

The RAINFOR database: monitoring forest biomass and dynamics

Peacock, J.*; Baker, T.R.; Lewis, S.L.; Lopez-Gonzalez, G. & Phillips, O.L.

Earth and Biosphere Institute, School of Geography, University of Leeds, Leeds LS2 9JT, UK;

**Corresponding Author; Tel. +44 1133433347; Fax +44 1133433308; E-mail j.peacock@leeds.ac.uk*

0 000 records of individual trees and over 300 000 total records of tree diameter measurements from successive censuses. The database is designed to enable linkages to existing soil, floristic or plant-trait databases. This database will be a useful tool for exploring the impact of environmental factors on forest structure and dynamics at local to continental scales, and long term changes in forest ecology. As an early example of its potential, we explore the impact of different methodological assumptions on estimates of tropical forest biomass and carbon storage.

Keywords: Amazon; Data management; Ecoinformatics; Long-term; Plot; Tropical.

Abbreviations: RAINFOR = The Amazon Forest Inventory Network; DBH = Diameter at breast height; POM = Point of measurement; CTFS = Centre for Tropical Forest Science; ABG = Above-ground biomass; AVDL = Applied Vegetation Dynamics Laboratory.

Introduction

The Amazon Forest Inventory Network (RAINFOR), an international project established to monitor large-scale patterns of forest structure and dynamics across Amazonia (Malhi et al. 2002), has collected and collated inventory data from a large number of forest plots. We have recently developed a database to integrate data from over 100 permanent sample plots across the Amazon which have typically been studied for 10-20 years each, some as long as 30 years.

A primary aim of RAINFOR is to monitor forest biomass and dynamics across Amazonia in a coordinated fashion in order to understand their relationships with soil and climate. The network examines the continental-scale relations between forest structure, dynamics and environmental factors, and explores how the ecology of these forests may change over time. For example, initial publications have concentrated on spatial and temporal patterns of biomass, wood productivity and tree mortality, recruitment and growth rates (Malhi et al. 2002; Baker et al. 2004a, b; Lewis et al. 2004; Malhi & Phillips 2004; Phillips et al. 2004).

It is clear that the projects aims cannot be fully achieved without a powerful database to collate, sort, and extract data for analysis. The database would need to be able to handle data from repeated measures through time and store information on the diameter at breast height (DBH) of individual trees, tree status and tree death. Other ecological databases exist, but none have a structure to suit the requirements of the RAINFOR project. For example, the AVDL (Applied Vegetation Dynamics Laboratory) (Le Duc et al. 2007, this issue) database is designed to handle data from repeated measures through time, but it is also designed to handle metadata generated from different complex experiments and structured surveys (Le Duc et al. 2007, this issue) – this level of information is not required in the RAINFOR database as the field protocol for each plot and re-census is standardised. Additionally these databases, and other available web-based phytosociological databases such as VEGBANK (Jennings et al. 2004) and TURBOVEG (Hennekens & Schaminée 2001), are

not suitable for records of tree diameter, status or death modes, information which is key to the RAINFOR research agenda. VEGBANK and TURBOVEG are more suitable for describing plant communities rather than recording individual trees.

