOSANDL, R. (eds.): *Gradients in a tropical mountain ecosystem of Ecuador*. Berlin, Heidelberg, 311–317. DOI: [10.1007/978-3-540-73526-7\\_30](https://doi.org/10.1007/978-3-540-73526-7_30)
- HOMEIER, J.; WERNER, F. A.; GRADSTEIN, S. R.; BRECKLE, S. W. and RICHTER, M. (2008): Potential vegetation and floristic composition of Andean forests in South Ecuador, with a focus on the RBSF. In: BECK, E.; BENDIX, J.; KOTTKE, I.; MAKESCHIN, F. and MOSANDL, R. (eds.): *Gradients in a tropical mountain ecosystem of Ecuador*. Berlin, Heidelberg, 87–100. DOI: [10.1007/978-3-540-73526-7\\_10](https://doi.org/10.1007/978-3-540-73526-7_10)
- HOMEIER, J.; BRECKLE, S.-W.; GÜNTER, S.; ROLLENBECK, R. T. and LEUSCHNER, C. (2010): Tree diversity, forest structure and productivity along altitudinal and topographical gradients in a species-rich Ecuadorian montane rain forest. In: *Biotropica* 42, 140–148. DOI: [10.1111/j.1744-7429.2009.00547.x](https://doi.org/10.1111/j.1744-7429.2009.00547.x)
- HUBBELL, S. P. (2001): *The unified neutral theory of biodiversity and biogeography*. Princeton.
- HUMBOLDT, A. v. and BONPLAND, A. (1805): *Essai sur la géographie des plantes*. Paris.
- HUTCHINSON, G. E. (1957): Concluding remarks. In: *Cold Spring Harbor Symposia on Quantitative Biology* 22, 415–427. DOI: [10.1101/SQB.1957.022.01.039](https://doi.org/10.1101/SQB.1957.022.01.039)
- ITOH, A.; YAMAKURA, T.; OHKUBO, T.; KANZAKI, M.; PALMIOTTO, P. A.; LAFRANKIE, J. V.; ASHTON, P. S. and LEE, H. S. (2003): Importance of topography and soil texture in the spatial distribution of two sympatric dipterocarp trees in a Bornean rainforest. In: *Ecological Research* 18, 307–320. DOI: [10.1046/j.1440-1703.2003.00556.x](https://doi.org/10.1046/j.1440-1703.2003.00556.x)
- JIMÉNEZ-VALVERDE, A. (2012): Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. In: *Global Ecology and Biogeography* 21, 498–507. DOI: [10.1111/j.1466-8238.2011.00683.x](https://doi.org/10.1111/j.1466-8238.2011.00683.x)
- JIMÉNEZ-VALVERDE, A.; LOBO, J. M. and HORTAL, J. (2008): Not as good as they seem: the importance of concepts in species distribution modelling. In: *Diversity and Distributions* 14, 885–890. DOI: [10.1111/j.1472-4642.2008.00496.x](https://doi.org/10.1111/j.1472-4642.2008.00496.x)
- JOHN, R.; DALLING, J. W.; HARMS, K. E.; YAVITT, J. B.; STALLARD, R. F.; MIRABELLO, M.; HUBBELL, S. P.; VALENCIA, R.; NAVARRETE, H.; VALLEJO, M. and FOSTER, R. B. (2007): Soil nutrients influence spatial distributions of tropical tree species. In: *Proceedings of the National Academy of Sciences of the United States of America* 104, 864–869. DOI: [10.1073/pnas.0604666104](https://doi.org/10.1073/pnas.0604666104)
- JONES, M. M.; TUOMISTO, H. and OLIVAS, P. C. (2008): Differences in the degree of environmental control on

