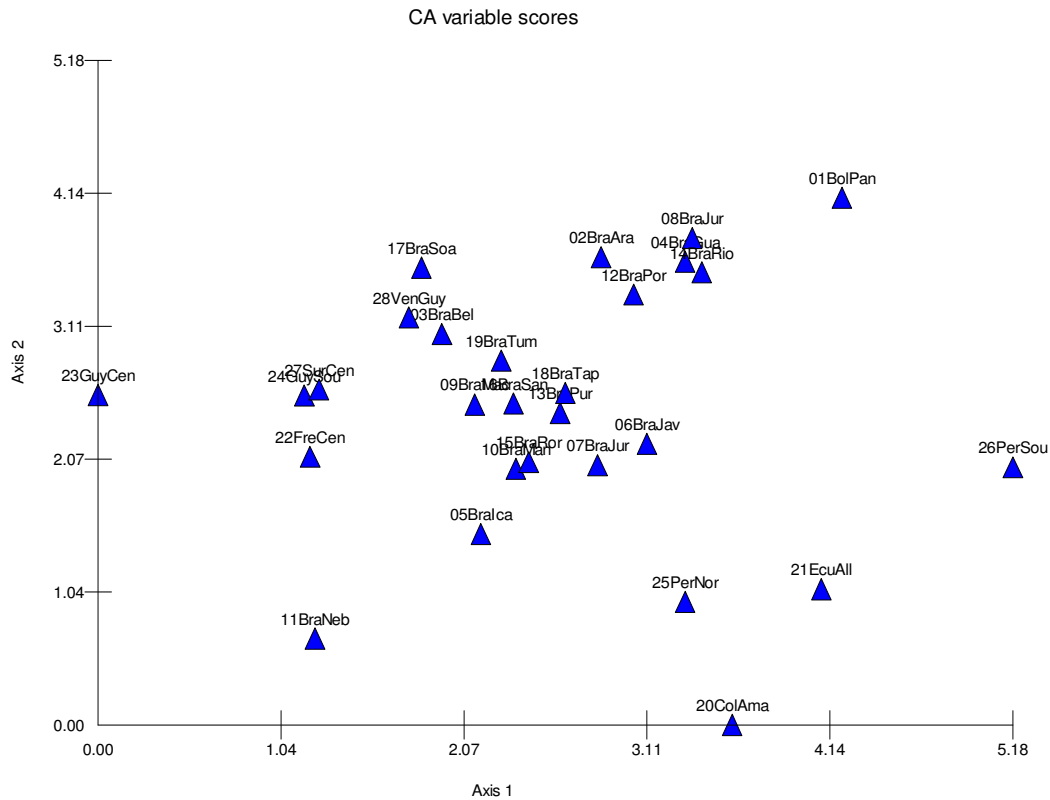


# Continental-scale patterns of canopy tree composition and function across Amazonia

Hans ter Steege et al 2006

## Supplementary information 1. Additional analyses.



**Supplementary Figure S1.** Biplot for the ordination with Detrended Correspondence Analysis (DCA) with 28 regions and 513 identified genera found in these regions.

The gradient, as described in the main text neatly orders the regions from east to west along the first axis of the DCA and from wet to dry along the second axis. The summary statistics for this analysis are given in supplementary table 1.

**Supplementary Table S1.** Summary statistics for DCA with 28 site and 513 genera.

	Axis 1	Axis 2	Axis 3
Eigenvalues	0.353	0.184	0.148
Percentage	15.829	8.265	6.624
Cum. Perc.	15.829	24.094	30.718

