https://doi.org/10.26525/jtfs2017.29.3.325333

CLUSTER AND DISCRIMINANT ANALYSES FOR STEM VOLUME MODELLING OF TREE SPECIES GROUPS IN AN AMAZON RAINFOREST

VC Cysneiros[1, *], AL Pelissari[1], SA Machado[1], A Figueiredo-Filho[2] & HC David[3]

[1]*Federal University of Parana, Av Prof Lothário Meissner, 632, Botanical Garden, Curitiba, Paraná State, Brazil, 80210-170*
[2]*Department of Forest Engineering, State University of the Central West, PR-153, Km 7, Irati, Paraná State, Brazil, 84500-000*
[3]*Rural Federal University of Amazonia, Travessa Pau Amarelo, Capitão Poço, Pará State, Brazil, 68650-000*

**vccysneiros.florestal@gmail.com*

Submitted June 2016; accepted November 2016

The diversity of species in native tropical forests causes difficulty in the interpretation of data that support their management and conservation. Species grouping, based on characteristics of interest, reduces significantly the number of volume equations and helps solve the problem of undersampling rare species. This study aims to group 32 Amazonian trees species of commercial interest based on regression coefficients of the Schumacher and Hall's model and their fit statistics. To accomplish this, we employ a two-stage approach, in which we first applied cluster analysis to classify species with higher sampling intensity (n > 30). This phase allowed us to allocate poorly sampled species (n < 30) to groups created by discriminant analysis, resulting in the second stage. This proposed approach has proven adequate for grouping timber species in the Amazon forest, and so the stem volume can be modelled on consistent groups of species. The grouping of Amazon rainforest commercial species, based on the regression coefficients and fit statistics, performs better in aggregation for the stem volume modelling, providing stabilisation of estimation error and supplying few equations for the evaluation of standing stock.

Keywords: Tropical forest, timber species, volume equations, multivariate analysis, sample intensity

INTRODUCTION

Tropical forests are the most diverse terrestrial ecosystem (Turner 2001), characterised by a large number of species with different growth patterns (Vanclay 1991). The Amazon rainforest possesses one of the richest sets of plant species in the world, i.e. approximately 16,000 tree species (ter Steege et al. 2013). However, due to intense shrinking of the rainforest as a result of illegal logging and expansion of agriculture, the sustainable management of these forests has become necessary (Gutierrez-Velez & Macdicken 2008).

Due to the structural complexity of tropical forests, modelling of the whole tree measurement data set can result in high estimation errors and hide important descriptors of these forests. Meanwhile, processing based on individual species involves greater complexity and hinders interpretation (Phillips et al. 2002, Akindele & LeMay 2006). Thus, species aggregation into groups tends to reduce the number of equations (Vanclay 1991) and avoids the requirement that specific equations be developed for species with low sample intensitiy.

Various methods have been used to group tropical wood species (Akindele & LeMay 2006). In ecological studies, species are often grouped according to common ecological characteristics such as life cycle, reproduction, propagation, growth rate, photosynthetic capacity and regeneration (Swaine & Whitmore 1988). However, ecological information on tropical species is scarce and the classification method can introduce subjectivity into group formation when fitting a volume equation. This subjectivity occurs especially because of large variability in the stem form, even among individuals of the same species.

Cluster analyses are often used for classification and discrimination of dendrometric data in native forests. These techniques generate similarity classifications that exclude the subjective aspects existing in other sorting methods (Chuman & Romportl 2010). In these assessments, the objects

© *Forest Research Institute Malaysia*

under study are interconnected in a hierarchy of levels, where the most similar objects are gathered to form groups and subgroups.

In forestry, examples of grouping application for data analysis are found in the fitting of stem volume models (Akindele & LeMay 2006), growth and yield studies (Vanclay 1991, Köhler & Huth 1998, Phillips et al. 2002), production stratification (Souza & Souza 2006), phytogeographic studies (Oliveira-Filho & Fontes 2000) and tropical species on distribution patterns (Plotkin et al. 2002). To fit stem volume models in native forests, species grouping using dendrometric characteristics can offer several advantages such as better results than individual equations, smaller number of equations, solving sampling problem related to species with low sample size and, by not following ecological standards in species classification, it avoids subjectivities in group formation (Akindele & LeMay 2006).

Improvement of commercial species grouping techniques can be promising for stem volume modelling in the Amazon rainforest, where this approach is still emerging. The aim of this study was to group 32 Amazonian commercial species based on the regression coefficients of the Schumacher and Hall's model and to combine them with their fit statistics. To accomplish this, we used a multivariate approach in two stages, with cluster and discriminant analyses for the formation and classification of species groups.

MATERIALS AND METHODS

Study area and data collection

The study area is the Jamari National Forest, located in the south-west of the Amazon rainforest, between the geographic coordinates 9° 0' to 9° 30' S and 62° 44' to 63° 16' W. This National Forest is a pioneer in native forest concessions in Brazil, covering an area of approximately 220,000 ha and dominated by tropical rainforest vegetation. According to the Köppen classification system, the climate is tropical rainy Aw, with a well-defined dry period in the winter season. Average annual rainfall is 2400 mm and average temperature is 25 °C.