- large and small tropical plants: just a sampling effect? In: *Journal of Ecology* 96, 367–377. DOI: [10.1111/j.1365-2745.2007.01340.x](https://doi.org/10.1111/j.1365-2745.2007.01340.x)
- JONES, M. M.; SZYSKA, B. and KESSLER, M. (2011): Microhabitat partitioning promotes plant diversity in a tropical montane forest. In: *Global Ecology and Biogeography* 20, 558–569. DOI: [10.1111/j.1466-8238.2010.00627.x](https://doi.org/10.1111/j.1466-8238.2010.00627.x)
- KAMINO, L. H. Y.; STEHMANN, J. R.; AMARAL, S.; MARCO, P. DE; RANGEL, T. F.; SIQUEIRA, M. F. DE; GIOVANNI, R. DE and HORTAL, J. (2012): Challenges and perspectives for species distribution modelling in the Neotropics. In: *Biology Letters* 8, 324–326. DOI: [10.1098/rsbl.2011.0942](https://doi.org/10.1098/rsbl.2011.0942)
- KHATCHIKIAN, C.; SANGERMANO, F.; KENDELL, D. and LIVDAHL, T. (2011): Evaluation of species distribution model algorithms for fine-scale container-breeding mosquito risk prediction. In: *Medical and Veterinary Entomology* 25, 268–275. DOI: [10.1111/j.1365-2915.2010.00935.x](https://doi.org/10.1111/j.1365-2915.2010.00935.x)
- LAN, G.; ZHU, H.; CAO, M.; HU, Y.; WANG, H.; DENG, X.; ZHOU, S.; CUI, J.; HUANG, J.; HE, Y.; LIU, L.; XU, H. and SONG, J. (2009): Spatial dispersion patterns of trees in a tropical rainforest in Xishuangbanna, southwest China. In: *Ecological Research* 24, 1117–1124. DOI: [10.1007/s11284-009-0590-9](https://doi.org/10.1007/s11284-009-0590-9)
- LANDIS, J. R. and KOCH, G. G. (1977): The measurement of observer agreement for categorical data. In: *Biometrics* 33, 159–174. DOI: [10.2307/2529310](https://doi.org/10.2307/2529310)
- LAUSCH, A.; FAHSE, L. and HEURICH, M. (2011): Factors affecting the spatio-temporal dispersion of *Ips typographus* (L.) in Bavarian Forest National Park: a long-term quantitative landscape-level analysis. In: *Forest Ecology and Management* 261, 233–245. DOI: [10.1016/j.foreco.2010.10.012](https://doi.org/10.1016/j.foreco.2010.10.012)
- LEDO, A.; BURSLEM, D. F.; CONDÉS, S. and MONTES, F. (2013): Micro-scale habitat associations of woody plants in a neotropical cloud forest. In: *Journal of Vegetation Science* 24, 1086–1097. DOI: [10.1111/jvs.12023](https://doi.org/10.1111/jvs.12023)
- LEGENDRE, P. and FORTIN, M. J. (1989): Spatial pattern and ecological analysis. In: *Plant Ecology* 80, 107–138. DOI: [10.1007/BF00048036](https://doi.org/10.1007/BF00048036)
- LENNON, J. J. (2000): Red-shifts and red herrings in geographical ecology. In: *Ecography* 23, 101–113. DOI: [10.1034%2Fj.1600-0587.2000.230111.x](https://doi.org/10.1034%2Fj.1600-0587.2000.230111.x)
- LIEBERMAN, M.; LIEBERMAN, D.; HARTSHORN, G. S. and PERALTA, R. (1985): Small-scale altitudinal variation in lowland wet tropical forest vegetation. In: *Journal of Ecology* 73, 505–516. DOI: [10.2307/2260490](https://doi.org/10.2307/2260490)
- LIESS, M.; GLASER, B. and HUWE, B. (2011): Functional soil-landscape modelling to estimate slope stability in a steep Andean mountain forest region. In: *Geomorphology* 132, 287–299. DOI: [10.1016/j.geomorph.2011.05.015](https://doi.org/10.1016/j.geomorph.2011.05.015)
- LIU, C.; WHITE, M. and NEWELL, G. (2013): Selecting thresholds for the prediction of species occurrence with presence-only data. In: *Journal of Biogeography* 40, 778–789. DOI: [10.1111/jbi.12058](https://doi.org/10.1111/jbi.12058)
- LOBO, J. M.; JIMÉNEZ-VALVERDE, A. and REAL, R. (2008): AUC: a misleading measure of the performance of predictive distribution models. In: *Global Ecology and Biogeography* 17, 145–151. DOI: [10.1111/j.1466-8238.2007.00358.x](https://doi.org/10.1111/j.1466-8238.2007.00358.x)
- LOBO, J. M.; JIMÉNEZ-VALVERDE, A. and HORTAL, J. (2010): The uncertain nature of absences and their importance in species distribution modelling. In: *Ecography* 33, 103–114. DOI: [10.1111/j.1600-0587.2009.06039.x](https://doi.org/10.1111/j.1600-0587.2009.06039.x)
- MCKINNEY, J. A.; HOFFMAYER, E. R.; WU, W.; FULLFORD, R. and HENDON, J. M. (2012): Feeding habitat of the whale shark *Rhincodon typus* in the northern Gulf of Mexico determined using species distribution modelling. In: *Marine Ecology Progress Series* 458, 199–211. DOI: [10.3354/meps09777](https://doi.org/10.3354/meps09777)
- MERCKX, B.; STEYAERT, M.; VANREUSEL, A.; VINCX, M. and VANAUWERBEKE, J. (2011): Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. In: *Ecological Modelling* 222, 588–597. DOI: [10.1016/j.ecolmodel.2010.11.016](https://doi.org/10.1016/j.ecolmodel.2010.11.016)
- MUENCHOW, J.; BRENNING, A. and RICHTER, M. (2012): Geomorphic process rates of landslides along a humidity gradient in the tropical Andes. In: *Geomorphology* 139–140, 271–284. DOI: [10.1016/j.geomorph.2011.10.029](https://doi.org/10.1016/j.geomorph.2011.10.029)
- MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; DA FONSECA, G. A. and KENT, J. (2000): Biodiversity hotspots for conservation priorities. In: *Nature* 403, 853–858. DOI: [10.1038/35002501](https://doi.org/10.1038/35002501)
- NEGRELLE, R. R. B. (1995): Sprouting after uprooting of canopy trees in the Atlantic rain forest of Brazil. In: *Biotropica* 27, 448–454. DOI: [10.2307/2388957](https://doi.org/10.2307/2388957)
- OESKER, M.; DALITZ, H.; GÜNTER, S.; HOMEIER, J. and MATEZKI, S. (2008): Spatial heterogeneity patterns – a comparison between gorges and ridges in the upper part of an evergreen lower montane forest. In: BECK, E.; BENDIX, J.; KOTITKE, I.; MAKESCHIN, F. and MOSANDL, R. (eds.): *Gradients in a tropical mountain ecosystem of Ecuador*. Berlin, Heidelberg, 267–274. DOI: [10.1007/978-3-540-73526-7\\_25](https://doi.org/10.1007/978-3-540-73526-7_25)
- PEARSON, R. G.; RAXWORTHY, C. J.; NAKAMURA, M. and TOWNSEND PETERSON, A. (2007): Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. In: *Journal of Biogeography* 34, 102–117. DOI: [10.1111/j.1365-2699.2006.01594.x](https://doi.org/10.1111/j.1365-2699.2006.01594.x)
- PÉLISSIER, R.; DRAY, S. and SABATIER, D. (2002): Within-plot relationships between tree species occurrences and hydrological soil constraints: an example in French Guiana investigated through canonical correlation analysis. In: *Plant Ecology* 162, 143–156. DOI: [10.1023/A:1020399603500](https://doi.org/10.1023/A:1020399603500)
- PETERSON, A. T. (2006): Uses and requirements of ecological niche models and related distributional models. In: *Biodiversity Informatics* 3, 59–72. DOI: [10.17161/bi.v3i0.29](https://doi.org/10.17161/bi.v3i0.29)
- PHILLIPS, S. J. and DUDIK, M. (2008): Modeling of species distributions with Maxent: new extensions and a com-

