Effects of sampling scale on patterns of habitat association in tropical trees

Garzon-Lopez, Carol X.; Jansen, Patrick A.; Bohlman, Stephanie A.; Ordonez Gloria, Alejandro; Olff, Han

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Early version, also known as pre-print

[Link to publication in University of Groningen/UMCG research database](#)

Effects of sampling scale on patterns of habitat association in tropical trees

Carol X. Garzon-Lopez, Patrick A. Jansen, Stephanie A. Bohlman, Alejandro Ordonez & Han Olff

Garzon-Lopez, C.X. (corresponding author c.x.garzon@gmail.com), **Jansen, P.A.** (patrick.jansen@wur.nl), **Ordonez, A.** (ordonezglori@wisc.edu) & **Olff, H.** (h.olff@rug.nl): Community and Conservation Ecology, Center of Life Sciences, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

Jansen, P.A. : Department of Environmental Sciences, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands

Jansen, P.A. & **Bohlman, S.A.** (sbohlman@ufl.edu): Smithsonian Tropical Research Institute, Apartado 0843–03092, Balboa, Ancon, Republic of Panama

Bohlman, S.A. : School of Forest Resources and Conservation, University of Florida, P.O. Box 110410, Gainesville, FL 32611-0410, USA

Ordonez, A. : Nelson Institute for Environmental Studies – Center for Climatic Research (CCR), 53706-1695 Madison, WI, USA

Abstract

Questions: Niche differentiation is a central explanation for the co-existence and distribution patterns of numerous tree species in tropical forests, but functional equivalence leading to neutral dynamics has been proposed as an alternative explanation. This niche vs neutral debate is fuelled by the highly variable results yielded by studies of the association between tree species distributions and environmental factors, where some studies find strong associations but others do not. Here, we ask how differences in sampling scale between studies contribute to this variation.

Location: Barro Colorado Island, Panama.

Methods: Using distribution maps of canopy-statured individuals, we evaluated patterns of habitat association in five tropical tree species on Barro Colorado Island across a wide range of sampling scales (from 50 to 1600 ha). We investigated the scale-dependency of species clumping patterns (Ripley's *K*) and the association of species distributions with important environmental variables (forest age, topography and geological formation) using point pattern analyses.

Results: Clump size and clump density had high variances within and among spatial scales. Significant habitat associations were found in all five species, with the number of habitat associations generally increasing with the sampling scale. Ignoring dispersal constraints inflated the number of significant habitat associations.

Conclusions: We found that patterns of habitat association (and hence conclusions on the importance of niche vs neutral processes) are strongly affected by the choice of sampling scale and location. Explicit inclusion of the effect of spatial scale is critical for studies of habitat association and the main processes that structure communities of tropical trees.

Introduction

Tropical forests are among the most diverse ecosystems in the world (Givnish 1999). Various hypotheses have been proposed to explain the high level of species co-existence that maintains this diversity (Wright 2002). One of the most important explanations is niche theory, which argues that species adaptation to specific conditions determines the distribution of different species along environmental gradients in space and time (Whittaker et al. 1975; Pulliam 2000; Hubbell 2001; Wright 2002). Thus, each species occupies a specific niche formed by a combination of environmental conditions (light, soil factors) that allow their establishment and survival (Pulliam 2000). However, recent studies indicate that dispersal limitation is an additional determinant of the spatial distribution of tree species (e.g. Svenning 2001; Vormisto et al. 2004; Svenning et al. 2006; John et al. 2007; Bohlman et al. 2008) that may contribute to the co-existence of species by preventing the local dominance of a single species. The degree to which dispersal limitation vs habitat specialization drives species spatial distributions along environmental gradients is currently an active subject of scientific investigation.

A simple indicator of the importance of niche differentiation is the percentage of species having a significant

association with different habitats or environmental factors. However, studies of habitat associations in tropical forest trees have yielded highly variable results in terms of the percentage of species with significant habitat associations (Appendix S1). Possible causes for this inconsistency include differences among studies in: (1) the set of environmental factors examined, which vary from soil chemistry to soil geologic origin and parental material (Appendix S1); (2) the statistical techniques used to detect associations, which range from Mantel tests and randomization tests to multivariate analyses (Itoh et al. 2010); (3) species-specific characteristics (e.g. body size, dispersal strategy; Nathan & Muller-Landau 2000); (4) the spatial scale of the study (Lam & Quattrochi 1992; Wu 2004; Cottenie 2005; McGill 2010), which varies from as little as 0.3 ha (Balvanera et al. 2002) to the 25- and 50-ha scale typically used for long-term forest dynamics plots (Harms et al. 2001; Russo et al. 2005) to as much as 158 ha (Phillips et al. 2003).

Spatial scale (i.e. the spatial extent of the area studied, rather than the level of detail, resolution or 'grain' of sampling; van Gemerden et al. 2005) is especially noteworthy because different environmental factors show heterogeneity at different spatial scales (Whittaker et al. 2001; Kneitel & Chase 2004; Snyder & Chesson 2004; Ricklefs 2008; Wu & Li 2009). At large scales (ca. 1000–10 000 ha), species distributions may reflect climatic gradients (Rhode 1992), whereas at small scales (<10), they may rather reflect soil-related heterogeneity with respect to nutrient or water availability (which may be partially driven by individual trees) and heterogeneity in canopy openness (Keddy 1982; Ceccon et al. 2003). Spatial heterogeneity at intermediate scales (ca. 10–1000 ha) may be caused by environmental factors (Clark et al. 1995), such as geology, topography and historical events (e.g. forest management history). To date, no study has evaluated how the sampling scale, ranging from local to intermediate spatial scales, affects the detection of habitat associations.

Differences in dispersal capacity of species can also strongly affect the spatial structure of communities, especially at intermediate scales. This can further confound the detection of environmental effects (Duque et al. 2002; Gilbert & Lechowicz 2004; Alonso et al. 2006). Seed dispersal patterns determine the initial potential distribution of individuals, but if habitat characteristics are also important for survival, the final distribution will resemble the potential distribution modified by the habitat requirements of the species. To discriminate between the effects of environmental factors and dispersal limitations, it is essential to incorporate them in appropriate statistical models (Nathan & Muller-Landau 2000).

In this paper, we consider the effect of sampling scale on the detection of habitat associations in the tropical moist forest of Barro Colorado Island (BCI), Panama. Four studies of habitat association have previously been conducted on BCI; these used different sampling scales and have drawn different conclusions. The two studies of the well-studied BCI 50-ha plot (Harms et al. 2001; John et al. 2007) found that 64% and 29% of 171 and 75 tree species, respectively, had significant habitat associations and that topography and soil chemistry were important factors. Two studies at the scale of the whole island (Svenning et al. 2004, 2006), which sampled 32 and 7 ha of forest, respectively, found that 25% and 68% of the 94 and 26 species studied, respectively, had significant habitat associations and that forest age was an important factor (Appendix S1). Focusing on five tree species, the work presented here not only encompasses the spatial scales of these previous studies, but also uses a much larger range of the size of forest sampled (i.e. 50, 100, 200, 400, 800 and 1600 ha) than in any previously published study. This is possible because we develop a map of all canopy-statured individuals for these five species across all of BCI using high-resolution aerial photographs.

We tested the hypothesis that the percentage of species with significant habitat associations varies with sampling scale. In particular, we expect the percentage of significant associations to increase with sampling scale (from 50 to 1600 ha) as greater heterogeneity in environmental variables is captured as well as a greater number of individual trees are sampled. We expect that the number of significant habitat associations detected will decrease across all spatial scales if species-specific clumping, likely arising from dispersal limitation, has been taken into account. This decrease should be highest at small spatial scales (≈50–200 ha) where such clumping is the highest. Support for our overall expectation that detection of habitat association varies with scale would indicate that sampling scale and species-specific aggregation patterns need to be explicitly considered when evaluating the relative importance of niche- vs neutral-based explanations for tropical forest biodiversity.

