

# Determinants and spatial modeling of tree $\beta$ -diversity in a tropical forest landscape in Panama

Chust, Guillem<sup>1,2</sup>; Chave, Jérôme<sup>1\*</sup>; Condit, Richard<sup>3</sup>; Aguilar, Salomon<sup>3</sup>;  
Lao, Suzanne<sup>3</sup> & Pérez, Rolando<sup>3</sup>

<sup>1</sup>Laboratoire Evolution et Diversité Biologique, CNRS – Université Paul Sabatier, Bâtiment 4R3, 118 route de Narbonne, 31062 Toulouse, France; <sup>2</sup>Present address: AZTI - Tecnalia / Marine Research Division, Herrera kaia portualdea z/g 20110 Pasaia (Gipuzkoa), Basque country, Spain; <sup>3</sup>Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002-0948, USA; \* Corresponding author; Fax +33 561557327; E-mail [chave@cict.fr](mailto:chave@cict.fr)

## Abstract

**Questions:** The objectives of this study were to clarify the extent to which environmental factors and geographical distance account for tropical floristic composition, and propose a methodology for delimiting the boundaries of floristic types based on species similarity.

**Location:** The Panama Canal watershed.

**Methods:** To assess which factors (climate, topography, geology and geographical distance) account for floristic composition, we performed Mantel tests on distance matrices and partitioned variation in species composition using canonical analysis. We used a permutation-based regression model computed on distance matrices and a hierarchical clustering of the tree composition to construct a predictive map of forest types of the Panama Canal Watershed.

**Results:** We found that spatial variation alone explained 22–27% of species variation, while the fraction of species variation explained by environmental variables was smaller (10–12%); 13–19% of the variation was accounted for by the joint effect of environmental variation and geographic distance. The similarity-based map emphasizes the principal division in tree flora between the drier Pacific side and the wetter Caribbean slopes.

**Conclusions:** The distribution of Panamanian tree species appears to be primarily determined by dispersal limitation, then by environmental heterogeneity. ‘Environmental segregation’ processes do play an important role. Maps of broad-scale vegetation patterns based on thorough tree inventories can be used in conservation planning in the tropics.

**Keywords:** Dispersal limitation; Neutral hypothesis; Panama; Remote sensing; Species similarity; Variation partitioning.

**Abbreviation:** PCW = Panama Canal Watershed.

## Introduction

Biogeographers and community ecologists have long debated whether the regional distribution of species is limited by dispersal or by environmental conditions (Ricklefs & Schluter 1994), but recent progress in Neotropical plant ecology has generated a great deal of attention on this issue (Duivenvoorden & Lips 1998; Pitman et al. 2001; Potts et al. 2002; Condit et al. 2002; Duivenvoorden et al. 2002; Ruokolainen & Tuomisto 2002; Slik et al. 2003; Tuomisto et al. 2003b; Cannon & Leighton 2004). In a neutral community, where all individuals are assumed to have the same prospects of reproduction and death, and with limited seed dispersal, floristic similarity among census plots is expected to decrease logarithmically with increasing geographic distance (Hubbell 2001; Chave & Leigh 2002). Condit et al. (2002) tested this theory of tree species turnover in three Neotropical forests, and they found that the data at hand were consistent with theory. Other studies, however, suggested that environmental factors should also be taken into account explain  $\beta$ -diversity of tropical trees, albeit with high collinearity with geographical distance (Duivenvoorden et al. 2002; Ruokolainen & Tuomisto 2002; Phillips et al. 2003).

In conservation biology, much effort has focused on preserving areas of exceptional local diversity, as measured by high values of species richness, rarity, or of endemism. Lesser attention has been paid on the issue of complementarity, that is, selecting areas such that the overlap in species composition across sites is minimized (Williams et al. 1996, 2000; Myers et al. 2000; ter Steege et al. 2003). This latter criterion is arguably more important in conservation planning than local diversity or endemism *per se*, as suggested by efforts to delimit complementary regions with optimization models (Cabeza & Moilanen 2001; Williams et al. 2004). Thus both empirical description and theoretical understanding of  $\beta$ -diversity are essential in the delimitation of priority

conservation areas (Rodrigues et al. 2004).

Our goals are to clarify the extent to which environmental factors and geographical distance account for floristic composition in the Panama Canal Watershed (PCW), and propose a general method to delimit the boundaries of floristic types based on tree species similarity. We assess the importance of different types of environmental factors (including topography, climate, and geologic types) and two predictors for dispersal limitation (geographic distance and habitat fragmentation) with Mantel tests and partitioned variation in species composition using canonical analysis. We then conducted a multiple regression of species similarity, adding remote sensing information as predictors, to map floristic types for the trees of the PCW. This map demonstrates the practical relevance of our study in conservation planning strategies, for instance in defining complementary areas. We also use our results to draw conclusions about the influence of niche-based versus dispersal-based mechanisms of diversity maintenance at the landscape scale.

## Methods

### *Study area, floristic data and environmental variables*

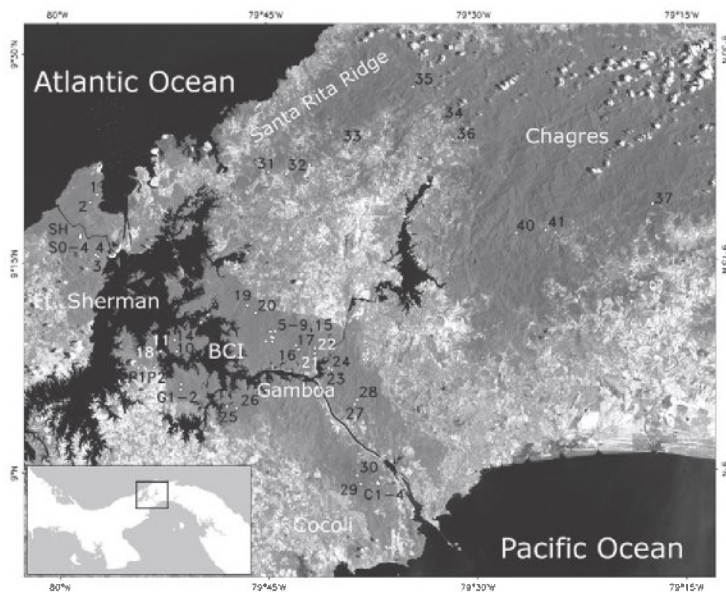
The Panama Canal Watershed is 3300 km<sup>2</sup> in area; it encompasses environmental gradients in climate, topography, and geology (Fig. 1). The forests near the Canal are < 200 m a.s.l., with rainfall up to 3100 mm.a<sup>-1</sup> on the Caribbean coast and 1600 mm.a<sup>-1</sup> in Panama City, on the Pacific side of the isthmus. Rainfall is also influenced by topography, and reaches 2400-4000 mm.a<sup>-1</sup> at

higher elevations. Soils are derived from a young and complex geologic terrain composed of either dense, relatively impermeable volcanic rocks or porous, chemically unstable sedimentary rocks and volcanic mud flow deposits (Dietrich et al. 1982).