We used the Smalian's method for stem volume calculation (Figueiredo-Filho 1983) from 5230 trees of 32 commercial species. Due to the high variability of data, outliers were detected using the Grubbs' test (Grubbs 1969) supported by graphic dispersion analysis among the variables. After the exclusion of outliers and separation of an independent sample for validation, the remaining 4366 sample trees formed the database used for the analyses.

Data analysis

We fit the Schumacher and Hall's (Akindele & LeMay 2006) model for each one of the 32 species and, subsequently for the groups formed by cluster and discriminant analyses. We used linear forms of this model, in which the volume was a function of the variables diameter at breast height and commercial height (Clutter et al. 1983).

$$\mathrm{Ln}(v) = \beta_0 + \beta_1 \times \mathrm{Ln}(d) + \beta_2 \times \mathrm{Ln}(hc)$$

where v = commercial volume (m^3), d = diameter at breast height (cm), hc = commercial height between the base of the tree and its morphological inversion point or to the first branch (m) and β_0, β_1 and β_2 = regression coefficients.

In the first stage, we applied the cluster analysis to constructed species groups based on regression coefficients (β_0, β_1 and β_2) of the Schumacher and Hall's model fitted for each one of the 21 species with the highest sample density ($n > 30$), according to the methodology applied by Akindele & LeMay (2006). As an alternative method for grouping species, we tested the combination of these regression coefficients with their fit statistics: standard error of the estimate and coefficient of determination (r^2).

We used the average method and Euclidean distance (Phillips et al. 2002, Akindele & LeMay 2006) and calculated cophenetic correlation coefficient to evaluate the degree of fit between the original matrix and the resulting matrix from the cluster process (Rohlf 1970). The cut point used in the cluster analyses was determined by the graphical method, plotting the fusion coefficients of the group and the respective similarity distances. The first stabilising trend indicated the cutting point in the dendrogram (Reis 1997, Albuquerque et al. 2005). The PROC CLUSTER in the SAS 9.0 software was used to perform the analyses.

$$d_{x,y} = \sqrt{\sum_{j=1}^{j} \left(x_j - y_j\right)^2}$$

where $d_{x,y}$ = Euclidean distance between groupings and x_j and y_j = analysed distance vectors.

$$rcof = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(c_{ij}-\bar{c})(d_{ij}-\bar{d})}{\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(c_{ij}-\bar{c})^2\right)^{1/2}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(d_{ij}-\bar{d})^2\right)^{1/2}}$$

where,

$$\bar{c} = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} c_{ij}$$

$$\bar{d} = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{ij}$$

and rcof = cophenetic correlation coefficient, c_{ij} = distance between i and j individuals in the cophenetic matrix, d_{ij} = distance between the same individuals in the original matrix and n = size of the matrix.

In the second stage, we used discriminant analysis in order to allocate the 11 species with low sample density (n < 30) to the pre-existing groups (Akindele & LeMay 2006). We used the Fisher's linear function in order to transform multivariate observations in univariate or linear combinations, which separate populations as much as possible (Johnson & Wichern 1992). In order to assess the effectiveness of the discriminant analysis, we applied the lambda Wilks's test (Rencher 2002).

$$D_m^2(x) = (x - \bar{x}_m)' COV^{-1}(x - \bar{x}_m)$$

where $D_m^2(x)$ = Fisher's linear discriminant function, x = average value of the vectors, $\bar{x}_m$ = grouping centroid and COV = covariance matrix.

$$\Lambda = \prod_{j=1}^{M}\left(\frac{1}{1+\widehat{\lambda_{(J)}}}\right)$$

where Λ = Wilks's lambda and $\widehat{\lambda_{(J)}}$ = square of the canonical correlation.

The fitted equations were compared with the coefficient of determination (r^2) and the standard error of the estimate (Syx%). The evaluation of the goodness of the fits was based on graphical analysis of residuals which was critical in choosing the regression model, even if the other statistical criteria suggested an alternative model (Draper & Smith 1998):

$$r^2 = 1 - \left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right)$$

$$Syx\,\% = \left[\frac{\left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-p)}\right)}{\bar{y}}\right] \times 100$$

where y_i = observed value, $\hat{y}_i$ = estimated value by the model, n = number of observations, p = number of model coefficients, and $\bar{y}$ = average observed values of the dependent variable.

RESULTS

Table 1 presents the summary statistics by species to describe the data set used in this study. The regression coefficients and fit statistics of Schumacher and Hall's model fitted for the 32 Amazon rainforest species of commercial interest are given in Table 2. These species are distributed in 13 botanical families, with greater representation by the Fabaceae comprising 11 commercial species, followed by Sapotaceae, Lecythidaceae and Moraceae, with three species each. The data showed high variation amplitudes, with diameters and commercial heights ranging from 50 to 245 cm and from 5.2 to 43.4 m respectively. The pronounced variability of the data reflected in the fit statistics generated estimation errors between 11.27 and 46.67% and r^2 between 0.166 to 0.96.

