SHORT COMMUNICATION

Two biases in estimating range sizes of Amazonian plant species

KALLE RUOKOLAINEN*, HANNA TUOMISTO*[1], JAANA VORMISTO*
and NIGEL PITMAN†

*Department of Biology, University of Turku, FIN-20014 Turku, Finland
†Center for Tropical Conservation, Box 90381, Duke University, Durham, NC 27708, USA

The Amazonian rain forests form one of the largest areas of continuous forest in the world, and they harbour a substantial part of the world's biodiversity (World Conservation Monitoring Centre 1992). The extravagant species richness of Amazonia makes collecting information on species distribution patterns there a formidable challenge. Relatively few studies have been able to provide data that cover both a large enough number of species and wide enough geographical areas to allow relevant biogeographical and ecological generalizations (such studies concerning plants include Balslev 1988, de Oliveira & Daly 1999, Pitman *et al*. 1999, Prance 1973, ter Steege *et al*. 2000, Terborgh & Andresen 1998, Tuomisto & Poulsen 1996). The scarcity of field observations is undoubtedly one of the reasons why very different ideas exist about the mechanisms that control plant distributions and patterns of species richness in Amazonia (Balslev 1988, Condit 1996, Gentry 1988, Nelson *et al*. 1990, Pitman *et al*. 1999, Prance 1973, 1982; ter Steege *et al*. 2000, Tuomisto *et al*. 1995).

Quantitative tree inventories, regional plant check-lists and taxonomic revisions are the main sources of information for documenting Amazonian plant distribution patterns, and luckily new information is continuously accumulating

---

[1] Corresponding author. Email: hantuo@utu.fi

along all these lines. One recent example is the paper of Pitman *et al.* (1999, see also the commentary by Ricklefs 2000). Their analysis of distribution patterns of species encountered in a large network of tree plots in southern Peru suggested that Amazonian trees might have much larger range sizes than previously thought. This is good news for conservation, as it makes the prospect of protecting Amazonian tree species much more hopeful than would be the case if each part of the basin had vast arrays of narrowly endemic species.

However, Pitman *et al.* (1999) also noted two potential biases that could have distorted their results, and that could affect subsequent generalizations. These biases are: (1) large plants may be generally more widespread than small plants and (2) unidentified morphospecies may have, on average, smaller geographic ranges than identified species. The first bias is relevant because the data set of Pitman *et al.*, like most quantitative plant inventories in Amazonia, only included trees with a stem diameter greater than 10 cm at breast height, thereby excluding from the analysis a high number of plant species that were present at the study sites but too small to be sampled. The second bias is relevant because it is impossible to estimate the geographic range size of a species whose name is not known. Pitman *et al.* consequently had to exclude from their analyses 237 morphospecies for which they had not managed to find a species name. This is a typical situation in Amazonian tree inventories; it is common for 20–30% of the species-level taxa to remain as unidentified morphospecies (for lists of Amazonian tree inventories see ter Steege *et al.* 2000, Terborgh & Andresen 1998). Pitman *et al.* (1999) did not dwell on these biases, and no one else seems to have done so either. The purpose of the present communication is to evaluate and quantify their potential importance.

To test the first potential bias, we extracted information from recent taxonomic treatments of palms in Amazonia (Henderson 1995, 2000). This is the most recent Amazon-wide data set available to us that includes, for each species and variety in an entire family, information on both the size of the plants themselves and the geographical distribution range of the taxa. A similar analysis conducted earlier (Ruokolainen & Vormisto 2000) treated only 45 Peruvian palm taxa, while here we include all 189 palm taxa known to occur in Amazonia (these taxa belong to 136 species; infraspecific taxa are treated separately following Henderson 1995). The maximum stem height indicated for each taxon was used as a measure of plant size. The number of $2.5° \times 2.5°$ latitude-longitude grid cells from which the taxon has been reported was used as a measure of geographical range size.

