Borneo : a quantitative analysis of botanical richness, endemicity and floristic regions based on herbarium records

Raes, N.

Citation

Raes, N. (2009, February 11). *Borneo : a quantitative analysis of botanical richness, endemicity and floristic regions based on herbarium records*. Retrieved from https://hdl.handle.net/1887/13470

CHAPTER 5

The floristic regions of Borneo inferred from species distribution models

Niels Raes, E. Emiel van Loon, Marco C. Roos
and Hans ter Steege

Submitted

Abstract

Aim
A quantitative floristic analysis of Borneo using herbarium records to justify the recognition of floristic regions.

Location
Borneo; between approx. 4°S-7°N and 109-119°E.

Methods
From the collection of vascular plants of Borneo stored at the National Herbarium of the Netherlands, we extracted records (N=44,106) belonging to families revised in Flora Malesiana (including revised genera of the Annonaceae, Euphorbiaceae, and Orchidaceae). To limit the influence of collection bias, we modelled each species' distribution using the presence-only modelling technique 'Maxent' in combination with 11 environmental independent variables at 5 arc-minute resolution (ca. 100 km^2 grid cells). The significance of each species distribution model (SDM) was tested against a bias corrected null-model. We constructed a presence/absence matrix based on 1439 significant SDMs (63.3 % of the 2273 species tested) for the 8577 grid cells of Borneo. This matrix was then analysed using a hierarchical cluster analysis, and the resulting cluster dendrogram was pruned using indicator species analysis (ISA) to partition floristic regions. The relationship between floristic regions and environmental conditions was then explored using a classification and regression tree (CART) analysis.

Results
The cluster analysis identified 11 floristic regions for Borneo. Many of the regions overlap with previous classifications based on informal expert opinion. Our analysis indicates, however, that the lowland rain forest region can be divided into at least four distinct floristic sub-regions. The CART analysis identified meaningful ecological thresholds defining each floristic region, largely in accordance with the known ecology of each floristic region.

Main conclusions
Our collections-based analysis identified 11 floristic regions of Borneo. Results largely confirm the floristic distinctiveness and extent of montane rain forest, kerangas, peat swamp, and fresh water swamp forest. The lowland rain forest, previously recognized as one floristic region is divided in at least four (and possibly six) distinct regions, viz. the lowlands of (i) Sabah and Sarawak, (ii) East Kalimantan, (iii) southern Borneo, and (iv) the Wet hill forest of Sarawak. The 'Kinabalu highlands', mangroves, and forests on lime-stone and ultramafic rock could not be distinguished due to the 100 km^2 resolution of our analysis.

Keywords
Phytogeography; Borneo; species distribution model; Maxent; significance test; null-model; bias; cluster analysis; indicator species analysis; classification and regression trees.

Introduction

Borneo is one of the most important biodiversity hotspot of the world (Myers *et al.*, 2000; Ashton, 2005; Kier *et al.*, 2005), and is severely threatened by deforestation and land use change (Curran *et al.*, 2004; Langner *et al.*, 2007; Stibig *et al.*, 2007). Although this is widely recognized, Borneo's internal floristic division, defined as areas that are relatively homogeneous in plant species composition, is still largely based on informal expert opinion. With ongoing global efforts to digitize herbarium records, an increasing amount of data on the spatial distribution of species' occurrences has become available (Graham *et al.*, 2004). This creates new opportunities to analyse quantitatively the floristic structure of biodiversity hotspots such as Borneo, and thereby support conservation efforts.

Between 6-8 floristic regions have previously been recognized for Borneo (van Steenis, 1958b; Whitmore, 1984b; MacKinnon, 1997; Wikramanayake *et al.*, 2002). The first attempt to define the floristic structure of Borneo was made by Van Steenis (1935). For the entire country of Indonesia he recognized nine floristic regions, six of which occurred on Borneo (van Steenis, 1935a). Since then, various additions and improvements to the floristic map of South-East Asia, including Borneo, have been made (Hannibal, 1950; van Steenis, 1958a, b; Whitmore, 1984b; MacKinnon, 1997) (Fig. S5.2). The most recent of these is the WWF ecoregion map of Borneo, described in detail by Wikramanayake *et al.* (2002) (Fig. S5.2), which recognizes seven ecoregions for Borneo, based on the general framework of MacKinnon's (1997) 'bio-units'. WWF departed from MacKinnon's bio-units in three ways: a) ecoregion delineations were more closely based on potential vegetation maps, whereas vertebrates were the dominant influence on MacKinnon's bio-units; b) mixing of natural communities that are characteristic of specific habitat types was avoided; and c) lowland forests were separated from montane forests (Wikramanayake *et al.*, 2002).

The emerging consensus from past efforts is that at least six distinct floristic regions can be distinguished on Borneo: montane rain forest, lowland evergreen rain forest, heath forest, freshwater swamp forest, peat swamp forest, and mangrove forest. Additionally, Whitmore (1984) and MacKinnon (1997) recognize limestone forest formations, which were not mapped by WWF due to uncertainties concerning location and geographical extent (Wikramanayake *et al.*, 2002). WWF also recognizes the 'Kinabalu montane alpine meadows' as an separate ecoregion (Wikramanayake *et al.*, 2002), and the ultramafic outcrops of eastern Sabah as a distinctive flora (WWF & IUCN, 1995). The only quantitative floristic analysis of lowland flora of Borneo based on data of 28 plots, recognized five major floristic lowland regions, i.e. Santubong/Bako cluster of western Sarawak, eastern Sabah, central Sarawak, the Kalimantan provinces, and a northern cluster covering northern Sarawak and Brunei with an extension into northern East Kalimantan (Slik *et al.*, 2003). These results illustrate the potential for quantitative analyses to uncover hitherto unseen patterns of geographic variation, even within ecoregions, but to date such attempts are limited to that of Slik *et al.* (2003).

The volume of digitized herbarium records now make it possible to undertake quantitative analyses of floristic variation across very large spatial scales (Linder *et al.*, 2005). Unfortunately, collection localities are generally not uniformly spatially distributed (Kadmon *et al.*, 2004; Moerman & Estabrook, 2006; Hortal *et al.*, 2007), as was shown for Borneo (Raes & ter Steege, 2007 - Chapter 3). Efforts to overcome problems of geographically biased collection localities in a data set to be used for predicting species presence/absence across areas where no collections have been made has lead to development of a suite of species distribution modelling techniques (Guisan & Zimmermann, 2000; Elith *et al.*, 2006; Peterson, 2006). Species' distribution models (SDMs) predict the potential distribution of species by describing relationships between species' presence/absence-, or presence-only data, and environmental predictors across an area of interest. If meaningful environmental predictors are available at the appropriate resolution, in combination with sufficiently accurate collection localities, SDMs can predict the presence and absence of species across the entire area of investigation at the spatial resolution of the environmental predictors.

Considering the threatened status of Bornean forests, the variety of past efforts and results to describe floristic divisions, and recent advances made in the development of SDM methods, we aimed to: a) construct species presence/absence maps at 5 arc-minute (ca. 100 km^2) resolution for a large number of plant species using SDMs; b) delimit quantitatively the floristic regions of Borneo, by analysing the complete set of significant species presence/absence maps using hierarchical cluster analysis; and c) identify the ecological factors that drive the formation of Borneo's floristic regions.

Materials and Methods

Species data and environmental predictors

We extracted all georeferenced species records from Borneo belonging to the families treated in Flora Malesiana (Anon., 1959-2007) from the BRAHMS database of the National Herbarium of the Netherlands. We added the georeferenced records of revised genera of the Annonaceae, Euphorbiaceae, and Orchidaceae. This dataset comprised 66,262 georeferenced records belonging to 102 plant families. Only species with records in five, or more, grid cells were included. The resulting data set included 2273 species representing 44,106 unique records, ranging from 5 to 202 records per species.