For the cluster analysis, graphical method using fusion coefficient indicated different cutting points, forming eight and four distinct groups respectively (Figure 1). In both dendrograms, there were groups with only one species and they were incorporated into the nearest group from each of them. This was done to facilitate implementation of the Fisher's linear function which requires variability within the groups.

In dendrogram 1 (Figure 1a), *Hymenaea intermedia* and *Caryocar glabrum* were joined together due to similarity distance, forming a new group. In dendogram 2 (Figure 1b), *Cedrelinga cateniformis* was incorporated into the nearest group. The cophenetic correlation coefficients calculated for the dendrograms (0.813 and 0.721, respectively) were greater than 0.7, indicating a good fit from the original matrix to the generated matrix using the cluster analysis.

The two-stage approach was used to group the species as shown in Table 3. For both dendrograms, the Wilks's test was significant at 5%, indicating discrimination between the resulting groups. By comparing the degree of discrimination by the Wilks's lambda value,

Table 1 Descriptive statistics of variables used for data processing

Species	DBH (cm)				Commercial height (m)				Commercial volume (m^3)			
	Min	Mean	Max	SD	Min	Mean	Max	SD	Min	Mean	Max	SD
Allantoma decandra	50.0	80.5	140.0	17.5	13.5	21.7	31.8	4.2	2.49	7.14	19.85	3.8
Apuleia leiocarpa	57.0	87.7	165.0	19.6	10.6	20.1	31.2	3.6	2.43	8.47	24.42	4.3
Astronium lecointei	50.3	75.6	146.4	14.1	7.3	26.7	43.4	3.9	1.60	7.72	28.75	3.3
Bagassa guianensis	57.0	84.6	124.0	21.4	16.8	19.6	24.0	2.3	2.55	7.12	15.10	3.0
Bowdichia nitida	50.3	64.6	89.0	9.0	13.6	20.7	26.9	3.6	2.81	4.98	9.03	1.5
Brosimum rubescens	51.0	75.9	134.0	13.7	9.4	17.8	23.5	2.9	2.29	5.53	12.18	2.2
Cariniana micrantha	62.0	114.4	188.0	31.0	16.8	22.8	29.2	2.6	3.18	16.59	37.62	8.5
Caryocar glabrum	51.3	79.9	150.0	16.6	6.1	14.6	21.9	2.9	2.28	5.87	18.98	3.0
Caryocar villosum	57.3	90.4	143.0	16.2	6.2	15.2	23.5	2.9	1.85	7.37	19.41	3.3
Cedrela fissilis	52.5	75.0	99.0	14.5	10.5	15.4	19.5	3.1	1.89	4.08	9.44	2.1
Cedrelinga cateniformis	57.0	97.7	205.0	27.1	5.2	19.4	29.8	4.2	2.15	10.17	29.60	6.0
Clarisia racemosa	51.0	70.5	105.7	9.8	5.2	15.7	25.2	3.0	1.44	4.43	10.22	1.5
Cordia goeldiana	57.0	77.5	105.0	12.7	20.6	25.7	36.1	4.1	3.84	7.21	15.74	2.7
Couratari stellata	60.0	88.3	210.0	19.1	13.9	25.8	39.9	3.9	2.52	10.64	35.25	5.1
Dinizia excelsa	55.7	106.1	245.0	27.6	8.0	19.2	34.2	4.1	2.62	14.65	55.74	8.2
Diplotropis rodriguesii	54.4	66.1	80.9	8.6	12.7	20.5	27.5	3.2	3.06	4.30	6.96	0.9
Dipteryx alata	52.0	66.3	76.0	7.4	15.1	18.2	21.8	2.2	2.10	5.83	7.89	1.9
Dipteryx odorata	50.3	75.6	207.0	17.9	7.8	17.0	31.3	3.5	1.73	5.57	18.13	2.8
Erisma bicolor	53.0	80.7	150.0	17.6	11.4	19.3	28.9	3.3	2.37	6.47	21.72	3.3
Erisma fuscum	51.0	74.9	134.0	15.6	10.0	20.2	28.9	3.7	2.31	5.87	15.88	2.3
Goupia glabra	53.0	83.2	143.2	17.3	5.4	14.7	29.7	3.1	1.80	6.00	22.24	3.1
Handroanthus impetiginosus	67.0	90.0	117.8	12.7	20.8	26.4	32.2	3.4	4.63	12.36	21.98	5.2
Handroanthus incanus	52.5	75.2	114.9	15.4	13.8	27.0	34.3	4.3	2.56	6.70	22.37	3.7
Hymenaea intermedia	52.0	72.3	132.0	14.1	13.4	21.4	30.0	3.3	2.19	6.32	17.54	2.5
Hymenolobium heterocarpum	52.0	95.2	216.0	27.1	7.6	20.5	36.2	3.8	1.90	11.06	49.23	7.2
Manilkara elata	50.9	70.3	103.1	13.8	18.1	21.9	25.8	2.7	2.66	5.23	8.25	1.6
Mezilaurus synandra	60.5	73.9	102.5	12.0	7.7	17.3	26.7	3.9	1.84	5.05	11.19	2.5
Peltogyne paniculata	50.0	68.3	178.0	11.3	7.8	16.5	33.9	3.3	1.41	4.11	10.80	1.4
Peltogyne venosa	57.0	77.7	114.6	19.1	11.9	18.8	24.7	4.9	1.88	5.26	9.14	2.2
Pouteria guianensis	50.3	64.6	104.1	9.4	11.9	18.4	28.4	3.3	2.04	3.96	7.99	1.1
Qualea paraensis	50.0	67.9	119.0	12.0	10.4	22.4	40.7	3.9	1.88	5.37	13.87	2.2
Simarouba amara	50.3	60.4	78.0	7.4	11.6	19.3	26.2	4.0	2.03	3.57	7.78	1.4