Methods

Study site

Barro Colorado Island (BCI), Panama (9.9°N, 79.51°W), is a 1560-ha island that became isolated from the surrounding mainland tropical forest during 1910–1914, when the Chagres River was dammed to form the central part of the Panama Canal (Leigh 1999). The island is covered with tropical moist forest and has a mean annual rainfall of 2623 mm, with a 4-month wet season from December to April (Leigh 1999). BCI is home to a 50-ha forest dynamics plot, established in 1982 (Condit 1998), and the island has been the focus of many stud-

ies, including vegetation structure, geology, hydrology, soil dynamics and tropical ecology (Losos & Leigh 2004).

The island is heterogeneous at various spatial scales and in several aspects including forest age, soil type and topography. About half of the forest is 90–135 yrs old while the remaining is old-growth forest (Enders 1935; Leigh 1999; Fig. 1). BCI has a variety of soil types that vary systematically with the type of underlying rock and the topography (Leigh 1999). The island hosts two main geologic formations dating from the Oligocene, known as Bohio and Caimito, where the latter is separated into marine and volcanic facies (Fig. 1). The top of the island is covered with igneous Andesite flows (Johnson & Stallard 1989; Barthold et al. 2008), the only non-sedimentary lithology. Moreover, there is a large variation in slope, and the elevation ranges from 27 to 160 m a.s.l. (Svenning et al. 2004), a type of variation that can be classified into five topography classes (Fig. 1).

Tree distributions

Across the entire study area, the spatial distributions of canopy-statured individuals of the five study species (three arborescent palms – *Attalea butyraceae, Astrocaryum standleyanum, Oenocarpus mapora* – and two large-canopy tree species – *Jacaranda copaia, Tabebuia guayacan*; Table 1) were assessed from high-resolution aerial photographs (Garzon-lopez et al. 2012). The species were chosen because they could be easily distinguished in the photos, were widely distributed across the island, and represented a wide spectrum of shade tolerance and dispersal characteristics, such as dispersal mode and seed size (Table 1).

The photographs were taken in April 2005 and April 2006. Flights were flown in overlapping north–south swathes at an altitude of 400 m in 2005, 700 m in 2006 and 800–1000 m in 2007. In 2005, each photo, on average, covered 8.6 ha, with a spatial resolution of 0.085 m·pixel^{-1}. In 2006, coverage and resolution averaged 15.9 ha and 0.114 m·pixel^{-1}. The aerial photographs were registered to a georeferenced Quickbird satellite image of BCI (DigitalGlobe, Longmont, CO, US), captured in March 2004, using the ERDAS IMAGINE v8.7 program (Leica Geosystems, Norcross, GA, US). Features visible in both aerial photos and satellite images, including large tree crowns and telemetry towers, were used as registration points for warping and georeferencing the individual photographs to produce an island-wide orthorectified mosaic.

Following the criteria of Trichon (2001), a key for identifying the crowns of the tree species was developed based on crown structure and known positions of trees of different species based on the stem map of the 50-ha forest dynamics plot (Condit 1998; Hubbell et al. 1999; Hubbell

2005). All aerial photos were surveyed to map canopy-statured individuals belonging to the five tree species. Validation conducted using 75 ha of ground-mapped distributions, including those in the 50-ha forest dynamics plot, confirmed that the distribution maps reflected the actual distributions of adult trees of each species (Garzon-lopez et al. 2012). The r^2 of the spatial regression between stem and crown maps was on average 0.67, increasing from 0.51 to 0.84 with grid sizes of 0.6–4 ha.
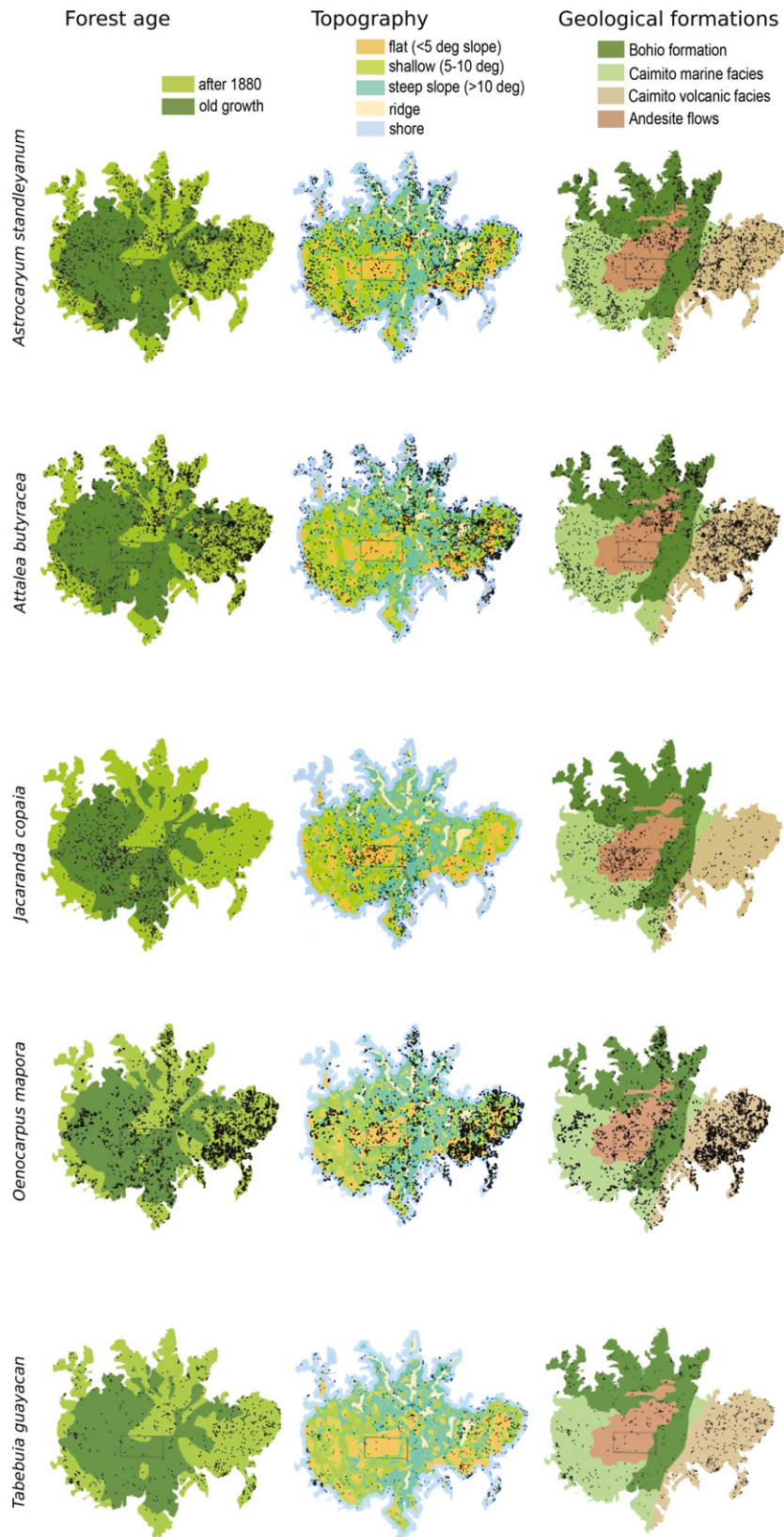
Environmental factors

Three environmental factors – geological formation, forest age and topography class – were used to test for associations with the distributions of canopy individuals (Table 2). These are a subset of the most common variables used in habitat association studies (Appendix S1). Geological formation (mapped at the 1:15000 scale) has a large impact on soil nutrient availability, as well as other soil-related characteristics (e.g. particle size, water retention capacity, pH). Forest age indicates different stages of succession, different canopy density and complexity, and liana coverage. Information on slope, elevation (mapped at a 1:25000 scale) and distance to shore were combined into one categorical topographic variable with five habitat types (Table 2) as a means to describe the biological response of evaluated species to topographical heterogeneity. By combining these factors, we were able to capture the variability in soil moisture, soil run-off and other hydrological factors, known to be collinear, using a single explanatory variable.