Initially, 37 environmental predictors were selected to model species distributions. We downloaded the digital elevation model (DEM (in m)) and the 19 bioclimatic predictors (~1950-2000) of the WORLDCLIM dataset (<http://www.worldclim.org>) for Borneo at 5 arc-minute (ca. 100 km^2) resolution (Hijmans *et al.*, 2005). Additionally, 15 soil property

Table 5.1. Pearson's correlation r values for the 11 environmental predictors used to model the species' distributions, and the uncorrelated quadratic terms additionally used in the CART. Legend: DEM - Altitude; Bio04 - Temperature seasonality; Bio07 - Temperature annual range; Bio12 - Annual precipitation; Bio15 - Precipitation seasonality; PCA01-05 - PCA soil axes 1-5; ENSO - El Nino drought impact.

	DEM	Bio04	Bio07	Bio12	Bio15	PCA01	PCA02	PCA03	PCA04	PCA05	ENSO	PCA02^2	PCA04^2
Bio04	-0.512												
Bio07	-0.361	**0.737**											
Bio12	0.207	-0.167	0.030										
Bio15	-0.412	0.514	0.507	-0.254									
PCA01	-0.216	0.194	0.142	-0.030	0.185								
PCA02	0.274	-0.078	0.094	0.214	0.040	0.138							
PCA03	0.226	-0.214	-0.202	0.066	-0.168	-0.306	-0.036						
PCA04	-0.064	0.026	0.081	-0.095	0.149	0.034	0.151	-0.018					
PCA05	-0.136	0.033	0.004	-0.062	0.048	0.068	-0.011	0.021	-0.010				
ENSO	0.372	-0.078	-0.099	0.274	-0.390	-0.053	-0.003	0.128	-0.083	0.034			
PCA02^2	0.078	-0.035	-0.023	-0.208	-0.026	-0.028	0.044	0.022	0.364	0.230	0.084		
PCA04^2	-0.085	0.134	0.193	-0.109	0.184	-0.230	0.323	0.285	0.464	0.119	-0.126	0.331	
PCA05^2	-0.062	0.149	0.112	-0.140	0.094	-0.170	0.060	-0.078	0.049	0.112	-0.058	0.334	0.292

Table 5.2. Pearson's correlation coefficients r values for the 15 FAO soil predictors and the five PCA soil axes. Values in italic r>0.4 or r<-0.4 (modest correlation) and in bold italic r>0.7 or r<-0.7 (strong correlation). CEC = Cation Exchange Capacity; C:N = Carbon:Nitrogen.

	PCA01	PCA02	PCA03	PCA04	PCA05
Base saturation % topsoil	-0.1013 ns	___-0.8429___ ***	0.0974 ns	0.2864 ns	0.2268 ns
CEC clay topsoil	_0.5712_ ***	-0.3342 *	_-0.5284_ ***	0.4161 *	-0.1366 ns
CEC soil topsoil	___0.7449___ ***	-0.1708 ns	0.2366 ns	0.0295 ns	0.0806 ns
C:N-ratio class topsoil	_0.5083_ ***	0.3100 ns	_-0.4183_ **	-0.2982 ns	_0.5314_ ***
Easy available water	___-0.7886___ ***	_0.4747_ **	0.0332 ns	-0.1926 ns	0.1242 ns
Effective soil depth	0.2428 ns	-0.3322 *	-0.1498 ns	___-0.8224___ ***	-0.2733 ns
Nitrogen % topsoil	___0.7360___ ***	0.2317 ns	0.3555 *	0.1245 ns	-0.1529 ns
Organic carbon % topsoil	_0.5523_ ***	_0.5221_ ***	0.3205 *	0.0227 ns	-0.2646 ns
Organic carbon pool	___0.7626___ ***	0.3883 *	0.2412 ns	-0.1427 ns	0.1172 ns
pH topsoil	_-0.4389_ **	_-0.6870_ ***	0.1403 ns	0.2410 ns	-0.0953 ns
Soil drainage class	___0.8323___ ***	-0.2111 ns	0.1628 ns	0.1071 ns	0.2241 ns
Soil moisture storage capacity	___-0.7108___ ***	_0.5545_ ***	-0.0116 ns	-0.0222 ns	0.1731 ns
Soil production Index	-0.0489 ns	___-0.8584___ ***	0.0836 ns	-0.1758 ns	-0.2444 ns
Textural class subsoil	-0.2891 ns	-0.0161 ns	___0.8747___ ***	-0.1090 ns	0.1550 ns
Textural class topsoil	-0.0382 ns	-0.1762 ns	___0.9153___ ***	-0.1784 ns	0.1307 ns

* p<0.05; ** p<0.01; *** p<0.001; ns = not significant.

predictors (FAO, 2002) and elevation ranges per grid cell based on the 90m resolution SRTM altitude data (<http://srtm.csi.cgiar.org>) were added. Finally, a data-layer reflecting the El Niño Southern Oscillation (ENSO) drought impact, defined as the relative average annual difference in Normalized Difference Vegetation Index (NDVI) values between months of a severe ENSO (07/1982-06/1983) and non-ENSO years (07/1981-06/1982) was added. We retained only grid cells with data for all environmental predictors, resulting in 8577 grid cells for Borneo. Records on the coastline falling just outside grid cells were shifted to their closest grid cell. Data-layer manipulations were performed with Manifold GIS (Manifold Net Ltd).

The inclusion of 37 environmental predictors introduces potential problems of multicollinearity (Graham, 2003), which results in SDM over-fitting (Peterson *et al.*, 2007). To prevent over-fitting, the number of environmental predictors was reduced. From the DEM, elevation range, ENSO and the 19 bioclimatic predictors, we selected the least correlated variables (Maximum Pearson's r = 0.737): DEM (correlated with and proxy for elevation range, and strongly negatively correlated with the mean annual temperature, maximum temperature warmest month and quarter, minimum temperature coldest month and quarter, mean temperature wettest and driest quarter); Bio04 - Temperature seasonality; Bio07- Temperature annual range (correlated with diurnal temperature range); Bio12 - Annual precipitation (correlated with and proxy for precipitation in the wettest month and quarter, driest month and quarter, warmest and coldest quarter); Bio15 - Precipitation seasonality; and ENSO (Table 5.1). To reduce the number of soil predictors we used a principal component analysis (PCA). The PCA was performed on the 41 unique combinations of the 15 soil predictors values observed for the 8577 grid cells of Borneo, to prevent overweighting of combinations overrepresented in the dataset. We selected the first five PCA-axes as our soil property predictors PCA01-05, together describing 83% of total variance in soil data. Pearson's correlation was used to determine which of the 15 FAO soil predictors were significantly correlated with the five PCA axes (Table 5.2). This reduced environmental predictors from 37 to 11 uncorrelated variables, which were used to construct SDMs (Table 5.1).

Presence/Absence matrix from significant SDMs

To model species distributions, we selected Maxent from the available suite of applications (ver. 3.0.4; <http://www.cs.princeton.edu/~shapire/maxent/>) (Elith *et al.*, 2006; Phillips *et al.*, 2006; Graham *et al.*, 2008). Maxent was set to use all species presence records for model building, by setting the 'random test percentage' to zero (see below). The following modelling rules were used: linear features for <10 records; adding quadratic features for 10-14 records; finally including hinge features for ≥15 records (Raes & ter Steege, 2007 - Chapter 3). For all 2273 species in the dataset, an SDM was developed based on presence records and the 11 environmental predictors.