DBH = diameter at 1.3 m and SD = standard deviation

dendrogram 1 ($\Lambda = 0.005$) showed a value lower than dendrogram 2 ($\Lambda = 0.083$), indicating greater discrimination between them. The resulting values showed that this analysis was suitable for aggregation of commercial species based on their stem form.

Regression coefficients of the fitted model and the fit statistics for the two dendrograms are given in Table 4. Both dendograms had higher estimation errors in groups with species of larger diameter such as *Dinizia excelsa*, *Cedrelinga catenaeformis* and *Hymenolobium heterocarpum*. The coefficient of determination ranged from 0.44 to 0.81 between formed groups, showing moderate correlation of variables, i.e. diameter at breast height and commercial height with stem volume.

Comparing the two classification methods employed, dendrogram 2 performed best,

Table 2 Regression coefficients and statistics of the Schumacher and Hall's model fitted for 32 commercial tree species of the Amazon rainforest

Species	n	β_0	p-value	β_1	p-value	β_2	p-value	Syx%	r^2
Astronium lecointei	640	-8.210	< 0.001	1.662	< 0.001	0.922	< 0.001	24.60	0.686
Peltogyne paniculata	623	-5.303	< 0.001	1.093	< 0.001	0.742	< 0.001	25.12	0.451
Dinizia excelsa	456	-6.738	< 0.001	1.538	< 0.001	0.738	< 0.001	36.01	0.498
Couratari stellata	420	-7.955	< 0.001	1.656	< 0.001	0.880	< 0.001	28.77	0.612
Hymenolobium heterocarpum	264	-8.092	< 0.001	1.644	< 0.001	0.975	< 0.001	31.82	0.702
Clarisia racemosa	245	-6.891	< 0.001	1.473	< 0.001	0.760	< 0.001	17.65	0.677
Dipteryx odorata	242	-7.801	< 0.001	1.668	< 0.001	0.798	< 0.001	27.51	0.721
Qualea paraensis	207	-7.366	< 0.001	1.633	< 0.001	0.679	< 0.001	22.30	0.688
Goupia glabra	156	-7.295	< 0.001	1.599	< 0.001	0.726	< 0.001	28.84	0.664
Apuleia leiocarpa	127	-7.684	< 0.001	1.625	< 0.001	0.827	< 0.001	32.32	0.608
Caryocar glabrum	121	-6.390	< 0.001	1.423	< 0.001	0.696	< 0.001	27.82	0.662
Brosimum rubescens	108	-7.658	< 0.001	1.525	< 0.001	0.948	< 0.001	24.32	0.652
Cariniana micrantha	104	-7.996	< 0.001	1.682	< 0.001	0.883	0.001	29.70	0.672
Erisma bicolor	84	-7.712	< 0.001	1.654	< 0.001	0.759	< 0.001	23.48	0.802
Erisma fuscum	74	-6.718	< 0.001	1.447	< 0.001	0.740	< 0.001	17.14	0.825
Hymenaea intermedia	63	-6.094	< 0.001	1.527	< 0.001	0.452	0.006	20.28	0.750
Allantoma decandra	60	-8.444	< 0.001	1.929	< 0.001	0.607	< 0.001	24.00	0.958
Caryocar villosum	59	-7.969	< 0.001	1.806	< 0.001	0.654	< 0.001	33.95	0.435
Pouteria guianensis	55	-5.448	< 0.001	1.047	< 0.001	0.841	< 0.001	20.96	0.961
Handroanthus incanus	51	-8.516	< 0.001	2.023	< 0.001	0.492	< 0.001	26.10	0.794
Cedrelinga cateniformis	47	-8.038	< 0.001	1.781	< 0.001	0.717	< 0.001	46.67	0.380
Mezilaurus synandra	25	-8.691	< 0.001	1.997	< 0.001	0.574	0.020	22.74	0.792
Bowdichia nitida	24	-5.970	< 0.001	1.478	< 0.001	0.459	0.033	19.12	0.681
Cordia goeldiana	20	-8.549	< 0.001	1.904	< 0.001	0.684	< 0.001	11.27	0.934
Simarouba amara	20	-9.637	< 0.001	1.952	< 0.001	0.969	< 0.001	15.31	0.869
Bagassa guianensis	12	-1.784	0.557	1.062	0.32	-0.366	0.689	32.04	0.289
Cedrela fissilis	12	-8.365	< 0.001	1.581	< 0.001	1.055	0.010	30.42	0.729
Diplotropis rodriguesii	12	-2.069	0.402	0.582	0.222	0.363	0.295	21.61	0.166
Manilkara elata	11	-7.629	< 0.001	1.565	< 0.001	0.840	0.064	13.86	0.843
Handroanthus impetiginosus	11	-6.144	0.177	1.421	0.082	0.716	0.349	33.23	0.259
Peltogyne venosa	8	-3.018	0.254	0.104	0.867	1.372	0.052	25.25	0.778
Dipteryx alata	5	-10.655	0.082	2.805	0.086	0.172	0.873	25.41	0.811

