ORIGINAL PAPER

Reassessing Neotropical angiosperm distribution patterns based on monographic data: a geometric interpolation approach

Claudia Raedig · Carsten F. Dormann · Anke Hildebrandt · Sven Lautenbach

Abstract Monographic data rely on specimens deposited in herbaria and museums, which have been thoroughly revised by experts. However, monographic data have been rarely used to map species richness at large scale, mainly because of the difficulties caused by spatially heterogeneous sampling effort. In this paper we estimate patterns of species richness and narrow endemism, based on monographic data of 4,055 Neotropical angiosperm species. We propose a geometric interpolation method to derive species ranges at a 1° grid resolution. To this we apply an inverse distance-weighted summation scheme to derive maps of species richness and endemism. In the latter we also adjust for heterogeneous sampling effort. Finally, we test the robustness of the interpolated species ranges and derived species richness by applying the same method but using a leave-one-out-cross-validation (LOOCV). The derived map shows four distinct regions of elevated species richness: (1) Central America, (2) the Northern Andes, (3) Amazonia and (4) the Brazilian Atlantic coast ('Mata Atlântica'). The region with the highest estimated species richness is Amazonia, with Central America following closely behind. Centers of narrow endemism are located over the entire Neotropics, several of them coinciding with regions of elevated species richness. Sampling effort has a minor influence on the interpolation of overall species richness, but it substantially influences the estimation of regions of narrow endemism. Thus, in order to improve maps of narrow endemism and resulting conservation efforts, more collection and identification activity is required.

C. Raedig (✉)
Department of Systematic Botany, Institute of Biology I, University of Leipzig, Johannisallee 21, 04103 Leipzig, Germany
e-mail: raedig@uni-leipzig.de

C. F. Dormann · S. Lautenbach
Department for Computational Landscape Ecology, UFZ—Helmholtz Centre for Environmental Research, Leipzig, Germany

C. F. Dormann
Agroecology, Georg-August-University Göttingen, Göttingen, Germany

A. Hildebrandt
Department Computational Hydrosystems, UFZ—Helmholtz Centre for Environmental Research, Leipzig, Germany

Introduction

Species distribution patterns enable scientists and conservation planners to estimate centers
of biodiversity (e.g. Williams et al. 1996; Kress et al. 1998; Barthlott et al. 2005) and to
identify priority areas for conservation actions (e.g. Davis et al. 1997; de Oliveira and Daly
1999; Schatz 2002; Tobler et al. 2007). Species confined to very small distribution areas,
so-called narrow endemic species (Williams et al. 1996; Andersen et al. 1997), pose
important conservation issues due to their vulnerability to extinction (Gentry 1986; Knapp
2002). Due to insufficient data collection and heterogeneous sampling effort, distribution
patterns in the Neotropics are still poorly described (Kress et al. 1998; Bates and Demos
2001; Hopkins 2007; Morawetz and Raedig 2007). Moreover, the number of Neotropical
angiosperm species is exceptionally large, estimated at up to 90,000 species (Raven 1988;
Thomas 1999; Smith et al. 2004), making compilation of all species distributions a
daunting task. Amazonia, the largest and least accessible part of the Neotropics, still
harbors many regions where no plants have been collected at all; Schulman et al. (2007)
reported 43% of Amazonia as devoid of botanical collections and an additional 28% as
poorly collected. Species with limited or low occurrence are more likely to remain
undiscovered, thus impeding the assessment of the distribution of narrow endemic species.

   Given the fact that large areas generally are under-sampled, different techniques have
been applied to map distribution patterns at large scale. The first essential steps toward
estimating plant biodiversity at the global scale have been made by Davis et al. (1997) and
Barthlott et al. (1999, 2005) using inventory-based data. These inventories are summary
data for geographic units of varying size, mainly based on floras, regional species accounts,
local checklists and plot-based data. Whereas Davis et al. (1997) collected information on
all of their 234 priority sites and created sub-maps centered on these sites, Barthlott et al.
(1999; 2005) estimated plant species richness for standardized units of area (10,000 km^2)
to derive global maps of plant species richness. In both studies, the Neotropics were
indicated to be species-rich, but it was also noted that underlying collection data are
lacking for vast parts of Amazonia (Kier et al. 2005; Kreft and Jetz 2007).

   As an alternative to inventory-based analyses of species richness, distribution patterns
can also be obtained by overlaying maps of geographic ranges of individual species,
henceforth referred to as species ranges. Basically, species ranges correspond to regions
where occurrences of individuals of the species have been recorded (Gaston 1991), but
various more sophisticated concepts of deriving species ranges from occurrence data exist
(Lomolino et al. 2006). For the Neotropics, two approaches to estimate angiosperm species
ranges and species richness patterns have been applied. These are exclusively based on
species occurrence records and do not rely on a summary of different data sources.
Hopkins (2007) studied ranges of 1,584 Amazonian species at 1° grid resolution. Here,
species ranges were generated by extrapolating from point occurrence data sets, if neighbor
occurrences were positioned within the maximum distance of roughly 500 km. The
superposition of the thus derived species ranges yielded a species richness map of known
species that recognized large parts of the Amazon basin as species-rich. At the same time it

displayed a bias for better collected areas. In addition to this approach based on species ranges, Hopkins (2007) modeled species richness based on species numbers, using the same maximum distance of roughly 500 km. In both approaches, this predefined limit can lead to overestimation of species ranges and of species numbers.

For the entire Neotropics, Morawetz and Raedig (2007) analyzed data of 3,715 angiosperm species to identify centers of diversity and narrow endemism. Species occurrences were overlaid onto a 1° grid and merged into the respective grid cells (quadrats). This point-to-grid conversion yielded species ranges with a high degree of range porosity. In contrast to the method applied by Hopkins (2007), this approach is prone to an underestimation of species ranges.

Point data, such as museum and herbarium specimen data, have proven useful for the generation of species ranges (Williams et al. 1996; Kress et al. 1998; Schatz 2002; Willis et al. 2003; Graham et al. 2004). However, there also exist some inherent drawbacks, such as heterogeneous sampling of space and taxa because of varying accessibility of areas and attractiveness of taxa to collectors (Nelson et al. 1990; Graham et al. 2004; Schulman et al. 2007; Sheth et al. 2008) and systematic inaccuracy (Meier and Dikow 2004; Hopkins 2007; Tobler et al. 2007). This problem can in part be avoided by using revised specimen data, which were reviewed by expert taxonomists and published in form of monographs, so-called monographic data (Thomas 1999; Knapp 2002; Hopkins 2007). After reviewing the available data, we found that monographic distribution data are the most promising—because of their taxonomic correctness and reference to large areas. Since survey data on angiosperm species do not cover such a large area, monographic data represent an alternative. However, these data are difficult to analyze, since standard methods used for abundance data cannot be applied.

Species ranges derived from point data are not only subject to uncertainty that originates from the underlying data but also from the construction method. Examples of techniques for the estimation of species ranges are the convex hull (Willis et al. 2003; Sheth et al. 2008), the minimum spanning tree (Hernández and Navarro 2007) or the minimum bounding box (Graham and Hijmans 2006). Generating species ranges by means of a convex hull often results in overestimation of species ranges (Burgman and Fox 2003) and ignores disjunct distribution patterns, particularly for widespread species. A refined method is the use of the alpha-hull (Edelsbrunner et al. 1983; Burgman and Fox 2003), which is based on a triangulation approach. When applying the alpha hull, first, the average distance between the occurrence points is calculated. For the resulting alpha hull, only those occurrences are considered which are connected by a line being a multiple (termed *a*) of this average line length. Subject to the selection of *a*, constructed ranges either resemble coarser (*a* being larger, maximum size: convex hull) or finer (*a* being smaller, minimum size: point) alpha hulls. Another widely used method for the estimation of species ranges is the ecological niche modeling approach. This approach relates species occurrences to site conditions such as climate variables (the predictor set) and predicts species ranges based on the pattern of these auxiliary variables.

So far, detailed species richness maps based on species ranges of large numbers of species cover only parts of the Neotropics or lack quantification of uncertainty due to heterogeneous sampling effort over area (Kress et al. 1998; Hopkins, 2007; Morawetz and Raedig 2007; Schulman et al. 2007). Here we introduce an interpolation approach, which can be applied for scant data, and which does not require more than the available pure species occurrence data. Our goal is to make the application of this approach independent of detailed knowledge of the ecological demands of the species. The resulting patterns are

only an approximation of 'real' distribution patterns, but produced in a standardized, reproducible way.

The aim of this study is (i) to present a method tailored to map distribution patterns of Neotropical angiosperm species based on scarce, yet taxonomically reliable monographic occurrence data, (ii) to estimate the distribution patterns of Neotropical angiosperm species and (iii) to explore whether the method presented is appropriate for the identification of centers of diversity and narrow endemism.

Methods

Our analysis is based on distribution data of angiosperm species taken from monographs or similar thoroughly revised treatments covering the Neotropical realm (see Appendix 1). The database was presented in a previous work (Morawetz and Raedig 2007) and since then has been complemented with a further 340 species. It now contains 4,055 species, in 230 genera and 66 families, with ~77% woody and 23% herbaceous species. Species occurrence data were taken from distribution maps and transferred to a grid with 1° grid resolution containing 2,519 quadrats sized ~100 km × 100 km (varying from 12,550 km^2 at the equator to 8,250 km^2 at Tierra del Fuego). The species recorded in the database represent about 5% of all Neotropical angiosperm species. It should be stressed that species richness numbers and patterns derived here are indices of species richness, not estimates of absolute numbers.