- prehensive evaluation. In: *Ecography* 31, 161–175. DOI: [10.1111/j.0906-7590.2008.5203.x](https://doi.org/10.1111/j.0906-7590.2008.5203.x)
- PHILLIPS, S. J.; ANDERSON, R. and SCHAPIRE, R. (2006): Maximum entropy modeling of species geographic distributions. In: *Ecological Modelling* 190, 231–259. DOI: [10.1016/j.ecolmodel.2005.03.026](https://doi.org/10.1016/j.ecolmodel.2005.03.026)
- PHILLIPS, S. J.; DUDIK, M.; ELITH, J.; GRAHAM, C. H.; LEHMANN, A.; LEATHWICK, J. and FERRIER, S. (2009): Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. In: *Ecological Applications* 19, 181–197. DOI: [10.1890/07-2153.1](https://doi.org/10.1890/07-2153.1)
- PITMAN, N. C. A.; STOCKS, G.; BRUNA, M.; WIDMER, J.; JENKINS, C. N.; SEALES, L. and PANIAGUA, F. (2011): Volume and geographical distribution of ecological research in the Andes and the Amazon 1995–2008. In: *Ecological Research* 4, 64–81. <http://hdl.handle.net/10161/5068>
- QUEENBOROUGH, S. A.; BURSLEM, D. F. R. P.; GARWOOD, N. C. and VALENCIA, R. (2007): Habitat niche partitioning by 16 species of Myristicaceae in Amazonian Ecuador. In: *Plant Ecology* 192, 193–207. DOI: [10.1007/s11258-007-9328-3](https://doi.org/10.1007/s11258-007-9328-3)
- RAES, N. (2012): Partial versus full species distribution models. In: *Natureza & Conservação* 10, 127–138. DOI: [10.4322/natcon.2012.020](https://doi.org/10.4322/natcon.2012.020)
- RAES, N. and STEEGE, H. ter (2007): A null-model for significance testing of presence-only species distribution models. In: *Ecography* 30, 727–736. DOI: [10.1111/j.2007.0906-7590.05041.x](https://doi.org/10.1111/j.2007.0906-7590.05041.x)
- ROLLENBECK, R. (2006): Variability of precipitation in the Reserva Biológica San Francisco/southern Ecuador. In: *Lyonia* 9, 43–51.
- SAGA Development Team (2008): System for automated geoscientific analyses (SAGA GIS). <http://www.saga-gis.org/>
- SÁNCHEZ-FERNÁNDEZ, D.; LOBO, J. M. and HERNÁNDEZ-MANRIQUE, O. L. (2011): Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. In: *Diversity and Distributions* 17, 163–171. DOI: [10.1111/j.1472-4642.2010.00716.x](https://doi.org/10.1111/j.1472-4642.2010.00716.x)
- SEGURADO, P.; ARAUJO, M. B. and KUNIN, W. E. (2006): Consequences of spatial autocorrelation for niche-based models. In: *Journal of Applied Ecology* 43, 433–444. DOI: [10.1111/j.1365-2664.2006.01162.x](https://doi.org/10.1111/j.1365-2664.2006.01162.x)
- SOBERÓN, J. M. (2010): Niche and area of distribution modeling: a population ecology perspective. In: *Ecography* 33, 159–167. DOI: [10.1111/j.1600-0587.2009.06074.x](https://doi.org/10.1111/j.1600-0587.2009.06074.x)
- SVENNING, J.-C.; HARLEV, D.; SØRENSEN, M. M. and BALSLEV, H. (2009): Topographic and spatial controls of palm species distributions in a montane rain forest, southern Ecuador. In: *Biodiversity and Conservation* 18, 219–228. DOI: [10.1007/s10531-008-9468-3](https://doi.org/10.1007/s10531-008-9468-3)
- SYFERT, M. M.; SMITH, M. J. and COOMES, D. A. (2013): The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. In: *PLoS ONE* 8, e55158. DOI: [10.1371/journal.pone.0055158](https://doi.org/10.1371/journal.pone.0055158)
- TAKYU, M.; AIBA, S.-I. and KITAYAMA, K. (2002): Effects of topography on tropical lower montane forests under different geological conditions on Mount Kinabalu, Borneo. In: *Plant Ecology* 159, 35–49. DOI: [10.1023/A:1015512400074](https://doi.org/10.1023/A:1015512400074)
- TANNER, E. V. J.; VITOUSEK, P. M. and CUEVAS, E. (1998): Experimental investigation of nutrient limitation of forest growth on wet tropical mountains. In: *Ecology* 79, 10–22. DOI: [10.2307/176860](https://doi.org/10.2307/176860)
- THULLER, W.; BROTONS, L.; ARAÚJO, M. B. and LAVOREL, S. (2004): Effects of restricting environmental range of data to project current and future species distributions. In: *Ecography* 27, 165–172. DOI: [10.1111/j.0906-7590.2004.03673.x](https://doi.org/10.1111/j.0906-7590.2004.03673.x)
- VALENCIA, R.; FOSTER, R. B.; VILLA, G.; CONDIT, R.; SVENNING, J.-C.; HERNÁNDEZ, C.; ROMOLEROUX, K.; LOSOS, E.; MAGARD, E. and BALSLEV, H. (2004): Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. In: *Journal of Ecology* 92, 214–229. DOI: [10.1111/j.0022-0477.2004.00876.x](https://doi.org/10.1111/j.0022-0477.2004.00876.x)
- VEDEL-SØRENSEN, M.; TOVARANONTE, J.; KLIT BØCHER, P.; BALSLEV, H. and BARFOD, A. S. (2013): Spatial distribution and environmental preferences of 10 economically important forest palms in western South America. In: *Forest Ecology and Management* 307, 284–292. DOI: [10.1016/j.foreco.2013.07.005](https://doi.org/10.1016/j.foreco.2013.07.005)
- VELOZ, S. D. (2009): Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. In: *Journal of Biogeography* 36, 2290–2299. DOI: [10.1111/j.1365-2699.2009.02174.x](https://doi.org/10.1111/j.1365-2699.2009.02174.x)
- VOLLAND-VOIGT, F.; BRÄUNING, A.; GANZHI, O.; PETERS, T. and MAZA, H. (2011): Radial stem variations of *Tabebuia chrysantha* (Bignoniaceae) in different tropical forest ecosystems of southern Ecuador. In: *Trees* 25, 39–48. DOI: [10.1007/s00468-010-0461-6](https://doi.org/10.1007/s00468-010-0461-6)
- VORPAHL, P.; ELSENBEER, H.; MÄRKER, M. and SCHRÖDER, B. (2012): How can statistical models help to determine driving factors of landslides? In: *Ecological Modelling* 239, 27–39. DOI: [10.1016/j.ecolmodel.2011.12.007](https://doi.org/10.1016/j.ecolmodel.2011.12.007)
- WANG, Z.; YE, W.; CAO, H.; HUANG, Z.; LIAN, J.; LI, L.; WEI, S. and SUN, I.-F. (2009): Species-topography association in a species-rich subtropical forest of China. In: *Basic and Applied Ecology* 10, 648–655. DOI: [10.1016/j.baae.2009.03.002](https://doi.org/10.1016/j.baae.2009.03.002)
- WEBB, E. L.; STANFIELD, B. J. and JENSEN, M. L. (1999): Effects of topography on rainforest tree community structure and diversity in American Samoa, and implications for frugivore and nectarivore populations. In: *Journal of Biogeography* 26, 887–897. DOI: [10.1046/j.1365-2699.1999.00326.x](https://doi.org/10.1046/j.1365-2699.1999.00326.x)

