

Estimating species richness at large spatial scales using data from small discrete plots

R. Krishnamani, A. Kumar and J. Harte

Krishnamani, R., Kumar, A. and Harte, J. 2004. Estimating species richness at large spatial scales using data from small discrete plots. – *Ecography* 27: 637–642.

Estimating species richness in large biomes is a central challenge in ecology and conservation biology. However, accurate census data is often available only from small discrete plots distributed within the biome. Using tree species richness data collected from 48 plots (0.25 ha each) widely distributed through 60 000 km² in the rainforests of the Western Ghats of southern India, we test the application of a proposed method for estimating species richness at large scales from measured species commonalities between pairs of censused plots. We show that the method allows extrapolation of species richness from a scale of 0.25 ha plots to that of the entire biome, or 10⁵ km².

R. Krishnamani (rkmani@rainforest-initiative.org) and A. Kumar. Div. of Conservation Biology, Salim Ali Centre for Ornithology and Natural History, Coimbatore 641 108, India. – J. Harte, Energy and Resources Group, 310 Barrows Hall, Univ. of California, Berkeley, CA 94720, USA. (Present address of R. K.: The Rainforest Initiative, 199, First St., Nalvar Nagar, Bharathiar Univ. PO, Coimbatore 641 046, India. Present address of A. K.: Centre for Wildlife Studies, 13th Cross, 7th Block, Bangalore 560 082, India.)

Knowledge of the total species richness in a region, for groups such as trees, birds, or insects, is a useful starting point for further research on ecosystem function, species extinctions under habitat loss, and design of conservation strategies, as well as being of inherent interest. Various methods claim to estimate species richness, taking into account species that are not actually recorded but whose presence can be inferred from the pattern of observed species occurrence (Burnham and Overton 1979, Palmer 1990, Bunge and Fitzpatrick 1993, Solow 1994, Colwell and Coddington 1994, Plotkin et al. 2000, Longino et al. 2002). In particular, statistical methods derived from the capture-recapture approaches used in population biology provide useful tools for estimating species richness but are seldom used in biodiversity studies (Boulinier et al. 1998) and are more suitable for animals rather than plants.

It is a well-known fact that the observed species richness could seriously underestimate the actual species

richness due to under sampling of rare species (Preston 1948, Lande et al. 2000). Also the number of species in many large regions is poorly known, but evidence exists that for some taxa actual species richness may substantially exceed currently catalogued levels (May 1988, 1990, Prance et al. 2000). Unfortunately, practical considerations generally prevent direct censuses of species at large spatial scales, and so methods for extrapolating data from small discrete plots, where accurate censuses are possible, to larger spatial scales are needed.

A potentially useful starting point for carrying out such extrapolations is the classic species-area relationship (Arrhenius 1921, Rosenzweig 1995)

$$S(A) = cA^z \quad (1)$$

where $S(A)$ is the average number of species in plots of area A .

Accepted 16 June 2004

Copyright © ECOGRAPHY 2004
ISSN 0906-7590

If z is independent of plot area, in other words if it is scale-independent, and its value is known, then extrapolation of species richness from small plots of area A to larger plots or an entire biome of area A' would be straightforward using eq. 1:

$$S(A') = (A'/A)^z S(A) \quad (2)$$

However, considerable evidence exists that z -values at small and large scales often differ (Rosenzweig 1995, Harte et al. 1999, Plotkin et al. 2000, Crawley and Harral 2001), and thus a method is needed to determine the value of z at scales larger than those at which z can be obtained readily from substitution of census data into eq. 1.

Recently a strategy was proposed for estimating the value of z at large spatial scales, such as that of an entire region, from species presence/absence data obtained on a subsample of small plots scattered throughout the region (Harte and Kinzig 1997, Harte et al. 1999). The conceptual basis for this approach originates with the observation that the larger the value of z , the greater the rate of increase of species richness with increasing area. Similarly, the greater the rate of increase of species richness with area, the greater the dissimilarity in species composition of two plots as the distance between these plots increases. Equivalently, the greater the value of z , the greater the rate of decrease of the similarity of species composition with increasing inter-plot distance.