Determining the level of spatial aggregation and habitat associations

Sampling

We used six different sizes of rectangular plots representing increasing sampling scales (i.e. 50-, 100-, 200-, 400 and 800-ha, and the entire island, 1560 ha). Each sampling scale was replicated by placing polygons at randomly generated points, with the exception of the two largest categories. Overlap among plots at the same sampling scale was not allowed so that the same configuration of habitats and species distributions was not replicated in multiple plots. Overlap was avoided by randomly removing one of any pair of overlapping plots. The number of replicates was ten for 50- and 100-ha plots, five for 200-ha plots, three for 400-ha and one for the 800-ha plot and the whole island (1560 ha). Plots had a range of heterogeneity in the levels of environmental variables. Not every plot had each level of each environmental variable present. Just three plots (all 50-ha in size), however, had no heterogeneity in a particular environmental variable (two for geology; one for forest age). Additionally, in one case (one 50-ha plot sam-

Fig. 1. Spatial distribution of five tropical tree species on Barro Colorado Island, Panama, in relation to environmental factors.

Table 1. Characteristics of the tree species studied. Crown diameter is the mean sun-exposed crown diameter calculated from the aerial photos.

Species	*Attalea butyracea*	*Astrocaryum standleyunum*	*Jacaranda copaia*	*Tabebuia guayacan*	*Oenocarpus mapora*
Dispersal mode	Animal	Animal	Wind	Wind	Animal
Height (m)	30	20	30	30	20
Characteristics	Palm tree	Palm tree	Deciduous	Deciduous	Palm tree
Shade tolerance	Low	Medium	Low	Low	High
Individuals mapped	3121	2456	977	688	7555
Seed Size (cm)	4×2	3×2	2×1.8	3×1	2×2
Crown diameter (m)	8.9	6.4	15.1	17.3	3.8
Density (ha^{-1})	2.05	1.73	0.67	0.45	5.41

Table 2. Environmental variables considered in this study and their levels.

Variable	Levels	Description	Source
Geologic formation	Bohio volcanic		Barthold et al. (2008)
	Caimito volcanic		
	Caimito marine		
	Andesite		
Forest age	Secondary growth	regenerated after 1880	Enders (1935)
	Old growth	regenerated before 1880	
Topography	Shore	<150 m to shore	Johnson & Stallard (1989)
	Flat	>150 m to shore; <5° slope; <63 m elevation	
	Ridge	>150 m to shore; <5° slope>63 m elevation	
	Shallow	>150 m to shore; 5–10° slope	
	Steep	>150 m to shore; >10° slope	

ple for *D. panamensis*), a subsample contained less than ten individuals of the target species and was excluded from the analysis. Each replicate was evaluated as an independent sample.

Spatial aggregation

We used Ripley's *K(d)* function (representing the aggregation of trees at distance *d*) as a measure of spatial aggregation (Ripley 1976). Ripley's *K* measures the second-order properties of a spatial point process, and describes the way that spatial interactions change through space (Baddeley & Silverman 1984). We used Ripley's *K* instead of alternative summaries (e.g. mean nearest-neighbour distance or cumulative density function of distances from random points to their nearest neighbours) as it allows the simultaneous evaluation of cross-scale patterns (e.g. clustering at large scales and regularity at small scales) by assessing the relationship between the observed intensity (mean number of observed trees per unit area) and the expected intensity (maximum number of points at a given distance *d* assuming a mean density) of a point process. The *K(d)* function was transformed to $L(d)/d = (\sqrt{K(d)/\pi} - d)/d$ and

plotted against distance to determine deviations from a random distribution (Besag & Diggle 1977). Values of $L(d)/d > 1$ indicate clumping, and <1 indicate even distributions. We measured aggregation for the entire island and at each of the seven smaller spatial scales in order to determine the effect of sampling scale on estimates of *K(d)*. Both clump density (ρ, defined as the number of clumps per unit area) and clump size (σ, defined as the SD of the mean distance of an observation to the centre of the cluster, with 66% of the observations being at σ metres from the cluster centre) were estimated to describe the clumping characteristics of each species at each sampled scale. We used the approach of Plotkin et al. (2000) to estimate *K(d)* by assuming that its distribution follows a Poisson point process described by the function:

$$K(d)^{\mathrm{PCP}} = \pi d^2 + \rho^{-1}\left(1 - \exp\left(\frac{-d^2}{4\sigma^2}\right)\right) \qquad (1)$$

Parameters σ and ρ were estimated at each spatial scale using the 'pcp' function in the R-package 'splancs' (R Foundation for Statistical Computing, Vienna, AT). Clump size (σ) is expressed in meters and clump density (ρ) in clumps per m^2.

Simulations

To determine the effect of the environmental and dispersal constraints on the observed spatial patterns, we used two simulation approaches (*i.e.* complete spatial randomness – CSR and a Poisson cluster process – PCP) to determine if habitat associations were significant even when clumping patterns were taken into account. For the PCP simulations, we simulated the spatial distributions of adult trees based on the aggregation characteristics (σ and ρ) of each species. This analysis reduces the likelihood that 'artificial' habitat associations are found, which in reality arise from large, dispersal-limited clumps of trees occurring in one or a few specific habitat patches, causing a type of pseudo-replication effect.

First, we simulated 1000 distributions for each one of the subsamples with complete spatial randomness (CSR),

which is determined only by the number of trees. Second, we generated 1000 spatially clumped distributions at each spatial scale, assuming a Thomas cluster process (a special case of a Poisson Cluster Process; De Smith et al. 2007) in which spatial patterns are determined by species-specific clumping parameters determined from the Ripley's K analysis. For this, we implemented the approach of Plotkin et al. (2000), using both the clump density (ρ, as defined above) and size (σ, as defined above) estimated form $K(d)$ at each scale as input parameters. PCP distributions were simulated by placing a number of clusters equal to $\rho_i*A + 0.5$, where A corresponds to the area of the sample for the ith species. The number of stems per cluster was given by n_i/ρ_i*A, where n corresponds to the total number of individuals in species i. Simulations based on PCP resembled the observed distributions more closely than simulations with CSR, showing the importance of dispersal limitation as a basic determinant of species distributions (Appendix S2).

Habitat associations

To test for species–habitat associations at each scale, we used the Gamma test (Plotkin et al. 2000), which is a modified version of a Chi-square test. This metric quantified the likelihood that habitat associations were found at a particular sampling scale (i.e. the frequency of observed occurrences at a particular habitat level was significantly higher or lower than expected based on a CSR or a PCP model). A standard Chi-square test is not suitable in this case due to its assumption of independence in the locations of conspecific trees. The Gamma statistic is defined as the proportion of trees found on habitat type jth from the J available habitats, which is defined as $(n_{i,j}/n_i)$; where n_i is the number of trees of species i in the sample, and $n_{i,j}$ is the number of trees of species i in the sample occurring on environmental factor jth. We first quantified the strength of the habitat associations on a given level of an environmental variable (e.g. 'young forest' is a level within the 'forest age' environmental variable) within a single plot as the proportion of the 1000 comparisons to the CSR and PCP expectations for which observed occurrences exceeded the expected occurrence of the species on that habitat level (Gotelli & Graves 1996). We then determined the direction of the occurrences–environment associations (positive or negative) by comparing observed Gamma values to the distributions of Gamma values generated in 1000 PCP and CSR simulations. If observed Gamma values fell outside the 95% confidence interval of simulated Gamma values, habitat associations were considered significant and either negative (Observed$_{Gamma}$ < Simulated$_{Gamma}$) or positive (Observed$_{Gamma}$ > Simulated$_{Gamma}$). Our Gamma test differs from a traditional Chi-square test in that the expected

frequency of observations within each category is not based on a theoretical Chi-square distribution, but rather drawn from the expected frequencies derived from a process-based null expectation (*i.e.* CSR of a PCP process). This approach was chosen, as opposed to a general linear model-based approach, because it allowed the comparison between the expectation based on a completely random model (CSR) and the expectation of a model constrained by clumping patterns due to dispersal (i.e. PCP). We performed all the analyses using the Splancs and grdevices package in the R program.