A positive correlation was found between range size and palm height (Pearson's $r = 0.29$, $P < 0.0001$, after normalizing both variables with Box–Cox transformation). One of the possible explanations for this pattern is that seeds of tall plants may get dispersed by highly mobile seed dispersers such as parakeets, macaws and monkeys, while seeds of short plants may get dispersed by shorter-ranging understorey birds or mammals. While we cannot prove that

this is the correct explanation, the pattern itself seems not to be confined just to palms: a similar correlation can be found among trees in general. For our own dataset of tree species native to western Amazonia (described in Pitman 2000), we found a positive relationship between the maximum recorded stature of a species and the product of its longitudinal and latitudinal ranges (Spearman's rho = 0.30, P < 0.0001, n = 269). The same data showed that species that occur in two well-inventoried western Amazonian forests separated by 1400 km have significantly higher stature than species found at only one (Pitman 2000). Whatever the actual explanation of these correlations, they suggest that one should be cautious if generalizations on species range sizes are extended from one plant group to another.

The second potential bias, that identified plant species have larger range sizes than unidentified morphospecies, appears to be a logical consequence from two tendencies typical of tropical rain forests and other poorly known biotas. First, if a species is widespread, the likelihood is increased that it has somewhere at some time attracted the attention of a taxonomist who has described it and given it a species name. Second, since widespread species can be collected in a large area, they tend to be better represented in herbaria than restricted-range species, which increases the likelihood that an identified specimen is found and can be matched with the material undergoing identification. This tendency is further strengthened if widespread species are locally more abundant than restricted-range species – a pattern that has been observed repeatedly among several types of organism (Brown 1984, Hanski 1982, Holt *et al.* 1997).

Here we ask whether such an identification bias exists in a large data set consisting of quantitative inventories of pteridophytes and Melastomataceae (a family of predominantly shrubs and small trees) in western Amazonian rain forests. The inventories have been made in four separate regions in the old-growth rain forests of Peru (Loreto and Madre de Dios), Ecuador (Yasuní) and Colombia (Caquetá). Within each region, the sampling of Melastomataceae and pteridophytes was conducted together using plots that usually were in the form of 5-m × 500-m transect lines. In some cases, Melastomataceae were sampled using a broader transect than pteridophytes to account for the lower density of individuals per unit area. All terrestrial Melastomataceae and those terrestrial and low-epiphytic pteridophytes that had at least one leaf exceeding 10 cm in length were recorded in the transects.

Voucher specimens were collected of all individuals not referable to a previous collection during the same field inventory trip. Most of the Melastomataceae collections from Loreto were identified by the late John Wurdack, and in addition they have been matched with the collections in US, MO and AAU (herbarium abbreviations from Holmgren *et al.* 1990) by the first author (KR). Collections from other regions have been matched with the samples from Loreto by KR. As species identifications of Melastomataceae samples from

Loreto have progressed further than those from elsewhere, we took Loreto as the focal region for comparing the distribution patterns of named and unnamed Melastomataceae.

With pteridophyte collections there is hardly any difference among regions in the quality of identifications. Initial identification was done in the herbaria AAU, U, Z, US, MO, NY and UC by the second author (HT), some problematic groups have been identified by John T. Mickel and Robbin C. Moran, and all species and morphospecies have been seen and commented on by Alan R. Smith. In our analyses we assume that both the species for which we have found a species name, and the unidentified morphospecies represent true biological species. In dealing with the morphospecies, we have tried to be conservative, e.g. we avoided basing a morphospecies on juvenile specimens only, and if in doubt on whether specimens from two regions belong to the same morphospecies or not, we would rather err towards lumping than splitting. Duplicates of all collections are deposited in TUR, and complete regional sets can be found in AMAZ, USM, QCA, CUZ, QCNE and COAH.

The Melastomataceae inventories totalled 102 ha, and pteridophyte inventories 80 ha (Table 1). On the basis of abundance data for those sites where individuals were counted it can be estimated that the total number of Melastomataceae individuals encountered was about 134 000, and the total number of pteridophyte individuals about 575 000. Comparisons with regional species lists (Brako & Zarucchi 1993, Duivenvoorden & Lips 1993, Jørgensen & León-Yánez 1999, Tryon & Stolze 1989–1994) indicate that a large proportion of the species known to occur in each region were actually found, in addition to numerous new species records. Altogether, we encountered 297 Melastomataceae species (the identification of 167 (56%) of these has only proceeded to the morphospecies stage), and 323 pteridophyte species (of which 53 (16%) have only been identified to the morphospecies stage). Species lists for many of our Loreto sites are available in Ruokolainen & Tuomisto (1998); details on the species composition of the other areas will be published elsewhere.