To express this quantitatively consider a region in which a complete species list of, say, trees is available from a set of small plots, each of area A , distributed over the region. For every pair of plots the similarity in species composition of the two plots can be expressed as a Sorensen index

$$\chi = 2(S_1 \cap S_2)/(S_1 + S_2) \quad (3)$$

where S_1 and S_2 are the numbers of species in the two plots and the symbol $S_1 \cap S_2$ means the number of species found in both plots. In general, this measure will be a function of both the plot area, A , and the inter-plot distance, d . If the value of z is constant over a scale range characterized by values of interplot distances, d , in the scale range $d_{\min} < d < d_{\max}$, and if $A/d_{\min}^2 \ll 1$, then we have argued from application of eq. 1 and dimensional analysis that the approximate relation

$$\chi(A, d) \sim (A/d^2)^z \quad (4)$$

should hold over that range of d -values (Harte et al. 1999). If plot area A is held fixed, then eq. 4 implies that $\chi \sim d^{-2z}$.

Thus a test of scale independence of z within some scale range encompassing large distances, and determination of its value within that range, consists of plotting $\ln(\chi)$ versus $\ln(d)$, implies that over a scale range in which z is constant, the slope of that plot will be constant and equal to $-2z$. Then the species richness

at scale d_{\max}^2 , expressed as a function of species richness at scale d_{\min}^2 , can be estimated from eq. 2, where z is obtained from the $\ln(\chi)$ versus $\ln(d)$ plot. We note that eq. 2 can be iterated if more than one z -value is needed to accurately represent the entire scale range of interest.

A test of this approach was carried out in subalpine meadow habitat (Harte et al. 1999); eq. 4 was used to extrapolate vascular plant species richness from 30 censused 16-m² plots to a total area of order 10² km². In that study, the average species richness in the 16-m² plots was 31.9, and application of eq. 4 predicted a total of 95 species in the larger area, in good agreement with results of plant surveys throughout the same region.

Here we test this strategy in the rainforests of the Western Ghats mountains of southern India over a much larger scale range. We estimate the tree species richness at a scale of 60 000 km², where tree species richness is approximately known so that a test is possible, using only tree census data gathered from 48 quarter-ha plots distributed throughout the region. Thus we will be using eqs 2 and 4 to extrapolate species richness over a scale range of 60 000/0.0025 = 2.4 × 10⁷.

Study area and methods

We used vegetation maps of the Western Ghats (Gausson et al. 1961, 1965; with additional information from Ramesh and Pascal 1997) to identify areas with rainforests and to select sites for sampling. The plots were randomly selected, the only constraint being that the plots were in the least disturbed parts. The surveys were done in the rainforests between the southern tip of the Western Ghats (southern India) at 8°30'N and Sharavathy River in Karnataka State, at 14°30'N, spanning a distance of almost 720 km (Fig. 1). The rainforest north of Sharavathy River is highly fragmented and degraded, often beyond recognition. The altitudinal range was between 100 m, the lowest altitude at which rainforest is found in the Western Ghats, and 1500 m above which the rainforest changes into montane rainforest and occurs often in association with montane grassland. The trees (>30 cm girth at breast height) were enumerated from 48 plots each of 250 × 10 m (0.25 ha). Each plot was further subdivided into 10 plots of 25 × 10 m. The vegetation was surveyed between 1994 and 2000. A floristic list of the trees of the Western Ghats is available as supplementary material at <http://www.rainforest-initiative.org/downloads/pdfs/wghats_trees.pdf>.