Results

We mapped a total of 15 209 crowns of the five tree species, where the density ranged from 0.45 (*T. guayacan*) to 5.4 trees·ha^{-1} (*O. mapora*) (Table 1). Of the five species, *O. mapora* had the highest $L(d)/d$ value, indicating the strongest aggregation (individuals per clump), while *T. guayacan* had the lowest aggregation in the range from 0 to 1500 m (Appendix S2). *J. copaia* had a peak $L(d)/d$ at the largest distance (12 m) of all species, which suggested it had the largest clump size (σ; Appendix S2). Visual inspection of the island-wide distribution maps suggested that all species had biased/clustered spatial distributions, with more individuals occurring in some habitat types than in others (Fig. 1). For example, *O. mapora* was associated with ridges, *J. copaia* with old-growth forest and *A. butyracea* with secondary forest.

The spatial clumping parameters showed several consistent trends among species (Fig. 2). While the size of clumps and the number of trees per clump increased with sampling scale, the clumps per ha decreased with plot size (Fig. 2). The variance in clumps per ha decreased with plot size for all species, but the variance in cluster size and trees per clump showed no clear pattern with plot size (data not shown).

Habitat associations with CSR model

The probability of detecting associations increased with sampling scale (Fig. 3). When disaggregated across levels for an environmental variable (Fig. 4), the level of the environmental variables with the highest number of habitat associations differed among species, but tended to be consistent for each species across different sampling scales. For example, *A. butyracea* showed a consistent association with young forest, and the strength of this association increased with sampling scale, while *J. copaia* showed a consistent association with old-growth forest that increased with sampling scale.

At the scale of the entire island, there were more significant habitat associations (higher percentage of positive or
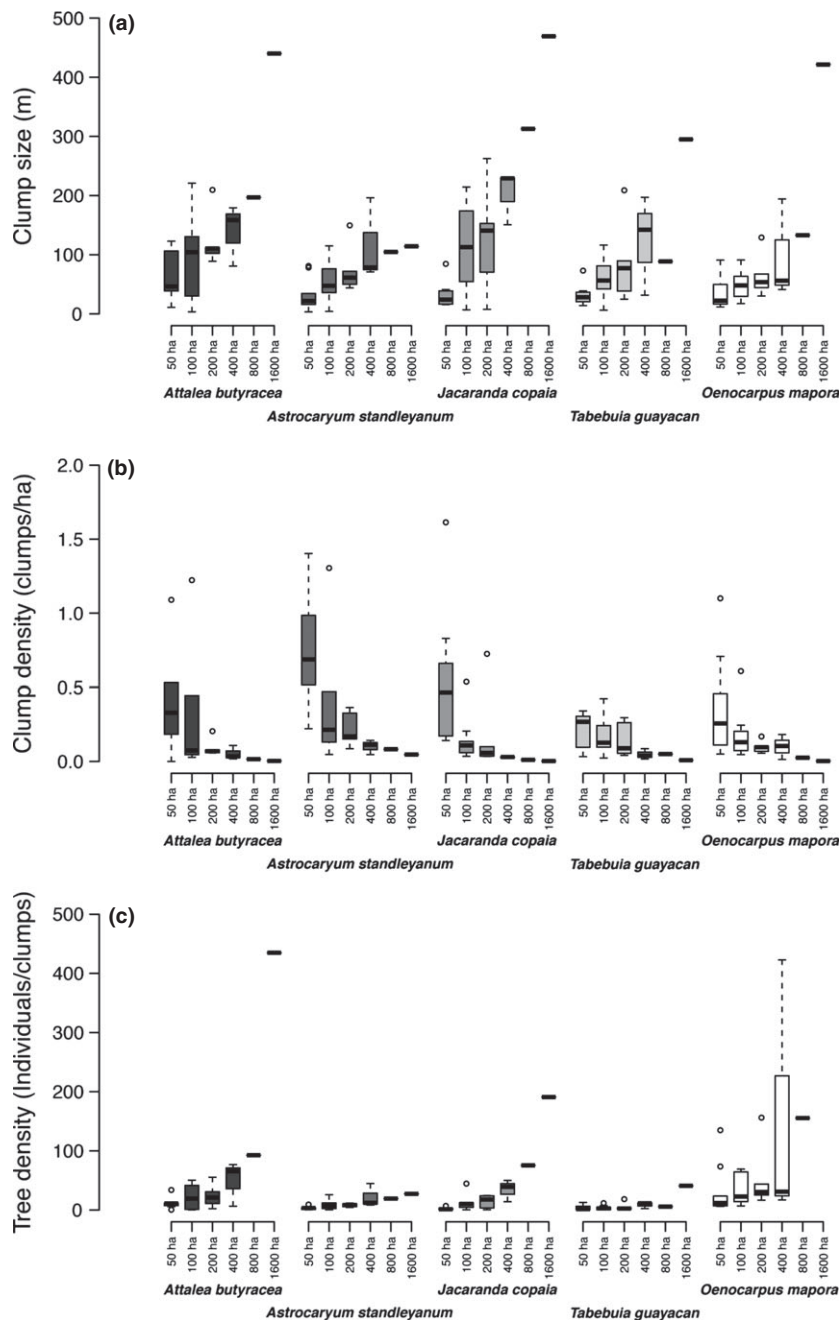
Fig. 2. Relationship with sampling scale of (**a**) clump size (m), (**b**) number of clumps per ha and (**c**) number of individuals per clump for five tropical tree species on Barro Colorado Island, Panama. Clumping parameters were determined by a Poisson cluster analysis. Box plots summarize the mean and interquartile variability across replicates at a particular sampling scale. Numbers of replicates per scale are 10 for 50 and 100 ha, six for 200 ha, three for 400 ha, and one for 800 and 1600 ha.

negative associations) than expected under complete spatial randomness (CSR). The percentage of samples that yielded significant habitat associations increased with sampling scale for all five species and for all environmental variables (Fig. 3; Appendix S3). At sampling scales ≥ 400 ha, all species had significant associations with

≥ 40% of the individual levels of the environmental variables (e.g. young vs old-growth forest for the forest age variable). The three palm species had significant associations, with ≥ 40% of levels within each environmental variables starting at plot sizes of 100 ha. *J. copaia* and *T. guayacan* had relatively few associations with
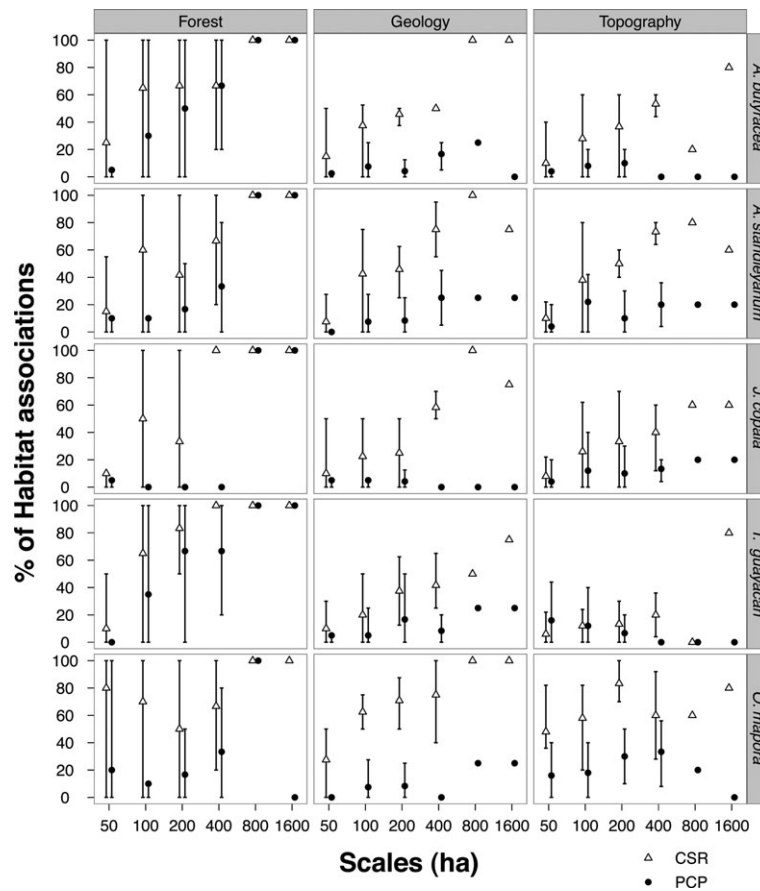
Fig. 3. Variation in habitat associations with different sampling scales assuming PCP (Poisson cluster process) and CSR (complete spatial randomness) in generating the simulated data sets for five tropical tree species on Barro Colorado Island, Panama. The y-axis (% habitat associations) is the mean number of levels (e.g. young forest within the forest age variable) within the given environmental variable (forest age, geology, topography) showing a significant association (either positive or negative) with the species distribution (Appendices S3 and S4). The error bars indicate the SE of significant associations among plots at the given sampling scale.

topography. For *T. guayacan*, the highest number of associations was with forest age.