- WERNER, F. A. and HOMEIER, J. (2015): Is tropical montane forest heterogeneity promoted by a resource-driven feedback cycle? Evidence from nutrient relations, herbivory and litter decomposition along a topographical gradient. In: *Functional Ecology* 29, 430–440. DOI: [10.1111/1365-2435.12351](https://doi.org/10.1111/1365-2435.12351)
- WERNER, F. A.; HOMEIER, J.; OESKER, M. and BOY, J. (2012): Epiphytic biomass of a tropical montane forest varies with topography. In: *Journal of Tropical Ecology* 28, 23–31. DOI: [10.1017/s0266467411000526](https://doi.org/10.1017/s0266467411000526)
- WILCKE, W.; YASIN, S.; SCHMITT, A.; VALAREZO, C. and ZECH, W. (2008): Soils along the altitudinal transect and in catchments. In: BECK, E.; BENDIX, J.; KOTTKE, I.; MAKESCHIN, F. and MOSANDL, R. (eds.): *Gradients in a tropical mountain ecosystem of Ecuador*. Berlin, Heidelberg, 75–85. DOI: [10.1007/978-3-540-73526-7\\_9](https://doi.org/10.1007/978-3-540-73526-7_9)
- WILCKE, W.; BOY, J.; GOLLER, R.; FLEISCHBEIN, K.; VALAREZO, C. and ZECH, W. (2011): Effect of topography on soil fertility and water flow in an Ecuadorian lower montane forest. In: BRUIJNZEEL, L. A.; SCATENA, F. N. and HAMILTON, L. S. (eds.): *Tropical montane cloud forests. Science for conservation and management*. Cambridge, 402–409. DOI: [10.1017/cbo9780511778384.045](https://doi.org/10.1017/cbo9780511778384.045)
- WILEY, E. O.; MCNYSET, K. M.; PETERSON, A. T.; ROBINS, C. R. and STEWART, A. M. (2003): Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. In: *Oceanography* 16 (3), 120–127. DOI: [10.5670/oceanog.2003.42](https://doi.org/10.5670/oceanog.2003.42)
- WILLIS, K. J. and WHITTAKER, R. J. (2002): Species diversity-scale matters. In: *Science* 295, 1245–1248.
- WISZ, M. S.; HIJMANS, R. J.; LI, J.; PETERSON, A. T.; GRAHAM, C. H.; GUIBAN, A. and NCEAS Predicting Species Distributions Working Group (2008): Effects of sample size on the performance of species distribution models. In: *Diversity and Distributions* 14, 763–773. DOI: [10.1111/j.1472-4642.2008.00482.x](https://doi.org/10.1111/j.1472-4642.2008.00482.x)
- WOLF, K.; VELDKAMP, E.; HOMEIER, J. and MARTINSON, G. O. (2011): Nitrogen availability links forest productivity, soil nitrous oxide and nitric oxide fluxes of a tropical montane forest in southern Ecuador. In: *Global Biogeochemical Cycles* 25. DOI: [10.1029/2010GB003876](https://doi.org/10.1029/2010GB003876)
- WOLLAN, A. K.; BAKKESTUEN, V.; KAUSERUD, H.; GULDEN, G. and HALVORSEN, R. (2008): Modelling and predicting fungal distribution patterns using herbarium data. In: *Journal of Biogeography* 35, 2298–2310. DOI: [10.1111/j.1365-2699.2008.01965.x](https://doi.org/10.1111/j.1365-2699.2008.01965.x)
- YAMADA, T.; NOOR, N. S. M. and OKUDA, T. (2010): Habitat association of trees in a 50-ha Malaysian rain forest plot. In: *Tropics* 19, 1–8. DOI: [10.3759/tropics.19.1](https://doi.org/10.3759/tropics.19.1)

## Authors

Daniel Kübler  
 Institute for World Forestry  
 University of Hamburg  
 Leuschnerstr. 91  
 21031 Hamburg  
 Germany  
[daniel.kuebler@uni-hamburg.de](mailto:daniel.kuebler@uni-hamburg.de)