Tropical forests do not extend continuously across the landscape of the PCW (Fig. 1). They are divided by the Canal, rivers, lakes artificially created in the early 20th century (Gatún, Alajuela), roadways and agricultural lands, mainly pasture. Land use practices have intensified since the beginning of the 20th century. From 1950 to 1990, the population in the PCW has increased fivefold (Sanjurjo et al. 1999).

Between 1996 and 2002, 53 permanent sampling forest plots were established across the PCW (Pyke et al. 2001; Condit et al. 2002, 2004; Chave et al. 2004) (Fig. 1). Plots vary in size, from 0.32 ha to 6 ha, but most of them ( $n = 35$ ) are 1 ha in size; in all analyses we divided the two largest plots of 6 ha and 4 ha into 1-ha squares and treated them as separate plots. Each of the 22 331 tree stems  $\geq 10$  cm DBH was mapped, tagged and identified to species or to morphospecies, for a total of 708 taxa; 154 individual trees remain unidentified. These plots were used to study the variation in floristic composition across the region (Pyke et al. 2001; Condit et al. 2004), spatial turnover in species composition (Condit et al. 2002), differential forest response to drought (Condit et al. 2000, 2004), and carbon storage in this area (Chave et al. 2004, 2005).

Each sampling site was characterized by a set of environmental variables: annual precipitation (mm.a<sup>-1</sup>), length of the dry season (in days), elevation (in m), geologic formation (four categories), and slope (Chave et al. 2004). We grouped the ten original geologic types



**Fig. 1.** The study area of the Panama Canal Watershed and location of the 53 floristic plots (white circles). The image was acquired by the Landsat TM satellite sensor on 27 March 2000. The mid infrared band shows water surfaces in black, urban and agricultural areas in light grey and brightest tones, respectively, and grey tones correspond to tropical forests. The inset image shows the location of the study area in Panama.

based on the United States Geological Survey map of the Panama Canal (Woodring et al. 1980) into four classes: Toro limestone, Chagres sandstone, basaltic formations, and ancient non-basaltic formations (including formations comprised of siltstone, sandstone, tuff and limestone; see App. 1). Elevation data were extracted from a digital elevation model at 90-m grid spacing, and at 1-m vertical intervals (source: U.S. Geological Survey, available online at <http://www.usgs.gov/>). These data were resampled to 30-m spatial resolution using a bilinear resampling method. Slope is an important parameter for soil formation since they affect soil depth. This field was derived from the elevation model with algorithms implemented in the GIS software Idrisi32 (Eastman 1999).

We acquired nine Landsat satellite images, from Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) sensors, all within the period of field sampling (1996-2002). The spatial resolution of TM and ETM+ sensors is 30 m for the visible and infrared wavelengths (bands 1-5 and 7). The images had systematically radiometric and geometric corrections (map projection: Universal Transverse Mercator, UTM), and were selected so as to provide the best comparable set of spectral values of all 53 plot sites. The only image providing unclouded spectral values for all plots was taken on 27.03.2000. This image was taken at the peak of the dry season, the optimum period to discriminate tropical vegetation types, as well as contrast between agricultural areas and forest (Condit et al. 2000; Bohlman 2004).

Spectral data for the six TM bands (1-5, 7) were averaged in a  $3 \times 3$  pixel square window (90 m  $\times$  90 m) centred on each tree plot, and they were used as independent predictors of diversity patterns. The study area is a complex matrix of old-growth forest and areas degraded by human activities (urban expansion and agricultural practices). We quantified the spatial extent of these cover classes by performing a supervised classification of land cover (Fig. 1). In the classification map, we discriminated forest (47% of the PCW area), shrubland (i.e. open forest, closed and open shrubland, 23%), pastures and agricultural land (10%), fresh water (12%), urban areas (1%), and clouds (5%). The reliability of our classification was assessed by ground truthing in two independent training sites. We used a maximum likelihood classification algorithm based on spectral bands, and a post-classification majority analysis to filter isolated pixels. As measures of classification error, we used the errors of commission and omission (Congalton 1991). The classification of forest cover had an estimated error of 1% of commission and 0.4% of omission (number of validation points for forest: 4753; total: 8867).

Published studies testing the relative importance of among-plot distance and environmental variation for predicting species similarity assumed plots were sepa-

rated by a continuous forest, and that straight-line distance is a suitable predictor of seed dispersal. Fig. 1 reveals that several of the Panamanian forest plots are surrounded by a mosaic of managed land. This is likely to reduce the exchange of tree propagules across plots, especially for animal-dispersed species, which make up most of the tree flora in the PCW (Croat 1978). The influence of habitat fragmentation on tree  $\beta$ -diversity was modelled by defining a measure of habitat contiguity between pairs of points in a landscape, called *cross-plot forest fraction*, or the fraction of forest present on a straight line separating any two plots.

#### *Statistical analysis of $\beta$ -diversity*

Given the set of  $n = 53$  plots and  $m = 708$  species, we computed a  $n \times n$  species similarity matrix. Each entry ( $i, j$ ) of the similarity matrix is an index that measures the pairwise compositional similarity between plot  $i$  and plot  $j$ . Two indices were used in this study. The Jaccard similarity index (for presence-absence data) is the number of species shared between the two plots, divided by the total number of species observed. The Steinhaus similarity index (for abundance data) is defined by  $2W/(A+B)$ , where  $W$  is the sum over all species of the minimum abundances (i.e. number of individuals) between the two plots of each species, and  $A$  and  $B$  are the sums of the abundances of all species at each of the two plots. Since plots vary in area, the number of individuals was divided by the plot area prior to calculating the Steinhaus index.