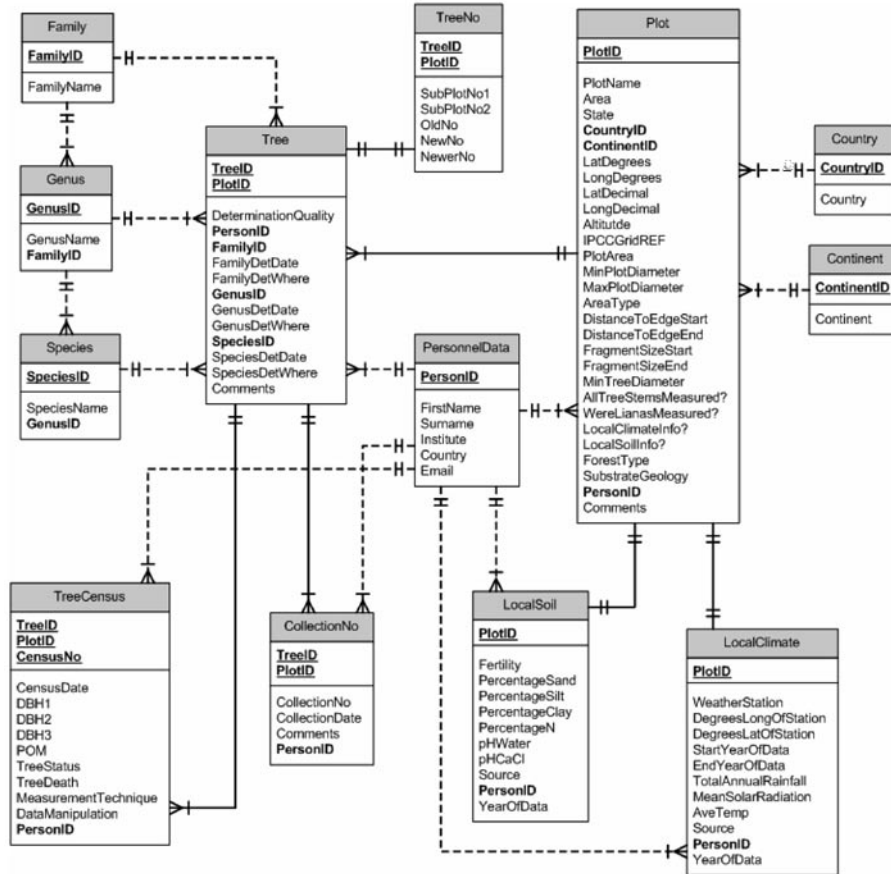


Fig. 1. Schema of the relations between the tables of the RAINFOR database, with a list of their respective fields. Primary keys are underlined, items in bold are required. Solid lines represent links between primary keys. The relationships between TreeNo and Tree, Plot and LocalSoil, and Plot and LocalClimate are one to one. All other relationships are one to many. The type of relationship is shown by the symbols: $\text{+} \leftarrow$ = many and $\text{+} \text{+}$ = one.

The database structure

The database consists of 13 linked tables (Fig. 1). The PersonnelData table is linked to most other tables. It contains the name and contact details, where appropriate, of the people involved in each step from the fieldwork, collection of samples, identification, providing environmental data, and data management in the UK.

The Plot table corresponds to each permanent sample plot in the field, each with a unique identification code. This table is designed so that data from different groups of plots such as, from a specified country or above a certain altitude, can be easily extracted. It also contains plot-level information, including plot size, substrate geology, and the dates of plot establishment and the most recent census, as well as the forest type, such as *terra firme*, seasonally flooded or swamp forest and the size of the fragment of forest where it is located, if appropriate.

In addition to the PersonnelData table, the Plot table is linked directly to two environmental tables. One contains information on the soil (LocalSoil) at each plot if

available and the other on the climatic (LocalClimate) data at each plot. The Plot table is also linked to the Tree table. Each tree in a plot is given an identification number, which, together with the plot identification code uniquely identifies each tree in the database. This table contains tree-level information including the codes for family, genus and species names for each tree and where the determination was carried out (in the field, in the herbarium or in the field and confirmed in the herbarium). Family, genus and species have all been coded and are contained in separate tables. Currently, in total there are > 2500 identified species in >100 families.

The TreeNo table contains all the original field numbers and subplots and makes comparison of field record sheets and the database efficient. This is important for correcting potential errors in the dataset.

The final table linked to the tree table is the TreeCensus table. The TreeCensus table has three primary keys – plot ID, tree ID and census number – which when used together identify a particular tree in a specific census. The tree census table contains the measured diameter

of each tree and the above-ground height at which it was measured, termed the point of measurement, POM. POMs sometimes have to be moved, as buttresses can develop on a stem and expand over time in many tropical trees. Hence, the database includes fields that standardise the diameter measurement to both the initial *and* the new POM (from the ratio of diameters at both POMs), which allows maximum flexibility in how the data can be subsequently analysed. Records where POM changes have occurred can be removed or utilised using a variety of methods to account for POM changes, for example, to remove spurious negative growth rates from individual stems when POMs have been changed, by utilising the field where growth is standardised to the initial or new POM. The tree status field codes for the state of the tree while it is alive. For example, these codes describe whether the tree is leaning or broken. Overall, there are 21 such codes which can be used in combination to describe a tree. The tree death field describes four aspects of tree death; how (e.g. broken trunk, or uprooted; whether it died alone or with other trees; whether it was the killed or the killer tree in a mortality event; and the most likely cause of death (e.g. disease, lightning strike).