To test the significance of the SDMs, we used the bias corrected null-model method of Raes & ter Steege (2007). This method tests whether an SDM's AUC value - a threshold independent and prevalence insensitive measure of model accuracy (Fielding & Bell, 1997; McPherson *et al.*, 2004; Raes & ter Steege, 2007 - Chapter 3) - is significantly different from expectations under random chance, taking into account the uneven distribution of collection localities. Advantages of this test are that a) it allows to make use of all presence records for SDM building and testing, b) the test corrects for bias in collection localities, and c) it allows significance testing of the SDM (Raes & ter Steege, 2007 - Chapter 3). One drawback of this methodology is that habitat generalists with a wide distribution will likely not deviate from random chance expectation. It is questionable, however, to what extend these species shape floristic regions. Only species with significant SDMs were retained in subsequent floristic analyses.

Constructing an extrapolated species presence/absence matrix from the continuous Maxent SDMs required setting a threshold for each significant SDM. Although species identifications, and georeferencing of the collection localities, were carried out with the greatest possible accuracy, we assumed that 10% of the records were either wrongly georeferenced or misidentified. Therefore, for all significant SDMs represented by ≥ 10 records, the fixed '10 percentile presence' threshold was used. For those with <10 records, we used either the 'sensitivity-specificity equality' or the 'sum maximization' threshold (Liu *et al.*, 2005), dependent on which of the two corresponding omission rate values was closest to 10%. This procedure allowed us to develop the presence/absence matrix of all species with a significant SDM for the 8577 grid cells covering Borneo.

Cluster analysis and delineation of the floristic regions

We selected a hierarchical cluster analysis to group the data in floristic regions with Sørensen's index as a distance measure (Proches, 2005) in combination with the 'flexible beta linkage method' (ß = -0.25). This combination is recommended because it is space-conserving, thereby avoiding distortion, and it has the least propensity to 'chain' (McCune & Grace, 2002; Perrin *et al.*, 2006), i.e. the tendency to link entities together through a series of intermediates to form large clusters and thereby fail to find distinct cluster groups.

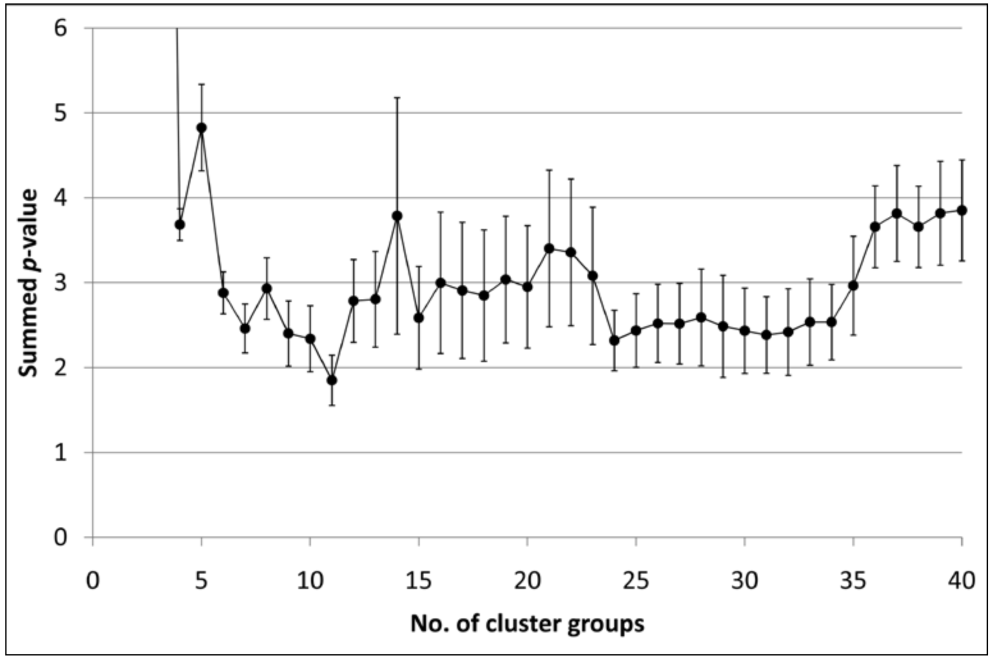

Figure 5.1. Average summed *p*-values (± SD) of the ISA for 2-40 cluster groups based on 5 x 1000 random sub-samples from the total of 8577 grid cells of Borneo.

An objective methodology to identify the ecologically most meaningful point to prune the cluster dendrogram in order to find the optimal number of final cluster groups is the indicator species analysis (ISA) (Dufrêne & Legendre, 1997) sensu McCune and Grace (2002). Although ISA is mostly applied to quantitative species data, it can also be used on presence/absence data. When applied to presence/absence data, the first part of the indicator value (IndVal) index is modified to become the ratio of the number of species presences in a sample group to the total number of species presences. The IndVal index is maximum when a species occupies all the sites of a single cluster group only (Dufrêne & Legendre, 1997). At any given level of clustering (here 2-40 cluster groups), a species is assigned to the cluster group for which its IndVal index is maximal. The significance of the assignment to a cluster group is determined with a Monte Carlo test using 1000 randomisations (Dufrêne & Legendre, 1997; Perrin *et al.*, 2006). Finally, as an objective criterion to select the optimum number of cluster groups, we used the summed ***p***-value over all species for 2-40 cluster groups. The number of cluster groups with the lowest summed *p*-value, is the level for which species are most indicative of the group to which they were assigned, hence the pruning point of the cluster dendrogram.

Following this procedure, randomization of 8577 spatially continuous grid cells with species presence-absence data derived from overlapping SDMs resulted in maximum significance at very low cluster levels (data not shown). We therefore randomly sub-sampled 1000 grid cells from the 8577 cells of Borneo, and repeated the analysis five times. For each of these five sub-samples, we calculated the summed *p*-values for 2-40 cluster groups using ISA. The number of cluster groups with lowest average summed *p*-value out of five sub-samples was taken as the most informative level of clustering, hence the pruning point to delimit the floristic regions. The indicator value (IndVal) for each species was calculated based on all 8577 grid cells to identify the most informative species per cluster group. Cluster analyses and ISA were performed in PC-Ord 5.0.

Relating floristic regions to environmental factors
Classification and Regression Trees (CART) were used as a classifier to relate environmental characteristics to resulting floristic regions derived from the cluster analysis. CART is known as a robust classifier, providing results that are relatively easy to interpret in the context of SDMs (Breiman *et al.*, 1984; Bell, 1999).

We supplied the CART algorithm with the 11 environmental variables and their uncorrelated quadratic terms to account for non-linear effects, which was the case for PCA02^2, PCA04^2 and PCA05^2. The range of values for all variables is given in the caption of Fig. 5.3. We used 90% of the data to derive the tree, and the remaining 10% to evaluate its predictive performance, measured by the error rate (Duda *et al.*, 2001). The cross-validation cycle of using 90% to derive a tree, and 10% to determine performance, was repeated 100 times. The full decision trees, as well as all pruned versions, were evaluated in this way, and the best model resulting from this process was retained. A given floristic region (cluster group) may appear on multiple terminal nodes of the CART, and each node thus indicates alternative environmental conditions under which that floristic region occurs (Urban *et al.*, 2002).

For each terminal node of the CART tree we assessed three aspects: 1) the percentage of the total surface of Borneo identified by the corresponding CART node; 2) the correct classification rate; and 3) the percentage of the surface area within a cluster group correctly classified by the corresponding CART node. For visual inspection, the geographical extent of each floristic region was superimposed on the areas of the corresponding terminal nodes of the CART (Urban *et al.*, 2002). This was performed for each floristic region separately. The CART analysis was conducted in MATLAB, using the function 'classregtree' from the statistics toolbox.

Results

From the 2273 modelled species, 1439 (63.3%) had a distribution pattern that differed significantly from random, and their maps were used to construct the extrapolated presence/absence matrix consisting of 1439 species x 8577 grid cells (Table S5.1).