We used ordinary least squares (OLS) regression of $\ln(\chi)$ on $\ln(d)$ to estimate, from eq. 4, z values at spatial scales larger than that of the 0.25 ha plots. Commonality data were binned into equi-distance intervals on a logarithmic scale. Because of non-independent data, we

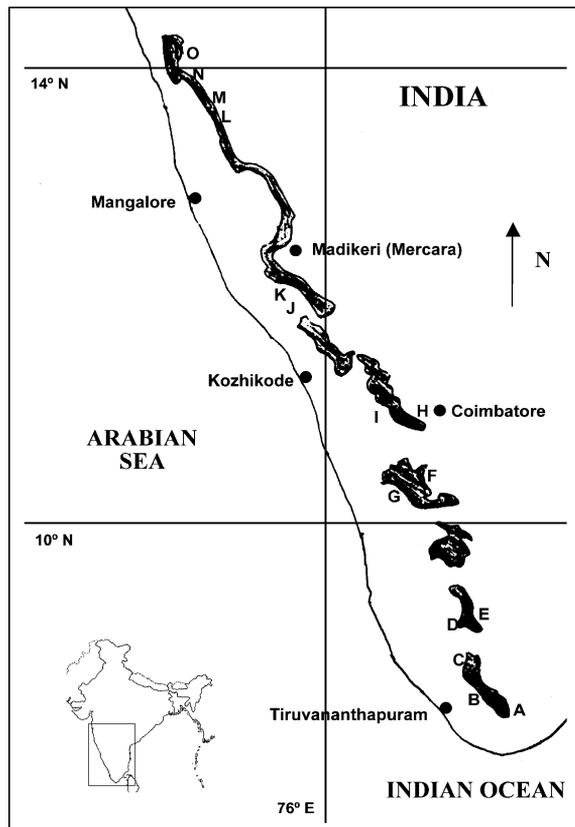


Fig. 1. Map showing the study areas where sampling was done. The distance between the study areas A and O is ca 720 km. A) Kalakkad-Mundanthurai Tiger Reserve (8), B) Peppara Wildlife Sanctuary (1), C) Shendurney Wildlife Sanctuary (3), D) Periyar Tiger Reserve (3), E) Megamalai Reserved Land (2), F) Indira Gandhi Wildlife Sanctuary (4), G) Parambikulam Wildlife Sanctuary (3), H) Bolampatti Reserved Forest (3), I) Silent Valley National Park (3), J) Aralam Wildlife Sanctuary (3), K) Brahmagiri Ghats Wildlife Sanctuary (3), L) Kudremukha National Park (3), M) Someshwara Wildlife Sanctuary (3), N) Mookambika Wildlife Sanctuary (3) and O) Sharavathy Valley Wildlife Sanctuary (3). The numbers in the parentheses denote the number of plots laid in each of the areas.

used a Mantel test (Sokal and Rohlf 1995) to estimate the significance of the regression results. Within-plot z values were estimated by OLS regressions of $\ln(S)$ on $\ln(A)$.

Results

A total of 341 tree species were recorded in the 48 (0.25 ha) plots, of which 8 remain unidentified. One plot had considerably lower species richness than all other plots (Fig. 2), due to the monodominance of one particular species (*Diospyros foliolosa* Wall.). Monodominant patches in rainforests may arise and disappear regularly,

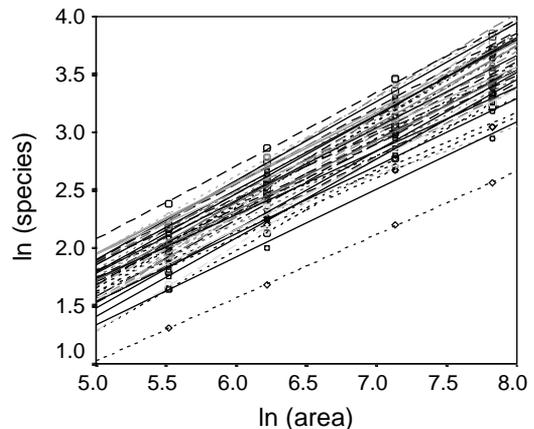


Fig. 2. Species accumulation in relation to area (in m^2) in the 48 plots. Each plot (250×10 m) was divided into 4 subplots of length 25, 50, 125 and 250 m.