Data entry and quality assurance

Data can be added in two ways: firstly bulk data from spreadsheets can be imported into the database. Secondly, forms have been designed which are useful for adding small amounts of data, or updating previously entered records. Both methods of data entry are simple and very little training is required to perform these tasks. Separate forms are used for updating tree (Fig. 2A) and plot data (Fig. 2B). Updating tree data is occasionally necessary, for example in cases where a tree is presumed dead during one census having not been found but is re-discovered in the next census.

In addition to in-field and data-entry quality control procedures, two database quality-control procedures further minimise the potential for errors to enter the data. Firstly, controls are set in each field (column) of every table so illegal data can not be entered (e.g. text in a numbers field, numbers outside a credible range, non-standard species names). Secondly after entering data several queries are run to check for errors. For example we calculate the total number of alive stems in a census, the number of stems that died and were recruited between census dates, the number of alive stems in the new census, and check that there is an internally consistent number of stems (Table 1).

Data extraction

Data can be extracted by one of two methods: firstly, using standard queries that have been pre-programmed, some static and some allowing user defined parameters, or secondly by user programmed queries. The first method requires little or no prior knowledge of the database structure, while the second requires greater understanding of the database but allows greater flexibility in the specific data extracted. Each pre-programmed query is accompanied by notes explaining the exact nature of the data that they extract. Extracted data can be imported into a spreadsheet or statistical package.

An example of a standard query with user defined parameters is one to calculate above ground biomass (AGB) per plot based on three different equations, the output from this query is shown in Table 2. The user can define which plots they wish to be included in the query and which census number or census date. The results from the query shown here cover the first census of each plot (Table 2 and Fig. 3). The calculation of biomass per plot is important when trying to quantify changes over time in carbon storage of forests, which is important for understanding potential impacts on global climate change (Prentice et al. 2001). This query calculates biomass using three different allometric equations. Equations 1 and 2 are based on the same tree harvest data (315 trees, harvested as part of the BIONTE project, near Manaus, Brazil); Eq. 1 calculates AGB on a per stand basis (Phillips et al. 1998), Eq. 2 calculates AGB on a per stem basis (Chambers et al. 2001). Eq. 3 was derived from an independent set of tree diameter and mass data of 378 trees (Chave et al. 2001). The equations used in the query are:

$$AGB = 0.6 * (66.92 + (16.85 * BA)) \quad (1)$$

$$AGB = \sum_1^n \exp(0.33(\ln D_i) + 0.933(\ln D_i)^2 - 0.122(\ln D_i)^3 - 0.37) \quad (2)$$

Table 1. Output of the standard query to check for errors in census data. A = number of trees alive in previous census; B = trees died by new census; C = trees recruited by new census; D = number of trees alive in new census; check, D = A+B+C, can return true or false, if false there is a mistake in the data.

QryCheckingStemCounts					
Plot ID	A	B	C	D	Check
CUZ-01	533	67	59	525	True
CUZ-02	539	56	75	558	True
CUZ-03	495	48	50	497	True
CUZ-04	599	91	83	591	True

$$AGB = \sum_1^n \exp(2.42(\ln D_i) - 2.00) \quad (3)$$

Where BA = basal area, calculated as $\pi(D_i/2)^2$

D_i is the diameter of a tree i and n is the number of trees per plot.

Three equations are used because, although equations that calculated AGB on a per stem basis are often presumed to be more accurate than stand-level basal area, sometimes the only published data is at stand-level, so it is useful to be able to derive biomass from stand level data.

Although the three equations give biomass estimates that are strongly correlated (r^2 varies from 0.854 for Eqs. 2 and 3, to 0.950 for Eqs. 1 and 2) there are differences in the estimated values (Fig. 3). For each possible set of two equations, biomass estimates were linearly regressed. The estimated slopes were tested to determine whether they significantly differed from one and the intercepts

Table 2. A sample of the output of a standard query that calculates biomass per plot (full output included 112 plots). Biomass estimates are in units of Mg.ha⁻¹.