The ISA for 2-40 cluster groups indicated that an 11-cluster group stage was the optimal pruning point of the hierarchical cluster dendrogram, as at this level the lowest average summed *p*-value was found (Fig. 5.1). The pruned dendrogram and the geographical presentation of cluster groups, representing the floristic regions, are shown in Fig. 5.2. Among these 11 groups, four 'meta-cluster' groups can be recognized: a) the central mountains covering 23.8% of Borneo's surface (cluster groups 1, 2 and 3; Fig. 5.2 and Table 5.3, '% Cluster'); b) the kerangas (4) and peat swamps (5), together covering 14.1%; c) lowland forests of South-, Central- and West Kalimantan, covering 29.9% (6, 7 and 8); and d) lowland forest of East Kalimantan (11), Sabah

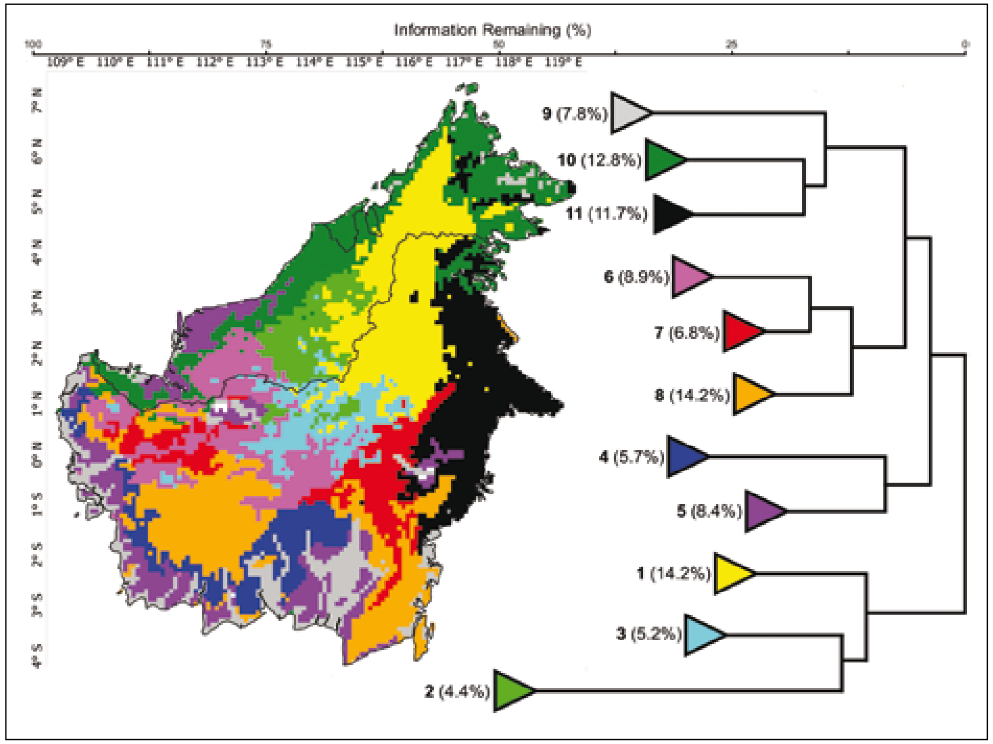

Figure 5.2. Results of the hierarchical cluster analysis representing the 11 cluster groups and their geographical distribution, indicating the different floristic regions of Borneo. Between brackets the percentage of the surface of Borneo covered by the corresponding floristic regions.

& Sarawak (10) and the fresh water swamp forests (9), together covering 32.3%. The IndVal for the 1439 significant SDMs based on the ISA of all 8577 grid cells are given in Table S5.1.

The CART, derived by selecting the tree with the smallest cross-validation error (0.076), had an overall correct classification rate of 93.1% of the cluster groups. Cluster group 3 was the less well classified, with 81.7% correct (Table 5.3). This group corresponds to the southern extrusions of the central mountain chain (Fig. 5.2; Fig. S5.1).

Generally, one terminal node -of all CART nodes classifying a given floristic region- correctly classified ≥ 50% of that region. These are given in italics in Table 5.3 and are indicated by arrows in Fig. 5.3. Floristic region 7 was an exception, with CART nodes 7.1 and 7.2 correctly classifying 33% and 40%, respectively. Most terminal nodes of the CART had a high percentage of classified cells falling within cluster groups ('% CART in Cluster'; Table 5.3). All variables were used in the CART, except PCA05^2. The geographic representation of each terminal CART node for each floristic region is given in Fig. S5.1, along with a summation.

Discussion

Our findings demonstrate that SDMs can be used to delineate floristic regions quantitatively.

The resulting floristic map of Borneo differs most notably from previous studies by distinguishing 11 floristic regions (Fig. 5.2), compared to the six to eight regions (Fig. S5.2) (van Steenis, 1958b; Whitmore, 1984b; Wikramanayake *et al.*, 2002). The 11 floristic regions were identified using a hierarchical cluster analysis on the presence/absence matrix derived from maps of 1439 significant SDMs. This is 10% of the extrapolated 14,423 plant species expected to occur on Borneo (Roos *et al.*, 2004), including trees, lianas, shrubs and herbs.

Some of the floristic regions correspond closely to those previously recognized, but others differ in important ways, discussed below (Fig. S5.2). Most notably, the WWF defined 'Borneo lowland rain forest' ecoregion, can be divided into at least four different floristic 'sub'-regions (see also Slik *et al.* 2003). The CART correctly classified 93.1% of the grid cells based on 13 environmental variables. The most important ecological thresholds that result in the correct classification of the corresponding floristic regions are discussed per region below.

Issues concerning possible circularity of reasoning

Use of the same environmental predictors to develop the SDMs and to explain the geographical extent of the floristic regions might seem to pose a risk of circular reasoning. We argue, however, that this is not true, because Maxent identifies correlations with environmental predictors independently for each of the 2273 species. Even if for two species the same environmental predictors are used by Maxent to predict their distribution, these two species can still have other optima in their response to these predictors, hence a different spatial distribution. We therefore argue that the 1439 significant SDMs used to develop the presence/absence matrix are essentially independent (within the limited space of 11 environmental variables).

The resulting classification into 11 different floristic regions based on a cluster analysis, does not consider the environmental predictors. The reason species are predicted to co-exist in the same floristic region is not that they all occupy the same ecological niche, but rather that different niche requirements are adequately met within the same floristic region. Additionally, although the cluster tree is pruned by the optimized ISA, and cells within a floristic region are characterized by the co-occurrence of many of the same species, this does not necessarily mean that the occurrence of a given species is fully restricted to one floristic region (see Table S5.1). The CART, in its turn, identified the relevant ecological thresholds to classify the 11 floristic regions. This is an analysis at community level, derived from the cluster analysis and based on individual SDMs with their own independent response to the environmental predictors.

Results of the CART also indicate that species in the same floristic region do not all share the same ecological niche. The majority (≥ 50%) of each floristic region is classified by one CART node, generally located at the geographical centre of the floristic regions (Figs 5.2 and S5.1). At the boundaries of each floristic region, areas are found that represent CART nodes that classify smaller percentages of the floristic regions (Fig. S5.1). The smaller CART areas belonging to neighbouring floristic regions often group on the same branches of the CART tree (e.g. 1.3, 2.2, and 3.5; Fig. 5.3). These areas possibly represent areas where floristic regions are less strictly separated at the scale of our analysis, and in fact can be regarded as gradients. Where the cluster analysis puts hard boundaries, the CART indicates that edges are fuzzy.