Due to the special characteristics of our database, we had to design a novel interpolation approach. Firstly, because our data set only includes presence data (not presence/absence data), the choice of suitable habitat quality models was already strongly limited (e.g. Graham et al. 2004; Phillips et al. 2006). Secondly, many species are represented in very few quadrats. Although ecological niche models have successfully been applied for species with only five records (Pearson et al. 2007), exclusion of species having less than five occurrences would exclude about 50% of the species of our data set. Thirdly, the rule of the thumb that each explanatory variable requires about ten data points (Harrell 2001; Reineking and Schröder 2006) would exclude 90% of the species in our database, even if we used a small predictor set of only three environmental variables. Therefore, ecological niche modeling is not suitable for our data set. Furthermore, the species richness pattern of the point-to-grid-data (Fig. 3a) shows a strong bias towards easily accessible areas. Fitting a generalized additive model (GAM; Wood 2006) with species richness as the response and distance to cities, distance to rivers and distance to coasts as explanatory variables explained a significant amount of the variance (Explained deviance 0.39 for the Neotropics and 0.51 for Amazonia). Thus, we opted for a geometric interpolation-based approach to deduce species richness patterns. A requirement for this approach was the possibility to correct for heterogeneous sampling effort. In the absence of an independent validation data set, a further requirement to be met was the validation of the resulting species richness patterns.

Interpolating species ranges

The species occurrences contained in our database were overlaid with a grid (Fig. 1a). However, this point-to-grid data set is incomplete as it only contains occurrences of species which actually have been found, in quadrats that have actually been visited. We expect the actual species ranges to be much larger. Thus, based on the centroids of these quadrats,

a conditional triangulation similar to the alpha hull approach was performed: if a point was less than a given interpolation distance d away from two other points, a triangle was created and added to the triangle set (Fig. 1b). If only two points were within the given interpolation distance d, and thus no triangle could be built, a line between these two points was created (Fig. 1c). Triangle and line sets as well as points (which could not be interpolated due to missing neighbor occurrences) were combined and the set of corresponding quadrats was identified as the interpolated species range for a given distance d (Fig. 1d). As an extension to the alpha-hull approach (Edelsbrunner et al. 1983; Burgman and Fox 2003), not only the polygons of the triangulation but also the lines and points were considered. Thereby we avoided the problem of exclusion of narrow endemic species from analysis.

From species ranges to weighted species richness

After processing all species for one distance class i, the interpolated species ranges were summed across all species, creating an estimate of species richness S_i. Results were cal-culated for the distance classes $i = 1, 2, 3,…,10$. These species richness grids S_i were combined by performing an inverse distance-weighted approach according to:

$$S_w = \sum_{i=2}^{10} \left(d_i^{-p} \cdot (S_i - S_{i-1}) \right) + S_1 \qquad (1)$$

with $p > 0$, $d \geq 1$. S_1 is the original point-to-grid species richness grid, S_w is the grid of the resulting weighted species richness and d_i is the distance ($d_2 = 2$, $d_3 = 3,…$) used as a threshold in the conditional triangulation.

   For each distance class, the increase in species richness relative to the next smaller distance class was calculated for each quadrat and multiplied by a weighting term d_i^{-p}. Thereby, p is a tuning parameter of the weighting procedure applied to the quadrats. For each $p > 0$ and $d \geq 1$, the corresponding weighting term lies between 0 and 1. The greater p becomes, the more relative weight is put on species richness calculated for smaller distances. The closer p is to 0, the more relative weight is put on species richness inter-polated for larger distances (see Appendix 2). For the present work, we selected $p = 0.5$, which resulted in a combination of high weights for small distances and relatively low weights for large distances. The weighted differences between the distance classes were then added to the original point-to-grid data (S_1), yielding the map of weighted species richness S_w. Species richness centers were identified as contiguous areas of quadrats with $S_w > 100$, i.e. more than 100 interpolated species.

Adjusting weighted species richness for sampling effort

We addressed the impact of uneven spatial sampling effort by incorporating an additional weighting factor. This factor is based on the ratio of the number of species recorded in a quadrat and the maximum number of species reported for each center of species richness C of the original point-to-grid map $[S_1/\max_C (S_1)]$. This relationship between the number of species in a quadrat to the respective reference quadrat is used as a proxy for sampling effort for each quadrat. The higher the relative sampling effort in a quadrat, the nearer it will be to 1, hence the smaller the weighting (1—relative sampling effort) for the respective quadrat will be (Eq. 2). The higher the weight (relative sampling effort close to 0), the larger is the fraction of the interpolated species richness that enters the final estimation of species richness for that specific quadrat. The application of this correction
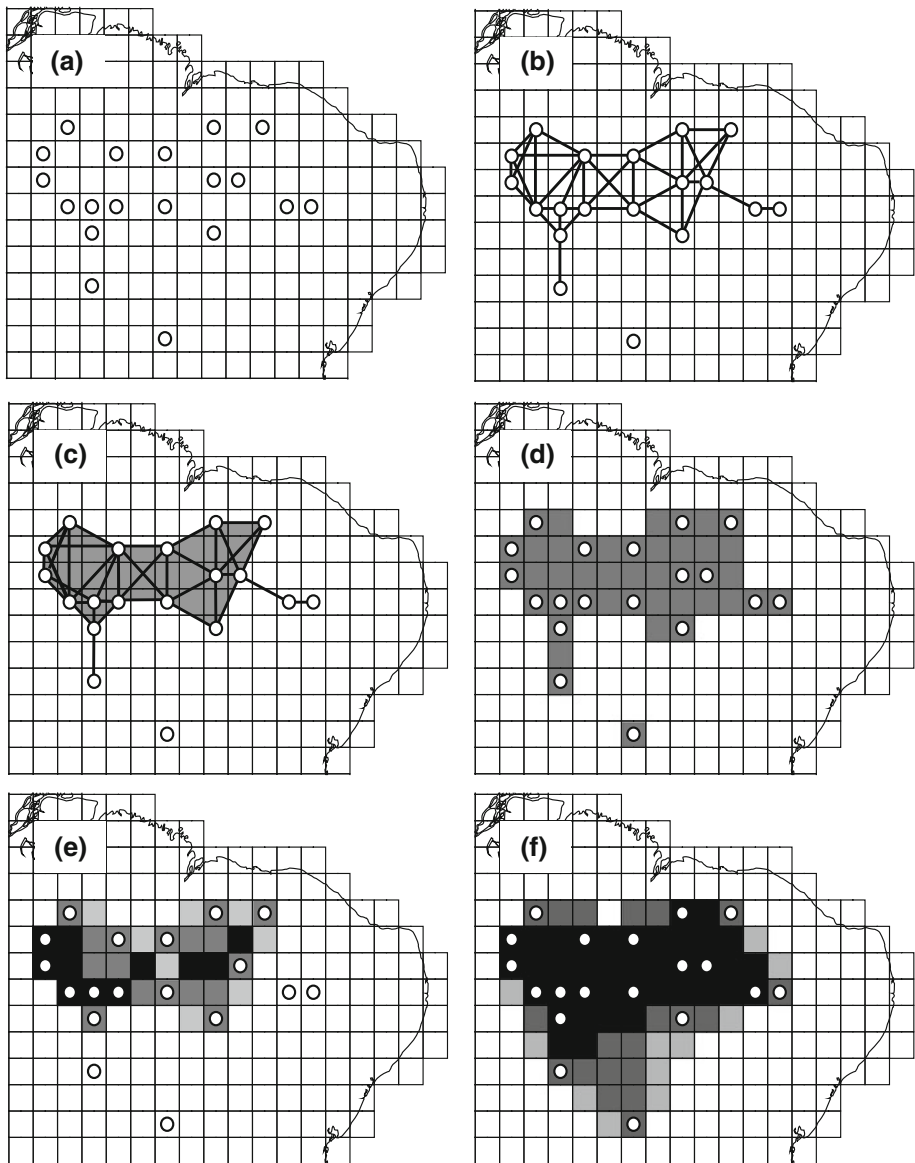
Fig. 1 Distance-weighted species range interpolation and LOOCV for *Parkia platycephala* Benth. (Hopkins 1986). **a–d** Interpolation using the distance of three quadrats (distance $i = 3$). **a** The point set as reported in the monograph. **b** Based on this point set and the given distance $i = 3$, a conditional polyline generation and **c** a conditional triangulation is performed. **d** The overlay of the three sets is then used to predict the species range (range$_i$) for the given distance in the underlying $1° \times 1°$ quadrats. **e–f** LOOCV. **e** For the interpolation distance of three quadrats, solo- and 2-point-occurences are not included into the resulting species range. **f** The interpolation distance of five quadrats yields a species range including all species occurrences

factor to the inverse distance-weighted sum of species richness at the distances 2–10, added to the observed point-to-grid species richness S_1 is henceforth referred to as adjusted species richness S_{adj}.

$$S_{\text{adj}} = \left(1 - \frac{S_1}{\max_C(S_1)}\right) * \sum_{i=2}^{10} (d_i^{-p} \cdot (S_i - S_{i-1})) + S_1 \qquad (2)$$

where $\max_C$ is a function which returns for each quadrat the maximum species richness for the diversity center the quadrat belongs to.

Estimating the interpolation robustness by cross-validation

In absence of a validation data set, we chose to estimate the robustness of the interpolation by performing a leave-one-out-cross-validation (LOOCV). Thereby, the interpolation steps were repeated on subsamples of the species points—leaving out each occurrence once—in order to cross-validate the interpolated species ranges (Efron and Gong 1983; Pearson et al. 2007). In contrast to the interpolation approach, this procedure generates floating point values indicating a robustness estimation for a species presence in a quadrat (Fig. 1e, f). For a detailed description of this approach, see Appendix 3. Dividing the resulting LOOCV-estimate by the weighted interpolation estimate S_w yielded the mean robustness of the weighted species richness estimation per quadrat.