QryComparingBiomassEquations (census one)			
PlotCode	Eq1	Eq2	Eq3
ACL-01	288.21	352.62	324.30
AGJ-01	342.26	354.07	233.13
ALM-01	160.83	181.65	148.14
ALP-11	289.22	324.89	295.42
ALP-12	252.29	291.77	283.71
ALP-21	269.51	306.78	299.44
ALP-22	252.65	279.40	241.07
ALP-30	227.87	264.11	260.67
BDF-03	288.47	332.00	338.72
BDF-04	276.15	307.96	332.33
BDF-05	251.97	280.38	298.02
BDF-06	255.65	285.44	299.49
BDF-07	290.14	334.58	366.99
BDF-08	282.92	325.53	337.30
BDF-09	303.53	356.91	389.15

A Tree Information

Plot: Tree Number: SpeciesComments:

Family: Genus: Species:

Census Information

TreeCensusID	TreeID	TreeNumber	CensusNo	CensusDate	dbh1	dbh2	dbh3	POM	Flag1	Flag2	Flag3	Flag4	Comments
92502	27688	6	1	1989.39	248	248	248	0 a	1	5	0		
92503	27688	6	2	1994.52	245	245	245	0 a	1	5	0		
92504	27688	6	3	1998.77	256	256	256	0 a	1	5	0		
92505	27688	6	4	2003.74	265	265	265	0 b	1	5	0		

(AutoNumber)

Collection Information

TreeID	Tree Number	Collection No	Collector	Date	Comments
27688	6	1720	Timana	1990	
27688	6	102	Phillips	1989	

Tree Numbers

TreeID	Tree Number	Subplot/transsect	Subplot/transsect2	Oldno	newno	newno2
27688	6	E-1		5	7	

B RAINFOR Plot Information

Plot Code: Plot Name: Area: State: Country: Continent:

Degrees **Decimal**

Latitude: Longitude: Latitude: Longitude: Altitude: IPCC Grid Reference:

Distance to Edge **Fragment Size**

Plot Area: MinDim: MaxDim: Area Type: Start: End: Start: End:

MinTreeDiam: All Tree Stems: Lianas: Local Climate: Local Soil: Forest Type: Substrate Geology:

Comments:

Fig. 2. Form for entering and updating. **A.** tree data. **B.** Plot data, within the RAINFOR database.

Table 3. *t*-distribution values and probabilities of significant differences, calculated following (Sachs 1984). * = $P < 0.02$, ** = $P < 0.001$; $df = 110$.

Equation pairs	Intercept different from 0	Slope different from 1
1 v 2	$t = 4.16^{**}$	$t = 10.73^{**}$
1 v 3	$t = 10.38^{**}$	$t = 1.78$
2 v 3	$t = 6.52^{**}$	$t = 2.43^{*}$

tested to determine if they were significantly different from zero (Sachs 1984) (Table 3).

Although the slope from the comparison of equations 1 and 3 does not depart significantly from one, the intercept is significantly different from zero (Table 3). This shows that the AGB estimates produced by the two equations are different by a constant value, with Eq. 3 providing lower estimates than Eq. 1 (Fig. 3B). The slopes produced from the comparison of Eqs. 1 and 2, and Eqs. 2 and 3 were significantly different from Eq. 1 (Table 3). This shows that the difference in output of these equations changes as biomass values change and not by an absolute value (Fig. 3A, C). Given these inconsistencies between the different models, including the three equations in the database allows users to assess the sensitivity of their conclusions concerning patterns of biomass and biomass change to uncertainty in the underlying allometric relationship (Chave et al. 2004).

Discussion and Conclusion

In developing the RAINFOR database it has been necessary to tackle several practical issues concerning the organisation of heterogeneous data: the plots contain tree records from two to 25 censuses, vary from 0.2 ha to 10 ha in size, and include multiple census dates from as early as the 1960s to 2006, over which time individual trees may be recruited, grow, have their point of measurement moved, and die. In addition, we need to track dynamic and changing nomenclature attributes, as well as harmonise the floristic data to common standards.