### *Testing the potential effect of misidentifications*

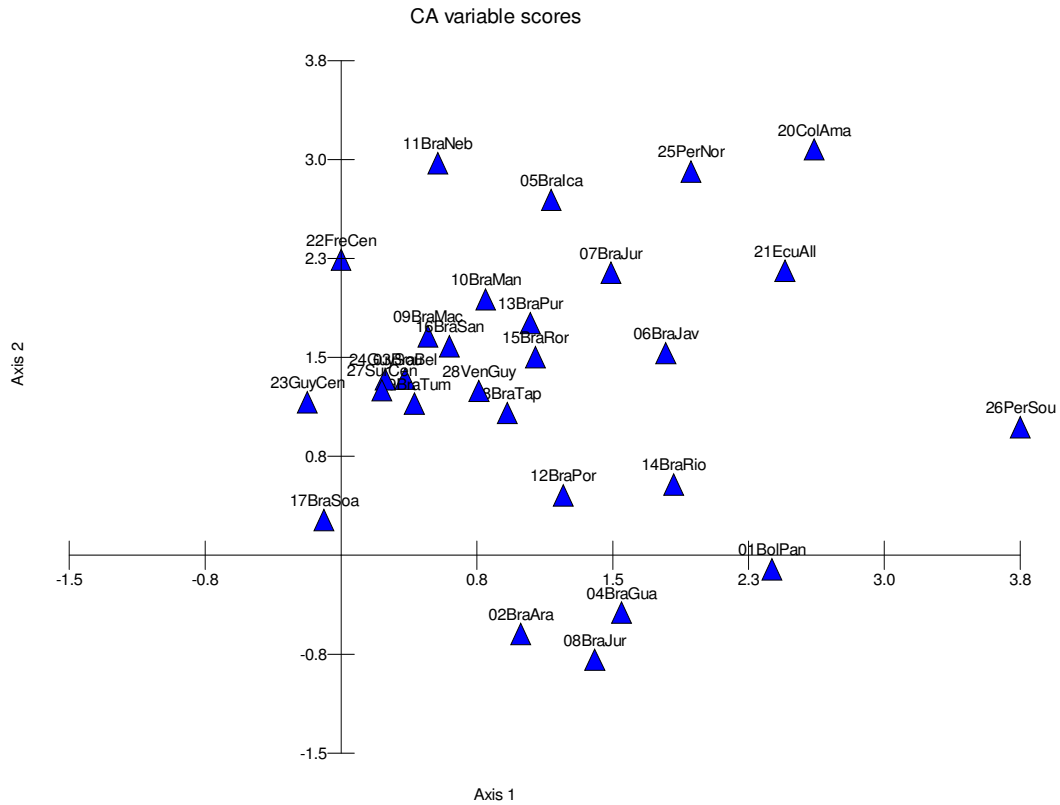
The quality of the patterns described in our paper obviously depends to a great extent on the quality of identification of the species (and genera) in the inventories. In most of the forest inventories, usually carried out by Forestry Commissions or another agency vested with the management of the resource base, identifications are routinely made by tree-spotters (or *mateiros*, as they are known in Brazil).

Our experience with tree spotters in the Guianas (Hans ter Steege pers. obs.) is very good. In Guyana and Suriname for our 1-ha plot studies we made use of several of the same tree-spotters who were also employed during the FAO Forest Inventories (see refs in supplementary information 2). In both countries the level of identification was high, even at the species level. At generic level, we estimate that more than 95% of all species are at least identified correctly at the genus level. Problems in identification are found mainly at the species level, mostly within a few very speciose genera (e.g. *Licania*, *Pouteria*). In Brazil, Fearnside<sup>20</sup> estimated, that during Radam over 80% of all individuals, of which wood density was also known, were at least identified correctly to the genus. We estimate that the number of all individuals of which the name can be relatively well coupled to a generic name now is well over 90% (see Supplementary Table 3). Since the rarest species are the most difficult to identify, the percentage of individuals correctly identified is higher than the percentage of species correctly identified.

To study the effect of individuals wrongly classified as to genus, one approach would be to lump genera, known to give problems in the identification and test the effect on the ordination results. At the generic level problems may be expected with particular genus pairs/groups such as: *Lecythis* – *Eschweilera*; *Chrysophyllum* – *Pouteria*; *Ocotea* – *Nectandra* – *Aniba*, and *Protium* – *Tetragastris*. We chose instead to take a more conservative approach, simply lumping all genera within families into a single taxon. This resulted in a gradient analysis at the family level. The ordination results are given in supplementary Fig. S2.