n = number of trees, β_0, β_1 and β_2 = regression coefficients, Syx% = standard error of the estimate and r^2 = coefficient of determination

as it used the combination of the regression coefficients with fit statistics as base for species grouping. This dendrogram had stable errors in the formed groups, with reduced error for some species such as *Cedrelinga cateniformis*, *Bagassa guianensis*, *Peltogyne venosa* and *Handroanthus incanus*. Another advantage offered by the proposed method was the use of only four volumetric equations for 32 commercial species. Dendrogram 1, reduced the estimate error for some species. However, the estimate error was substantially increased in two groups, i.e. > 40%. Dendrogram 1 generated eight groups and required larger number of volume equations eiight equations, one equation for each group). The graphical analysis of residuals (Figure 2)

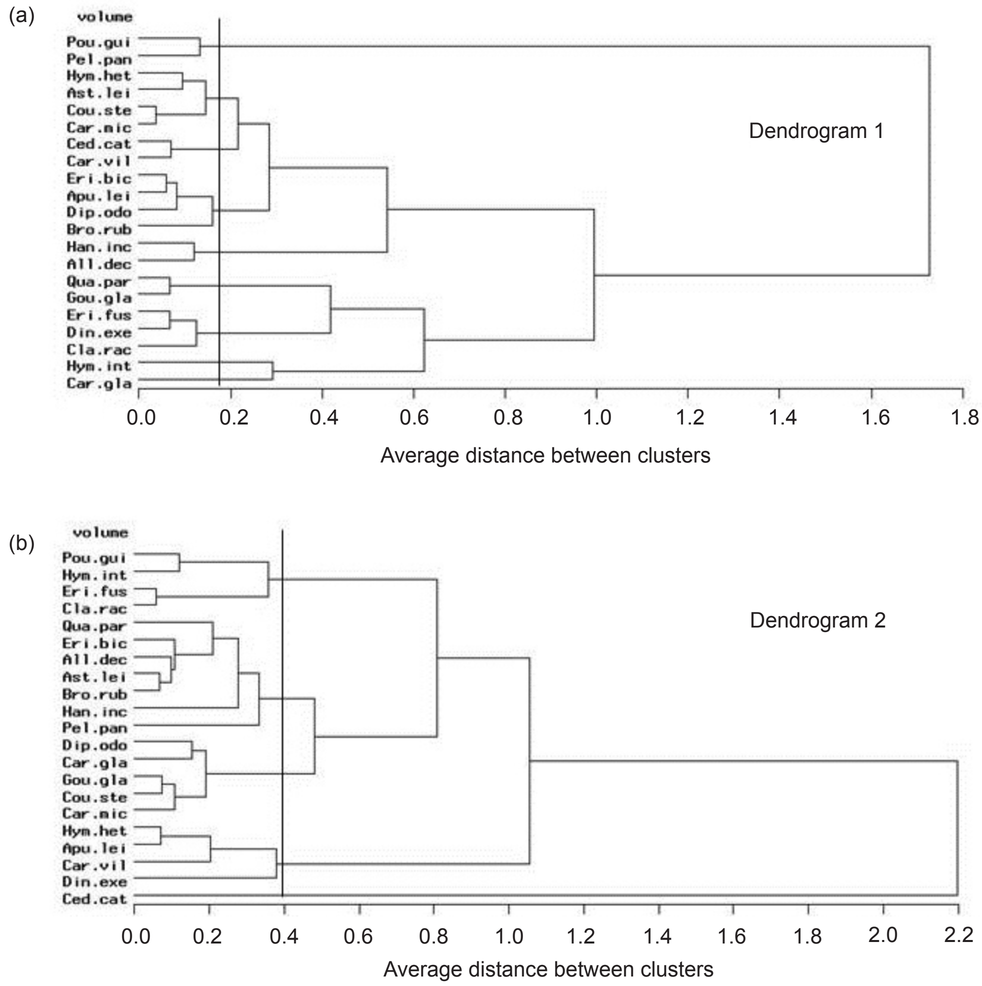

Figure 1 Dendrograms 1 and 2 of Amazonian tree species groups obtained using (a) regression coefficients of the Schumacher and Hall's model and (b) their combination with the fit statistics respectively

corroborated the decision to use equations generated from dendrogram 2.