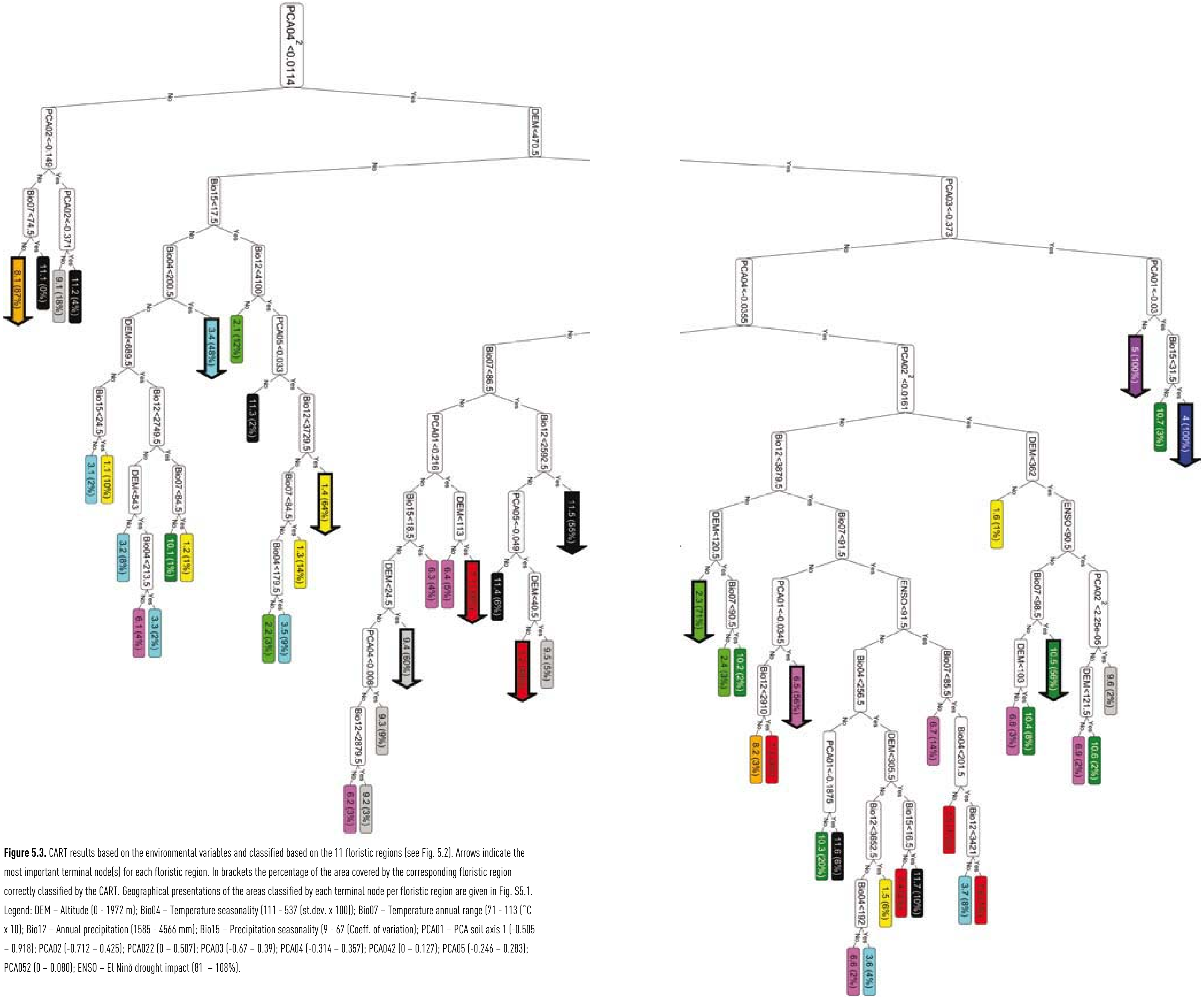

Figure 5.3. CART results based on the environmental variables and classified based on the 11 floristic regions (see Fig. 5.2). Arrows indicate the most important terminal node(s) for each floristic region. In brackets the percentage of the area covered by the corresponding floristic region correctly classified by the CART. Geographical presentations of the areas classified by each terminal node per floristic region are given in Fig. S5.1. Legend: DEM – Altitude (0 - 1972 m); Bio04 – Temperature seasonality (111 - 537 (st.dev. x 100)); Bio07 – Temperature annual range (71 - 113 (°C x 10); Bio12 – Annual precipitation (1585 - 4566 mm); Bio15 – Precipitation seasonality (9 - 67 (Coeff. of variation); PCA01 – PCA soil axis 1 (-0.505 – 0.918); PCA02 (-0.712 – 0.425); PCA022 (0 – 0.507); PCA03 (-0.67 – 0.39); PCA04 (-0.314 – 0.357); PCA042 (0 – 0.127); PCA05 (-0.246 – 0.283); PCA052 (0 – 0.080); ENSO – El Ninõ drought impact (81 – 108%).

Group	% Cluster	% CART	% correct classified	% CART in Cluster
01	**14.2**	**15.0**	**96.3**	**91.0**
01.1		1.4	9.8	99.2
01.2		0.1	0.7	100.0
01.3		2.4	14.0	84.2
01.4		9.5	64.4	96.1
01.5		1.4	5.9	61.5
01.6		0.3	1.4	73.9
02	**4.4**	**4.4**	**89.2**	**89.4**
02.1		0.6	11.6	91.7
02.2		0.2	3.4	81.3
02.3		3.5	70.7	89.0
02.4		0.2	3.4	100.0
03	**5.2**	**4.7**	**81.7**	**91.0**
03.1		0.2	2.5	84.6
03.2		0.5	8.1	87.8
03.3		0.1	2.5	100.0
03.4		2.6	47.7	95.9
03.5		0.6	8.5	79.2
03.6		0.3	4.5	87.0
03.7		0.5	8.1	83.7
04	**5.7**	**5.7**	**99.8**	**100.0**
05	**8.4**	**8.7**	**100.0**	**95.7**
06	**8.9**	**9.4**	**92.5**	**87.2**
06.1		0.4	3.7	82.4
06.2		0.3	3.4	86.7
06.3		0.4	4.3	86.8
06.4		0.5	4.7	76.6
06.5		5.4	55.8	92.2
06.6		0.3	1.7	56.5
06.7		1.4	14.3	90.1
06.8		0.3	2.9	88.0
06.9		0.3	1.6	42.9
07	**6.8**	**7.0**	**90.5**	**87.7**
07.1		2.5	33.0	88.1
07.2		3.0	40.0	90.3
07.3		0.2	3.1	100.0
07.4		0.2	1.7	76.9
07.5		1.0	11.3	80.5
07.6		0.1	1.4	66.7
08	**14.2**	**13.1**	**90.5**	**98.6**
08.1		12.6	87.1	98.5
08.2		0.5	3.4	100.0
09	**7.8**	**7.6**	**97.0**	**98.8**
09.1		1.4	17.6	100.0
09.2		0.2	2.7	100.0
09.3		0.7	8.7	100.0
09.4		4.7	60.5	100.0
09.5		0.5	5.1	81.0
09.6		0.2	2.4	100.0
10	**12.8**	**13.4**	**92.9**	**88.4**
10.1		0.2	1.0	61.1
10.2		0.3	2.4	89.7
10.3		2.9	20.3	90.6
10.4		1.1	8.1	93.7
10.5		7.5	56.2	95.1
10.6		0.9	1.6	21.8
10.7		0.4	3.3	97.3
11	**11.7**	**11.0**	**90.4**	**96.5**
11.1		0.0	0.0	0.0
11.2		0.5	4.7	100.0
11.3		0.3	2.4	92.3
11.4		0.8	6.7	100.0
11.5		7.1	59.7	97.9
11.6		0.7	6.2	100.0
11.7		1.5	10.8	87.2

Table 5.3. Results of the cluster analysis and the CART. '% Cluster' gives the percentage of the surface of Borneo covered by each cluster group. '% CART' gives the percentage of the surface of Borneo covered by each CART group, and for all terminal nodes separately. '% correct classified' gives the percentages of the Cluster area correctly classified by the CART group, and the percentages for all terminal nodes separately. '% CART in Cluster' gives the percentage of the surface of Borneo belonging to one CART group falling within the surface indicated by the cluster analysis, and the percentages for all terminal nodes separately.