We used Mantel tests (Legendre & Legendre 1998) to determine the correlation between species similarity matrices and matrices of environmental and spectral similarity, geographic distance, and cross-plot forest fraction. This test computes a statistic  $r_M$  which measures the correlation between two matrices and results from the cross product of the matrix elements after normalizing. The statistic  $r_M$  is bounded between  $-1$  and  $+1$ , and behaves like a correlation coefficient. Since similarity or distance matrix entries are not independent, the Mantel statistic is tested by a non-parametric permutation test. Before performing the Mantel test, we converted both similarity measures (Jaccard and Steinhaus indices) to a distance measure ( $1 - \text{similarity}$ ). Distance matrices for spectral and environmental variables were defined as the Euclidean distance between values at two plots (i.e. absolute value of the difference if the quantity is a scalar), except for geologic types where a binary index was used for measuring similarity (taking a value 0 when the two plots present different geologic types, and 1 otherwise). Geographical distance among sampling sites was also measured by Euclidean distance.

We partitioned the variance of the floristic composition across sites to determine the relative contribution of environmental factors and spatial pattern (measured as polynomial of geographic coordinates). The total intersite variation in species abundances was decomposed into four components: pure effect of environment, pure effect of spatial pattern, combined variation due to the joint effect of environment and spatial pattern, and unexplained variation. The Mantel correlation, or its multivariate form (the multiple correlation statistic,  $R^2$ , computed for distance matrices) does not estimate a proportion of the variance of the species abundance table (Dutilleul et al. 2000). We therefore used a canonical analysis (ter Braak & Šmilauer 1998) to estimate a proportion of the variance of the original floristic table of abundances (plots by species); contrary to the Mantel test that predicts species similarity (plots by plots). According to Lapointe & Legendre (1994), Mantel test and canonical analysis provide similar results, with minor differences due to the loss of information in the transformation from raw data to distances.

The variance partitioning analysis proceeds in two steps. For each type of explanatory variable (i.e., environmental variables and spatial variables), we selected the significant variables by the forward procedure using Monte-Carlo permutations. Then, a partial canonical analysis was carried out to determine the relative contribution of environmental and spatial variables in accounting for species variation. More precisely, we used two canonical methods to partition species abundances: partial Correspondence Canonical Analysis (CCA, ter Braak & Šmilauer 1998); and partial distance-based Redundancy Analysis (db-RDA Legendre & Anderson 1999) with Hellinger-transformed data (Legendre & Gallagher 2001). We used two different canonical methods because in CCA, species are assumed to have unimodal response surfaces with respect to compound environmental gradients, while in db-RDA, species are assumed to have linear response surfaces. The unduly large influence of rare species in unimodal analysis has been reduced by downweighting species abundance values (ter Braak & Šmilauer 1998). In addition, we also performed these tests after log-transforming species abundance values. Spatial effects were modelled with third-degree polynomial terms of latitude ( $X$ ) and longitude ( $Y$ ):  $X$ ,  $Y$ ,  $X*Y$ ,  $X^2$ ,  $Y^2$ ,  $X^2*Y$ ,  $Y^2*X$ ,  $X^3$  and  $Y^3$  (cubic trend surface analysis, Legendre 1993). Each categorical geologic type was entered individually as a dummy variable. Cross-plot forest fraction was directly calculated from pairwise comparisons, so it could not be considered on the canonical analysis. The analysis was run with the canonical community ordination software CANOCO 4.0 (ter Braak & Šmilauer 1998).

To evaluate the effect of plot size, the variation

partitioning analysis was also carried out on a rarefied data set, by considering only data available from plots of 0.25 ha, or of 0.5 ha. We subdivided 1-ha plots into non-overlapping subsamples (four subsamples of 0.25 ha, and two of 0.5 ha), and used these as pseudo-replicates.

#### *Clustering the landscape into forest types*

We constructed a map of forest types of the PCW using the tree similarity values. Forest types were defined by hierarchical agglomerative clustering based on tree species abundance or occurrence, using a proportional-link linkage algorithm (Sneath 1966; connectedness level: 0.5). We then sought to predict the floristic composition of any given pixel in the landscape based on its environmental similarity to all the network of plots. To construct a predictive model of  $\beta$ -diversity, we performed a multiple regression on resemblance matrices (Legendre & Legendre 1998) that relates species similarity ( $S$ ) with the cross-site similarity in explanatory variables. Only those explanatory variables that contain raster spatial information across the tropical forest of the study area were considered, i.e. six spectral bands, elevation, slope, geologic types, habitat fragmentation and logarithm of geographical distance. The model was as follows:

$$\hat{S}(\text{plot } i, \text{plot } j) = B_0 + \sum_{l=1}^k B_l X_l \quad \forall i \neq j$$

where  $B_l$  are estimates of the regression coefficients,  $X_l$  are the similarity in explanatory variables, and  $k$  is the number of explanatory variables. The significance of the parameters in multiple regressions was tested by 999 permutations. Backward elimination was applied to select only statistically significant terms ( $P < 0.05$ ). The backward procedure tests different models by a stepwise elimination of explanatory variables, and selects the model where the multiple correlation coefficient ( $R$ ) is the highest and where only the significant partial regression coefficients are retained. Multiple regression analyses were run with Permute! 3.4 ( $\alpha$  version), available online at <http://www.fas.umontreal.ca/biol/casgrain/en/labo/permute/index.html>.

For each 30-m pixel of raster information of the study area, a value of the species similarity was estimated from the above model comparing the information of the pixel and each one of the  $n$  reference plots ( $n = 53$  in our case). Thus, we obtained  $n$  likelihood values  $\hat{a}$  (pixel  $i$ , plot  $j$ ), of species similarity for each pixel in the landscape. The maximum among these indicated which plot was the most similar to the pixel under study, and we decided that the pixel under study belonged in the same forest type as this plot. This procedure could allocate any forest pixel in the PCW to a forest type.

**Results**

*Floristic patterns*

The Jaccard index ranged from 0 to 0.67 among pairs of sites (mean = 0.18), and the Steinhaus index from 0 to 0.76 among pairs of sites (mean = 0.17), emphasizing the low similarity among plots (Condit et al. 2002). The Jaccard similarity was significantly correlated with all explanatory variables at the 5% level (Table 1). The most correlated variable was the logarithm of geographic distance ( $r = 0.69$ ). Climate and elevation also correlated highly with species similarity, especially precipitation ( $r = 0.57$ ), whereas slope and geology showed weak correlations ( $r = 0.32$ ,  $r = 0.13$ , respectively). All spectral bands correlated significantly with floristic similarity, with bands 1 and 4 showing the highest correlations ( $r = 0.31$ ,  $r = 0.26$ , respectively). Cross-plot forest fraction was also significantly correlated with floristic similarity ( $r = 0.32$ ).