Habitat associations on the basis of the PCP model

We found that the probability of detecting habitat associations also increased with sampling scale under the PCP model (Fig. 3), but the variables driving this pattern changed between scales (Fig. 5). The simulations under PCP, which incorporated species-specific aggregation patterns, yielded fewer habitat associations than CSR simulations (Fig. 3; Appendix S4). The decrease in associations from CSR to PCP simulations was highest for geological formations and topography (Fig. 3). The number of significant associations was nearly the same under CSR and PCP for forest age. For simulations under PCP, forest age had the most significant associations for nearly all the species, especially at large plot sizes. For association with geologic formation, and to a lesser degree for association with

topography, the divergence between PCP and CSR simulations increased at larger plot sizes. For *A. butyracea* and *O. mapora*, the PCP simulations showed a drop in the percentage of significant correlations at the largest spatial sizes. *O. mapora* showed significant reductions in the percentage of significant habitat associations between PCP and CSR simulations at all spatial scales. At the 800-ha to 1600-ha scales, *A. butyracea, A. standleyanum* and *J. copaia* showed the largest decreases in the percentage of habitat associations under PCP simulations.

With the CSR model, there was a high variability in the direction of habitat associations (positive or negative) within and among sampling scales (Appendix S3). Between the 50- and 400-ha sampling scales, there were 62 examples in which the same species showed both a positive and negative association with the same level of environmental variable between two plots of the same size. For example, there were both significantly positive and negative associations at the 100-ha and 200-ha scales
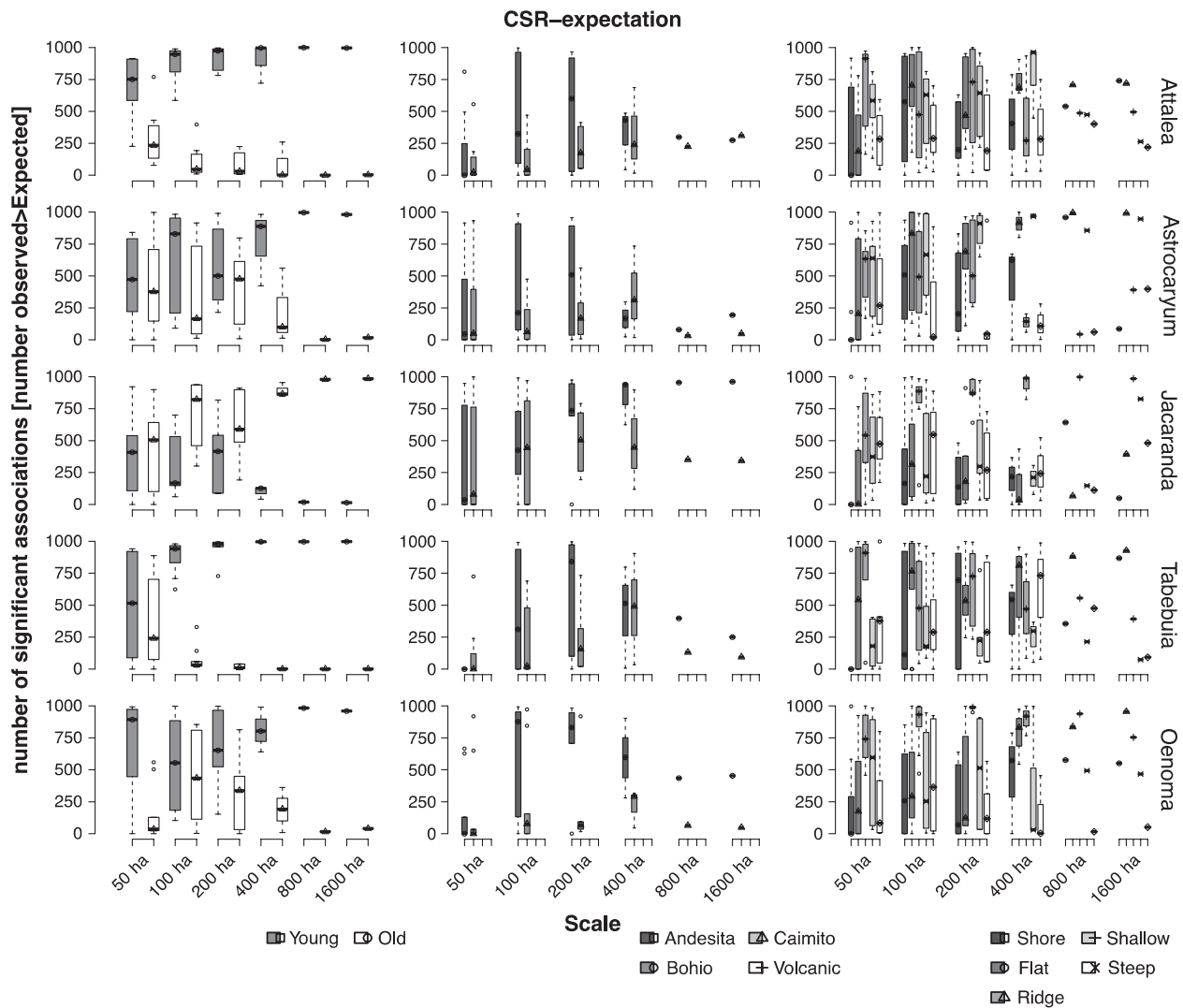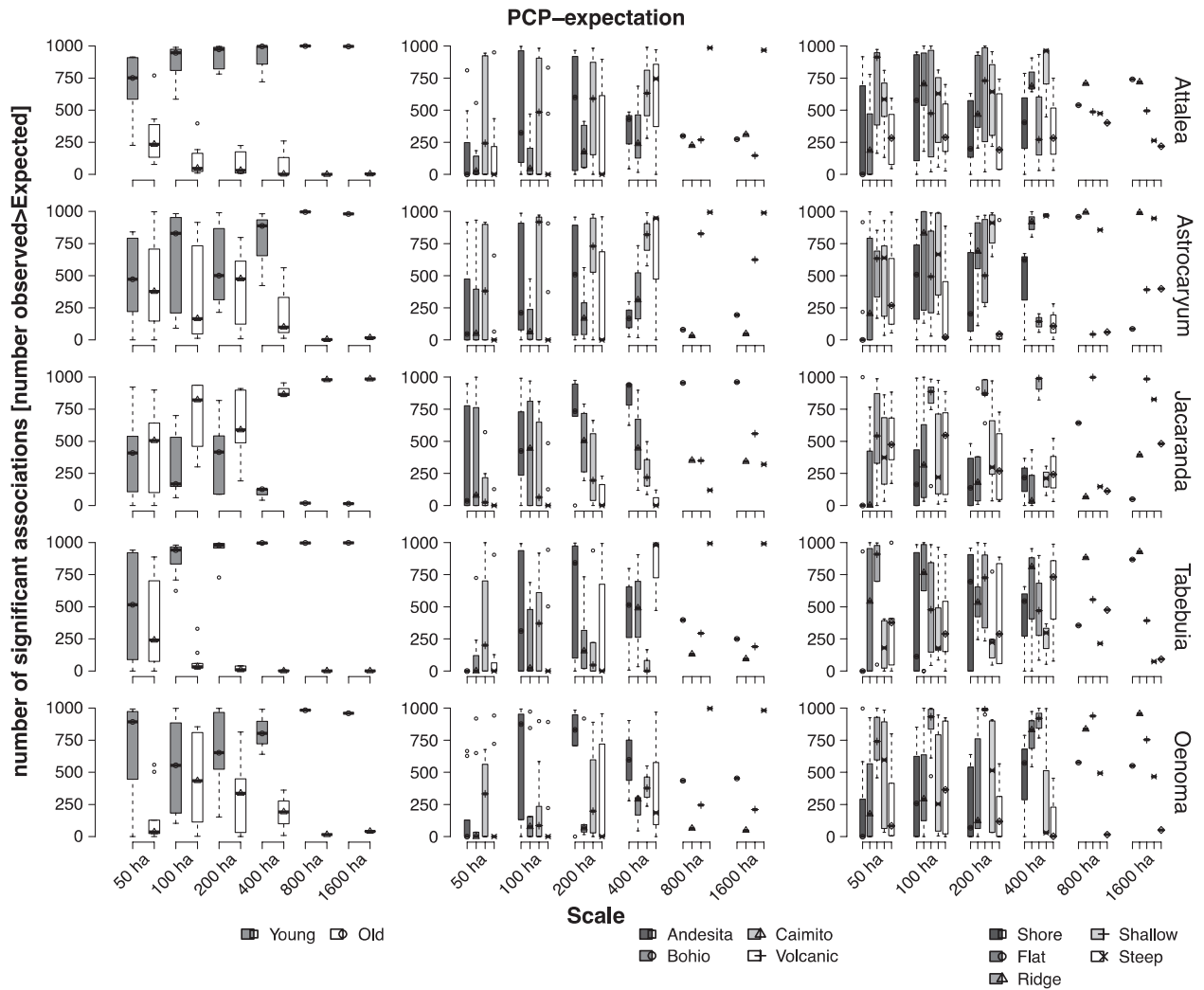
Fig. 4. Strength of habitat associations based on comparison to a complete spatial randomness (CSR) model. Habitat associations were determined using an observed vs expected approach based on comparison to 1000 CSR simulations for each level of the evaluated environmental variable. Strength of the relationship was measured as the number of the 1000 contrasts where observed occurrence on a certain habitat was higher than the expected occurrence on that habitat. Box plots summarize the median and interquartile variability across replicates at a particular sampling scale. Numbers of replicates per scale are 10 for 50 and 100 ha, six for 200 ha, three for 400 ha, and one for 800 and 1600 ha.