and there are many possible reasons for their occurrence (see Torti et al. 2001). The mean species richness of the individual plots was 32.9 (SD = 7.53). Within each plot, a value for z at plot scale was obtained by fitting eq. 1 to nested census data for nested subplots with a width of 10 m and lengths of 25 m ($n=10$), 50 m ($n=5$), and 125 m ($n=2$), and the entire plot with dimensions of 10×250 m. Eq. 1 provided a good model for intraplot scaling of species richness for all 48 plots. The fitted z values ranged from 0.42 to 0.81, and were clustered around a mean of $z=0.65$ (SD = 0.07) (Fig. 2). The R^2 values for linear fits to $\ln(S)$ vs $\ln(\text{area})$ ranged from 0.976 to 1.000, with an average of $R^2=0.996$.

The R^2 values for the species-area regressions and for the commonality-distance regressions characterize the fraction of explained variance but because of non-independent data, significance levels cannot be assigned to these regressions. A Mantel test, which overcomes this problem (Sokal and Rohlf 1995), on the whole data set gave significant correlations; in fact, the Pearson "r" and the Mantel "r" values were very similar and had similar p-values.

These z values are higher than is typical in species-area analyses (Rosenzweig 1995). This is in part because the nested subplots of increasing area all have the same width and are increasingly elongated. As shown previously (Kunin 1997, Harte et al. 1999), this shape dependence leads to a larger value for z than would be obtained by censusing patches that retain their shape as area increases. Moreover, the effect on z of patch elongation can be approximately compensated for, and a corrected value obtained which should be approximately valid for shape-preserving patches (Harte et al. 1999). Applying the method of Harte et al. (1999) to this data set, we obtain a mean z -value of 0.43 for shape-preserving plots at the scale of 0.25 ha.

Recognizing that the shape of the plots (10×250 m) is roughly similar to the shape of the rainforest region of the Western Ghats (40×1500 km), we can estimate the total species richness in the region if these z values, applicable at the scale of the 0.25 ha plots, remained constant with scale for the entire Western Ghats. Using the average corrected z of 0.43, eq. 2 implies that the total tree species richness for the region would be

$$S(60\,000\text{ km}^2) = (60\,000\text{ km}^2 / 0.0025\text{ km}^2)^{0.43} (32.9) \\ \cong 49\,000\text{ species.}$$

This is certainly incorrect as it is enormously higher than the known tree species richness of the well-studied Western Ghats, which is currently 857 species (Gamble 1928, Pascal and Ramesh 1987, Sasidharan 1997, Ramesh and Pascal 1997). Thus we deduce that at regional scale, z must be smaller than its directly measured value at plot scale. To determine z at large scale we use commonality data in conjunction with eq. 4.

To estimate the value of z at the scale of an area A , we require commonality data for $d \sim A^{1/2}$. Hence, to estimate z up to region scale, we require commonality data for interplot distances up to $d \sim 240$ km. Application of eq. 4 to the binned commonality data calculated from eq. 3 yields a good fit to a power-law function over the scale range from $d \sim 1$ to $d \sim 200$ km. The slope of the fit of $\ln(\chi)$ vs $\ln(d)$ gives a z -value of 0.12 (Fig. 3). Note that the condition that eq. 4 be applicable, $A/d_{\min}^2 = 0.0025/1 \ll 1$, is satisfied. The value of the commonality index defined by eq. 3 will generally depend on sample size. In our data set, each plot is completely censused and therefore sample size here refers to plot size. Scaling

theory (Harte et al. 1999) predicts the dependence of commonality on plot area, A , to be $\chi \sim (A/d^2)^z$, a result that was successfully tested in a previous study (Harte et al. 1999).