DISCUSSION

High variability of tree measurements data is common in the study of tropical forests, since it is characterised by structural and flora diversity (Akindele & LeMay 2006, ter Steege et al. 2013) and the presence of individual trees of large size (Worbes & Junk 1999). Furthermore, the heterogeneity in tree composition and structure, even within a small area, the abundance of species and the variability of age are challenges in estimation of tree volume in natural forests. The high variability makes it difficult to use average form factor and equations for tree species (Figueiredo-Filho 1983, Akindele & LeMay 2006).

Variation in the number of sample trees per species (Table 2) reflected typical structural features of multi-aged and multi-species forests due to the occurrence of low frequency and locally rare species groups (Condit et al. 2000) as well as widely distributed common species (ter Steege et al. 2013). This variation allowed the formation of two dendrograms with respect to sample density (Figure 1 and Table 3).

Grouping method in this study was based purely on statistical procedures, ensuring the absence of subjectivity in taxonomic and ecophysiological categorisations (Akindele & LeMay 2006). This analysis is appropriate and effective for grouping tropical species of commercial value. If the cophenetic correlation coefficient is closer to 1, distortion in the groups will be lower due to good representation of dissimilarity matrices in the form of dendrograms

Table 3 Amazonian tree species groups formed by cluster and discriminant analyses

Dendrogram 1			
Group 1	Group 3	Group 5	Group 7
Pouteria guianensis	*Cedrelinga cateniformis*	*Handroanthus incanus*	*Erisma fuscum*
Peltogyne paniculata	*Caryocar villosum*	*Allantoma decandra*	*Dinizia excelsa*
Peltogyne venosa		*Mezilaurus synandra*	*Clarisia racemosa*
Diplotropsis rodriguesii		*Dipteryx alata*	
		Cordia goeldiana	
Group 2	Group 4	Group 6	Group 8
Hymenolobium heterocarpum	*Erisma bicolor*	*Qualea paraensis*	*Hymenaea intermedia*
Astronium lecointei	*Apuleia leiocarpa*	*Goupia glabra*	*Caryocar glabrum*
Couratari stellata	*Dipteryx odorata*	*Manilkara elata*	*Bagassa guianensis*
Cariniana micrantha	*Brosimum rubescens*		*Bowdichia nitida*
Simarouba amara			*Handroanthus impetiginosus*
Cedrela fissilis			
Dendrogram 2			
Group 1	Group 2	Group 3	Group 4
Pouteria guianensis	*Qualea paraensis*	*Dipteryx odorata*	*Hymenolobium heterocarpum*
Hymenaea intermedia	*Erisma bicolor*	*Caryocar glabrum*	*Apuleia leiocarpa*
Erisma fuscum	*Allantoma decandra*	*Goupia glabra*	*Caryocar villosum*
Clarisia racemosa	*Astronium lecointei*	*Couratari stellata*	*Dinizia excelsea*
Bagassa guianensis	*Brosimum rubescens*	*Cariniana micrantha*	*Cedrelinga cateniformis*
Bowdichia nitida	*Handroanthus incanus*	*Cedrela fissilis*	*Handroanthus impetiginosus*
Cordia goeldiana	*Peltogyne paniculata*		
Diplotropsis rodriguesii	*Dipteryx alata*		
Manilkara elata	*Mezilaurus synandra*		
Peltogyne venosa	*Simarouba amara*		

Table 4 Regression coefficients and fit statistics of Amazonian tree species groups formed by cluster and discriminant analyses

Group	n	S	β_0	p-value	β_1	p-value	β_2	p-value	Syx%	r²
Dendrogram 1										
1	698	4	-5.173	< 0.001	1.065	< 0.001	0.735	< 0.001	24.95	0.446
2	1459	6	-8.356	0	1.816	0	0.778	< 0.001	30.47	0.719
3	106	2	-7.995	< 0.001	1.792	< 0.001	0.687	< 0.001	41.11	0.450
4	561	4	-7.844	< 0.001	1.669	< 0.001	0.807	< 0.001	28.98	0.716
5	161	5	-8.549	< 0.001	1.999	< 0.001	0.542	< 0.001	23.10	0.816
6	374	3	-7.335	< 0.001	1.660	< 0.001	0.634	< 0.001	25.36	0.680
7	775	3	-8.990	< 0.001	1.934	< 0.001	0.854	< 0.001	42.99	0.629
8	231	5	-6.619	< 0.001	1.472	< 0.001	0.701	< 0.001	31.08	0.629
Dendrogram 2										
1	523	10	-6.595	< 0.001	1.412	< 0.001	0.745	< 0.001	21.31	0.725
2	1823	10	-7.577	0	1.573	0	0.838	< 0.001	25.99	0.755
3	1055	6	-8.061	< 0.001	1.700	< 0.001	0.853	< 0.001	31.42	0.776
4	964	6	-7.809	< 0.001	1.726	< 0.001	0.773	< 0.001	37.97	0.552