Correlation results with the Steinhaus index showed similar patterns. Correlations with the Jaccard index were slightly higher than with the Steinhaus index, except for band 3, 5, 7 and geologic types. The two similarity indices were highly correlated with one another ( $r = 0.91$ ).

The two variants of canonical analysis (with untransformed and log-transformed abundances) produced similar results (Table 2). The fraction of unexplained variance was 45-50%, depending on the type of analysis. All nine spatial terms were significant and overall they explained 22-27% of variation in species abundances ( $p < 0.001$ ) when controlling for environmental variables. The fraction of variation in species abundances explained by environmental variables alone was signifi-

**Table 1.** Mantel's matrix correlation coefficients ( $r_M$ ) as measured with the Mantel test between the species similarity (both for Jaccard and Steinhaus indices) and the environmental similarity, geographical distance and cross-plot forest fraction. Statistical significance of each correlation coefficient was calculated using 999 permutations. The Jaccard and Steinhaus indices were converted to a distance measure (using  $1 - \text{similarity}$ ) before performing the Mantel test.

	Jaccard	P	Steinhaus	P
Steinhaus	0.906	0.001		
Dispersal-related factors				
Geographical Distance (GD)	0.637	0.001	0.502	0.001
ln(GD)	0.696	0.001	0.610	0.001
Cross-plot forest fraction	0.323	0.001	0.272	0.001
Environmental factors				
Elevation	0.424	0.001	0.348	0.001
Slope	0.318	0.001	0.227	0.001
Precipitation	0.572	0.001	0.499	0.001
Dry season	0.461	0.001	0.437	0.001
Geologic types	0.126	0.007	0.134	0.003
Spectral data				
Band 1	0.305	0.001	0.258	0.001
Band 2	0.117	0.042	0.116	0.019
Band 3	0.127	0.036	0.179	0.003
Band 4	0.258	0.001	0.224	0.001
Band 5	0.148	0.012	0.157	0.001
Band 7	0.160	0.016	0.186	0.002

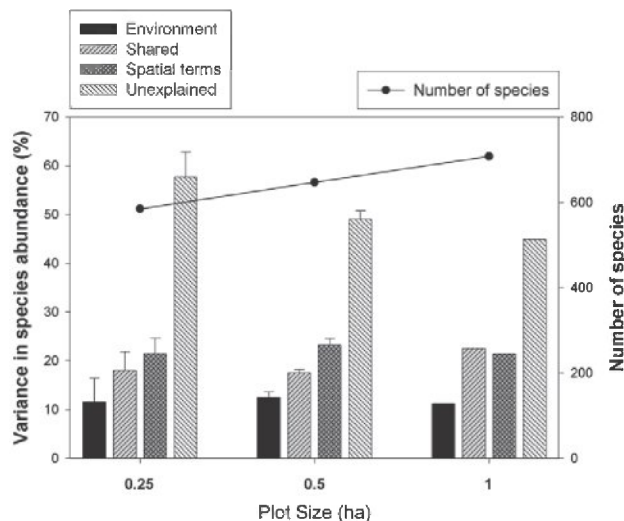
cant but less important than that of spatial terms (10-12%,  $p < 0.001$ ). The significance of environmental variables was fairly consistent and coincided in the different analyses: elevation, slope, and basaltic formations were identified in all four canonical methods; precipitation and Chagres sandstone were identified in three. On the other hand, the remaining terms were significant in less than three analyses; 13 - 19% of species variation was accounted by the joint effect of environmental and spatial variables.

**Table 2.** Variation partitioning of species abundances using two different canonical analyses: correspondence canonical analysis (CCA), and distance-based redundancy analysis (db-RDA) based on Hellinger transformation. Abundance and Hellinger distances data were analysed without transforming and logarithmically transformed. Explanatory variables were selected with forward selection using permutation tests (4999 permutations) for each group of variables separately (environmental variables –environment–, and 3rd-order polynomial spatial terms). The order of significant variables corresponds to their weight in explaining variation of species abundances.

	CCA No transform	CCA Log transform	Db-RDA Hellinger No transform	Db-RDA Hellinger Log transform
Environment alone	12.13	10.28	9.80	9.80
Shared	16.64	18.63	13.10	13.30
Spatial terms alone	21.95	25.25	27.30	27.00
Unexplained	49.28	45.85	49.80	49.90
Significant variables	Elevation, Dry days, Slope, Toro limestone, Chagres sandstone, Ancient non-basaltic formations Basaltic formations	Elevation, Precipitation, Slope, Basaltic formations, Ancient non-basaltic formations	Elevation, Precipitation, Slope, Chagres sandstone, Basaltic formations	Elevation, Precipitation, Slope, Chagres sandstone, Basaltic formations
	*	*	*	*

\* All spatial terms: X, Y, X\*Y, X<sup>2</sup>, Y<sup>2</sup>, X<sup>2</sup>\*Y, Y<sup>2</sup>\*X, X<sup>3</sup> and Y<sup>3</sup> were significant at  $p < 0.05$ .





**Fig. 2.** Influence of plot size in variance of species abundance decomposed in four different components: pure effect of environment alone, pure effect of spatial terms alone, combined variation due to the joint effect of environment and spatial variables, and variation not accounted for by the variables included in the analysis. Variation partitioning was carried out at three plot sizes with correspondence canonical analysis (CCA) where abundances were logarithmically transformed. Explanatory variables were selected with forward selection using permutation tests (4999 permutations) for each group of variables separately (environmental variables – environment – and 3rd-order Polynomial spatial terms). Taking subsamples as pseudo-replicates, we carried out CCA analysis several times (four subsamples of 0.25 ha, two subsamples of 0.5 ha, and one 1-ha subsample). Bars represent the minimum value of the percentage explained by each component for all subsamples, and error bars represent the maximum value.