The RAINFOR database is a product of the large collaborative efforts of researchers working across Amazonia and beyond. It is intended to permit efficient and accurate storage of inventory data to provide information to these researchers to address questions about the structure and functioning of Amazonian forests. Developing collaborative networks of researchers is an increasingly important theme of research in tropical forests (e.g. Corlett & Primack 2006). The spatially distributed network of many long-term monitoring sites that RAINFOR maintains has led to insights into the changing

Fig. 3. The relationship between estimates of AGB calculated per stem and per plot. Dotted lines show a one to one relationship, through the origin. **A.** Results from Eq. 1 vs. Eq. 2. **B.** Results from eq. 1 vs. Eq. 3. **C.** Results from Eq. 2 vs. Eq. 3.

ecology of tropical forests (Baker et al. 2004a; Lewis et al. 2004; Phillips et al. 2004). The Centre for Tropical Forest Science (CTFS) large-scale plot network has led to fundamental insights into the processes maintaining tropical forest diversity (Losos & Leigh 2004), and the Amazon Tree Diversity Network (http://www.bio.uu.nl/~herba/Guyana/Amazon_plot_network/Introduction.htm; ter Steege et al. 2003) has drawn together data from several hundred single census plots to improve our understanding of patterns of diversity and composition of Amazonian forests. Development of the RAINFOR database is indeed itself indebted to the existence of the TROPICOS database at the Missouri Botanical Garden (<http://mobot.mobot.org/W3T/Search/vast.html>), the SALVIAS project (<http://www.salvias.net/pages/index.html>) and the CTFS Neotropical Tree Species Checklist (<http://ctfs.si.edu/neotropicaltree/>), themselves all products of major research collaborations. By uniquely drawing together numerous long-term inventory data from a wide geographical area, including plots in a wide range of habitats, and some of the longest running monitoring sites in Amazonia, the RAINFOR database will enable further collaborative research and provide an novel resource for addressing pressing questions about the current and future ecology of tropical forests in a rapidly changing world (Lewis 2006).

Acknowledgements. This work was funded with support from the EU Framework VI programme (project PAN-AMAZONIA), and a NERC grant (NE/B503384/1) and a Leverhulme Trust Research Fellowship to Oliver Phillips. Simon Lewis is supported by a Royal Society University Research Fellowship. We also thank Alison Manson, School of Geography Graphics Unit, University of Leeds, for her help in preparing Fig. 1.