An inspection of the distribution patterns of fully identified species and unidentified morphospecies of both plant groups show very clearly that the fully identified species have much wider ranges than the morphospecies (Tables 2

Table 1. Summary table of quantitative Melastomataceae and pteridophyte inventories in four regions in western Amazonian old-growth rain forests.

Region	Area sampled (ha)		Estimated number of individuals		Number of species		Maximum distance between samples (km)
	Melast.	pterid.	Melast.	pterid.	Melast.	pterid.	
Loreto	47.46	37.00	73 000	237 000	188	195	290
Yasuní	17.84	15.94	16 000	151 000	120	197	90
Caquetá	24.97	14.74	40 000	101 000	133	127	50
Madre de Dios	12.00	12.00	5000	86 000	109	173	240
Total	102.27	79.68	134 000	575 000	297	323	1500

Table 2. *Distribution of those Melastomataceae species that were found in Loreto in four regions (Loreto, Yasuní, Caquetá and Madre de Dios) in western Amazonian old-growth rain forests.*

Regions where a species is recorded	Number of all species	Number of fully identified species	Number of morphospecies
only Loreto	51 (27%)	17 (14%)	34 (50%)
Loreto + 1	56 (30%)	33 (28%)	23 (34%)
Loreto + 2	57 (30%)	49 (41%)	8 (12%)
Loreto + 3	24 (13%)	21 (18%)	3 (4%)

and 3). With Melastomataceae, most (59%) of the identified species that were recorded in our Loreto sample were also found in at least two of the other regional samples, whereas most morphospecies (84%) were recorded either only in Loreto or just in one other region. With pteridophytes, the percentage of identified species that were only recorded in a single region was 31%, while 75% of the morphospecies were restricted to a single region.

In our Melastomataceae data set, 69% of the identified species found in Madre de Dios were also found in Yasuní. This is the same percentage that Pitman *et al.* (1999) found for their 506 identified tree species for which geographical range-size information was available. However, when we repeated the comparison for all Melastomataceae species (including the unidentified morphospecies), the percentage of shared species fell to 50% because the morphospecies had, on average, smaller ranges than the identified species. If the same tendency exists in the data set of Pitman *et al.* (1999), the percentage of species shared between the two regions may be significantly lower than 69%. In the extreme case that none of the 319 species that were excluded from that analysis occur in Ecuador the percentage would drop to 42%.

The identification bias is powerful, and it is independent of individual botanist's taxonomic idiosyncrasies. A person with a narrow species concept will have a greater number of species to deal with than a person with a wider species concept, but both will observe the same relative bias between identified species and morphospecies as long as they use good taxonomic characteristics for separating taxa.

Unidentified plant specimens (and consequently, morphospecies) can be divided into two main categories: (1) specimens representing new species that have not yet been described, and (2) specimens belonging to already described species that have not yet been identified.

Table 3. *Distribution of pteridophyte species in four regions (Loreto, Yasuní, Caquetá and Madre de Dios) in western Amazonian old-growth rain forests.*

Number of regions where a species is recorded	Number of all species	Number of fully identified species	Number of morphospecies
1	125 (39%)	85 (31%)	40 (75%)
2	78 (24%)	72 (27%)	6 (11%)
3	69 (21%)	67 (25%)	2 (4%)
4	51 (16%)	46 (17%)	5 (9%)

Given enough time and effort, the morphospecies of category 2 would eventually be identified even if no new taxonomic revisions were published. Therefore, the importance of the identification bias ultimately depends on what proportion of the Amazonian flora is currently undescribed. No one knows, of course, but there is no indication that the accumulation of new plant species is going to stop or even slow down in the near future. The accumulation of new angiosperm species in the world showed no sign of levelling off during the years 1989–1997 (Prance *et al.* 2000). In Ecuador, which is botanically among the best-known areas in the neotropics, the rate of description of both new species and new endemic species has been almost constant over the past few decades (Fig. 2 in Jørgensen & León-Yánez 1999). No statistics of this kind exist specifically for Amazonia, but given the low collection density in the area it is unlikely to be better known floristically than the rest of the tropics.