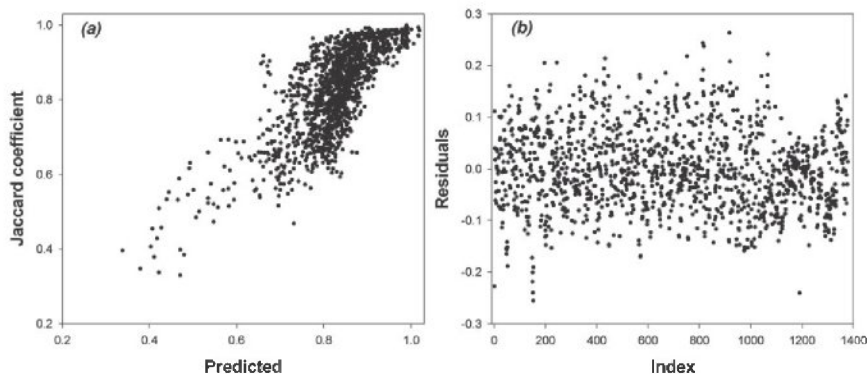
**Table 3.** Multiple regression model of  $\beta$ -diversity (Jaccard index expressed as dissimilarity) considering only those significant variables that contain raster spatial information along the tropical forest of the study area as entry variables, i.e. elevation, slope, geologic types, six spectral bands, habitat fragmentation and logarithm of geographical distance. The coefficient of determination  $R^2 = 0.573$ ,  $P < 0.001$ .

Explanatory variables	Standardized Regression coefficient	$P$
Band 1	-0.147	0.004
Band 2	-0.171	0.001
Band 3	0.231	0.001
Band 4	0.173	0.001
Elevation	0.286	0.002
ln(GD)	0.635	0.001

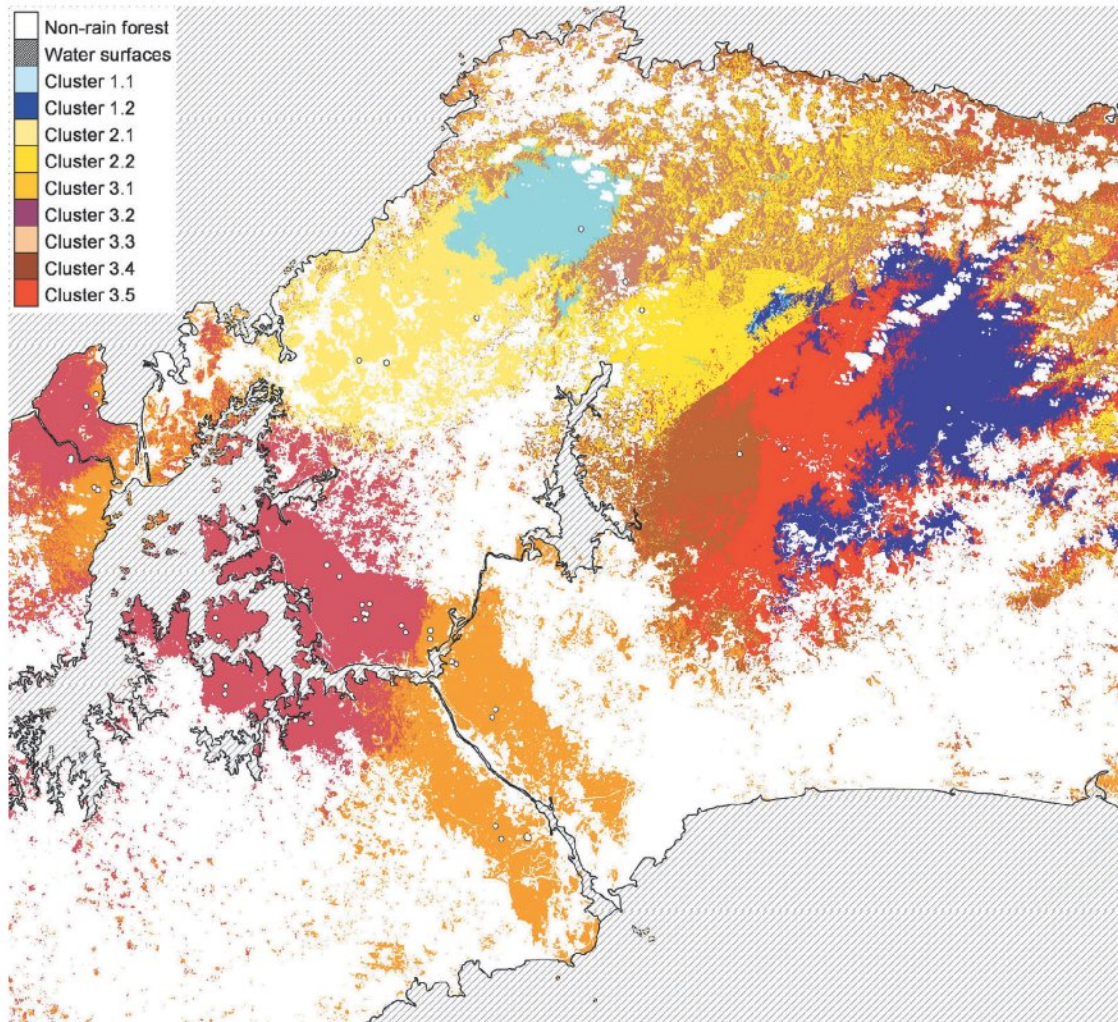
The effect of plot size on variation partitioning of floristic composition was carried out only with correspondence canonical analysis to log-transformed abundances, since this method explained the most variation in the abundance matrix (Fig. 2). When varying plot size, the most significant effect was an increase of 7% of overall variation explained when moving from 0.25 ha to 0.5 ha, and a further increase of 3% moving from 0.5 ha to 1 ha. The fraction of species variation explained by each component showed little difference with plot size. Environmental factors did not explain more variation at the 1-ha scale than at the 0.25-ha scale.

#### *Clustering the landscape into forest types*

We selected the Jaccard index to construct a forest type map because its correlation with explanatory variables was higher in most cases. The multiple regression model of species similarity was based on six selected variables: bands 1, 2, 3, 4, elevation, and the logarithm of geographical distance (Table 3). The relatively high



**Fig. 3. a.** Scatterplot of observed Jaccard coefficient versus predicted values from the multiple regression model using dissimilarities in spectral bands 1, 2, 3, and 4, elevation and logarithm of geographic distance (see Table 3 for details). **b.** Residuals of the model of the Jaccard coefficient presented in panel a, calculated as observed minus predicted values.