Species ranges

So far we focused on species richness, originating from an overlay of species ranges. To detect the effort of interpolation on the species ranges of each species, we calculated the weighted range size range_w by combining the interpolated species ranges for each distance (range_i) for each species (Eq. 3, derived from Eq. 1).

$$\text{range}_w = \sum_{i=2}^{10} (d_i^{-p} \cdot (\text{range}_i - \text{range}_{i-1})) + \text{range}_1 \qquad (3)$$

Results are depicted as range size frequency distribution for the weighted range sizes (range_w) and are compared to the range size frequency distribution for individual distance classes.

Species richness of narrow endemic species

We used the same approximate definition for narrow endemic species as Gentry (1986): narrow endemic are those species for which the maximum interpolated range size was five quadrats (ca. 50,000 km^2, but the respective area varies with latitude between 41,250 and 62,750 km^2). While the LOOCV was useful in validating the interpolated species ranges and derived species richness centers, it was not used for the validation of narrow endemism centers because it would exclude too many species (at least 80.5% of narrow endemic species).

Results

Species ranges

The range size frequency distribution of the original point-to-grid ranges (Fig. 2a) is highly right-skewed (skewness = 4.8), with a mean of 12.3 (±22.4 SD) and a maximum of 327
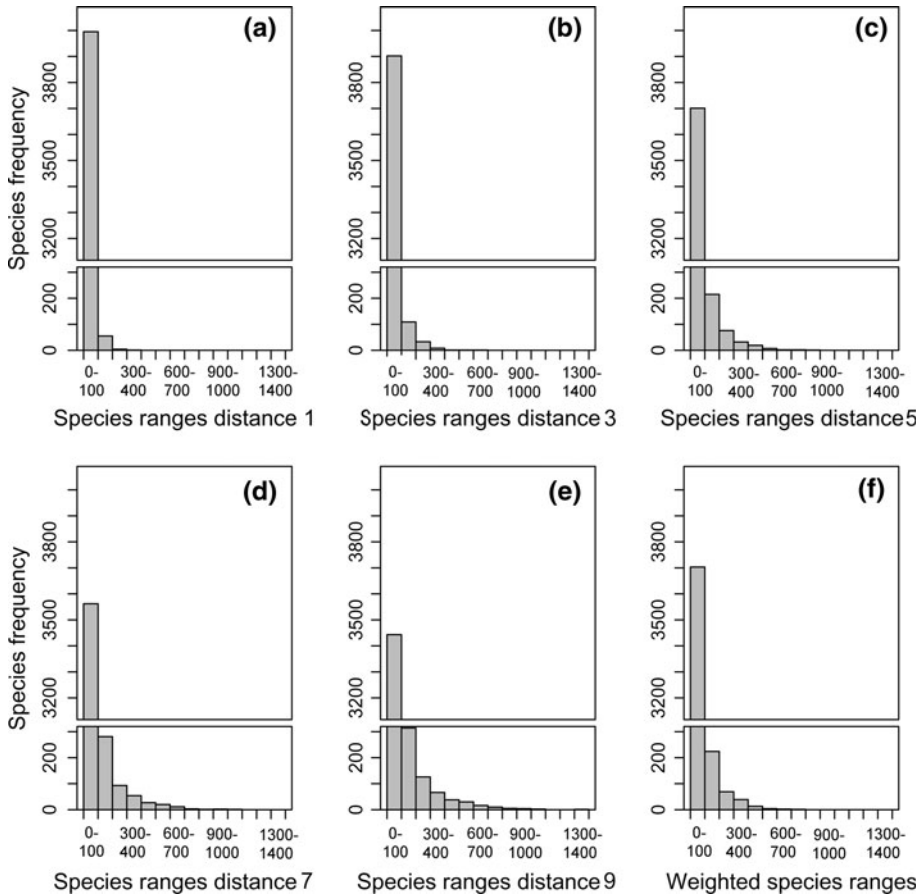
Fig. 2 Range size frequency distributions for all species. **a** Range size frequency distributions of the point-to-grid data. **b–e** Range size frequency distributions for selected interpolation distances. **f** Distance-weighted range size frequency distributions. The *y*-axis extends to 3,800, including a gap for *y*-values between 320 and 3,100

quadrats per species. Most species (3,995 = 99%) occur in less than 100 quadrats. With increasing interpolation distance *d* (see Eq. 1), both the mean and the maximum number of quadrats per species increase to 59.6 ± 123.2 and 1,378 quadrats for distance 10 (Fig. 2b–e). The combined inverse-distance weighted range size frequency distribution (Fig. 2f, according to Eq. 3) results in a mean of 32.6 ± 65.3, a maximum of 750.8 quadrats per species and a skewness of 4.1. While the mean value for *d* = 5 (Fig. 2c) is rather similar (33.3 ± 69.2), its range size frequency distribution has a higher skewness (4.5) and a higher maximum (831).

Species richness

Although our original point-to-grid species richness map (Fig. 3a) contains more species than the species richness map of a previous study (Morawetz and Raedig 2007) it identifies rather similar biodiversity centers. Point-to-grid species richness centers lie in Guatemala and
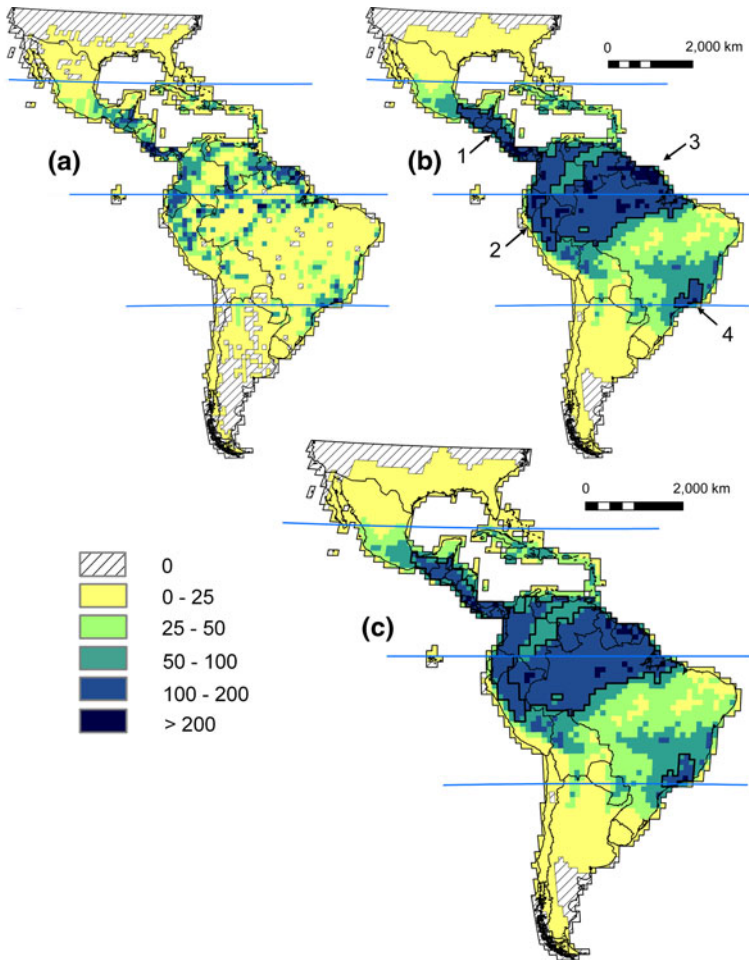
Fig. 3 Species richness of Neotropical angiosperms per quadrat. **a** Point-to-grid species richness (maximum number of species per quadrat: 331). **b** Weighted species richness (maximum number of species per quadrat: 391). **c** Species richness adjusted for sampling effort (maximum number of species per quadrat: 331) with delineation of the four largest species richness centers. 1—Central American, 2—Andean, 3—Amazonian, 4—Mata Atlântica species richness center. *Projection*: Aitoff, Central Meridian 60°W

adjacent regions, in Costa Rica and Panama reaching into the Chocó, in the Guyanas and at the border triangle of Venezuela, Colombia and Brazil. Moreover they stretch along the Andes (with peaks in the Ecuadorian and Peruvian Andes), along the Amazon with peaks close to Iquitos, Manaus, Santarém and Belém, and at the Brazilian Atlantic coast (Fig. 3a). The combination of the species richness grids over all distances according to Eq. 1 yields the map of weighted species richness (Fig. 3b) and results in four prominent species richness centers: one in Central America (1), crossing into the Andean species richness center (2), one Amazonian center (3) and one center in coastal Brazil (4). The final species richness map (Fig. 3c) adjusts for sampling effort according to these centers of species richness. It turned out that the reference quadrats with the maximum number of species chosen for each of the four centers are all located close to cities and rivers, i.e. easily accessible and therefore related

Table 1 Mean and standard deviation values of angiosperm species richness in the four centers identified in Fig. 3b for original point-to-grid species richness and for interpolated species richness

	No. of quadrats	Point-to-grid species richness Fig. 3a	Interpolated species richness Fig. 3b	Adjusted species richness Fig. 3c
Central America	60	91.8 ± 56.6	155.7 ± 52.5	136.8 ± 42.2
Andes	100	75.3 ± 33.8	152.7 ± 31.9	121.0 ± 18.0
Amazonia	333	50.7 ± 49.5	158.3 ± 44.0	143.5 ± 32.9
Mata Atlântica	21	75.8 ± 46.1	135.9 ± 33.0	119.2 ± 30.6

to higher sampling effort: the quadrat at Iquitos (Peru) for Amazonia, the quadrat north from San José (Costa Rica) for Central America, the quadrat at Cali (Colombia, Valle de Cauca) for the Andes, and the quadrat at Rio de Janeiro (Brazil) for the Mata Atlântica.