The new species waiting to be described are likely to have smaller geographic ranges than the already named species. In western Amazonian tree communities, we find that species with wide ranges have generally been described much earlier than species with restricted ranges. For example, in our inventory of trees in eastern Ecuador, 16 of the 17 species believed to be endemic to that region were described by taxonomists later than 1980. By contrast, tree species that were present both in Ecuador and in a similar inventory in southeastern Peru, had an average date-of-description of 1871 (Pitman 2000). More generally, among 277 tree species recorded in these inventories, we find a strong negative correlation between the date a species was described and the size of its geographic range (the product of its latitudinal and longitudinal extent; Spearman's rho -0.57, $P < 0.0001$).

In conclusion, it can be noted that both biases mentioned by Pitman *et al.* (1999) exist and can significantly distort conclusions based on even carefully collected species distribution data. Since plants of different statures seem to have different average range sizes, a study of just one stature class may give a biased estimate of the average range size of all plants. When range size estimates are based on fully identified species only (as usually is the case out of necessity), there is a significant bias towards overestimating the range size of all plants, as most of the undescribed and difficult-to-identify species seem to have narrower ranges than the well-known species.

We have focused on plants, but the biases we describe are probably just as disruptive to the study of tropical animal communities. The identification bias in particular has the potential to mislead macroecological research on any poorly known group of organisms, and the distortion will not necessarily be confined to distributional patterns. We have shown that the named species in the forests we study are neither an ecologically nor a morphologically unbiased subset of the communities they inhabit. The lesson for ecologists who work in communities with a non-trivial proportion of undescribed taxa is clear: data from a subset of reliably identified species should not be assumed to provide a representative picture of the community at large.

For Amazonian plants, the identification bias can be overcome only by directly comparing the unidentified specimens collected by different research teams. At present this is made difficult by the common practice that once the specimens are identified to a level considered sufficient for publishing the results, they are deposited among the general collections of different herbaria. Thereafter, it is difficult to compare a full set of specimens collected at one site with those collected at another site, and even if new identifications are made, they seldom reach the inventory data base. Furthermore, the initial identification to different morphospecies may be lost in the process, as all unidentified specimens that belong to the same genus are stored together.

One solution to these problems would be to maintain the vouchers from floristic inventories as separate collections in herbaria. This could be achieved either physically or, if the herbarium manages a rigorously updated database of its holdings, virtually. In this way, the specimens will be readily available for cross-checking or other future needs. Even better accessibility to the data could be obtained by displaying good quality photographs of representative voucher specimens on the Internet.

By now, at least 330 one-hectare tree inventory plots have been inventoried in Amazonia (Rankin de Merona *et al*. 1992, ter Steege *et al.* 2000). A careful comparison of the combined vouchers from these plots would go a long way towards clarifying the picture of Amazonian tree distributions. No complicated techniques are needed to achieve this, but many working hours and some additional funds will be necessary. We believe that these costs are well justified, because they will help focus efforts to understand Amazonian plant species distributions in a more efficient way than before.

LITERATURE CITED

BALSLEV, H. 1988. Distribution patterns of Ecuadorean plant species. *Taxon* 37:567–577.

BRAKO, L. & ZARUCCHI, J. L. 1993. Catalogue of the flowering plants and gymnosperms of Peru. *Monographs in Systematic Botany from the Missouri Botanical Garden* 45:1–1286.

BROWN, J. H. 1984. On the relationship between abundance and distribution of species. *American Naturalist* 124:255–279.

CONDIT, R. 1996. Defining and mapping vegetation types in mega-diverse tropical forests. *Trends in Ecology and Evolution* 11:4–5.

DE OLIVEIRA, A. A. & DALY, D. C. 1999. Geographic distribution of tree species occurring in the region of Manaus, Brazil: implications for regional diversity and conservation. *Biodiversity and Conservation* 8:1245–1259.

DUIVENVOORDEN, J. F. & LIPS, J. M. 1993. Landscape ecology of the Middle Caqueta Basin. Explanatory notes to the maps. *Studies on the Colombian Amazonia*, III A:1–301.

GENTRY, A. H. 1988. Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Annals of the Missouri Botanical Garden* 75:1–34.

HANSKI, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* 38:210–221.

HENDERSON, A. 1995. *The palms of the Amazon*. Oxford University Press, Oxford. 362 pp.