n = number of trees; S = number of species; $\beta_{0,}$ β_1 and β_2 = regression coefficients; Syx% = standard error of the estimate; and r^2 = coefficient of determination

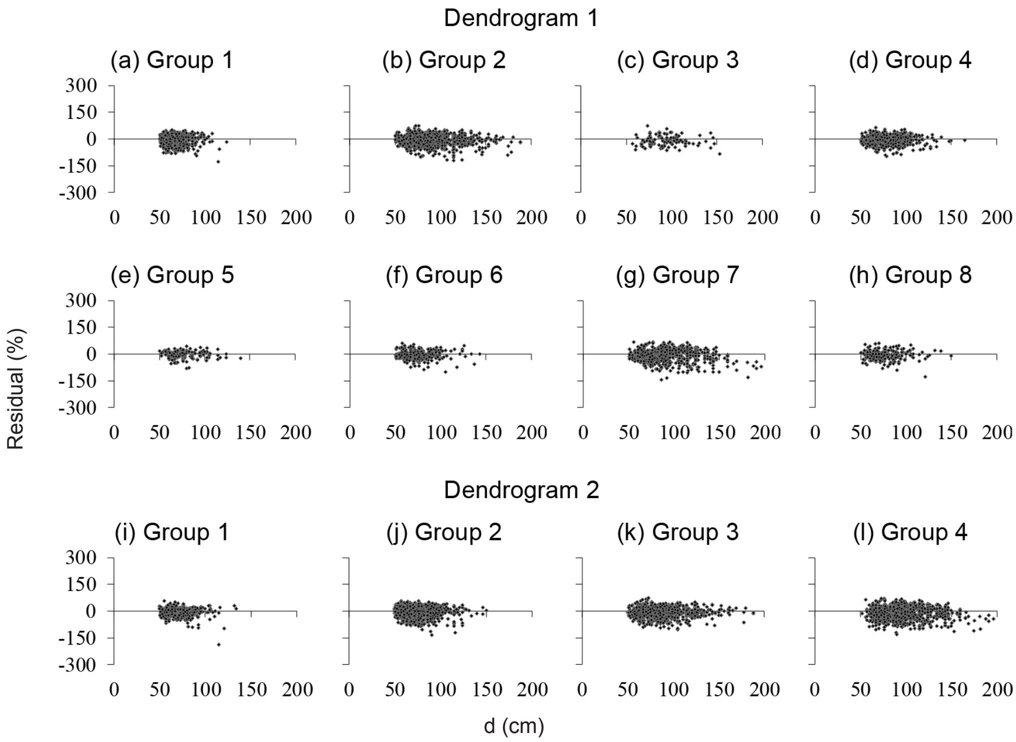

Figure 2 Dispersion of the residuals obtained by Schumacher and Hall's model fitted for Amazonian tree species groups; d = diameter at 1.3 m (cm)

(Albuquerque et al. 2005). Dendrogram 1 shows aggregation of the two species of the genus *Peltogyne* in the same group, indicating possible influence of taxonomic classification (Figure 1 and Table 3), although this contradicts findings of Akindele and LeMay (2006).

The estimation errors in larger trees were typical of heterogeneous natural forests (Brandeis et al. 2006). The Schumacher and Hall's model is often cited as one of the most appropriate to estimate volume of trees in tropical forests (Akindele & LeMay 2006, Igbinosa & Amoo 2014). Thus, the classification approach based on the regression coefficients would seem appropriate, since it effectively reflects tree taper. The selection of dendrogram 2 as the most suitable was refuted by cophenetic correlation coefficient value and Wilks's test, since both indicated dendrogram 1 as the most consistent and with better discrimination of groups. Thus, evaluation statistics of the multivariate analyses failed to ensure optimal grouping for fitting stem volume models according to groups. Results of multivariate methods should therefore be carefully interpreted because, regardless of the selection criteria, there is no guarantee that the result is the best for a particular purpose (Johnson & Wichern 1992).

For decisive regression model selection, graphical analysis of the residuals (Draper & Smith 1998) seemed to support the decision to use equations generated for the groups within dendrogram 2. Groups 1, 2 and 3 showed residuals distributed homogeneously throughout the regression line. However, the fitted model for group 4 overestimated residuals due to the influence of variability in large trees. In dendrogram 1, this behaviour was evident in groups 3, 7 and 8.