We also have to estimate the value of z in the interval between plot scale (0.0025 km^2) and the smallest scale at which interplot commonality data provide an estimate (1 km^2). To this end, we use intraplot commonality data as measured by species overlap between pairs of 10×25 m subplots. We consider inter-subplot distances ranging from 100 to 225 m, corresponding to areas ranging from $d^2 = 0.01$ to $d^2 = 0.05\text{ km}^2$. We did not attempt to use data with $d < 100$ m because of the criterion that $A/d^2 \ll 1$. Regression of $\ln(\chi)$ against $\ln(d)$ for each plot yields $z = 0.34$; for about half the plots, the regression was insignificant at $p < 0.05$, but if only those plots for which the regression was significant are included in the analysis, the same value of z is obtained, with a 95% confidence interval of 0.03. We note that this z -value of 0.34 at intermediate areas is intermediate between $z = 0.12$ at larger scales ($d^2 \geq 1\text{ km}^2$) and $z = 0.43$ at smaller scales ($A < 0.0025\text{ km}^2$). We also note that our data set leaves us with two ‘‘blind spots’’ where we lack the data that would permit use of eq. 4: these are the intervals between 0.0025 and 0.01 km^2 and between 0.05 and 1 km^2 .

Now we can use eq. 2 to estimate the tree species richness of the entire region. We emphasize that although z is not constant across the huge range of scales considered here, eqs 2–4 nevertheless permit scale extrapolation. We first calculate the result assuming that in the two ‘‘blind spots’’, scale-breaks in the value of z occur at the arithmetic means (0.0063 and 0.53 km^2) between the bounding areas. In that case,

$$S(60\,000) = 32.9(0.0063/0.0025)^{0.43} \\ \times (0.53/0.0063)^{0.34} (60\,000/0.53)^{0.12} \\ = 893\text{ species} \quad (5)$$

If, in the intermediate region where z has a 95% confidence interval of ± 0.03 , we use a value for z of 0.31 or 0.37 then we get answers either $\sim 10\%$ smaller or larger than the above. If, as an alternative assumption about the ‘‘blind spots’’, we either take the geometric mean between the bounding areas or if we take the z -values to be given by their averages ($z = (0.43 + 0.34)/2$ in the first interval; $z = (0.34 + 0.12)/2$ in the second) then we get our lowest estimate of $S(60\,000\text{ km}^2) = 723$ species. Thus our estimates are bounded by ca 700 and 1000 species. Of course if we make the extreme assumptions that $z = 0.43$ applies throughout all of the first ‘‘blind spot’’ and $z = 0.34$ applies throughout all of the second, then we get a still larger result, and if we assume $z = 0.34$ applies throughout the first ‘‘blind spot’’ and $z = 0.12$ applies throughout the second, and then we get a still smaller result. Such assumptions are highly unlikely to be correct, however.

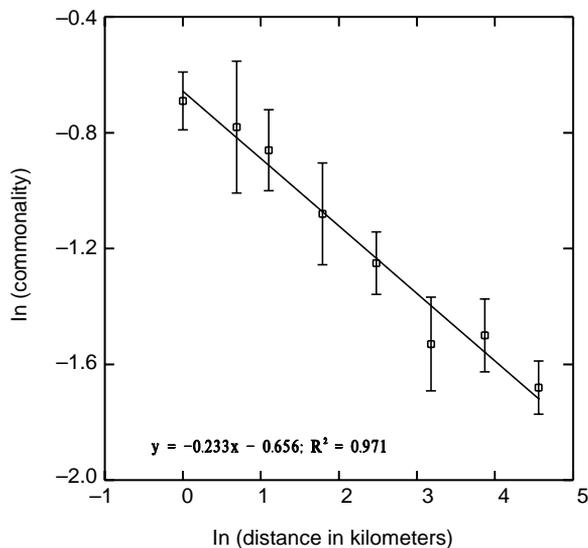


Fig. 3. Dependence of \ln of commonality on the \ln of interplot distance for binned commonality data. Error bars are 95% confidence intervals.