**Fig. 4.** Predicted forest types of the Panama Canal Watershed based on a multiple regression model on distance matrices and a hierarchical agglomerative clustering method. The model considers elevation, four spectral bands, and logarithm of geographical distance ( $R^2 = 0.573$ ,  $P < 0.001$ ), see Table 3. The model is extrapolated to rain forests (young secondary forest, old secondary forest and old-growth forest). The colours represent forest types at two hierarchical levels of similarity: at 8% with three main clusters (blues, yellows and browns) and at 20% with nine low-level clusters, see App. 2. In white, non-rain forested land: agricultural and urban land, mangroves, shrub land and clouds.

value of the coefficient of determination ( $R^2 = 0.57$ ), the absence of outliers, and the homogeneity in residuals, indicated a robust relationship (Fig. 3).

The hierarchical clustering of the similarity matrix (App. 2) and extrapolation with the multiple regression model produced a map of forest types (Fig. 4). Our assignment method correctly placed all 53 plots to the forest type they belonged to. Along the Panama Canal, two different forest types dominated (3.1 and 3.2), splitting the region into Pacific and Atlantic vegetation. But fine details within the two main areas indicate that factors other than geographic distance affect species distributions. In the upper Chagres area, the variety of forest types reflects the high variability in species composition among Chagres plots.

## Discussion

We found that the best predictor of floristic similarity among sites of the Panama Canal was the logarithm of geographic distance. Climate variables and elevation were also important predictors, especially precipitation. These results were supported by a partition of variance using canonical analysis, which also identified spatial variation as the best predictor of landscape-scale structuring of species composition, although environmental variables also contributed to explaining floristic variation. Geographic distance explained 35-46% of the variation, 13-19% of which was shared with environmental variables. This suggests that the distribution of Panamanian tree species is consistent with the dispersal limitation



hypothesis, supplemented by an 'environmental segregation' process.

The importance of distance in shaping spatial patterns of diversity of Panamanian tree species suggests that dispersal limitation enables species to escape their superior competitors (Hurt & Pacala 1995; Mouquet & Loreau 2003). The less important role played by environmental factors in predicting variation in species abundances and the large amount of unexplained variation (lack of fit in the modelling) support the view that deterministic factors alone cannot explain which species are found in any one plot. This is surprising given the broad range of environmental variation in this landscape. One explanation is that we are missing key explanatory factors of the abiotic environment, such as details of soil texture, soil chemistry, waterlogging. It would be important to collect more information to quantify these dimensions of the niche. However, we showed that soil types were not a major factor limiting plant distribution in the PCW, and therefore they do not constitute an environmental filter in most of this area, save the scattered podzolized limestone formations, that generally harbour a distinct dry forest flora (Pyke et al. 2001; Brewer et al. 2003), and an island of andesite soils toward the dry side of the isthmus supporting wet-forest species (Condit et al. 2002). We also found that rainfall does act as an environmental filter: many species are unable to withstand the prolonged dry season of the southern PCW.

Several recent reports have also compared the role of environmental factors and geographic distance in explaining the spatial variation of plant diversity in Amazonia. Tuomisto et al. (2003a) used data on all individuals in the *Melastomataceae* family (a diversified family of bird-dispersed understorey plants, more rarely canopy trees) as well as ferns (*Pteridophyta*) censused along a long transect in the Iquitos region, Peru. They showed that geographic distance accounted for only a small fraction of variance in similarity for both groups ( $r = 0.39$  for ferns and melastoms combined, see Table 2 in Tuomisto et al. 2003a), while environmental variation accounted for a much larger fraction ( $r = 0.55$ ). Phillips et al. (2003) used tree censuses from Southeast Peru (Rio Madre de Dios and Rio Tambopata), and found that geographical distance was less important than environmental variation ( $r = 0.31$  for distance, explaining 10% of the variance in similarity; four environmental variables explained 40% of the variance, see Table 7 in Phillips et al. 2003). Finally, Vormisto et al. (2004) performed a similar study with palms in the Western Amazon basin (Iquitos region and Amazonian Ecuador), and found that most of the variance in similarity was explained by geographic distance ( $r = 0.77$ , 40% of the variance explained by distance alone), while environment alone was less important (8%). All

three studies were located in high-rainfall areas; rainfall should then contribute less to the environmental variability than in the PCW, even if this hypothesis was not tested in these studies.

How can we reconcile these findings with our results? First, variation across sites should be expected given the variety of life forms, evolutionary and ecological histories of these plant groups. For instance, understorey plants and trees of  $> 2.5$  cm DBH respond differently to environment and distance than trees  $> 10$  cm DBH (Duque et al. 2002; Phillips et al. 2003). Second, while the role of distance was assessed relatively consistently across these studies, different environmental variables were found to be important: soils in Amazonia, rainfall in Panama. Third, these studies were performed over different areas and with different sampling intensities, from a single 43-km long transect (Tuomisto et al. 2003a), to multiple plots scattered across the landscape (21 in Vormisto et al. 2004; 88 in Phillips et al. 2003; 53 in our study). All of this should contribute to the differences observed across studies because dispersal, speciation and environmental segregation act at different spatial and temporal scales. This comparison emphasizes that groups of small-sized and well-dispersed plants such as melastomes and ferns are more dependent on environmental variation than groups of large and poorly dispersed plants (most tropical tree species, palms).

We found little influence of habitat fragmentation on tree species composition in central Panama. Several of the forest plots in the Panama Canal Watershed have been separated by human-transformed land since the beginning of the twentieth century, changes in tree species composition might still be ongoing (Leigh et al. 1993).

Our method for spatializing  $\beta$ -diversity enabled us to create a map of forest types based on tree composition (Fig. 4). We identified the forests of Chagres National Park and Santa Rita as highly distinctive floristically, relative to forests near the canal. Nearly half of all the tree species recorded in central Panama are restricted to these wet Caribbean forests (Condit et al. in press). Because the Santa Rita and upper Chagres areas have been poorly surveyed, they should be considered as priority research areas. Closer to the Canal, along the well sampled forest tract bordering the Canal, the map shows two main forest types, one Caribbean and one Pacific, supporting our general impressions based on our experience here. The section of Pacific coast forest type near the Caribbean is on a conspicuous outcrop of Toro limestone, and is easy to spot in the dry season. The boundary between a Pacific type and a Caribbean type near Gamboa is also well known to us, and easy to see from hill tops near Gamboa: there is indeed a fairly abrupt transition, associated with different geological formations underlying the various hills around Gamboa.