A visual inspection reveals that the DCA with families produces essentially the same bi-plot as the DCA with genera. The order of the regions along the axes is very similar. Comparing the site scores for the two ordinations (the coordinates in the bi-plots) shows that the two analyses are indeed very alike:  $R^2$  between site scores on axis 1 for the generic and family ordination is 0.85; that for axis 2: 0.83.

Hence, when individuals are grouped by family (as a conservative measure to avoid misidentifications at the genus-level) the same composition gradients are extracted. The explained variation is somewhat greater than for the ordination with family data due to the fact that most families are widespread, whereas many genera have localized ranges or less well-documented occurrence within our dataset.



**Supplementary Figure S2.** Biplot for Detrended Correspondence Analysis (DCA) with 28 regions and 89 families found in these regions.

**Supplementary Table S2.** Summary statistics for DCA with 28 regions and 513 genera.

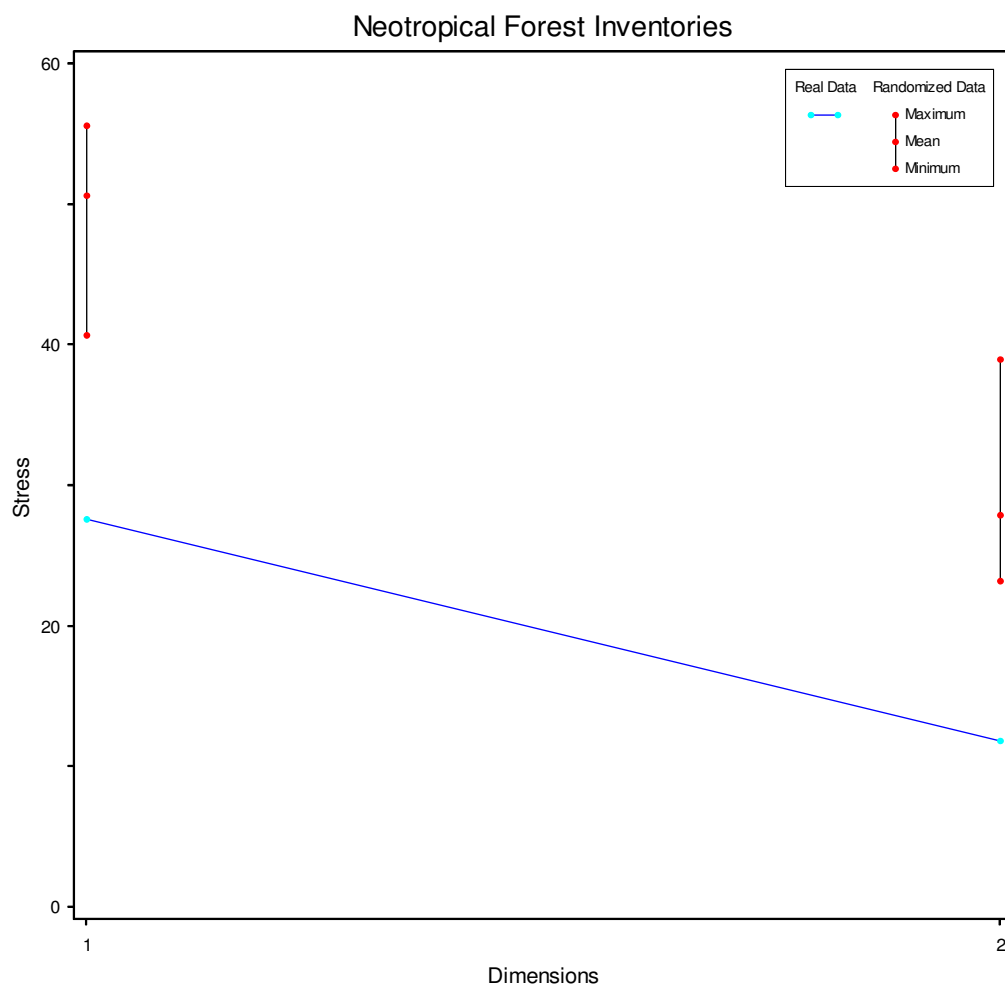
	Axis 1	Axis 2	Axis 3
Eigenvalues	0.172	0.054	0.034
Percentage	32.743	10.293	6.561
Cum. Perc.	32.743	43.036	49.597

### *Testing the effect of Ordination method*

Choice of ordination technique is frequently controversial among ecologists. In order to test the extent to which our correspondence analysis results were technique-dependent, we also carried out an analysis with Nonmetric Multidimensional Scaling (NMDS, PCORD, MjM Software, USA). We ran this analysis with 50 runs with the real data and 99 runs with randomized data to test whether our real data performed better than randomized data.