PD Dr. Sven Günter  
 Thünen Institute of International  
 Forestry and Forest Economics  
 Leuschnerstr. 91  
 21031 Hamburg  
 Germany

Dr. Patrick Hildebrandt  
 Dr. Bernd Stimm  
 Prof. Dr. Dr. Michael Weber  
 Prof. Dr. Dr. Reinhard Mosandl  
 Institute of Silviculture  
 TU München  
 Hans-Carl-von-Carlowitz-Platz 2  
 85354 Freising  
 Germany

Johana Muñoz  
 Dr. Nikolay Aguirre  
 Área Agropecuaria y de Recursos  
 Naturales Renovables  
 Universidad Nacional de Loja  
 Ciudad Universitaria Guillermo  
 Falconi Espinosa La Argelia  
 Ecuador

Omar Cabrera  
 Departamento de Ciencias Naturales,  
 Universidad Técnica Particular de Loja,  
 San Cayetano Alto,  
 Calle París,  
 Ecuador

Jörg Zeilinger  
 Dr. Brenner Silva  
 University of Marburg  
 Faculty of Geography  
 Deutschhausstraße 12  
 35032 Marburg  
 Germany

## Appendix

### Supplementary material

#### A 1 Data partitioning and randomization

AUC and TSS were calculated for final models (created with the complete dataset) and different models obtained from either null-distributions or k-fold data partitioning.

Null-models were used to test if the final models differ significantly from what would be expected by chance alone (RAES and STEEGE 2007). For this, presence points were randomly sampled from all locations within the modeled area. Maxent and ENFA models were then fitted to these presence points and the AUC and TSS were calculated for each model. In order to account for differences in the number of occurrences per species, we created null-models for different quantity ranges of randomly sampled locations, which were defined by intervals of 10 (30–39, 40–49, ... 190–199). For each quantity range, this process was repeated 100 times. The AUC and TSS values of the final models generated with actual presence data were then compared with those of the null-models of the respective quantity range. If the metrics for the actual presence data was higher than the one-sided 95 % confidence interval of the frequency distribution obtained from the null-models, the actual species model was interpreted as performing significantly better than expected by chance (RAES and STEEGE 2007).

For each species, we additionally used a fivefold cross-validation, where available presence points were randomly split into five equally sized data partitions. In turn, five different ENFA and Maxent models were then trained with 4/5 of the data, while the withheld 1/5 of the data was used to calculate AUC and TSS values for each model. In order to obtain a more robust estimation of test statistics, we repeated this process 10 times for each species.

At local scales, the distribution of tree species in tropical forests depends not only on environmental factors, but also on dispersal limitations, biotic interactions, historical events and stochastic processes (SVENNING et al. 2009; JONES et al. 2011; HARMS et al. 2001; BALDECK et al. 2013). All these factors might result in spatial autocorrelation in tree occurrences, which in turn might falsely inflate measures of model performance (VELOZ 2009). In order to examine the effect of spatial autocorrelation on model performance, we separated training and test data into spatially separated partitions (VELOZ 2009). This was

achieved by using a pre-existent system of subplots installed in 2003, subdividing the permanent sample plots into 52 quadrats with dimensions of approximately 50 x 50 m (GÜNTER et al. 2008). Considering that spatial autocorrelation in many tropical forests is strongest at scales < 20 m (CONDIT et al. 2000), we assumed that the size of the pre-existent subplots of 50 x 50 m was sufficiently large to account for spatial autocorrelation. For each species, subplots were randomly sampled from the study area until the cumulative number of presence points within these subplots exceeded 20 % of the total presence points of the respective species. All presence points within those randomly sampled subplots were assigned to a test group, and the remaining presence points to a training group. ENFA and Maxent models were generated with the training group and test statistics were calculated for the test group. This process was repeated 50 times for each species.



## A 2 Supplementary tables and figures

Tab. A-1: Correlation matrix of the topographic variables used for the modelling process. Only variables with absolute values of the Pearson's correlation coefficient  $r \leq 0.45$  were retained. All correlations are significant

	Eastness	Elevation	Slope	TPI 100 m	Wetness Index
Eastness	1.00				
Elevation	-0.03	1.00			
Slope	0.07	-0.17	1.00		
TPI 100m	-0.03	0.45	-0.21	1.00	
Wetness Index	-0.05	-0.33	-0.26	-0.39	1.00

Tab. A-2: Coefficients of the marginality factor of the ENFA for 15 tree species and 5 topographic variables. Positive coefficients of marginality indicate that the species prefers higher-than-mean values with respect to the study area, while negative values indicate the preference of lower-than-mean values. The column "Cluster group" shows the result from a hierarchical cluster analysis (Fig. A-4). Column "Forest type" lists the forest type(s) each species was assigned to by HOMEIER et al. (2010)

Species	Eastness	Elevation	Slope	TPI 100 m	Wetness Index	Cluster group	Forest type
<i>Tapigunia</i>	-0.84	0.03	0.46	0.03	-0.27	A	I
<i>Cecrandi</i>	0.29	-0.59	0.16	-0.45	0.58	B	I
<i>Cecrangu</i>	0.15	-0.82	-0.03	-0.42	0.35	B	I
<i>Cedrmont</i>	0.31	-0.92	0.13	-0.08	0.18	B	
<i>Guarkunt</i>	0.34	-0.48	0.21	-0.72	0.30	B	I
<i>Heliamer</i>	0.05	-0.79	-0.17	-0.41	0.43	B	I
<i>Hyeraspe</i>	0.10	-0.83	-0.16	-0.43	0.31	B	I
<i>Merifran</i>	0.11	-0.73	0.43	-0.48	0.22	B	I
<i>Nectmemb</i>	-0.03	-0.92	0.21	-0.29	-0.14	B	I
<i>Sapiglan</i>	0.31	-0.73	0.11	-0.44	0.41	B	I
<i>Tabechry</i>	0.27	-0.79	0.04	-0.47	0.28	B	I
<i>Alchgran</i>	0.12	0.73	-0.29	0.53	-0.30	C	II, III
<i>Alzavert</i>	-0.42	0.66	-0.03	0.56	-0.27	C	II, III
<i>Clusducu</i>	-0.10	0.74	-0.22	0.51	-0.37	C	II, III
<i>Grafemar</i>	-0.49	0.72	0.12	0.38	-0.28	C	I, II, III