found for *A. butyracea* with the ridge habitat type, depending on plot location. In contrast, we found only six examples of this under the PCP model (Appendix S4). This indicates that clumping, most likely related to dispersal constraints, not only increases the number of significant habitat associations, but also does so in an arbitrary manner with respect to habitat type. This also indicates that the actual spatial placement of plots had an influence (via environmental heterogeneity, clustering characteristics, species density, etc.) on the observed association patterns. Forest age had consistently positive or consistently negative associations for both PCP and CSR at nearly all spatial scales and for all species. Geologic substrate and topogra-

phy had more instances of having both negative and positive associations for the same species within and between spatial scales, especially with the CSR simulations.

Under PCP simulations, several combinations of species and environmental variables showed some consistency within or among spatial scales in terms of direction and significance of association (Appendices S3, S5). All palm species plus *T. guayacan* were positively associated with the secondary forest and negatively associated with old-growth forest (Fig. 3). In contrast, *J. copaia* was positively associated with the old-growth forest and negatively associated with secondary forest. These associations with forest age were most consistent at the largest plot sizes

Fig. 5. Strength of habitat associations based on contrast to a Poisson cluster process (PCP) model. Habitat associations were determined by using an observed vs expected approach based on comparison to 1000 PCP simulations for each level of the environmental variables. Strength of the relationship was measured as the number of the 1000 contrasts where observed occurrence on a certain habitat was higher than the expected occurrence on that habitat. Box plots summarize the median and interquartile variability across replicates at a particular sampling scale. Numbers of replicates per scale are 10 for 50 and 100 ha, six for 200 ha, three for 400 ha, and one for 800 and 1600 ha.

(800–1600 ha). *O. mapora* had the most numerous habitat associations. It was significantly and positively associated with the Andesite and Caimito volcanic types, regrowth forest and the shore and ridge habitats, but negatively associated with the old-growth forest and steep habitats.

Discussion

Habitat association is a distinctive feature of niche specialization (Harms et al. 2001). However, comparisons across previous studies yielded many inconsistencies in the number and strength of habitat associations (Appendix S1; Cottenie 2005). The results of our study strongly suggest that an important part of these inconsistencies may be

attributed to the variation in the sampling scale across studies and/or to failure to account for species-specific aggregation patterns that are typically due to dispersal limitation. For the five tree species evaluated in this work, the percentage of detected habitat associations varied substantially with the sampling scale. However, differences among plot sizes for a particular species decreased when species-specific clumping patterns had been taken into account. Changes in the percentage of habitat associations depended on the environmental factor considered. A consistently large number of associations were observed across spatial scales for forest age (especially spatial scales ≥ 400 ha), both with and without incorporating dispersal limitation, but not for geological formation and topogra-

phy. This result is consistent with Svenning et al. (2004), who found that forest age had the highest number of significant relationships with plant distributions on BCI of all environmental variables considered. The fact that for some species and environmental factors, no associations were found at particular spatial scales (especially smaller sampling scales where placement of plots is most critical) does not necessarily mean that no relationship exists. Rather, consistent results showing habitat associations for canopy-statured tropical forest species may require large sampling scales (>200 ha). This indicates the scale of the study should match the scale at which the variables of interest vary across space (in this case, the largest plot sizes).

Across plots with a 50-ha size, the standard in studies of tropical tree diversity and distribution patterns, there was much variation in environmental and species arrangement. The mean percentage of significant habitat associations was lowest for the 50-ha plot scales for nearly all combinations of species and environmental factors. Variance in percentage of significant habitat associations among 50-ha plots was also high. This variability indicates that the particular placement of a single plot can determine the results of a study. Therefore, 50 ha may not be a suitable scale for identifying consistent trends of habitat associations that could be applicable at landscape scales for many species and environmental variables, especially for the canopy-statured species and the environmental variables (forest age, geology and topography, which vary on local to intermediate spatial scales) studied here. Additionally, different plot sizes may yield entirely opposite conclusions on the importance of niche differentiation, especially if clumping characteristics are not adequately characterized. Analyses at larger spatial scales may reveal habitat associations that are not detectable at smaller scales, simply because the chance for many factors to vary at small scale is limited.
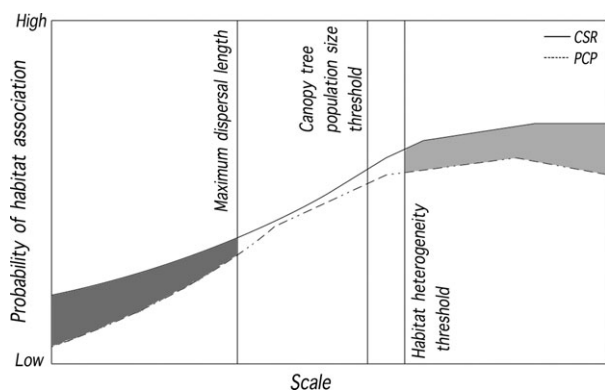
The point pattern analysis used in this study identified pronounced species-specific aggregation patterns, in agreement with Condit et al. (2000) and Plotkin et al. (2002). However, the calculated clumping characteristics changed when larger sampling scales were considered. In general, larger sample scales yielded larger clump sizes. This probably occurred because large clumps are relatively rare, and thus more likely picked up at the largest scales. If there were a single set of 'best' clump characteristics (clump size, clumps per ha, individuals per clump; Fig. 2) for these species, we would expect clumping characteristics to remain the same for at least several adjacent sampling scales, which was clearly not the case. Also at smaller plot sizes, large clumps may be quantified inaccurately due to boundary effects. The variation in clumping characteristics with sampling size suggests different mechanisms are causing spatial aggregation depending on the spatial scale. For example, different dispersal agents may cause recruits to aggregate at different scales. For animal-dispersed seeds, seeds may initially clump near parent trees. For species with multiple animal dispersers, such as *A. butyracea*, different dispersers (agoutis, squirrels) may move seeds different distances from the parent tree. Also, tree-fall gap formation may generate clumps of characteristics sizes, especially for light-demanding species. Four of the five species in this study have low shade tolerance, and thus recruitment is elevated in tree-fall gaps (Table 1). Previous studies (Plotkin et al. 2002; Wiegand et al. 2007, 2009) have quantified clustering properties at different spatial scales, but all within plots of ≤ 50 ha. Our results suggest different critical clustering sizes continue to develop at larger spatial scales and imply that it would be useful in future studies to apply techniques such as those of Plotkin et al. (2002) and Wiegand et al. (2007) to landscape-scale data sets, so as to analyse a larger range of spatial patterns and their underlying causes.