The lack of data for consistent model generation of some tropical species is mainly due to the presence of many rare species (ter Steege et al. 2013). Species grouping in tropical forests is an advantage to estimate commercial volume thus causing reduction in the number of equations to a manageable amount, facilitating processing and data analysis (Vanclay 1991).

CONCLUSIONS

The grouping of Amazon rainforest commercial species, based on the regression coefficients and fit statistics, performed better in aggregation for stem volume modelling, providing stabilisation of the standard error of estimate and supplying smaller number of equations for the evaluation of standing stock. The two stages multivariate approach with cluster and discriminant analyses based on the regression coefficients was appropriate for the composition of commercial species groups in the Amazon rainforest. Besides reducing the number of volume equations for individual species, this method minimises the problem of low-density data of certain species while forming consistent groups for the regression analysis.

ACKNOWLEDGEMENTS

We thank the National Council for Scientific and Technological Development for the research scholarship received and to Amata Company for providing the dataset.

REFERENCES

Akindele SO & LeMay VM. 2006. Development of tree volume equations for common timber species in the tropical rain forest area of Nigeria. *Forest Ecology and Management* 226: 41–48.

Albuquerque MA, Ferreira RLC, Silva JAA, Santos ES, Stosic B & Souza AL. 2005. Estabilidade em análise de agrupamento: estudo de caso em ciência florestal. *Revista Árvore* 30: 257–265.

Brandeis TJ, Delaney M, Parresol BR & Royer L. 2006. Development of equations for predicting Puerto Rican subtropical dry forest biomass and volume. *Forest Ecology and Management* 233: 133–142.

Chuman T & Romportl D. 2010. Multivariate classification analysis of cultural landscapes: an example from the Czech Republic. *Landscape and Urban Planning* 98: 200–209.

Clutter J, Fortson JC, Pienaar LV, Brister GH & Bailey RL. 1983. *Timber Management: A Quantitative Approach*. John Wiley and Sons, New York.

Condit R, Ashton P, Baker P et al. 2000. Spatial patterns in the distribution of tropical tree species. *Science* 288: 1414–1418.

Draper NR & Smith H. 1998. *Applied regression analisys*. Third edition. John Wiley and Sons, New York.

Figueiredo-Filho A. 1983. *Estudos de modelos matemáticos para estimar o volume por unidade de área em uma floresta tropical úmida na Amazônia brasileira*. MSc thesis, Universidade Federal do Paraná, Curitiba.

Grubbs FE. 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11: 13–14.

Gutierrez-Velez VH & Macdicken K. 2008. Quantifying the direct social and governmental costs of illegal logging in the Bolivian, Brazilian, and Peruvian Amazon. *Forest Policy and Economics* 10: 248–256.

Igbinosa AH & Amoo OB. 2014. Appropriate volume functions for Leguminosae family in two tropical rainforests in Cross River State, Nigeria. *Journal of Environment and Ecology* 5: 206–221.

Johnson RA & Wichern DW. 1992. *Applied Multivariate Statistical Analysis*, Third edition. Prentice Hall, New Jersey.

Köhler P & Huth A. 1998. The effects of tree species grouping in tropical rainforest modelling: simulations with the individual based model FORMIND. *Ecological Modelling* 109: 301–321.

Oliveira-Filho AT & Fontes MAL. 2000. Patterns of floristic differentiation among Atlantic forests in Southeastern Brazil and the influence of climate. *Biotropica* 32: 793–810.

Phillips PD, Yasman I, Brash TE & Van Gardingen PR. 2002. Grouping tree species for analysis of forest data in Kalimantan (Indonesia Borneo) *Forest Ecology and Management* 157: 205–216.

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

Reis E. 1997. *Estatística Multivariada Aplicada*, Edições Silabo, Lisboa.

Rencher AC. 2002. *Methods of Multivariate Analysis*. Second edition. Wiley-Interscience, New York.

Rohlf FJ. 1970. Adaptative hierarchical clustering schemes. *Systematic Zoology* 19: 58–82.

Souza AL & Souza DR. 2006. Análise multivariada para estratificação volumétrica de uma floresta ombrófila densa de terra firme, Amazônia Oriental. *Revista* Árvore 30: 49–54.

Swaine MD & Whitmore TC. 1988. On the definition of ecological species groups in tropical rain forests. *Vegetation* 75: 81–86.

Ter Steege H, Pitman NCA, Sabatier D et al. 2013. Hyperdominance in the Amazonian tree flora. *Science* 342: 324–334.

Turner IM. 2001. *The Ecology of Trees in the Tropical Rainforest*. Cambridge University Press, Cambridge.

Vanclay JK. 1991. Aggregating tree species to develop diameter increment equations for tropical rainforests. *Forest Ecology and Management* 42: 143–168.

Worbes M & Junk WJ. 1999. How old are tropical trees: the persistence of a myth. *IAWA Journal* 20: 255–260.