Transferring the outlines of these centers of species richness to the maps of point-to-grid (Fig. 3a) and adjusted species richness (Fig. 3c), the Amazonian point-to-grid center of species richness has the lowest mean value (50.7 ± 49.5 species per quadrat, Table 1), whereas the mean value for the Amazonian center of adjusted species richness is highest (143.5 ± 32.9). Although the sizes of the species richness centers vary between 21 and 333 quadrats, the mean values of adjusted species richness per center are within a close range (119.2 ± 30.6–143.5 ± 32.9). The high standard deviation decreases from the point-to-grid towards the adjusted species richness map (Table 1), the standard deviation values for the Andean species richness center notably being the lowest.

Whereas the effect of interpolation on range sizes is shown in Fig. 2f, the effect on point-to-grid species richness is shown in Fig. 4. This effect varies according to the centers of species richness (Fig. 4, ①–④) and to the quadrats not assigned to any of these centers (⑤, 'unassigned quadrats'). While it has little effect on the unassigned quadrats ⑤, the interpolation effect is highest for Amazonia ① and the Andes ②. For the smallest center of species richness, the Mata Atlântica ④, the effect is heterogeneous and also the lowest out of the four centers.

The results of the cross validation are high for most quadrats, but the four species richness centers are reflected by slightly higher LOOCV values than the unassigned quadrats (Table 2).The mean robustness per quadrat ranges between 0.777 ± 0.073 and 0.832 ± 0.043, with highest LOOCV values for the Amazonian center of species richness (Table 2).

According to the World Database on Protected Areas 2007 (WDPA Consortium 2008), most Neotropical quadrats are without any protection status (1,253; Fig. 5a) or with low protections status (986; Fig. 5b). The 160 quadrats with highest protection status (Fig. 5d) show maximum levels of species richness at comparably high human population density (Ciesin and Ciat 2005). Better protected quadrats (Fig. 5c, d) show varying correlation with population density, whereas quadrats without or with low protection status (Fig. 5a-b) consistently exhibit lower levels of species richness over all population density classes.

Narrow endemic species

Of the 4,055 species present in the database, 40% (1,573 species) were considered to be narrow endemic Neotropical species. The reference quadrats with the largest numbers of narrow endemic species chosen for each of the centers of species richness to adjust for sampling effort were the quadrats north of Manaus (Amazonia), east of San José (Central
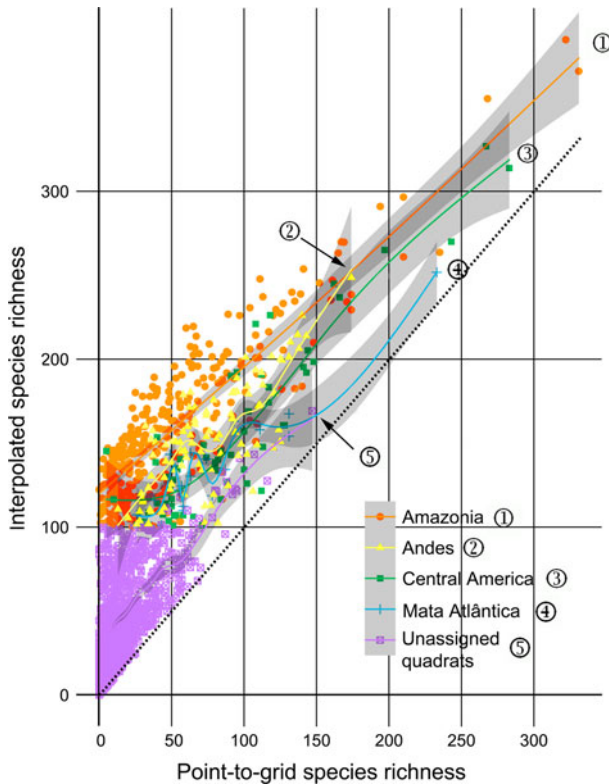
Fig. 4 Effect of inverse distance-weighted interpolation on the distribution patterns of angiosperm species. ①–④: centers of species richness; ⑤: quadrats not assigned to a center of species richness. *Symbols* above the *dotted equity line* indicate that the interpolated species richness variable of the *y*-axis outnumbers the point-to-grid species richness of the *x*-axis. Non-linear regressions (*trend lines* and *shaded standard error envelope*) using Generalized Additive Models indicate different effects of interpolation for the different centers

Table 2 Ratio between the species richness estimate by leave-one-out cross-validation (2,549 species) and by weighted interpolation (4,055 species) of the species richness centers identified in Fig. 3b

	LOOCV
Central America	0.813 ± 0.046
Andes	0.768 ± 0.054
Amazonia	0.833 ± 0.043
Mata Atlântica	0.780 ± 0.070
Unassigned quadrats	0.730 ± 0.307

Similar species richness estimates (ratios near 1) indicate that the interpolation results in an area are less influenced by the LOOCV and therefore robust

America), at Rio de Janeiro (Mata Atlântica), and at Cali (Andes). The map of centers of narrow endemism adjusted for sampling effort (Fig. 6a) did not differ much from the original point-to-grid map (*Kendall's τ*: 0.96). Salient centers of adjusted species richness of narrow endemic angiosperms are situated in Costa Rica and Panama, along the Andes
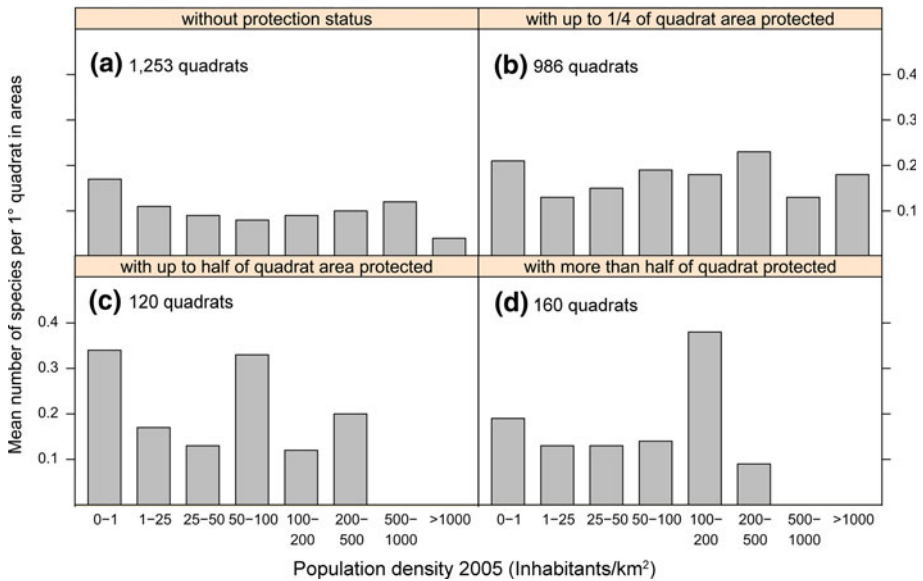
Fig. 5 Distribution of species on quadrats classified by protection status according to the World Database on Protected Areas 2007 (WDPA Consortium 2008) and estimated population density for 2005 (Ciesin and Ciat 2005). Species to be found in quadrats **a** without protection status, **b** with a proportion up to 25% of protected area, **c** with a proportion of 25–50% of protected area, and **d** with a proportion of more than 50% of protected area. The title of the *y*-axis continues above each panel of the *graph*

(from western Colombia to northern Peru) and at the Brazilian Atlantic coast close to Bahia and close to Rio de Janeiro, but a mosaic of quadrats containing up to five narrow endemics extends over the whole Neotropical region. Less prominent, but equally coherent areas of narrow endemism are located in the south of Mexico, the Caribbean islands, the southern Peruvian and the Bolivian Andes, parts of the Amazon basin, southeastern Cerrado and along the Pacific, the Atlantic and the Caribbean mainland coast. In combination, these areas exceed the areas suggested by Gentry (1992), who restricted Neotropical local endemism mainly to cloud forests ridges, inter-Andean valleys, Cuba and Hispaniola and isolated patches with specific habitat conditions especially in Amazonia. With the exception of the Amazonian species richness center, species richness centers identified in Fig. 3c are well reflected by the centers of narrow endemism. The 276 quadrats holding narrow endemic species and without protection status according to the categories Ia–IV (WDPA Consortium 2008) are highlighted in Fig. 6b.

Discussion

Methods interpolating species richness: spoiled for choice?

In this research we developed a new method for generating species ranges, which we used later to derive maps of species richness and centers of narrow endemism. At first glance it seems that we could have chosen between various approaches for generating species ranges (see section "Introduction"), why should we add yet a new one? The answer is that most methods were inappropriate, considering the characteristics of our data set, and thus also
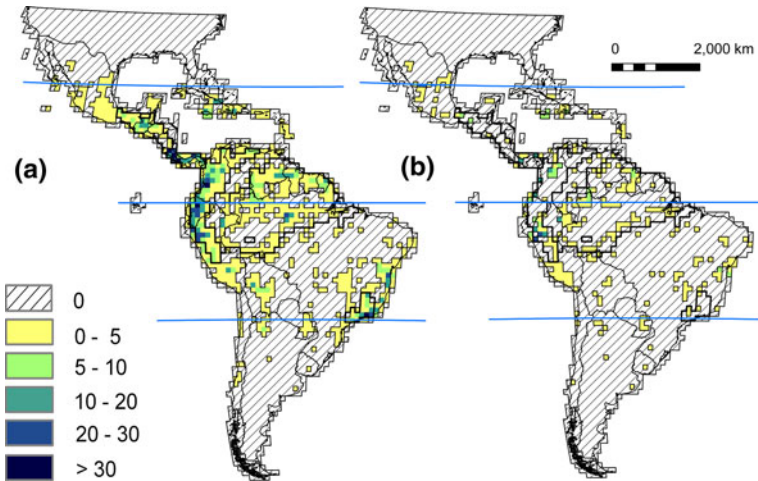
Fig. 6 Centers of narrow endemism of Neotropical angiosperm species (species richness per quadrat). **a** Adjusted species richness (Maximum number of narrow endemic species is 50). **b** Narrow endemic species not covered by a protection status according to categories Ia–IV (WDPA Consortium 2008). Maximum number of narrow endemic species per unprotected quadrat is 23. *Projection*: Aitoff, Central Meridian 60°W

for many similar situations. The proportion of 1,324 species in our database with fewer than three occurrences drastically reduced the number of applicable methods. Also, we found no justification to extrapolate beyond the outmost occurrences of our species. This is due to the fact that every species' range estimation is uncertain since it integrates over areas wherein the species in question has not been sampled. Uncertainty increases with distance to known species occurrences. Extrapolating our data beyond the outer species occurrences would therefore especially overestimate narrow-ranging species and include peripheral areas not belonging to the species range.