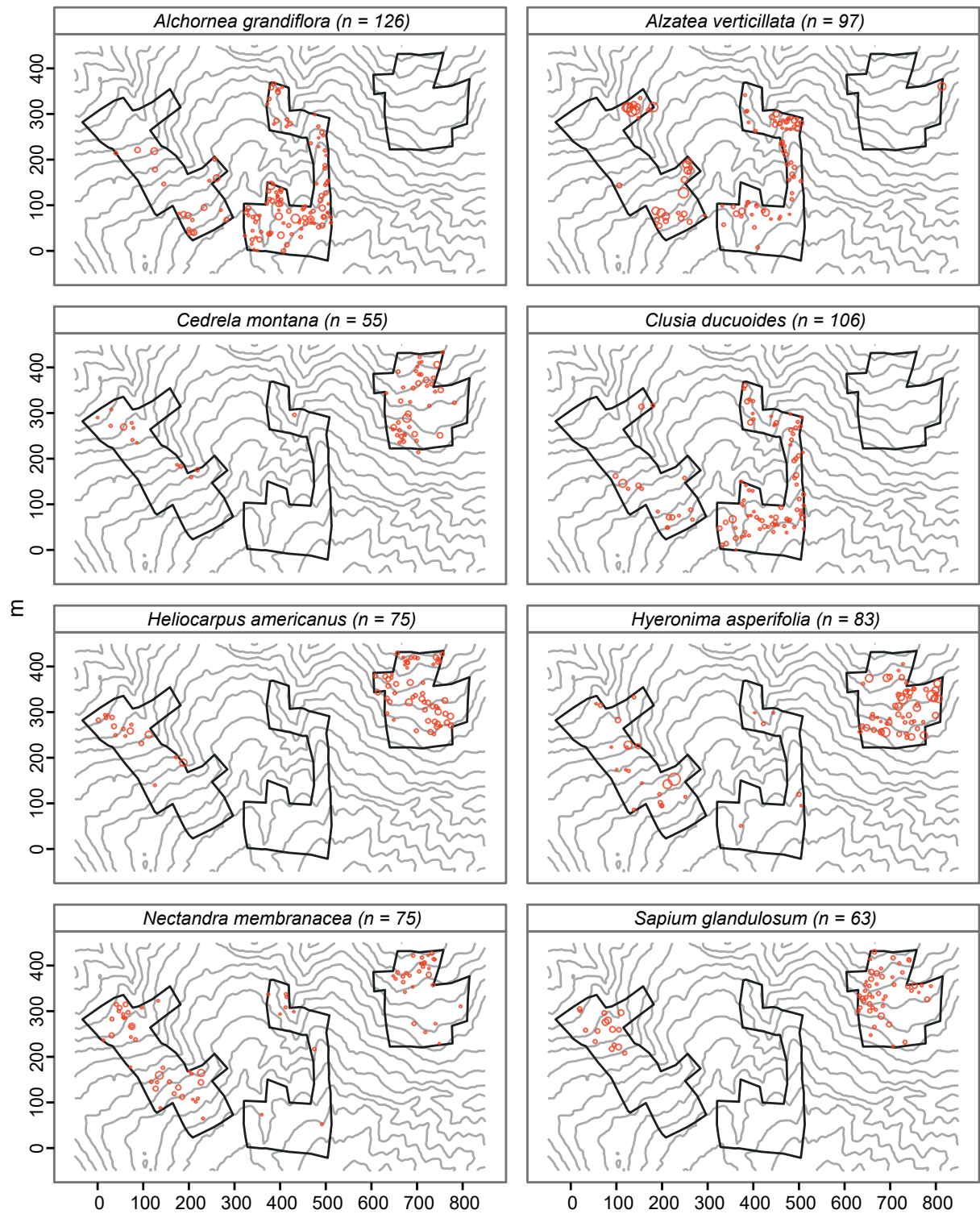
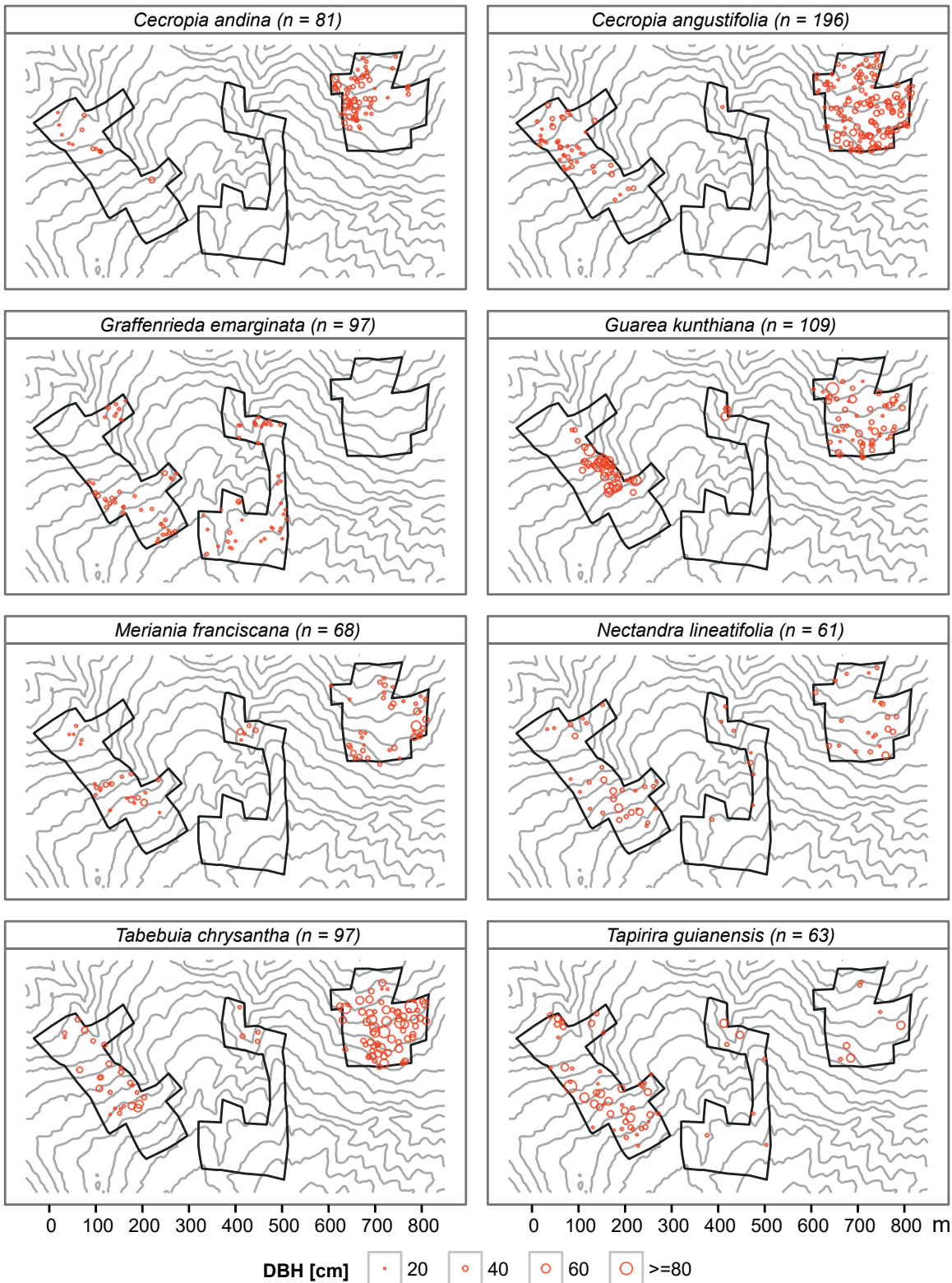