References

- Baker, T.R., Phillips, O.L., Malhi, Y., Almeida, S., Arroyo, L., Di Fiore, A., Killeen, T., Laurance, S.G., Laurance, W.F., Lewis, S.L., Lloyd, J., Monteagudo, A., Neill, D.A., Patiño, S., Pitman, N.C.A., Silva, N. & Vásquez Martínez, R. 2004a. Increasing biomass in Amazonian forest plots. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359: 353-365.
- Baker, T.R., Phillips, O.L., Malhi, Y., Almeida, S., Arroyo, L., Di Fiore, A., Killeen, T., Laurance, S.G., Laurance, W.F., Lewis, S.L., Lloyd, J., Monteagudo, A., Neill, D.A., Patiño, S., Pitman, N.C.A., Silva, N. & Vásquez Martínez, R. 2004b. Variation in wood density determines spatial patterns in Amazonian forest biomass. *Global Change Biol.* 10: 545-562.
- Boyle, B.L. 2004. *TaxonScrubber Version 1.2*. The SALVIAS Project, <http://www.salvias.net>. (Accessed [05/04/2005]).
- Chambers, J.Q., dos Santos, J., Ribeiro, R.J., Higuchi, N. 2001. Tree damage, allometric relationships, and above-ground net primary production in a central Amazon forest. *For. Ecol. Manage.* 152: 73-84.
- Chave, J., Riera, B. & Dubois, M.A. 2001. Estimation of biomass in a neotropical forest of French Guiana: spatial and temporal variability. *J. Trop. Ecol.* 17: 79-96.
- Chave J., Condit, R., Aguilar, S., Hernandez, A., Lao, S. & Perez, R. 2004. Error propagation and scaling for tropical forest biomass estimates. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359: 409-420.
- Corlett, R.T. & Primack, R.B. 2006. Tropical rainforests and the need for cross-continental comparisons. *Trends Ecol. Evol.* 21: 104-110.
- Hennekens, S.M. & Schamineé, J.H.J. 2001. TURBOVEG, a comprehensive data base management system for vegetation data. *J. Veg. Sci.* 12: 589-591.
- Jennings, M., Faber-Langendoen, D., Peet, R., Loucks, O., Glenn-Lewin, D., Damman, A., Barbour, M., Pfister, R., Grossman, D., Roberts, D., Tart, D., Walker, M., Talbot, S., Walker, J., Hartshorn, G., Waggoner, G., Abrams, M., Hill, A. & Rejmánek, M. 2004. *Guidelines for describing associations and alliances of the U.S. National Vegetation Classification*. The Ecological Society of America, Vegetation classification Panel. (Download available at: http://vegbank.org/vegdocs/panel/NVC_guidelines_v4.pdf)
- Le Duc, M.G., Yang, L. & Marrs, R.H. 2007. A database application for long-term ecological field experiments. *J. Veg. Sci.* 18: 509-516.
- Lewis, S.L. 2006. Tropical forests and the changing Earth System. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 361: 195-210.
- Lewis, S.L., Phillips, O.L., Baker, T.R., Lloyd, J., Malhi, Y., Almeida, S., Higuchi, N., Laurance, W.F., Neill, D., Silva, N., Terborgh, J., Torres Lezama, A., Vásquez Martínez, R., Brown, S., Chave, J., Kuebler, C., Núñez Vargas, C. & Vinceti, B. 2004. Concerted changes in tropical forest structure and dynamics: evidence from 50 South American long-term plots. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359: 421-436.
- Losos, E. & Leigh, E.G. 2004. *Tropical forest diversity and dynamics: findings from a large scale plot network*. University of Chicago Press, Chicago, IL, US.
- Malhi, Y. & Phillips, O.L. 2004. Tropical forests and global atmospheric change: a synthesis. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359: 549-555.
- Malhi, Y., Phillips, O.L., Lloyd, J., Baker, T.R., Wright, J., Almeida, S., Arroyo, L., Frederiksen, T., Grace, J., Higuchi, N., Killeen, T., Laurance, W.F., Leão, C., Lewis, S., Meir, P., Monteagudo, A., Neill, D., Núñez Vargas, P., Panfil, S., Patiño, S., Pitman, N., Quesada, C.A., Rudas-Ll. A., Salomão, R., Saleska, S., Silva, N., Silveira, M., Sombroek, W.G., Valencia, R., Vásquez Martínez, R., Vieira, I.C.G. & Vinceti, B. 2002. An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). *J. Veg. Sci.* 13: 439-450.
- Phillips, O.L., Malhi, Y., Higuchi, N., Laurance, W.F., Núñez Vargas P., Vásquez Martínez, R., Laurence, S.G., Ferreira, L.V., Stern, M., Brown, S. & Grace, J. 1998. Changes in the carbon balance of tropical forests: evidence from long term plots. *Science.* 282:439-442.

- Phillips, O.L., Baker T.R., Arroyo, L., Higuchi, N., Killeen, T., Laurance, W.F., Lewis, S.L., Lloyd, J., Malhi, Y., Monteagudo, A., Neill, D., Núñez Vargas, P., Silva, N., Terborgh, J., Vásquez Martínez, R., Alexiades, M., Almeida, S., Brown, S., Chave, J., Comiskey, J.A., Czimczik, C.I., Di Fiore, A., Erwin, T., Kuebler, C., Laurance, S.G., Nascimento, H.E.M., Palacios, W., Patiño, S., Pitman, N., Olivier, J., Quesada, C.A., Saldias, M., Torres Lezama, A. & Vinceti, B. 2004. Pattern and process in Amazon tree turnover, 1976-2001. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359: 381-407.
- Prentice, I.C. (and 60 others) 2001. The carbon cycle and atmospheric carbon dioxide. In: IPCC (ed.) *Climate change 2001: the scientific basis*, pp. 183-197. Cambridge University Press, Cambridge, UK.
- Sachs L. 1984. *Applied Statistics: A handbook of techniques*. 2nd. wd. Springer-Verlag, New York, NY, US.
- Ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., Van der Hout, P., Daly, D.C., Silveira, M., Phillips, O.L., 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. *Biodivers. Conserv.* 12: 2255-2277.

Received 6 February 2006;
Accepted 6 December 2006;
Co-ordinating Editor: R. Bekker.