Interpolating species ranges

One challenge when applying our interpolation method to generate species ranges was to choose the right interpolation distance. To tackle this problem, we used the inverse-distance summation scheme described above. This approach ensures that the results of *all* interpolation distances are included, while the weighting favors smaller distances. Thereby, the risk of overestimation of species richness due to the generation of large and coherent species ranges for widespread, but locally scarce species is lowered. It has been shown, that particularly widespread species dominate distribution patterns (Jetz and Rahbek 2002; Kreft et al. 2006). If species with medium or large number of occurrences are interpolated with too much weight on long distances, the resulting large ranges will further aggravate this effect on species distribution patterns. Moreover, the risk of overestimation is reduced by putting a constraint on the largest possible interpolation distances, $d_{max} = 10$. Avoiding even larger distances (>1000 km) is in accordance with Hopkins (2007) who modeled ranges of Amazonian angiosperm species considering interpolation distances between one and nine quadrats (corresponding to 100 and 900 km).

Another important step for our species richness estimation was the adjustment for sampling effort. It is difficult to quantify the influence of overall sampling effort, yet we can apply some adjustment for heterogeneous spatial sampling effort. We did this by defining reference quadrats for the centers of species richness. As a result, quadrats with low species numbers are assigned higher weights than quadrats with high observed species numbers. Thereby, quadrats with high observed species richness acquire fewer additional species from interpolation while quadrats with a low number of observed species could acquire a larger fraction of additional species—if the unadjusted interpolation results predict additional species. We accepted overestimating species richness in some quadrats, knowing that vast areas of the Neotropics are under-sampled (Prance et al. 2000; Ruokolainen et al. 2002; Tobler et al. 2007). Although detailed maps of botanical sampling effort are available for some areas within the Neotropics (e.g. for Amazonia by Schulman et al. 2007), they are not available everywhere and therefore not used in the present work. Also, the procedure to adjust for sampling effort proposed here has the advantage of only requiring information inherent in the available point-to-grid data.

Species richness

Areas of elevated levels of species richness are the result of multiple overlapping species ranges. Most species occupy small ranges (Fig. 2a). Weighting of the species ranges (Eq. 3) demonstrates that the range sizes increase when applying our interpolation approach (Fig. 2f), but with a lower skewness and a lower maximum number of species compared to a medium interpolation distance of five quadrats (Fig. 2c), thus avoiding overestimation of ranges of widespread species.

The 'smoothed' increase of the range sizes due to the interpolation approach is reflected in the species richness maps (Fig. 3b, c). Whereas the inclusion of 340 more species (Fig. 3a) showed no major differences to the point-to-grid species richness map presented in Morawetz and Raedig (2007), considerable distinctions are evident in both maps of species richness (Fig. 3b, c). For the weighted interpolation, these differences are plotted in Fig. 4. For all centers of diversity as well as for the unassigned quadrats, interpolated species richness is above the equity line. The different effect of interpolation on the species richness according to diversity center is particularly revealing for Amazonia. Even for small distances, the interpolation of species ranges here is consistently high.

Comparison of maps 3b and 3c reveals the effect of adjusting species richness for sampling effort: the range of species richness is reduced, whereas the peaks of species richness found in Fig. 3b are retained in Fig. 3c. This effect is also apparent in the lower mean and standard deviation values for the centers of adjusted species richness, and in their closer range (Table 1). The Andean species richness center (Fig. 3c, polygon 2) shows the lowest standard deviation relative to the mean values (Table 1), suggesting more equal species richness and sampling effort of these Andean quadrats. The most obvious difference is that the Amazonian species richness center is by far the largest.

Amazonia contains the largest part of today's remaining contiguous rainforest area (e.g. Davis et al. 1997; Bates and Demos 2001). It has been suggested to be exceptionally species-rich (e.g. Kress et al. 1998; Ruokolainen et al. 2002; Schulman et al. 2007; Saatchi et al. 2008), which has been explained by habitat heterogeneity in combination with historical events (de Oliveira and Daly 1999; de Oliveira and Mori 1999) such as river dynamics and geological history.

In a global overview on species richness within ecoregions, Kier et al. (2005) suggested that the majority of ecoregions from the Andes to the Brazilian coast are very species-rich,

but they placed the Chocó and parts of the northern Andes along with the entire Cerrado as the most species-rich zones. This contrasts with the patterns we detected for Amazonia, where we identified highest species richness, and for the Cerrado, where we identified high species richness only in the peripheral zones. The diversity zones of a global comparison of vascular plants (Barthlott et al. 2005) differ from ours mainly in that they are much less pronounced for southwestern Amazonia.

In comparison with a plot-based model of Amazonian tree diversity (ter Steege et al. 2003), the Amazonian diversity center we found is spatially more uniform and includes parts of lower Amazonia as well. Our species richness map (Fig. 3c) also differs from the maps of Amazonia presented by Hopkins (2007) and ranges in between his overall species richness map (generated by a bootstrap approach based on species occurrences) and the species richness map generated by the overlay of extrapolated species ranges. The latter method is comparable to the one applied here, but some differences exist: (1) our approach is more conservative seeking to avoid overestimation and avoiding disproportionate influence of widespread species on distribution patterns, (2) we applied a weighed interpolation approach (as opposed to using only one interpolation distance), (3) we used a larger number of species and we also were able to consider a larger area.

The species richness estimates were validated by LOOCV to specify the robustness of the species ranges and therefore the robustness of the derived species richness map. Thus, the differences in the robustness depicted in Table 2 are due the spatial distribution of the species occurrences and give an indication of how heavily the prediction relies on information from single points. Observations from single points are important (1) when only few observations exist, and the information from one point represents a larger area, (2) for species that are widespread and only loosely connected and (3) for species with restricted distribution. In all cases leaving out single observations might lead to considerably smaller species ranges, and consequently to lower predicted species richness in the quadrats affected. With this in mind, the ratio between species estimates derived from LOOCV and from weighted interpolation will be smaller, indicating a lower robustness. We can thus re-interpret the higher robustness found for Amazonia: it suggests a high proportion of more uniformly distributed species with medium and larger numbers of species occurrences, and a low proportion of small-clustered species and species with few occurrences.

The LOOCV approach does not account for errors due to heterogeneous data quality or sampling effort. Whereas we integrated a strategy to adjust for heterogeneous spatial sampling effort at the level of species richness, we did not include an adjustment for the fact that more recent monographs will be more complete in terms of both taxa and occurrences considered. For the future, the interpolation process could be altered to include an additional weighting at species level. Furthermore, our maps will improve if more data based on future monographs were to be included in the analysis.

The results identified here are not absolute estimates of species richness per quadrat. To obtain a rough estimate of the absolute figures, the numbers per quadrat found need to be multiplied by the factor 20, since our data set represents approximately about 5% of the angiosperm flora occurring in the Neotropics. Following this estimation, our uppermost results would lie in close proximity to the uppermost results of Barthlott et al. (2005) suggesting more than 5,000 vascular plant species in the most species-rich 10,000 km^2 units, and that of Kreft and Jetz (2007), modeling 6,500 species at maximum per most species-rich 1° quadrats. Although our species richness map can only approximate 'real patterns', this consistency broadly supports our estimation of distribution patterns.

Narrow endemic species

Compared with previous work (Morawetz and Raedig 2007), in spite of considering more species, a similar number of species is identified as narrow endemic species. Previously, all species occurring in three or fewer quadrats were defined as narrow endemic species irrespective of distance between species occurrences, while in the present work only those species that occurred in five or less quadrats *after* interpolation with the maximum distance of five quadrats qualified as narrow endemic. Although the threshold of five quadrats appears more generous, the method is more rigorous in that it considers spatial distance. The main differences seen between Morawetz and Raedig (2007) and the present study are the absences of some species in southeastern Amazonia and in the Cerrado and Caatinga (two Brazilian floristic provinces) whose recorded occurrences were too geographically distant to be considered narrow endemic.

The analysis of narrow endemic species revealed two shortcomings of our interpolation method: first, if quadrats hold no species after interpolation, no adjustment of sampling effort can be applied. Considering the large number of empty quadrats, the map of narrow endemism (Fig. 6a) might reflect sampling effort more than distribution patterns. Second, we are interpolating species ranges, but not species richness per quadrat. Thus, narrow endemic species that have never been collected are absent from our analysis. We can hypothesize that quadrats near to well-collected quadrats with many narrow endemic species (Fig. 6a) might also hold more narrow endemic species. Considering the low levels of collecting and taxonomic activity in Amazonia in combination with the shortcomings of our method, the question remains elusive, whether narrow endemic species are a common phenomenon in Amazonia. Clarification in this matter can only be achieved by sampling of quadrats which have not been sampled appropriately (Bates and Demos 2001; Hopkins 2007), by taxonomical classification of the unidentified specimens already deposited in herbaria (Ruokolainen et al. 2002) and by publishing of these results as well as constant complementing and updating of databases with this information. Accordingly, our long-time objective is the complementing and updating of our database in combination with the integration of topographic or satellite-based or species-related information in the process of interpolating (e.g. inclusion of detailed soil data in combination with knowledge of the edaphic demands of species).