The 11 floristic regions of Borneo

1. Montane rain forest

This largest floristic region (14.2%) covers almost the entire central mountain range, including the Crocker range and the high mountains of east Sabah, but excluding the Müller mountains in the south, which belong to region 3 (Figs 5.2 and S5.2). This region is recognized by most previous floristic classifications (Fig. S5.2). Our analysis indicates, however, that the extent is much larger than shown by Whitmore (1984) and WWF (Wikramanayake *et al.*, 2002). WWF used the 1,000 m elevation contour to delimit the lower bounds of the 'Borneo montane rain forests' ecoregion, whereas our data suggest that this floristic region extends down to 224m. One partial explanation for the lower boundary in our method is that 224m is the average altitude for a ca. 100 km^2 grid cell, but maximum altitude is much higher.

Associated areas of this region are found in valleys between the northern Crocker range, with Mt. Kinabalu, and the central mountain range and the high mountains of Sabah. Our analysis does not recognize the WWF 'Kinabalu montane alpine meadows' ecoregion as a distinct floristic region. This likely reflects the fact that in order to be modelled species had to be represented in at least five unique grid cells, whereas the 'Kinabalu montane alpine meadows' as defined by WWF occurs above 2,600 m has a surface area of only 54 km^2. Considering the spatial resolution of our analysis (ca. 100 km^2), it is not surprising that Mt. Kinabalu was not distinguished. Furthermore, the total area of the 'Kinabalu montane alpine meadows' ecoregion as reported by WWF is 4,300 km^2, yet this vastly overstates the area above the lower limit of 2600m, suggesting that more than ca. 90% of the 'Kinabalu' ecoregion should, in fact, be assigned to the 'Borneo montane rain forests' ecoregion, as we have done. We do recognize, however, that Mt. Kinabalu is a separate floristic region and discuss this further below under point '12 – Miscellaneous'.

The majority of this ecoregion (64%) is characterized by areas with an altitude above 470 m, low precipitation seasonality, intermediate to low soil C:N-ratio, and less than 3,730 mm precipitation per year (Figs 5.3 - node 1.4, and S5.1 - Group 1). The intermediate to low C:N-ratio seems to contradict with the findings that montane soils have high C:N-ratio's compared to lowland (Soethe *et al.*, 2008). The low elevation areas of Borneo harbour many peat swamps, however, which have an even higher C:N-ratio. The CART indicates that the Crocker Mountains along with Mt. Kinabalu (Fig. 5.3 - node 1.1) are characterized by different ecological conditions. The precipitation seasonality is higher than most of the region and altitude is above 690m. Similar conditions are found on Gunung Mejapa, Kong Kemoel, and the northern parts of Müller Mountains in East Kalimantan (see Fig. S5.2 A).

Many of the species with their maximum IndVal for the 'Montane rain forest' floristic region (Table S5.1) belong to plant families typical of the montane rain forest, such as Clethraceae, Cunoniaceae, Ericaceae, Fagaceae, Nepenthaceae, Orchidaceae, Podocarpaceae and Symplocaceae (Whitmore, 1984a; Wikramanayake et al., 2002).

2. Wet hill forest of Sarawak

The 'Wet hill forest of Sarawak' (Fig. 5.2 - cluster 2), located south-west from the 'Montane rain forest' region (1), was not recognized as separate floristic region by WWF, MacKinnon (1997), Whitmore (1984b) or Van Steenis (1958). This region was distinguished, however, by Slik *et al.* (2003), who classified it as the 'central Sarawak' cluster. Furthermore, the Asian Regional Centre for Biodiversity Conservation (ARCBC: <http://www.arcbc.org.ph/wetlands/>) classified this region as 'The upper basin of the Baram, Tinjar and Rajang Rivers' wetland area.

The largest part of this region (Fig. 5.3 and Table 5.3 - node 2.3, 71 %) is characterized by an altitude between 120-471 m (Fig. 5.3), shallow soils (Table 5.2 - PCA04) with an intermediate to finely structured top- and subsoil (Table 5.2 - PCA03), low base saturation and soil productivity index (Table 5.2 - PCA02^2), and annual precipitation exceeding 3,880 mm yr^{-1}, hence the classification 'wet'. Few species are fully restricted to this region, and many occur in the other two montane floristic regions (1 & 3), and in the 'Lowland rain forest of Sabah and Sarawak' (10) region (table S5.1). Species with a maximum IndVal for this region belong to montane plant families such as Rosaceae, Ericaceae, and Fagaceae; but also include species typical of lowland rain forest families such as Dipterocarpaceae, Euphorbiaceae and Annonaceae. This region can be regarded as a transition between the true montane rain forest and lowland rain forest, characterized by very wet conditions.

3. Montane rain forest of the Müller- and Upper Kapuas Mountains (incl. Gunung Saran and Gunung Niut)

The 'Montane rain forest of the Müller- and Upper Kapuas Mountains' floristic region can be regarded as an extension plus outliers of the 'Montane rain forest' region (1). The cluster analysis indicates that this region is most closely related to the 'Montane rain forest' (Fig. 5.2). The classification as a separate region largely relates to the absence of many species present in region 1 (Table S5.1). The absent species are those with the lower maximum IndVal's for region 1, meaning that many species characteristic of the 'Montane rain forest' region also occur in region 3 (Table S5.1). Only nine species had their maximum IndVal for this floristic region, largely belonging to montane plant families.

The absence of the species present in region1 may reflect the much lower collection density on southern Borneo compared to northern parts of the mountain range. When a species is modelled with Maxent on data not fully covering a species' ecological niche due to lack of collections, Maxent will not predict the occurrence of that species under those conditions. In this case, that species may be predicted absent from southern parts of the mountain chain. The opposite is probably the case for the Schwaner Mountains, with Bukit Baka and Raja, which are excluded from floristic region 3, whilst Gunung Saran and Gunung Niut are included. The Schwaner Mountains were very extensively sampled by an expedition led by H. Nooteboom in 1982-1983 including many lowland species. Considering the size of the mountains compared to the resolution of our analysis this probably resulted in the clustering of the Schwaner Mountains with lowland region 6, whilst it more likely belongs with region 3.

Five of the seven CART nodes are found on the same branches as the nodes of region 1, mainly differing in a higher precipitation seasonality and lower temperature seasonality (Fig. 5.3 - node 3.4, 48 %). This further supports the interpretation that this region should be included in the 'montane rain forest' region (1).

4. Kerangas or heath forest

The kerangas floristic region, according our study, is found in central and west Kalimantan, often located between lowland floristic regions 6, 7 and 8, and the peat swamp region (5). The kerangas, together with the peat swamp forest, are the only two regions characterized by one CART node (Figs 5.3 and S5.1) that correctly classified the entire floristic region (Table 5.3). Both kerangas and peat swamps have small values for variable PCA03, which is related to coarse textured, organic soils (Table 5.2). Kerangas differ from peat swamps in that they have lower values for PCA01, which is related to a low cation exchange capacities (CEC) and nitrogen concentration of topsoil, a small organic carbon pool, well-drained soils with a low moisture storage capacity, and little easily available water. These results accord well with descriptions of both Mackinnon *et al.* (1996) and Whitmore (1984a) that heath forests are found on 'white sand' soils derived from siliceous parent materials which are inherently poor in bases (related to the low CEC), highly acidic, commonly coarsely textured, free-drained and often covered with a thin layer of peat or humus.