Discussion

Application of eqs 2 and 4 results in an estimate of between 700 and 1000 tree species in the sample space of 60 000 km². Our largest estimate is ca 10% greater than the catalogued total of 857 native tree species reported from the entire Western Ghats (Gamble 1928, Pascal and Ramesh 1987, Sasidharan 1997, Ramesh and Pascal 1997), while our lower estimate is ca 20% lower than the catalogued total. Tree species that are either new to science or have an extended range are currently being discovered at the rate of around four or five per year in the Western Ghats (Krishnamani 2002), and hence the already catalogued tree species richness of 857 is a lower limit to the actual value.

Our approach appears to provide a reasonable set of bounding estimates for total tree species richness in the rainforests of the Western Ghats. The analysis suggests that the use of commonality data to estimate values of the species-area exponent at large spatial scales provides a practical way to estimate species richness in regions too large to census directly. As shown here, our method can be applied whether or not z is constant over the scale range from small censused plots to an entire region. Our work also points to the importance of designing plot census surveys within large regions in such a way as to allow a reasonably uniform coverage of interplot distances, thereby avoiding the kinds of "blind spots" that resulted in the major ambiguity in our result.

We know of no alternative method for extrapolating species richness to large spatial scales from census data obtained from small discrete plots scattered throughout a landscape or region. Kempton (1979) felt that accurate estimation of actual species richness in a community requires a large sample size and an assumption about the form of the species abundance distribution. Again, Engen (1978), Chao and Lee (1992), Bunge and Fitzpatrick (1993) are of the opinion that the "actual species richness" of an area cannot be accurately extrapolated by any method using small samples containing a minority of species in a community. Even "nonparametric" estimators of actual species richness (Bunge and Fitzpatrick 1993) make implicit assumptions about species abundance distributions, as in principle there could be almost any number of extremely rare, unobserved species in a community. Non-parametric statistical methods (Colwell and Coddington 1994, Chazdon et al. 1998) can be used for analysis of species accumulation or collector's curves when some species are likely to be missed during collections, but they were developed for estimating species richness data in a contiguous partially censused plot or for samples collected over time. Considering all these, the method we have tested here, using discrete plots, has the potential to greatly improve our knowledge of the species richness at large spatial scale for many taxonomic

groups. In particular, for taxa such as arthropods where detailed census data are available from small plots or fumigated canopies scattered throughout a region, there is currently an order-of-magnitude uncertainty in the total biome-scale species richness (May 1990).

Acknowledgements – We thank the Forest Depts of Tamil Nadu, Kerala and Karnataka States for the necessary permits; H. Ramachandra Swamy, N. Sasidharan and K. Ravikumar for plant identification; Viola, Joya Thapa, Ronnie and Appu for other assistance; Chicago Zoological Society, Primate Conservation, Inc., International Primatological Society, Wildlife Conservation Society and National Geographic Society for funding the first author.