Protection status

In the Neotropics, almost 90% of the quadrats are without or with low protection status according to the WDPA 2007 (WDPA Consortium 2008; Fig. 5a, b). This figure is worryingly high, and reveals the size of many protected areas to be rather small. Species richness in better protected quadrats (Fig. 5c, d) in populated regions is low, which hints at the conflict between species diversity and human settlement; the existence of large cities in a quadrat excludes the establishment of large protected areas.

Bearing in mind the limitations of our approach, the large number of endemic-rich quadrats lacking protection status (Fig. 6b) demonstrates the urgency of the situation. Such quadrats were found in all parts of the Neotropical region. Since our database probably excludes many as yet undescribed narrow endemic species, the picture could be substantially worse. Many quadrats in particular in north-eastern Amazonia are empty in our map, and rather poorly provided with protected areas. In comparison to a previous analysis based on the WDPA 2005 (Morawetz and Raedig 2007), some quadrats containing many narrow endemic species but lacking protection status are now protected. However, as shown in Fig. 5, the proportion of the respective quadrats under protection is often small

(Grenyer et al. 2006). Our map of protection status of narrow endemic species (Fig. 6b) could serve as s a first step towards prioritizing the creation of protected sites, while better resolution of endemism data would greatly improve the results. In summary, the distribution patterns found here, although based on incomplete data and therefore preliminary, advocate the establishment of further protected areas in the Neotropics.

Conclusion

In the light of increasing data availability and ever growing distribution data sets, methods need to be tailored to their analysis. Although distribution modeling approaches are available, their applicability for monographic data and for presence-only data in general is often compromised by data scarcity, poor data quality and lack of knowledge of the environmental correlates of species. Our method is precisely targeted at such data and can also be adjusted to accommodate different taxonomical groups by changing the weighting of interpolation distances for species range generation.

Using this new method, we identified and validated centers of Neotropical angiosperm species richness and compared them to the current protection status and human population density. In addition, we identified areas where insufficient data do not allow for reliable estimates of distribution patterns. This is due to the sensitivity of the distribution patterns of the narrow endemic species towards sampling effort. In particular, our method might underestimate the numbers and the ranges of narrow endemic species in poorly collected areas. Our maps also indicate areas for further sampling activity, because the available data do not yet allow for robust estimation of species richness patterns. To permit pinpointing of species-rich areas for conservation priorities, a robust estimate of total species richness and narrow endemic species richness is necessary. Therefore, future collection activity should focus on under-sampled areas and under-sampled taxa. Further taxonomic identification of both new and already collected, unidentified specimens is necessary, which requires additional training and support of expert taxonomists. New and reliable data will enable the scientific community to further clarify Neotropical angiosperm distribution and in particular endemism patterns to improve response to conservation needs.

Appendices

Appendix 1

Literature consulted for compilation of the Neotropical angiosperm species distribution database.

Anderson C (1982) A monograph of the genus *Peixotoa* (Malpighiaceae). Contr Univ Michigan Herb 15:1–92

Anderson WR (1982) Notes on Neotropical Malpighiaceae—I. Contr Univ Michigan Herb 15:93–136

Andersson L (1977) The genus *Ischnosiphon* (Marantaceae). Opera Bot 43:1–114

Andersson L (1985) Revision of *Heliconia* subgen. *Stenochlamys* (Musaceae-Heliconioideae). Opera Bot 82:5–123

Areces-Mallea AE (1992) *Leptocereus santamarinae* (Cactaceae) a new species from Cuba. Brittonia 44:45–49

Arroyo MTK (1976) The systematics of the legume genus *Harpalyce* (Leguminosae: Lotoideae). Mem N Y Bot Gard 26:1–80

Ayers TJ (1990) Systematics of *Heterotoma* (Campanulaceae) and the evolution of nectar spurs in the New World Lobelioideae. Syst Bot 15:296–327

Barfod A (1991) A monographic study of the subfamily Phytelephantoideae (Arecaceae). Opera Bot 105:1–73

Barringer K (1991) A revision of *Epidendrum* subgenus *Epidanthus* (Orchidaceae). Brittonia 43:240–252

Berg CC (1972) Olmedieae, Brosimeae (Moraceae). Flora Neotrop 7

Berg CC, Akkermans RWAP, van Heusden ECH (1990) Cecropiaceae: *Coussapoa* and *Pourouma*, with an introduction to the family. Flora Neotrop 51

Bolick MR (1991) Systematics of *Salmea* (Compositae: Heliantheae). Syst Bot 16:462–477

Breckon GJ (1979) Studies in *Cnidoscolus* (Euphorbiaceae) 1. *Jatropha tubulosa, Jatropha liebmanni* and allied taxa from Central Mexico. Brittonia 31:125–148

Bricker JS (1991) A revision of the genus *Crinodendron* (Elaecarpaceae). Syst Bot 16: 77–88

Casper SJ (1966) Once more: the Orchid-flowered butterworts. Brittonia 18:19–28

Clark LG (1990) *Chusquea* sect. *Longiprophyllae* (Poaceae: Bambusoideae): A new Andean section and new species. Syst Bot 15:617–634

Cowan RS (1967) *Swartzia* (Leguminosae, Caesalpinoideae, Swartzieae). Flora Neotrop 1

da Silva MF (1976) Revisão taxonômica do gênero *Peltogyne* Vog. (Leguminosae-Caesalpinioideae). Acta Amazonica 6 (Suplemento):1-61

da Silva MF (1986) *Dimorphandra* (Caesalpiniaceae). Flora Neotrop 44

Dressler RL (1965) Notes on the genus *Govenia* in Mexico (Orchidaceae). Brittonia 17:266–277

Eckenwalder JE (1989) A new species *Ipomoea* sect. *Quamoclit* (Convolvulaceae) from the Caribbean and a new combination for a Mexican species. Brittonia 41:75–79

Ehrendorfer F, Silberbauer-Gottsberger I, Gottsberger G (1979) Variation on the population, racial, and species level in the primitive relic angiosperm genus *Drimys* (Winteraceae) in South America. Plant Syst Evol 132:53–83

Elias TS (1976) A monograph of the Genus *Hamelia* (Rubiaceae). Mem N Y Bot Gard 26(4):81–144

Forero E (1976) A revision of the American species of *Rourea* subgenus *Rourea* (Connaraceae). Mem N Y Bot Gard 26(1):1–119

Forero E (1983) Connaraceae. Flora Neotrop 36

Gates B (1982) A monograph of *Banisteriopsis* and *Diplopterys*, Malpighiaceae. Flora Neotrop 30

Gentry AH (1980) Bignoniaceae Part l (Crescentieae and Tourrettieae). Flora Neotrop 25

Gentry AH (1992) Bignoniaceae Part 2 (tribe Tecomae). Flora Neotrop 25

Grear JW (1984) A revision of the New World species of *Rhynchosia* (Leguminosae–Faboideae). Mem N Y Bot Gard 31:1–168

Hekking WHA (1988) Violaceae. Part l—*Rinorea* and *Rinoreocarpus*. Flora Neotrop 46

Henderson A (2000) *Bactris* (Palmae). Flora Neotrop 79

Henderson A, Galeano G (1996) *Euterpe*, *Prestoea* and *Neonicholsonia* (Palmae). Flora Neotrop 72

Henderson A (1990) Arecaceae. Part 1. Introduction and the Iriarteinae. Flora Neotrop 53

Hopkins CF (1986) *Parkia* (Leguminosae: Mimosoideae). Flora Neotrop 43

Judd WS, Beaman RS (1988) Taxonomic studies in the Miconieae (Melastomataceae). 2. Systematics of the *Miconia subcompressa* complex of Hispaniola, including the description of two new species. Brittonia 40:368–391

Kaastra RC (1982) A monograph of the Pilocarpinae (Rutaceae). Flora Neotrop 33

Kallunki JA (1992) A revision of *Erythrochiton* sensu lato (Cuspariinae, Rutaceae). Brittonia 44:107–139

Knapp S (1989) A revision of the *Solanum nitidum* group (section Holophylla pro parte): Solanaceae. Bull Br Mus (Nat Hist) Bot 19:63–102

Knapp S (1989) Six new species of *Solanum* sect. *Geminata* from South America. Bull Br Mus (Nat Hist) Bot 19:103–112

Knapp S (1992) Five new species of *Solanum* section *Geminata* (Solanaceae) from South America. Brittonia 44:61–68

Kubitzki K (1989) The ecogeographical differentiation of Amazonian inundation forest. Plant Syst Evol 162:285–304

Kubitzki K, Renner SS (1982) Lauraceae: *Aniba* and *Aiouea*. Flora Neotrop 31

Landrum LR (1986) *Campomanesia*, *Pimenta*, *Blepharocalyx*, *Legrandia*, *Acca*, *Myrrhinium*, and *Luma* (Myrtaceae). Flora Neotrop 45

Lee YS, Seigier DS, Ebinger JE (1989) *Acacia rigidula* (Fabaceae) and related species in Mexico and Texas. Syst Bot 14:91–100