Two of the hills have large stands of *Cavanillesia platanifolia*, along with *Calycophyllum candidissimum* and *Bursera simaruba*, three species characteristic of Pacific type forest; the other two hills have no individuals of those species, but instead carry a southernmost stand of Caribbean forest, with *Tapirira guianensis* and *Vantanea depleta* both common. The two forest types bordering the Canal correspond to environmental and geologic factors, whereas the Canal itself did not represent a natural boundary.

This approach for mapping floristic diversity at the landscape scale based on extensive tree inventory, satellite imagery, and statistical extrapolation is strongly supported by our field experience of the area. Beyond the local interest in constructing a floristic map of the PCW, this method has the potential to be easily implemented in other regions and other vegetation types for conservation purposes and carbon accounting projects (Chave et al. 2005). Contrary to current methods that attempt to maximize the representation of regional diversity (Williams et al. 2004), our approach should be particularly suited to mapping projects in poorly surveyed or inaccessible areas, such as tropical regions (for an overview, see Tuomisto 1998). Thus, it might be the basis for a regional gap analysis (Jennings 2000), guiding the scarce resources available for protecting biodiversity toward maximum effectiveness.

**Acknowledgements.** This paper is dedicated to the memory of Charles S. ReVelle. We thank S. Bohlman, R. Martínez, E. Mariscal and L. Valencia for their technical support in Panama. We benefited from discussions with P. Legendre. GC was supported by a grant from the Spanish government. JC was supported by a grant from the Institut Français de la Biodiversité. The US Department of Defense Legacy Program supported the census of the plots at Fort Sherman and Cocoli, both on former US military land. The authors thank the many field workers who tagged and measured the trees, and the Smithsonian Tropical Research Institute for logistical and financial support. This is a scientific contribution from the Center for Tropical Forest Science.

## References

- Bohlman, S.A. 2004. *The relationship between canopy structure, light dynamics and deciduousness in a seasonal tropical forest in Panama: A multiple scale study using remote sensing and allometry*. Ph.D. Thesis. University of Washington, Seattle, WA, US.
- Brewer, S.W., Rejmánek, M., Webb, M.A.H. & Fine, P.V.A. 2003. Relationships of phytogeography and diversity of tropical tree species with limestone topography in southern Belize. *J. Biogeogr.* 30: 1669-1688.
- Cabeza, M. & Moilanen, A. 2001. Design of reserve networks and the persistence of biodiversity. *Trends Ecol. Evol.* 16: 242-248.
- Cannon, C.H. & Leighton, M. 2004. Tree species distributions across five habitats in a Bornean rain forest. *J. Veg. Sci.* 15: 257-266.
- Chave, J. & Leigh, E.G. 2002. A spatially explicit neutral model of beta-diversity in tropical forests. *Theor. Pop. Biol.* 62: 153-168.
- Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S. & Perez, R. 2004. Error propagation and scaling for tropical forest biomass estimates. *Phil. Trans. R. Soc. B* 359: 409-420.
- Chave, J., Chust, G., Condit, R., Aguilar, S., Perez, R. & Lao, S. 2005. Error propagation and scaling for tropical forest biomass estimates. In: Phillips, O.L. & Malhi, Y. (Eds.) *Tropical forests and global atmospheric change*, pp 155-163. Oxford University Press, Oxford, UK.
- Condit, R., Watts, K., Bohlman, S.A., Perez, R., Foster, R.B., & Hubbell, S.P. 2000. Quantifying the deciduousness of tropical forest canopies under varying climates. *J. Veg. Sci.* 11: 649-658.
- Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Nunez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E. & Hubbell, S.P. 2002. Beta-diversity in tropical forest trees. *Science* 295: 666-669.
- Condit, R., Aguilar, S., Hernandez, A., Perez, R., Lao, S., Angehr, G., Hubbell, S.P. & Foster, R.B. 2004. Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. *J. Trop. Ecol.* 20: 51-72.
- Condit, R., Pérez, R., Lao, S., Aguilar, S. & Somoza, A. In press. Geographic ranges and  $\beta$ -diversity: discovering how many tree species there are where. *Biol. Skr.*
- Congalton, R.G. 1991. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sens. Environ.* 37: 35-46.
- Croat, T.R. 1978. *Flora of Barro Colorado Island*. Stanford University Press, Stanford, CA, US.
- Dietrich, W., Windsor, D. & Dunne, T. 1982. Geology, climate, and hydrology of Barro Colorado Island. In: Leigh Jr, E.G., Rand, A. & Windsor, D. (eds.) *The ecology of a tropical forest*, pp. 21-46. Smithsonian Institution Press, Washington, DC, US.
- Duivenvoorden, J.F. & Lips, J.M. 1998. Mesoscale patterns of tree species diversity in Colombian Amazonia. In: Dallmeier, F. & Comiskey, J.A. (eds.) *Forest biodiversity in North, Central and South America, and the Caribbean. Research and monitoring*. MAB Series Vol. 21, pp. 535-549. UNESCO, Paris, FR.
- Duivenvoorden, J.F., Svenning, J.C. & Wright, S.J. 2002. Beta diversity in tropical forests. *Science* 295: 636-637.
- Duque, A., Sánchez, M., Cavelier, J. & Duivenvoorden, J.F. 2002. Different floristic patterns of woody understorey and canopy plants in Colombian Amazonia. *J. Trop. Ecol.* 18: 499-525.
- Dutilleul, P., Stockwell, J.D., Frigon, D. & Legendre, P. 2000. The Mantel test versus Pearson's correlation analysis: Assessment of the differences for biological and environmental studies. *J. Agri. Biol. Env. Stat.* 5: 131-150.
- Eastman, J.R. 1999. Idrisi32. Worcester, Clark University. URL: <http://www.clarklabs.org>.