Contrary to the accordance in the description of the soils on which kerangas occur, is the extent of the range of this floristic region. Generally, all authors (Whitmore, 1984b; MacKinnon *et al.*, 1996; MacKinnon, 1997; Wikramanayake *et al.*, 2002) recognize the large kerangas area in Central Kalimantan (Fig. 5.2). Additionally, they recognize smaller areas around the 'Mahakam' lakes area of East Kalimantan and northeast hereof, and around the 'Danau Sentarum' lakes of West Kalimantan. Our analysis does not classify these areas as kerangas, but as lowland forest. Again, the scale of our analysis might play a role here. Although kerangas are generally strikingly different in flora, structure and physiognomy from other forest types, under the most favourable conditions there is considerable similarity with evergreen rain forest, with dipterocarps prominent among the larger trees, and a canopy height of 27-31m (Whitmore, 1984a; Newbery, 1991; MacKinnon *et al.*, 1996). Moreover, kerangas often form part of a mosaic with other forest formations (MacKinnon *et al.*, 1996). The presence of lowland species has likely resulted in clustering of these cells with lowland floristic regions. For the same reasons our analysis does not recognize the small patches of kerangas in north-east Sarawak, indicated by Whitmore (1984b) and studied by Newbery (1991), nor the areas fringing the coast in north-east East Kalimantan indicated by WWF (Wikramanayake *et al.*, 2002) and MacKinnon (1997).

These findings are supported by the ISA (Table S5.1), which shows that many species occurring in kerangas are also found in the lowland regions. The large area indicated by MacKinnon (1997) west of the 'Danau Sentarum' lakes area in West Kalimantan was, besides for some small patches, not recognized by other authors, and is classified in our analysis as lowland rain forest region 7 (see below). Unfortunately, this region of Borneo is largely deforested (Whitmore, 1984b; Stibig *et al.*, 2007). Whether this region should be classified as kerangas therefore remains unresolved. Our study does predict additional kerangas areas inwards along the coast line of West Kalimantan. Considering the very low collection density in the area (Raes & ter Steege, 2007 - Chapter 3), this warrants further research.

5. Peat swamp forest

The 'Peat swamp forest' floristic region 5 is found along the coast of southern Borneo, between the kerangas (4) and freshwater swamp forests (9), along large parts of the coast of Sarawak, and surrounding the 'Mahakam'- and 'Danau Sentarum' lakes. Besides these areas, peat swamp forests are also found along the coast of Brunei and northeast East Kalimantan (Whitmore, 1984b; MacKinnon *et al.*, 1996; Wikramanayake *et al.*, 2002), not recognized by our study. Peat swamp forest share many species with kerangas, as well as with lowland forests (Whitmore, 1984a). For reasons similar to those provided for why some kerangas areas were not identified as such in our study, the peat swamp areas of Brunei and East Kalimantan also were not recognized. Furthermore, these two areas are located close to the northern Borneo centre of biodiversity (Ashton, 2005; Kier *et al.*, 2005), because of which our models predict the presence of many lowland species in the swamp- and lowland rain forest mosaic, hence the clustering with the lowland rain forest floristic regions.

The CART (Fig. 5.3 -node 5) characterized the occurrence of peat swamp forest by a coarse textured top- and subsoil, a high organic carbon pool, a low pH, very poorly drained, and a low soil moisture storage capacity; all in accordance with the descriptions of other authors (Whitmore, 1984a; MacKinnon *et al.*, 1996).

Many species found to be characteristic for peat swamp forests had their maximum IndVal for this floristic region, such as *Shorea albida, Copaifera palustris, Gonystylus bancanus, Combretocarpus rotundatus*; and *Dactylocladus stenostachys* and *Campnosperma coriaceum* with a maximum IndVal for lowland rain forest region 10, but also frequently found in the peat swamps (Table S5.1) (Whitmore, 1984a; Wikramanayake *et al.*, 2002).

6, 7, 8: Lowland rain forests of southern Borneo

The delineation of the lowland rain forest floristic regions 6, 7 and 8 are the least clear-cut. The clustering of these three regions in one meta-cluster group (Fig. 5.2) is supported by the plot-based inventory study of lowland dipterocarp forest of Borneo by Slik *et al.* (2003) that classified the entire region as the 'southern cluster'. Other researchers classified these regions together with our regions 10 and 11 as lowland rain forest (Whitmore, 1984b; MacKinnon, 1997; Wikramanayake *et al.*, 2002) (Fig. S5.2). Cluster groups 6, 7 and 8 are mainly characterized by the absence of species which have their highest IndVal (Table S5.1) for lowland rain forest regions 10 and 11, and the wet hill forest of Sarawak (2) . These findings could support the suggested presence of the savanna corridor at the last glacial maximum (LGM ~21,000 BP) (Heaney, 1991; Gathorne-Hardy *et al.*, 2002; Bird *et al.*, 2005), when obligate rain forest species were retracted to refugia, probably present in northern and eastern Borneo (Ashton, 1992; Gathorne-Hardy *et al.*, 2002). For many species occurring in the Australian Wet Tropics (Graham *et al.*, 2006), and Europe (Svenning *et al.*, 2008), it was shown that they still have not occupied their potential range, and many are restricted to their Pleistocene refugia even today. The same may account for the lowland species of Borneo. An alternative explanation is the much lower collection density for southern Borneo compared to the north, as was explained for region 3.

Region 6

The majority of region 6 (Figs 5.3 - node 6.5, 56% and S5.1) is closely related to the 'Wet hill forest of Sarawak' (2), differing in that it receives less than 3,880 mm precipitation yr^{-1}. On the same branches of the CART tree is the most important node characterizing region 10 (10.5) found, which is different from regions 6 and 2 in that it has intermediate values for PCA02 (Table 5.2), related to base saturation, organic carbon content, pH, soil moisture storage capacity, and the soil production index. The ISA (Table S5.1) indicates that indeed many species are found in all three floristic regions. Only nine species had their maximum IndVal for region 6, and all these species had part of their distribution in other floristic regions as well.

Region 7

This region is distributed over two large geographically disjunct regions (a property shared with region 8). It is the only region, however, that is characterized by two equally imported CART nodes (Fig. 5.3 - node 7.1 and 7.2). The eastern part, node 7.2 is ecologically most closely related to the 'Lowland rain forest of East Kalimantan' (11) (Fig. 5.3 - node 11.5), different in that it receives more than 2,593 mm precipitation yr^{-1}. The western part, node 7.1, is ecologically most similar to the 'Fresh water swamp forest' (9) different in that it has lower values for PCA01 (Fig. 5.3 - node 9.4). Low PCA01 values were also the condition separating kerangas from peat swamps. This could indicate that this region has similarities with the kerangas, possibly explaining why MacKinnon (1997) characterized the area as such (Fig. S5.2 C). The threshold for PCA01<0.216 is higher than for the kerangas (PCA01<-0.03), however. This western region might be a distinct floristic unit with elements of 'Kerangas' (4) and 'Fresh water swamp forest' (9); some isolated kerangas elements on sandstone cuesta formations have been reported (G. Paoli, pers. comm.). Unfortunately, only very small patches remain forested today (Stibig *et al.*, 2007).

Region 8

Although MacKinnon (1997) characterized the region as lowland rain forest, he did recognize the 'Meratus Mountains of South Kalimantan' and the 'southern lowland planes' as separate biogeographic units (Fig. S5.2 C - 25i and 25b, respectively). The CART indicates that region 8 is characterized by partly shallow, and partly deep soils (Fig. 5.3 - PCA04^2>0.0114), low in base saturation, soil production index and pH (Fig. 5.3 - PCA02<-0.149). Although the floristic similarity of the two parts of region 8 is supported by similarities in our soil data (FAO, 2002), there are indications that the geological origin of both areas is different. The western part of the region 8, located in Central Kalimantan, consists of 'Schwaner Cretaceous granites & tonalites' (Hall *et al.*, 2008), and was characterized by MacKinnon (1996) as floristically relatively poor based on lithology. The eastern 'Meratus Mountains' area of region 8, together with the southern extension of the eastern part of region 7, is believed to be a cretaceous accretionary complex (Wakita & Metcalfe, 2005), characterized by Hall *et al.* (2008) as a mixture of the 'Meratus Cretaceous volcanic arc & ophiolite' and 'Ophiolite basic/ultrabasic rocks'. Whether these recent discoveries are correctly reflected in our soil property variables (Table 5.2) is not known.