Lleras E (1978) Monograph of the family Trigoniaceae. Flora Neotrop 19

Lorence DH, Nee M (1987) *Randia retroflexa* (Rubiaceae) a new species from Southern Mexico. Brittonia 39:371–375

Luteyn JL (1983) Ericaceae Part l *Cavendishia*. Flora Neotrop 35

Luteyn JL (1984) Revision of *Semiramisia* (Ericaceae: Vaccinieae). Syst Bot 9:359–367

Luther HE, Sieff E (1991) An alphabetical list of bromeliad binomials. Selby Botanical Gardens, Orlando

Maas PJM (1977) *Renealmia* (Zingiberaceae—Zingiberoideae) Costoideae (additions) (Zingiberaceae). Flora Neotrop 18

Maas PJM, Maas van de Kamer H, van Bethem J, Snelders HCM, Rübsamen T (1986) Burmanniaceae. Flora Neotrop 42

Maas PJM, Rübsamen T (1986) Triuridaceae. Flora Neotrop 40

Maas PJM, Ruyters P (1986) *Voyria* and *Voyriella* (saprophytic Gentianaceae). Flora Neotrop 41

Maas PJM, Westra LYT (1984) Studies in Annonaceae. II. A monograph of the genus *Anaxagorea* A. St. Hil. (Annonaceae). Bot Jahrb Syst 105:73–134

Maas PJM, Westra LYT (1992) *Rollinia*. Flora Neotrop 57

Miller JS (1989) A revision of the New World species of *Ehretia* (Boraginaceae). Ann Mo Bot Gard 76:1050–1076

Molau U (1988) Scrophulariaceae, Part 1. Calceolarieae. Flora Neotrop 47

Molau U (1990) The genus *Bartsia* (Scrophulariaceae-Rhinanthoideae). Opera Bot 102: 1–99

Moraes MR (1996) *Allagoptera* (Palmae). Flora Neotrop 73

Moraes RM, Henderson A (1990) The genus *Parajubaea* (Palmae). Brittonia 42:92–99

Morawetz W (1995) Unpublished confirmed records for the genus *Annona* (Annonacae)

Morawetz W (1982) Morphologisch-ökologische Differenzierung, Biologie, Systematik und Evolution der neotropischen Gattung *Jacaranda* (Bignoniaceae). Austrian Academy of Sciences, Denkschriften 123, pp 1–184

Morawetz W (1984) Karyological races and ecology of *Duguetia furfuracea* as compared with *Xylopia aromatica* (Annonaceae). Flora 175:195–209

Mori SA (1981) New species and combinations in neotropical Lecythidaceae. Brittonia 33:357–370

Mori SA, Prance GT (1981) The "*Sapucaia*" group of *Lecythis* (Lecythidaceae). Brittonia 33:70–80

Murray NA (1993) Revision of *Cymbopetalum* and *Porcelia* (Annonaceae). Syst Bot 40: 1–121

Pennington TD (1990) Sapotaceae. Flora Neotrop 52

Pennington TD, Styles BT (1981) A monograph of Neotropical Meliaceae. Flora Neotrop 28

Pennington TD (1997) The genus *Inga*—Botany. Roy. Bot. Gard. Kew

Poppendieck HH (1981) Cochlospermaceae. Flora Neotrop 27

Powell AM (1965) Taxonomy of *Tridax*. Brittonia 17:47–96

Prance GT (1972) A monograph of the neotropical Dichapetalaceae. Flora Neotrop 10

Prance GT (1972) A monograph of the Rhabdodendraceae. Flora Neotrop 11

Prance GT (1989) Chrysobalanaceae. Flora Neotrop 9S

Prance GT, da Silva MF (1973) A monograph of Caryocaraceae. Flora Neotrop 12

Prance GT, Mori SA (1979) Lecythidaceae—Part I. The actinomorphic-flowered New World Lecythidaceae (*Asteranthos*, *Gustavia*, *Grias*, *Allantoma* and *Cariniana*). Flora Neotrop 21

Rainer H (1995): *Annona*. In: Steyermark JA, Berry PE, Holst BK (eds) Flora of the Venezuelan Guayana, vol 2. Missouri Botanical Garden and Timber Press, Saint Luis, Portland

Renner SS (1989) Systematic studies in the Melastomataceae: *Bellucia*, *Loreya*, and *Macairea*. Mem N Y Bot Gard 50:1–112

Renner SS (1990) A revision of *Rhynchanthera* (Melastomataceae). Nord J Bot 9:601–630

Rodrigues WA (1980) Revisão taxonômica das espécies de *Virola* Aublet (Myristicaceae) do Brasil. Acta Amazonica 10(Suplemento):1–127

Roe KE (1967) A revision of *Solanum* sect. *Brevantherum* (Solanaceae) in North and Central America. Brittonia 19:353–373

Rogers GK (1984) *Gleasonia*, *Henriquezia* and *Platycarpum* (Rubiaceae). Flora Neotrop 39

Rueda R (1994) Systematics and evolution of the genus *Petrea* (Verbenaceae). Ann Mo Bot Gard 81:610–652

Silverstone-Sopkin PA, Graham SA (1986) Alzateaceae, a plant family new to Colombia. Brittonia 38:340–343

Sleumer HO (1984) Olacaceae. Flora Neotrop 38

Smith LB, Downs RJ (1983) Tillandsioideae (Bromeliaceae). Flora Neotrop 14(2)

Stahl B (1991) A revision of *Clavija* (Theophrastaceae). Opera Bot 107:1–77

Stahl B (1992) On the identity of *Jacquinia armillaris* (Theophrastaceae) and related species. Brittonia 44:54–60

Taylor CM (1989) Revision of *Palicourea* (Rubiaceae) in Mexico and Central America. Syst Bot 26:1–102

1543

Tebbs MC (1989) Revision of *Piper* (Piperaceae) in the New World—I: review of characters and taxonomy of *Piper* section *Macrostachys*. Bull Br Mus (Nat Hist) Bot 19:117–158

Tebbs MC (1990) Revision of *Piper* (Piperaceae) in the New World—II: The taxonomy of *Piper* sect. *Churumayu*. Bull Br Mus (Nat Hist) Bot 20:193–236

Thomas WW (1984) The systematics of *Rhynchospora* section *Dichromena*. Mem N Y Bot Gard 37:1–116

Till W (1985) Sippendifferenzierung innerhalb *Tillandsia* subgenus *Diaphoranthema* in Südamerika mit besonderer Berücksichtigung des Anden-Ostrandes und der angrenzenden Gebiete. Dissertation, University Vienna

Todzia CA (1988) Chloranthaceae: *Hedyosmun*. Flora Neotrop 48

Todzia CA (1989) A revision of *Ampelocera* (Ulmaceae). Ann Mo Bot Gard 76:1087–1102

Wallnöfer B (1997) A revision of *Styrax* L. section *Pamphilia* (Mart. ex A.DC.) B.Walln. (Styracaceae). Ann Naturhist Mus Wien B 99:681–720

Webster GL (1984) *Jablonskia*, a new genus of Euphorbiaceae from South America. Syst Bot 9:229–235

Weiner G (1992) Zur Stammanatomie der Rattanpalmen. Dissertation, University of Hamburg

Wessels Boer JG (1968) The Geonomoid palms. Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Afd. Natuurkunde, Tweede Reeks 58:1–202

Wheeler GA (1990) Taxonomy of the *Carex atropicta* complex (Cyperaceae) in South America. Syst Bot 15:643–659

Zona S (1996) *Roystonea* (Arecaceae: Arecoideae). Flora Neotrop 71

Zuloaga FO, Judziewicz EJ (1991) A revision of *Raddiella* (Poaceae: Bambusoideae: Olyreae). Ann Mo Bot Gard 78:928–941
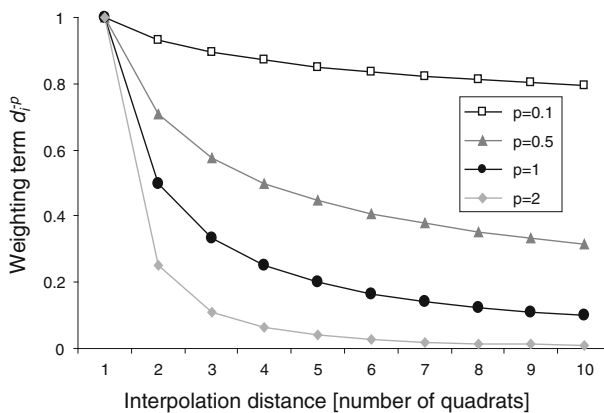
Appendix 2

Fig. 7 Effects of varying factor p (Eqs. 1–3) on the inverse-distance weighting term d_i^{-p} over all distances. A small factor p results in a rather consistent weighting term d_i^{-p} over all distances. The greater p becomes, the more weight is put on the smaller distances when interpolating

Appendix 3

Leave-one-out-cross-validation in detail.

Short of an independent validation dataset, we decided to use a cross-validation similar to an approach introduced by Pearson et al. (2007). The interpolation steps (according to our Eq. 1) were repeated on subsamples of the species points in order to cross-validate the interpolated species ranges and therefore to estimate the robustness of the derived weighted species richness map. For each species, n subsamples were selected, with n being the number of occurrences of the species. Subsequently, each species occurrence was left out once for interpolation, resulting in $(n - 1)$ occurrences per subsample.

We calculated a LOOCV-weight of robustness for each species and quadrat, as the number of times the species occurrences have been estimated to be part of the species range derived from the n subsamples, divided by the number of subsamples n. In contrast to the interpolation approach, this procedure generates floating point values in the interval [0,1] indicating a robustness estimation for a species presence in a quadrat. Quadrats which were frequently belonging to the estimated species range were assigned a value close to 1, and those which were rarely part of the estimated species range received a value close to 0.