- Hubbell, S.P. 2001. *A unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ, US.
- Hurt, G.C. & Pacala, S.W. 1995. The consequences of recruitment limitation: reconciling chance, history and competitive differences between plants. *J. Theor. Biol.* 176: 1-12.
- Jennings, M.D. 2000. Gap analysis: concepts, methods, and recent results. *Landscape Ecol.* 15: 5-20.
- Lapointe, F.-J. & Legendre, P. 1994. A classification of pure malt Scotch whiskies. *Appl. Stat.* 43: 237-257.
- Legendre, P. & Anderson, M.J. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* 69: 1-24.
- Legendre, P. & Legendre, L. 1998. *Numerical ecology*. Elsevier, Amsterdam, NL.
- Legendre, P. & Gallagher, E. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129: 271-280.
- Leigh, E.G., Wright, S.J., Putz, F.E. & Herre, E.A. 1993. The decline of tree diversity on newly isolated tropical islands: A test of a null hypothesis and some implications. *Evol. Ecol.* 7: 76-102.
- Mouquet, N. & Loreau, M. 2003. Community patterns in source-sink metacommunities. *Am. Nat.* 162: 544-557.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858.
- Phillips, O.L., Vargas, P.N., Monteagudo, A.L., Cruz, A.P., Zans, M.E.C., Sanchez, W.G., Yli-Halla, M. & Rose, S. 2003. Habitat association among Amazonian tree species: a landscape-scale approach. *J. Ecol.* 91: 757-775.
- Pitman, N.C.A., Terborgh, J.W., Silman, M.R., Nuñez, V.P., Neill, D.A., Ceron, C.E., Palacios, W.A. & Aulestia, M. 2001. Dominance and distribution of tree species in upper Amazonian terra firme forests. *Ecology* 82: 2101-2117.
- Potts, M.D., Ashton, P.S., Kaufman, L.S. & Plotkin, J.B. 2002. Habitat patterns in tropical rain forests: a comparison of 105 plots in northwest Borneo. *Ecology* 83: 2782-2797.
- Pyke, C.R., Condit, R., Aguilar, S. & Lao, S. 2001. Floristic composition across a climatic gradient in a neotropical lowland forest. *J. Veg. Sci.* 12: 553-566.
- Ricklefs, R.E. & Schluter, D. 1993. *Species diversity in ecological communities: Historical and geographical perspectives*. University of Chicago Press, Chicago, IL, US.
- Rodrigues, A.S.L., Andelman, S.J., Bakarr, M.I., Boitani, L., Brooks, T.M., Cowling, R.M., Fishpool, L.D.C., da Fonseca, G.A.B., Gaston, K.J., Hoffmann, M., Long, J.S., Marquet, P.A., Pilgrim, J.D., Pressey, R.L., Schipper, J., Sechrest, W., Stuart, S.N., Underhill, L.G., Waller, R.W., Watts, M.E.J. & Yan, X. 2004. Effectiveness of the global protected area network in representing species diversity. *Nature* 428: 640-643.
- Ruokolainen, K. & Tuomisto, H. 2002. Beta-diversity in tropical forests. *Science* 297: 1439.
- Sanjur, A., Gonzalez, F., Prieto, C. & Heckadon-Moreno, S. 1999. Las poblaciones humanas. In: Heckadon-Moreno, S., Ibanez, R. & Condit, R. (eds.) *La cuenca del Cala: deforestacion, urbanizacion y contaminacion*, pp. 83-100. Smithsonian Tropical Research Institute, Balboa (Panama).
- Slik, J.W.F., Poulsen, A.D., Ashton, P.S., Cannon, C.H., Eichhorn, K.A.O., Kartawinata, K., Lanniari, I., Nagamasu, H., Nakagawa, M., van Nieuwstadt, M.G.L., Payne, J., Saridan, A., Sidiyasa, K., Verburg, R.W., Webb, C.O. & Wilkie, P. 2003. A floristic analysis of the lowland dipterocarp forests of Borneo. *J. Biogeogr.* 30: 1517-1531.
- Sneath, P.H.A. 1966. A comparison of different clustering methods as applied to randomly-spaced points. *Classif. Soc. Bull.* 1: 2-18.
- ter Braak, C.J.F. & Šmilauer, P. 1998. *CANOCO reference manual and user's guide to Canoco for Windows: Software for canonical community ordination (version 4)*. Microcomputer Power, Ithaca, NY, US.
- ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., van der Hout, P., Daly, D.C., Silveira, M., Phillips, O., Vasquez, R., van Andel, T., Duivenvoorden, J., De Oliveira, A.A., Ek, R., Lilwah, R., Thomas, R., van Essen, J., Baider, C., Maas, P., Mori, S., Terborgh, J., Vargas, P.N., Mogollon, H. & Morawetz, W. 2003. A spatial model of tree alpha-diversity and tree density for the Amazon. *Biodiv. Conserv.* 12: 2255-2277.
- Tuomisto, H. 1998. What satellite imagery and large-scale field studies can tell about biodiversity patterns in Amazonian forests. *Ann. Mo. Bot. Garden* 85: 48-62.
- Tuomisto, H., Ruokolainen, K., Aguilar, M. & Sarmiento, A. 2003a. Floristic patterns along a 43-km long transect in an Amazonian rain forest. *J. Ecol.* 91: 743-756.
- Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. 2003b. Dispersal, environment, and floristic variation of western Amazonian forests. *Science* 299: 241-244.
- Vormisto, J., Svenning, J.-C., Hall, P. & Balslev, H. 2004. Diversity and dominance in palm Arecaceae communities in terra firme forests in the western Amazon basin. *J. Ecol.* 92: 577-588.
- Williams, P., Gibbons, D., Margules, C., Rebelo, A., Humphries, C. & Pressey, R. 1996. A comparison of richness hotspots, rarity hotspots and complementary areas for conserving diversity using British birds. *Cons. Biol.* 10: 155-174.
- Williams, P., Humphries, C., Araujo, M., Lampinen, R., Hagemeyer, W., Gasc, J.-P. & Mitchell-Jones, T. 2000. Endemism and important areas for conserving European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian J. Entomol.* 2: 21-46.
- Williams, J.C., ReVelle, C.S. & Levin, S.A. 2004. Using mathematical optimization models to design nature reserves. *Frontiers Ecol. Environm.* 2: 98-105.
- Woodring, W., Stewart, R. & Stewart, J. 1980. *Geologic map of the Panama Canal*. U.S. Geological Survey, Reston, CA, US.

Received 11 May 2005;

Accepted 22 November 2005.

Co-ordinating Editor: J. Franklin.

For Apps. 1 and 2, see JVS/AVS Electronic Archives;  
www.opuluspress.se