Some of the clump sizes were quite large at the largest spatial scales, ranging from 120 m for *A. standleyanum* to 500 m for *J. copaia* at the whole-island scale (1560 ha). It is debatable whether these large clump sizes represent dispersal limitation ('self-organized clumping') or clumping due to habitat associations ('imposed clumping'). However, it must be remembered that the trees studied here are canopy-sized individuals with large crowns that exist at low densities interspersed mainly with crowns of other species. Therefore, a clump of 20 trees of one of the target species may cover many hectares. However, *J. copaia* at the plot size ≥ 400 ha had clump sizes exceeding 200 m and more than 40 trees per clump. In these cases, the PCP method may be characterizing clumping patterns related to habitat associations rather than dispersal.

Once the effect of spatial aggregation (predominantly due to dispersal limitation) was accounted for, habitat associations that emerged in tests under CSR often disappeared. While some previous studies of habitat association have treated trees as independent units without accounting for the natural clumping that may arise from dispersal limitation (e.g. Baillie et al. 1987; Clark et al. 1999; Dalle et al. 2002; Blundell & Peart 2004; Eilu et al. 2004; Costa et al. 2005), recent studies that do include dispersal limitation find that aggregation patterns have a strong effect on the number of plant–habitat associations found for a species (Plotkin et al. 2000; Svenning et al. 2006; Bohlman et al. 2008; Leithead et al. 2009). Unique to this study is evidence of how clumping patterns operate at multiple spatial scales. In fact, we found the largest reduction in habitat associations between CSR and PCP occurred at the largest spatial scales, which were rarely used in previous studies.

We propose the following conceptual model (Fig. 6) that integrates the probability of detecting habitat associations based on (1) species-specific scales of dispersal and population density; (2) scale-dependent heterogeneity of the environmental variable being studied; and (3) whether species- and scale-dependent clumping patterns are included in the analysis of habitat associations (in this case, CSR vs PCP model). The probability of detecting habitat associations increases with sampling scale as the amount of habitat heterogeneity and number of individuals of the target species increases until thresholds are reached for both habitat heterogeneity (specific to the spatial arrangement of the environmental variable under study) and number of individuals evaluated (dependent of the spatial arrangement and density of the tree species studied). Above these thresholds, the probability of detecting habitat associations levels off. Below the maximum dispersal length of the target species, CSR will have a higher probability of detections than PCP because CSR will be erroneously attributing some clumping patterns due to dispersal to habitat associations. Above the habitat heterogeneity threshold, some of the clumping patterns quantified by PCP may not be related to dispersal, but rather to habitat association, thus decreasing the probability of detecting habitat associations for PCP vs CSR. Additional studies are needed to distinguish whether clumping patterns, especially those at large spatial scales, are due to dispersal characteristics or habitat heterogeneity.

Our study supports the emerging view that the spatial distribution of tropical tree species involves both niche differentiation and dispersal limitation, where relative importance of both factors varies with spatial scale. Rather than being two contrasting processes, niche differentiation and dispersal limitation interact with each other. This implies that dispersal limitation needs to be accounted for when niche differences are studied and *vice versa*.

Acknowledgements

References

Alonso, D., Etienne, R.S. & McKane, A.J. 2006. The merits of neutral theory. *Trends in Ecology & Evolution* 21: 451–457.

Baddeley, A.J. & Silverman, B.W. 1984. A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics* 40: 1089–1093.

Baillie, I.C., Ashton, P.S., Court, M.N., Anderson, J.A.R., Fitzpatrick, E.A. & Tinsley, J. 1987. Site characteristics and the distribution of tree species in mixed Dipterocarp forest on tertiary sediments in Central Sarawak, Malaysia. *Journal of Tropical Ecology* 3: 201–220.

Balvanera, P., Lott, E., Segura, G., Siebe, C. & Islas, A. 2002. Patterns of β-diversity in a Mexican tropical dry forest. *Journal of Vegetation Science* 13: 145–158.

Barthold, F.K., Stallard, R.F. & Elsenbeer, H. 2008. Soil nutrient–landscape relationships in a lowland tropical rainforest in Panama. *Forest Ecology and Management* 255: 1135–1148.

Besag, J.E. & Diggle, P.J. 1977. Simple Monte Carlo tests for spatial pattern. *Applied Statistics* 26: 327–333.

Blundell, A.G. & Peart, D.R. 2004. Density-dependent population dynamics of a dominant rain forest canopy tree. *Ecology* 85: 704–715.

Bohlman, S., Laurance, W., Laurance, S.G., Nascimento, H.E.M., Fearnside, P.M. & Andrade, A. 2008. Importance of soils, topography and geographic distance in structuring central Amazonian tree communities. *Journal of Vegetation Science* 19: 863–874.

Ceccon, E., Huante, P. & Campo, J. 2003. Effects of nitrogen and phosphorus fertilization on the survival and recruitment of seedlings of dominant tree species in two abandoned tropical dry forests in Yucatán, Mexico. *Forest Ecology and Management* 182: 387–402.

Fig. 6. Conceptual model for the detection of habitat associations in relation to sampling scale. The probability of detecting habitat associations increases with sampling scale, as the amount of habitat heterogeneity and population size of trees increases, until a threshold is reached for both tree population size and habitat heterogeneity, at which the probability of detecting habitat association levels off. Below the maximum dispersal length (dark grey area), CSR will yield a higher probability of detections than PCP because CSR will attribute some clumping patterns to habitat associations rather than dispersal limitation. Above the habitat heterogeneity threshold (light grey area), PCP will start to attribute some habitat associations to dispersal-related clumping, thus decreasing the probability of detecting habitat associations for PCP vs CSR.

Clark, D.A., Clark, D.B., Sandoval, R. & Castro, M.V. 1995. Edaphic and human effects on landscape-scale distributions of tropical rain forest palms. *Ecology* 76: 2581–2594.

Clark, D.B., Palmer, M.W. & Clark, D.A. 1999. Edaphic factors and the landscape-scale distributions of tropical rain forest trees. *Ecology* 80: 2662–2675.

Condit, R. 1998. *Tropical forest census plots: methods and results from Barro Colorado Island*. Springer, Berlin, DE.

Condit, R., Ashton, P.S., Baker, P.S., Buyavejchewim, S., Gunatilleke, S., Gunatilleke, N., Hubbell, S.P., Foster, R.B., Itoh, A. (...) & Yamakura, T. 2000. Spatial patterns in the distribution of tropical tree species. *Science* 288: 1414–1418.

Correa, M.D.A., Galdames, C. & Stapt, M. 2004. *Catálogo de las plantas vasculares de Panamá*. Quebecor World Bogota, S.A., Bogota, Colombia.