Fig. A-1: Stem distributions of all trees with DBH  $\geq$  20 cm on the 11.1 ha permanent sample plot in the RBSF for the 16 most abundant tree species. Each red circle represents one tree; the size of the circle shows the DBH of trees. The grey lines are elevation contour lines at 20-m intervals. The black lines are limits of the permanent sample plots



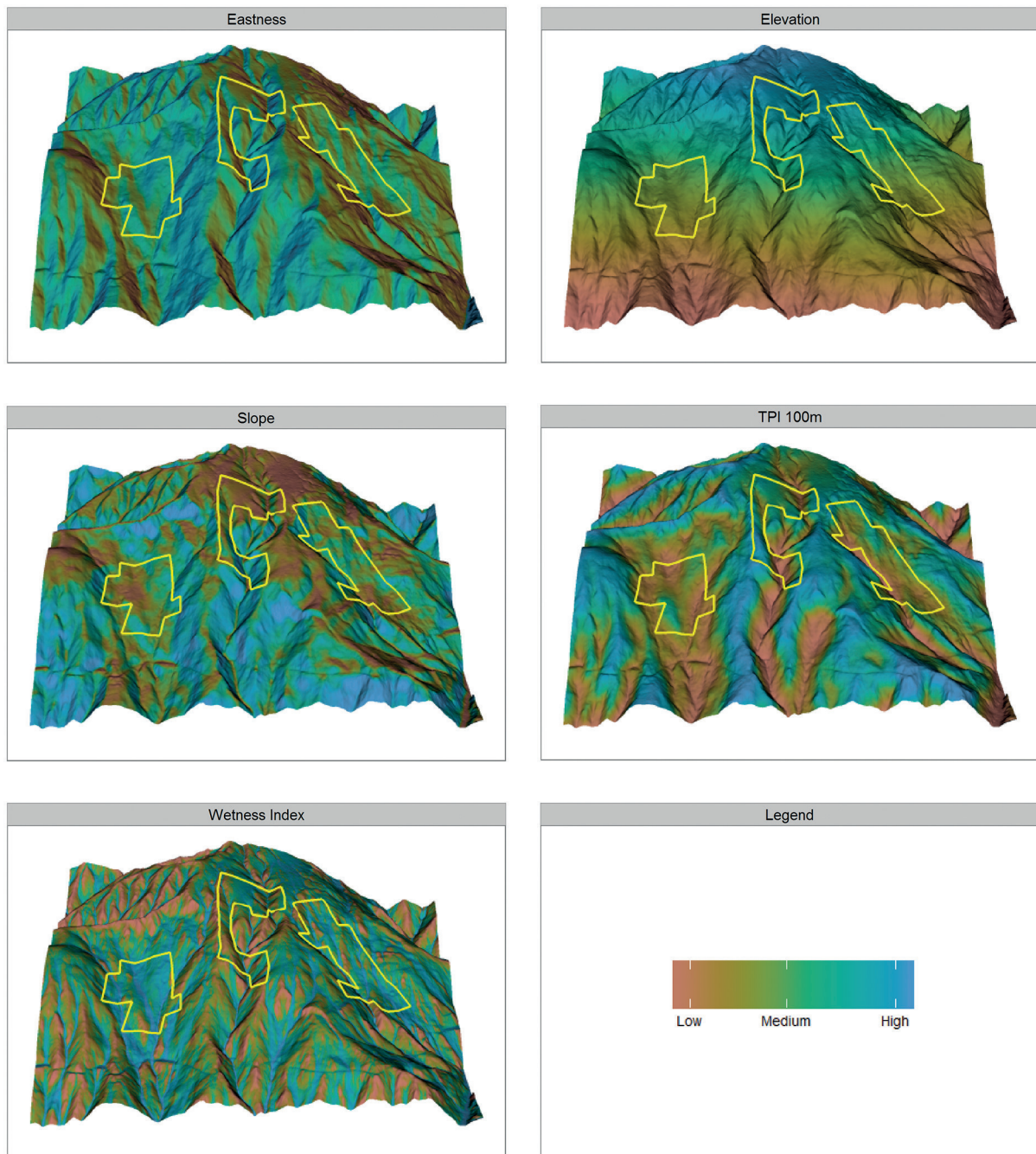


Fig. A-2: Maps of the five topographic variables used for the modelling process. Yellow lines show the position of the permanent sample plots