References

- Arrhenius, O. 1921. Species and man. – *J. Ecol.* 9: 95–99.
- Boulinier, T. et al. 1998. Estimating species richness: the importance of heterogeneity in species detectability. – *Ecology* 79: 1018–1028.
- Bunge, J. and Fitzpatrick, M. 1993. Estimating the number of species: a review. – *J. Am. Stat. Assoc.* 88: 364–373.
- Burnham, K. P. and Overton, W. S. 1979. Robust estimation of population size when capture probabilities vary among animals. – *Ecology* 60: 927–936.
- Chao, A. and Lee, S.-M. 1992. Estimating the number of classes via sample coverage. – *J. Am. Stat. Assoc.* 87: 210–217.
- Chazdon, R. L. et al. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of northeastern Costa Rica. – In: Dallmeier, J. A. and Comiskey, J. A. (eds), *Forest biodiversity research, monitoring and modeling: conceptual background and Old World case studies*. Man and Biosphere Ser. 20, UNESCO and Parthenon Publ. Group, Paris and New York, pp. 285–309.
- Colwell, R. K. and Coddington, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. – *Philos. Trans. R. Soc. Lond. B* 345: 101–118.
- Crawley, M. J. and Harral, J. E. 2001. Scale dependence in plant biodiversity. – *Science* 291: 864–867.
- Engen, S. 1978. Stochastic abundance models. – Chapman and Hall.
- Gamble, J. S. 1928. *Flora of the Presidency of Madras*. – Adlard, London.
- Gaussen, H., Legris, P. and Viart, M. 1961. International map of the vegetation: Cape Comorin. Scale 1/1,000,000. – French Inst., Pondicherry, India.
- Gaussen, H. et al. 1965. International map of the vegetation: Mysore. Scale 1/1,000,000. – French Inst., Pondicherry, India.
- Harte, J. and Kinzig, A. P. 1997. On the implications of species-area relationships for endemism, spatial turnover, and food web patterns. – *Oikos* 80: 417–427.
- Harte, J. et al. 1999. Estimating species-area relationships from plot to landscape scale using species spatial-turnover data. – *Oikos* 86: 45–54.
- Kempton, R. A. 1979. The structure of species abundance and measurement of diversity. – *Biometrics* 35: 307–321.
- Krishnamani, R. 2002. Phyto-ecology of the lion-tailed macaque (*Macaca silenus*) habitats in southern India. – Ph.D thesis, Bharathiar Univ., Coimbatore.
- Kunin, W. E. 1997. Sample shape, spatial scale, and species counts: implications for reserve design. – *Biol. Conserv.* 82: 369–377.
- Lande, R., DeVries, P. J. and Walla, T. R. 2000. When species accumulation curves intersect: implications for ranking diversity using small samples. – *Oikos* 89: 601–605.

- Longino, J. T., Coddington, J. and Colwell, R. K. 2002. The ant fauna of a tropical rain forest: estimating species richness three different ways. – *Ecology* 83: 689–702.
- May, R. M. 1988. How many species are there on Earth? – *Science* 241: 1441–1449.
- May, R. M. 1990. How many species? – *Philos. Trans. R. Soc. Lond. B* 330: 293–304.
- Palmer, M. W. 1990. The estimation of species richness by extrapolation. – *Ecology* 71: 1195–1198.
- Pascal, J.-P. and Ramesh, B. R. 1987. A field key to the trees and lianas of the evergreen forests of the Western Ghats (India). – French Inst., Pondicherry, India.
- Plotkin, J. B. et al. 2000. Predicting species diversity in tropical forests. – *Proc. Natl. Acad. Sci. USA* 97: 10850–10854.
- Prance, G. T. et al. 2000. The tropical flora remains under-collected. – *Ann. Mo. Bot. Gard.* 87: 67–71.
- Preston, F. W. 1948. The commonness, and rarity, of species. – *Ecology* 29: 254–283.
- Ramesh, B. R. and Pascal, J.-P. 1997. Atlas of endemics of the Western Ghats (India): distribution of tree species in the evergreen and semi-evergreen forests. – French Inst., Pondicherry, India.
- Rosenzweig, M. L. 1995. Species diversity in space and time. – Cambridge Univ. Press.
- Sasidharan, N. 1997. Forest trees of Kerala. – Kerala Forest Research Inst, Peechi, India.
- Sokal, R. R. and Rohlf, F. J. 1995. *Biometry*. – W. H. Freeman.
- Solow, A. R. 1994. On the Bayesian estimation of the number of species in a community. – *Ecology* 75: 2139–2142.
- Torti, S. D., Coley, P. D. and Kursar, T. A. 2001. Causes and consequences of monodominance in tropical lowland forests. – *Am. Nat.* 157: 141–153.