HENDERSON, A. 2000. *Bactris* (Palmae). *Flora Neotropica Monograph* 79:1–181.

HOLMGREN, P. K., HOLMGREN, N. H. & BARNETT, L. C. 1990. *Index Herbariorum. Part I: The Herbaria of the World*. New York Botanical Garden, New York. 693 pp.

HOLT, R. D., LAWTON, J. H., GASTON, K. J. & BLACKBURN, T. M. 1997. On the relationships between range size and local abundance: back to basics. *Oikos* 78:183–190.

JØRGENSEN, P. M. & LEÓN-YÁNEZ, S. (eds). 1999. *Catalogue of the vascular plants of Ecuador*. Missouri Botanical Garden Press, St. Louis. 1181 pp.

NELSON, B. W., FERREIRA, C. A., DA SILVA, M. F. & KAWASAKI, M. L. 1990. Refugia, endemism centers and collecting density in Brazilian Amazonia. *Nature* 345:714–716.

PITMAN, N. C. A. 2000. *A large-scale inventory of two Amazonian tree communities*. Ph.D. thesis, Duke University, USA. 226 pp.

PITMAN, N. C. A., TERBORGH, J., SILMAN, M. & NUÑEZ, P. V. 1999. Tree species distributions in an upper Amazonian forest. *Ecology* 80:2651–2661.

PRANCE, G. T. 1973. Phytogeographic support for the theory of Pleistocene forest refuges in the Amazon Basin, based on evidence from distribution patterns in Caryocaraceae, Chrysobalanaceae, Dichapetalaceae and Lecythidaceae. *Acta Amazonica* 3:5–28.

PRANCE, G. T. (ed.). 1982. *Biological diversification in the tropics*. Columbia University Press, New York. 714 pp.

PRANCE, G. T., BEENTJE, H., DRANSFIELD, J. & JOHNS, R. 2000. The tropical flora remains undercollected. *Annals of the Missouri Botanical Garden* 87:67–71.

RANKIN DE MERONA, J. M., PRANCE, G. T., HUTCHINGS, R. W., DA SILVA, M. F., RODRIGUES, W. A. & UEHLING, M. E. 1992. Preliminary results of a large-scale tree inventory of upland rain forest in the Central Amazon. *Acta Amazonica* 22:493–534.

RICKLEFS, R. E. 2000. Rarity and diversity in Amazonian forest trees. *Trends in Ecology and Evolution* 15:83–84.

RUOKOLAINEN, K. & TUOMISTO, H. 1998. Vegetación natural de la zona de Iquitos. In Kalliola, R. & Flores, S. (eds.). Geoecología y desarrollo Amazónico: estudio integrado en la zona de Iquitos, Perú. *Annales Universitatis Turkuensis Serie A II* 114:253–365.

RUOKOLAINEN, K. & VORMISTO, J. 2000. The most widespread Amazonian palms tend to be tall and habitat generalists. *Basic and Applied Ecology* 1:97–108.

TER STEEGE, H., SABATIER, D., CASTELLANOS, H., VAN ANDEL, T., DUIVENVOORDEN, J., ADALARDO DE OLIVEIRA, A., EK, R., LILWAH, R., MAAS, P. & MORI, S. 2000. An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. *Journal of Tropical Ecology* 16:801–828.

TERBORGH, J. & ANDRESEN, E. 1998. The composition of Amazonian forests: patterns at local and regional scale. *Journal of Tropical Ecology* 14:645–664.

TRYON, R. M. & STOLZE, R. G. 1989–1994. Pteridophyta of Peru. Parts I–VI. *Fieldiana Botany, new series*, numbers 20, 22, 27, 29, 32, 34.

TUOMISTO, H. & POULSEN, A. D. 1996. Influence of edaphic specialization on pteridophyte distribution in neotropical rain forests. *Journal of Biogeography* 23:283–293.

TUOMISTO, H., RUOKOLAINEN, K., KALLIOLA, R., LINNA, A., DANJOY, W. & RODRIGUEZ, Z. 1995. Dissecting Amazonian biodiversity. *Science* 269:63–66.

WORLD CONSERVATION MONITORING CENTRE. 1992. *Global biodiversity*. Chapman & Hall, London. 585 pp.