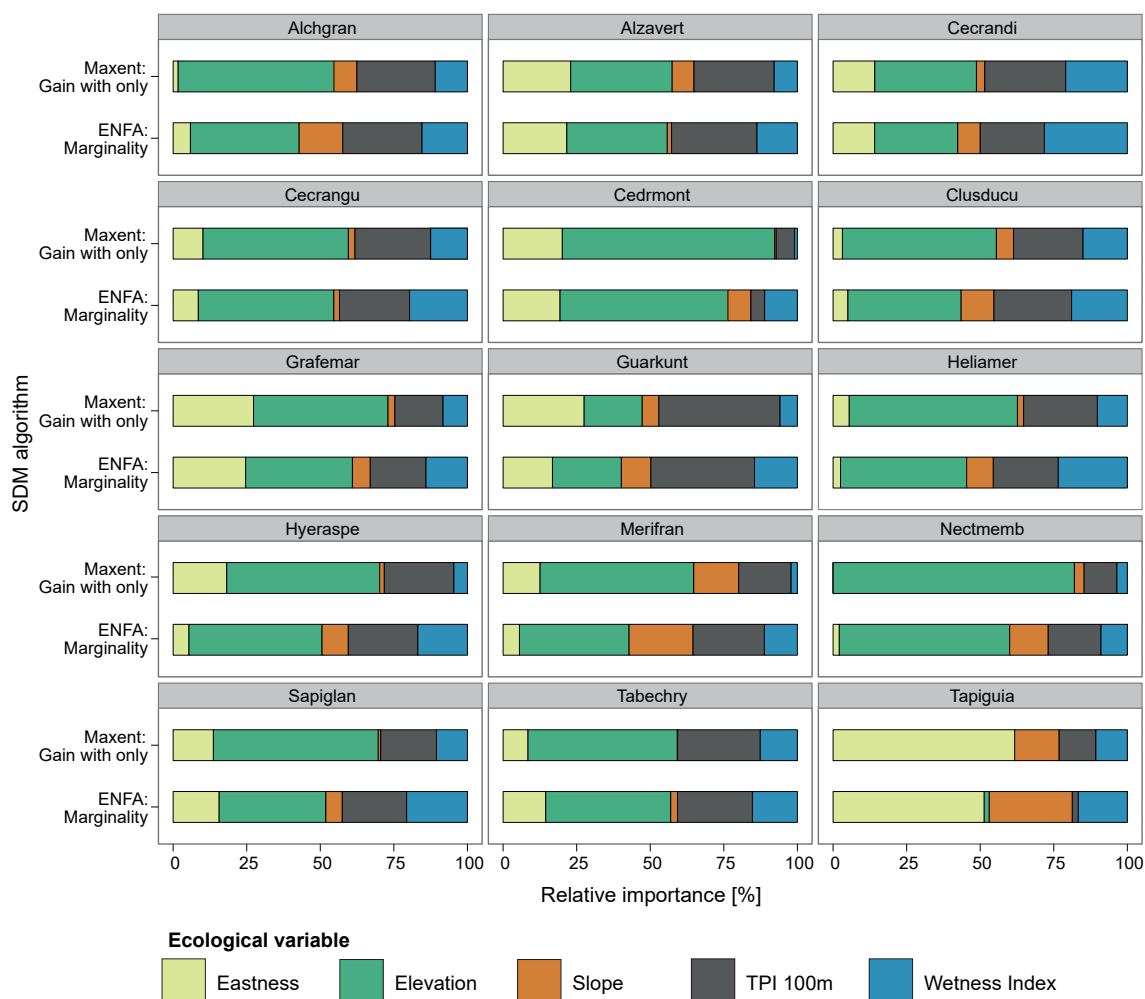


Fig. A-3: Contribution of ecological variables for the distribution of each of the 15 retained species after model evaluation. The upper bar on each panel shows results from the jackknifing test in Maxent, where models were created with only one variable in turn and regularized training gain of these models was subsequently compared with models created with all predictor variables. The lower bar shows absolute values of the marginality factor of the ENFA models. In order to allow a better comparison between Maxent and ENFA models, absolute values for importance were transformed to relative values

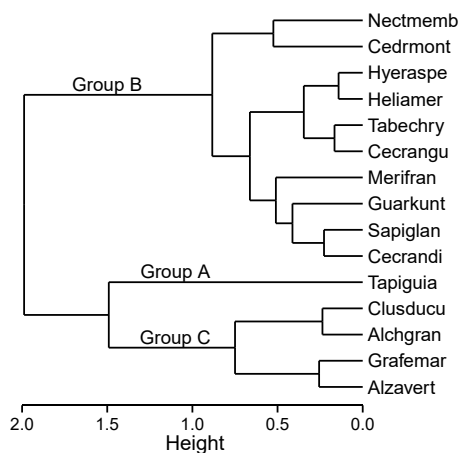


Fig. A-4: Dendrogram of 15 tree species based on the values of the marginality factor of the ENFA for each topographic variable (Tab. A-2). A dissimilarity matrix was calculated using Euclidean distance and subsequently, three groups (Group A, B and C) were identified with the complete linkage method