In the process of cross-validation, the number of neighboring occurrences was considered, and only occurrences having at least two neighbors within the interpolation distance were included for interpolation (Fig. 1e, f), thus reducing the total number of species for LOOCV to the 2,549 species with more than two records. To compare the results of the cross-validation with the species richness map from Eq. 1 we combined the species richness maps from the cross-validation by the following inverse distance weighted approach:

$$S_{w,\text{LOOCV}} = \sum_{i=3}^{10} \left(d_i^{-p} \cdot \left(S_{i,\text{LOOCV}} - S_{i-1,\text{LOOCV}} \right) \right) + S_{2,\text{LOOCV}} \tag{4}$$

Dividing the resulting LOOCV-estimate $S_{w,\text{LOOCV}}$ by the weighted interpolation estimate S_w (for the distances 3–10, otherwise identical to Eq. 1) yielded the mean robustness of the weighted species richness estimation per quadrat.
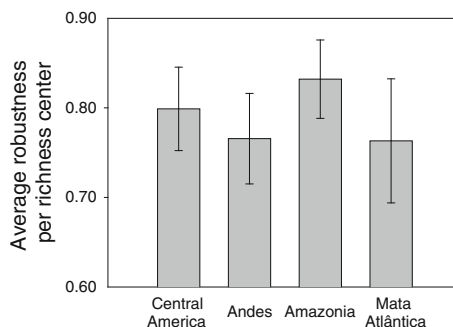

Fig. 8 Ratio between the species richness estimate by LOOCV and by weighted interpolation of the species richness centers identified in Fig. 3b. Similar richness estimates (ratios near 1) indicate that the interpolation results in an area are less influenced by the leave-one-out cross-validation and therefore robust

References

Andersen M, Thornhill AD, Koopowitz H (1997) Tropical forest disruption and stochastic biodiversity losses. In: Laurance WF, Bierregaard RO (eds) Tropical forest remnants: ecology, management, and conservation of fragmented communities. University of Chicago Press, Chicago

Barthlott W, Biedinger N, Braun G, Feig F, Kier G, Mutke J (1999) Terminological and methodological aspects of the mapping and analysis of the global biodiversity. Acta Bot Fenn 162:103–110

Barthlott W, Mutke J, Rafiqpoor MD, Kier G, Kreft H (2005) Global centers of vascular plant diversity. Nova Acta Leopold 92:61–83

Bates JM, Demos TC (2001) Do we need to devalue Amazonia and other large tropical forests? Divers Distrib 7:249–255

Burgman MA, Fox JC (2003) Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. Anim Conserv 6:19–28

Center for International Earth Science Information Network (Ciesin), Centro Internacional de Agricultura Tropical (Ciat) (2005) Gridded population of the world, version 3 (GPWv3) data collection. http://sedac.ciesin.columbia.edu/gpw/index.jsp. Cited 12 Feb 2008

Davis SD, Heywood VH, Herrera-Macbryde O, Villa-Lobos J, Hamilton AC (eds) (1997) The Americas. In: Centres of plant diversity: A guide and strategy for their conservation, vol. 3. IUCN Publications Unit, Cambridge

de Oliveira AA, Daly DC (1999) Geographic distribution of tree species occurring in the region of Manaus, Brazil: implications for regional diversity and conservation. Biodivers Conserv 8:1245–1259

de Oliveira AA, Mori S (1999) A central Amazonian terra firme forest. I. High tree species richness on poor soils. Biodivers Conserv 8:1219–1244

Edelsbrunner H, Kirkpatrick DG, Seidel R (1983) On the shape of a set of points in the plane. IEEE Trans Inform Theory IT 29:551–559

Efron B, Gong G (1983) A leisurely look at the bootstrap, the jackknife, and cross-validation. Am Stat 37:36–48

Gaston KJ (1991) How large is a species' geographic range? Oikos 61:434–438

Gentry AH (1986) Endemism in tropical versus temperate plant communities. In: Soulé ME (ed) Conservation biology: the science of scarcity and diversity. Sinauer Associates Inc., Sunderland

Gentry AH (1992) Tropical forest biodiversity: distributional patterns and their conservational significance. Oikos 63:19–28

Graham CH, Hijmans RJ (2006) A comparison of methods for mapping species ranges and species richness. Glob Ecol Biogeogr 15:578–587

Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol Evol 19:497–503

Grenyer R, Orme CDL, Jackson SF, Thomas GH, Davies RG, Davies TJ, Jones KE, Olson VA, Ridgely RS, Rasmussen PC, Ding T, Bennett PM, Blackburn TM, Gaston KJ, Gittleman JL, Owens IPF (2006) Global distribution and conservation of rare and threatened vertebrates. Nature 444:93–96

Harrell FEJ (2001) Multivariable modeling strategies. In: Regression modeling strategies–with applications to linear models logistic regression, and survival analysis. Springer, New York

Hernández HM, Navarro M (2007) A new method to estimate areas of occupancy using herbarium data. Biodivers Conserv 16:2457–2470

Hopkins CF (1986) *Parkia* (Leguminosae: Mimosoideae). Flora Neotrop 43

Hopkins MJG (2007) Modelling the known and unknown plant biodiversity of the Amazon Basin. J Biogeogr 34:1400–1411

Jetz W, Rahbek C (2002) Geographic range size and determinants of avian species richness. Science 297:1548–1551

Kier G, Mutke J, Dinerstein E, Ricketts TH, Küper W, Kreft H, Barthlott W (2005) Global patterns of plant diversity and floristic knowledge. J Biogeogr 32:1107–1116

Knapp S (2002) Assessing patterns of plant endemism in Neotropical uplands. Bot Rev 68:22–37

Kreft H, Jetz W (2007) Global patterns and determinants of vascular plant diversity. Proc Natl Acad Sci USA 104:5925–5930

Kreft H, Sommer JH, Barthlott W (2006) The significance of geographic range size for spatial diversity patterns in Neotropical palms. Ecography 29:21–30

Kress WJ, Heyer WR, Acevedo P, Coddington J, Cole D, Erwin TL, Meggers BJ, Pogue M, Thorington RW, Vari RP, Weitzman MJ, Weitzman SH (1998) Amazonian biodiversity: assessing conservation priorities with taxonomic data. Biodivers Conserv 7:1577–1587

Lomolino MV, Riddle BR, Brown JH (2006) Biogeography, 3rd edn. Sinauer Associates Inc., Sunderland

Meier R, Dikow T (2004) Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. Conserv Biol 18:478–488

Morawetz W, Raedig C (2007) Angiosperm biodiversity, endemism and conservation in the Neotropics. Taxon 56:1245–1254

Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. Nature 345:714–716

Pearson RG, Raxworthy CJ, Nakamura M, Townsend PA (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J Biogeogr 34:102–117

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259

Prance GT, Beentje H, Dransfield J, Johns R (2000) The tropical flora remains undercollected. Ann Mo Bot Gard 87:67–71

Raven P (1988) Tropical floristics tomorrow. Taxon 37:549–560

Reineking B, Schröder B (2006) Constrain to perform: regularization of habitat models. Ecol Model 193:675–690

Ruokolainen K, Tuomisto H, Vormisto J, Pitman N (2002) Two biases in estimating range sizes of Amazonian plant species. J Trop Ecol 18:935–942

Saatchi S, Buermann W, ter Steege H, Mori SA, Smith TB (2008) Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. Remote Sens Environ 112:2000–2017

Schatz GE (2002) Taxonomy and herbaria in service of plant conservation: lessons from Madagascar's endemic families. Ann Mo Bot Gard 89:145–152

Schulman L, Toivonen T, Ruokolainen K (2007) Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. J Biogeogr 34:1388–1399

Sheth SN, Lohmann LG, Consiglio T, Jimenez I (2008) Effects of detectability on estimates of geographic range size in Bignonieae. Conserv Biol 22:200–211

Smith N, Mori SA, Henderson A, Stevenson DW, Heald SV (eds) (2004) Introduction. In: Flowering plants of the Neotropics. Princeton University Press, Princeton

ter Steege H, Pitman N, Sabatier D, Castellanos H, van der Hout P, Daly DC, Silveira M, Phillips O, Vasquez R, van Andel T, Duivenvoorden J, de Oliveira AA, Ek R, Lilwah R, Thomas R, van Essen J, Baider C, Maas P, Mori SA, Terborgh J, Núñez VP, Mogollón H, Morawetz W (2003) A spatial model of tree α-diversity and tree density for the Amazon. Biodivers Conserv 12:2255–2277

Thomas WW (1999) Conservation and monographic research on the flora of Tropical America. Biodivers Conserv 8:1007–1015

Tobler M, Honorio E, Janovec J, Reynel C (2007) Implications of collection patterns of botanical specimens on their usefulness for conservation planning: an example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. Biodivers Conserv 16:659–677

WDPA Consortium (2008) WDPA World Database on Protected Areas 2007. World Conservation Union (IUCN) and UNEP-World Conservation Monitoring Centre (UNEP-WCMC). http://www.unep-wcmc.org/wdpa/index.htm. Cited 12 Feb 2008

Williams PH, Prance GT, Humphries CJ, Edwards KS (1996) Promise and problems in applying quantitative complementary areas for representing the diversity of some Neotropical plants (families Dichapetalaceae, Lecythidaceae, Caryocaraceae, Chrysobalanaceae and Proteaceae). Biol J Linn Soc 58:125–157

Willis F, Moat J, Paton A (2003) Defining a role for herbarium data in Red List assessments: a case study of Plectranthus from eastern and southern tropical Africa. Biodivers Conserv 12:1537–1552

Wood SN (2006) Generalized additive models: an introduction with R. Chapman & Hall/CRC Press, Boca Raton