Costa, F.R.C., Magnusson, W.E. & Luizao, R.C. 2005. Mesoscale distribution patterns of Amazonian understory herbs in relation to topography, soil and watersheds. *Journal of Ecology* 93: 863–878.

Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters* 8: 1175–1182.

Dalle, S.P.B., Lopez, H., Diaz, D., Legendre, P. & Potvin, C. 2002. Spatial distribution and habitats of useful plants: an initial assessment for conservation on an indigenous territory, Panama. *Biodiversity and Conservation* 11: 637–667.

De Smith, M.J., Goodchild, M.F. & Longley, P.A. 2007. *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Troubador Publishing, Leicester, UK.

Duque, A., Sanchez, M., Cavelier, J. & Duivenvoorden, J.F. 2002. Different floristic patterns of woody understorey and canopy plants in Colombian Amazonia. *Journal of Tropical Ecology* 18: 499–525.

Eilu, G., Hafashimana, D.L.N. & Kasenene, J.M. 2004. Density and species diversity of trees in four tropical forests of the Albertine rift, western Uganda. *Diversity and Distributions* 10: 303–312.

Enders, K. 1935. Mammalian life histories from Barro Colorado Island, Panama. *Bulletin of the Museum of Comparative Zoology* 78: 385–502.

Garzon-lopez, C.X., Bohlman, S.A., Olff, H. & Jansen, P.A. 2012. Mapping tropical forest trees using high-resolution aerial digital photographs. *Biotropica*. 45: 308–316.

Gilbert, B. & Lechowicz, M.J. 2004. Neutrality, niches, and dispersal in a temperate forest understory. *Proceedings of the National Academy of Sciences of the United States of America* 101: 7651–7656.

Givnish, T.J. 1999. On the causes of gradients in tropical tree diversity. *Journal of Ecology* 87: 193–210.

Gotelli, N.J. & Graves, G.R. 1996. *Null models in ecology*. Smithsonian Institution, Washington, DC, US.

Harms, K.E., Condit, R., Hubbell, S.P. & Foster, R.B. 2001. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology* 89: 947–959.

Hubbell, S.P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ, US.

Hubbell, S.P. 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology* 19: 166–172.

Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J. & de Lao, S. 1999. Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* 283: 554–557.

Itoh, A., Ohkubo, T., Nanami, S., Tan, S. & Yamakura, T. 2010. Comparison of statistical tests for habitat associations in tropical forests: a case study of sympatric dipterocarp trees in a Bornean forest. *Forest Ecology and Management* 259: 323–332.

John, R., Dalling, J.W., Harms, K.E., Yavitt, J.B., Stallard, R.F., Mirabello, M., Hubbell, S.P., Valencia, R., Navarrete, H., Vallejo, M. & Foster, R.B. 2007. Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences of the United States of America* 104: 864–869.

Johnson, M.J. & Stallard, R.F. 1989. Physiographic controls on the composition of sediments derived from volcanic and sedimentary terrains on Barro Colorado Island, Panama. *Journal of Sedimentary Petrology* 59: 768–781.

Keddy, P.A. 1982. Population ecology on an environmental gradient: *Cakile edentula* on a sand dune. *Oecologia* 52: 348–355.

Kneitel, J.M. & Chase, J.M. 2004. Trade-offs in community ecology: linking spatial scales and species coexistence. *Ecology Letters* 7: 69–80.

Lam, N.S. & Quattrochi, D.A. 1992. On the issues of scale, resolution, and fractal analysis in the mapping sciences. *The Professional Geographer* 44: 88–98.

Leigh, E.G. 1999. *Tropical forest ecology: a view from Barro Colorado Island*. Oxford University Press, Oxford, UK.

Leithead, M.D., Anand, M. & Deeth, L. 2009. A synthetic approach for analyzing tropical tree spatial patterns through time. *Community Ecology* 10: 45–52.

Losos, E. & Leigh, E.G. 2004. *Tropical forest diversity and dynamism*. University of Chicago Press, Chicago, IL, US.

McGill, B.J. 2010. Matters of scale. *Science* 328: 575–576.

Nathan, R. & Muller-Landau, H.C. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution* 15: 278–285.

Phillips, O.L., Nuñez, P., Monteagudo, A.L., Cruz, A.P., Chuspe, M., Sanchez, W.G., Yli-Halla, M. & Rose, S. 2003. Habitat association among Amazonian tree species: a landscape scale approach. *Journal of Ecology* 91: 757–775.

Plotkin, J.B., Potts, M.D., Leslie, N., Manokaran, N., LaFrankie, J. & Ashton, P.S. 2000. Species–area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology* 207: 81–99.

Plotkin, J.B., Chave, J. & Ashton, P.S. 2002. Cluster analysis of spatial patterns in Malaysian tree species. *The American Naturalist* 160: 629–644.

Pulliam, H.R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3: 349–361.

Rhode, K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* 65: 514–527.

Ricklefs, R.E. 2008. Disintegration of the ecological community. *The American Naturalist* 172: 741–750.

Ripley, B. 1976. The second order analysis of stationary point processes. *Journal of Applied Probabilities* 13: 255–266.

Russo, S.E., Davies, S.J., King, D.A. & Tan, S. 2005. Soil-related performance variation and distributions of tree species in a Bornean rain forest. *Journal of Ecology* 93: 879–889.

Snyder, R.E. & Chesson, P. 2004. How the spatial scales of dispersal, competition and environmental heterogeneity interact to affect coexistence. *The American Naturalist* 164: 633–650.

Svenning, J.C. 2001. Environmental heterogeneity, recruitment limitation and the mesoscale distribution of palms in a tropical montane rain forest (Maquipucuna, Ecuador). *Journal of Tropical Ecology* 17: 97–113.

Svenning, J.C., Kinner, D.A., Stallard, R.F., Engelbrecht, B.M.J. & Wright, S.J. 2004. Ecological determinism in plant community structure across a tropical forest landscape. *Ecology* 85: 2526–2538.

Svenning, J.C., Engelbrecht, B.M.J., Kinner, D.A., Kursar, T.A., Stallard, R.F. & Wright, S.J. 2006. The relative roles of environment, history and local dispersal in controlling the distributions of common tree and shrub species in a tropical forest landscape, Panama. *Journal of Tropical Ecology* 22: 575–586.

Trichon, V. 2001. Crown typology and identification of rainforest trees on large-scale aerial photographs. *Plant Ecology* 153: 301–312.

van Gemerden, B.S., Etienne, R.S., Olff, H., Hommel, P.W.F.M. & van Langevelde, F. 2005. Reconciling methodologically different biodiversity assessments. *Ecological Applications* 15: 1747–1760.

Vormisto, J., Svenning, J.C., Hall, P. & Balslev, H. 2004. Diversity and dominance in palm (Aracaceae) communities in *terra firme* forests in the western Amazon basin. *Journal of Ecology* 92: 577–588.

Whittaker, R.H., Levin, S.A. & Root, R.B. 1975. Reasons for distinguishing niche, habitat and ecotope. *The American Naturalist* 109: 479–482.

Whittaker, R.J., Willis, K.J. & Field, R. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography* 28: 247–254.

Wiegand, T., Gunatilleke, S. & Gunatilleke, N. 2007. Species associations in a heterogeneous Sri Lankan Dipterocarp forest. *The American Naturalist* 170: 77–95.

Wiegand, T., Martinez, I. & Huth, A. 2009. Recruitment in tropical tree species: revealing complex spatial patterns. *The American Naturalist* 174: 106–140.

Wright, J.S. 2002. Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia* 130: 1–14.

Wu, J. 2004. Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecology* 19: 125–138.

Wu, H. & Li, Z.-H. 2009. Scale issues in remote sensing: a review on analysis, processing and modeling. *Sensors* 9: 1768–1793.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Methods and outcomes of published studies on habitat associations of tropical forest trees.

Appendix S2. Ripley's *K* functions for five species on Barro Colorado Island, Panama.

Appendix S3. Table of habitat associations in CSR-generated simulations.

Appendix S4. Table of habitat associations for PCP-generated simulations.