Like for region 6 and 7, the ISA indicated that few species had their maximum IndVal for region 8 (Table S5.1). That the Meratus Mountains are not recognized as a distinct floristic region might be for reasons similar to those offered for Mt. Kinabalu, namely study resolution and low number of species with a minimum of five presence records (see region 1, above). According to MacKinnon (1996), the Meratus Mountains are floristically distinct and especially rich in orchids.

9. Fresh water swamp forest

The 'Fresh water swamp forest' is, not surprisingly, found in the same cluster as the lowland rain forest regions 10 and 11 (Fig. 5.2), since this floristic region has many species in common with the surrounding lowland rain forest (Whitmore, 1984a; Wikramanayake *et al.*, 2002), as can be concluded for the ISA (Table S5.1). Many species belonging to the families of the Cyperaceae and Rhizophoraceae have their maximum IndVal for this floristic region. The occurrence of the rain- and river fed swamp forest up to the coast has resulted in the lumping of the 'Fresh water swamp forest with the mangrove forest. The areas identified as 'Fresh water swamp forest' are largely in accordance with the findings of other authors (Fig. S5.2). The only areas not recognized by our study are located in northern East Kalimantan.

The CART (Fig. 5.3) indicates that the most important node 9.4, classifying the fresh water swamps, is found on the same branch system as the node 11.5 classifying the 'Lowland rain forest of East Kalimantan' which supports the floristic similarities between the regions. Node 9.4 is related to high CEC's, high nitrogen percentage, and poorly drained soils with high soil moisture capacity (Table 5.2 - PCA01>0.216), all in accordance with the descriptions of Whitmore (1984a).

10. Lowland rain forest of Sabah and Sarawak

Contrary to previous floristic analyses of Borneo (Whitmore, 1984b; MacKinnon, 1997; Wikramanayake *et al.*, 2002) that classified the Borneo lowlands as one region, our analysis indicates that the 'Lowland rain forest of Sabah and Sarawak' are distinct from lowland regions 6, 7 and 8, and 11 (Fig. 5.2). These results are supported by Ashton (1992), who labelled the western, and north-western parts of Sarawak as areas of exceptional richness and endemicity, the so-called Riau-pocket (Corner, 1960; Ashton, 1992). The north-eastern part of region 10 was identified by both Slik *et al.* (2003) and Ashton (1992) as the separate 'East coastal Sabah sub-province', however. Our study shows that the north-eastern part of Borneo is a mosaic of lowland rain forest regions 10 and 11, fresh water swamp forest, and montane rain forests, and can be described as an area where the 'Lowland rain forest of Sarawak and Sabah' and the 'Lowland rain forest of East Kalimantan' mix.

The CART characterized the area by deep soils (Fig. 5.3 - PCA04<-0.0355), which is supported by the presence of the Crocker Fan consisting of Tertiary basin fills in north-western Sarawak, and Tertiary and Quaternary sediments in west Sarawak and in Sabah (Hall *et al.*, 2008). The CART further indicates that this region is least struck by ENSO droughts and is found below 362 m altitude (Fig. 5.3 - node 10.5). The ISA (Table S5.1) found most species with a maximum IndVal for this floristic region, which further support the findings of Ashton (1992) that the region is characterized by exceptional richness and endemicity.

11. Lowland rain forest of East Kalimantan

Although the 'Lowland rain forest of East Kalimantan' is most closely related to the 'Lowland rain forest of Sabah and Sarawak', they are two distinct floristic regions and not one, as suggested by Slik *et al.* (2003). The ISA (Table S5.1) indicates that there are many species characteristic for this region, which is further supported by the presence of Ulin (*Eusideroxylon zwageri*) in the lowland forests of East-, South- and Central Kalimantan but not in Brunei, Sabah and Sarawak (van Steenis, 1958b). That the lowland forests of the Kalimantan provinces are different from those of Sabah and Sarawak was also recognized by Ashton (1992), who classified the entire region as 'The generalized inland flora on udult ultisol soils'.
The CART indicated that the majority of the region has shallower soils than region 10 (Fig. 5.3 - PCA04>-0.0355), has a small annual temperature range (Bio07<8.7 °C) and receives less than 2,593 mm precipitation year^{-1}. The many species with their highest IndVal (Table S5.1) for this region support the separate floristic status. Although the CART paths to this region did not include the ENSO drought predictor, it is this region of Borneo which is most severely struck by ENSO droughts, an important reason why large parts of this region are deforested (Langner *et al.*, 2007).

12. Miscellaneous – Not distinguished

There are several additional floristic regions present on Borneo not classified as such by our analysis. These include: the Kinabalu montane meadows, discussed under 1; the Meratus Mountains, discussed under 8; the mangroves (Whitmore, 1984a, b), classified in our analysis with freshwater swamps (9); and the limestone and ultramafic floristic regions not included because of a lack of data. The exact extent and location of these latter two regions is not clear, a reason why WWF did not recognize these in their ecoregions (Wikramanayake *et al.*, 2002). We do recognize the existence of these floristic regions, but mainly due to resolution of our analysis we were not able to distinguish them.

Conclusions

For the first time, the floristic division of Borneo is quantitatively analysed based on 1439 significant SDMs and a cluster analysis that has resulted in the recognition of 11 floristic regions. Some doubts remain, however, especially about the southern Borneo lowland rain forest regions. These regions are largely characterized by the absence of indicator species, which could support the presence of a savanna corridor present during the LGM, forcing obligate rain-forest species to the northern refugia from where they still not have recovered their potential distribution range, a phenomenon known as the postglacial migration lag (Svenning *et al.*, 2008). Another issue that could have resulted in a different floristic delineation of southern Borneo is related to the fact that the most common species are excluded from our analysis based on statistical grounds. Common species, which show no niche preference, cannot be distinguished from a random null-model.
The presence of common species throughout all lowland areas results in more similar Sørensen's index values, potentially resulting in the aggregation of cluster groups 6, 7 and 8. On the other hand, the deficiency of collections from eastern, and especially southern Borneo (Raes & ter Steege, 2007 - Chapter 3), potentially has resulted in less recognized endemic species for these regions, since the limit for modelling was set at five records. The chance that an endemic from these regions is represented by five records is therefore much lower. Recognition of potential endemic species could have resulted in more clearly distinct cluster groups with their own indicator species. The same arguments could account for the relative absence of indicator species for the 'Montane rain forest of the Müller- and Upper Kapuas Mountains' (3) compared to the 'Montane rain forest of the central mountain chain' (1). To clarify these issues additional collections are required.

Nonetheless, our study quantitatively supports most of the previously recognized floristic regions, while at the same time indicates that the lowland rain forest of Borneo, often regarded as one floristic region (Fig. S5.2), should be divided in at least four (and potentially six) different regions: The lowland rainforests of 'Sabah and Sarawak'(10), 'East Kalimantan' (11), and 'southern Borneo' (6, 7, 8), and the 'Wet hill forest of Sarawak' (2).

Acknowledgements - We like to thank Jaap Vermeulen for sharing his orchid database, Ben Kieft for digitizing all maps, Gary Paoli for many constructive comments, the Alberta Mennega Stichting for funding the colour images, and the NWO-Groot 'Building databases for life'-project (Grant# 175.010.2003.010) for funding the digitization of all Borneo collections used for this research.