Two axes reduced the stress to 12.4% (Supplementary Fig. 3). A third axis only added 4% and was rejected.

The pattern extracted from the real data performed significantly better than the randomized data (Supplementary Fig. 3) and we concluded that the NMDS extracted meaningful results.



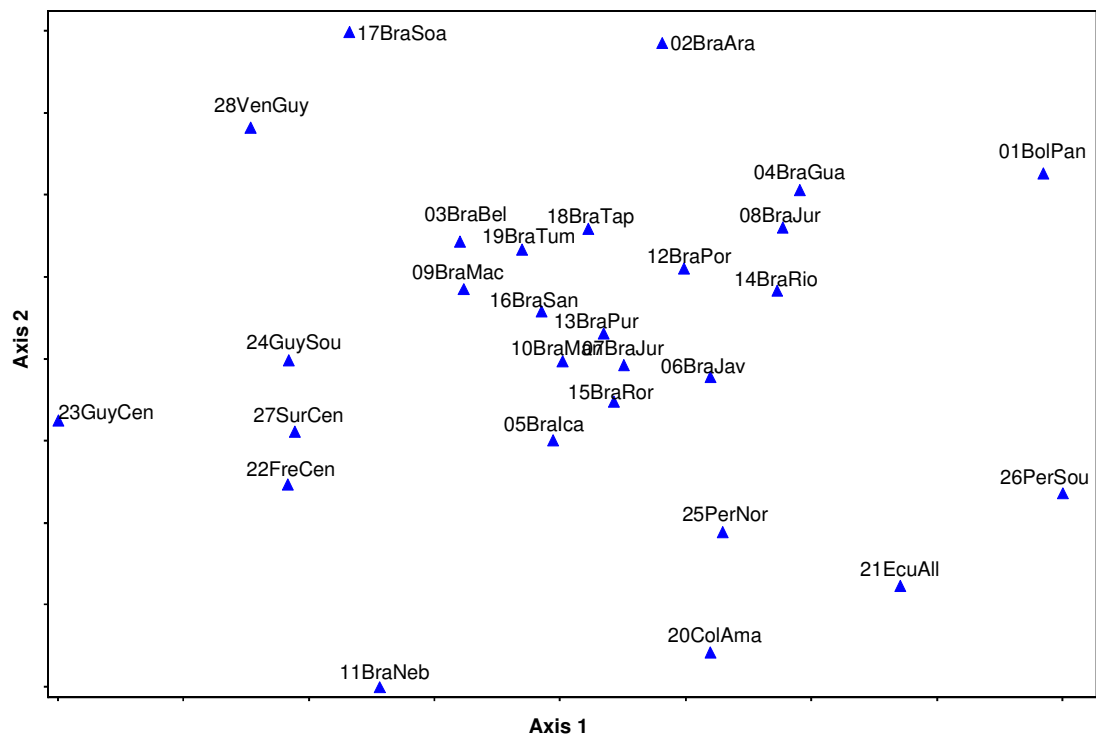
**Supplementary Figure S3.** Scree plot for NMDS analysis with 28 regions and 513 genera found in these regions.

As NMDS by definition does not extract ecological gradients, but rather maximizes rank-order correlation between distance measures and distance in ordination space, a relationship

between the configuration of points and the ecological gradients has to be found by visual examination and rotation of the axes.

The ordination with NMDS produces very similar results to the DCA with genera (Supplementary Fig. S4). After a rotation of -64 degrees the correlation between the site scores for the first axes of the two ordinations is high ( $R^2 = 0.95$ ); for the second axis  $R^2 = 0.80$ .

Hence NMDS and DCA are in nearly identical in their output. The data are robust with regard to the analytical technique chosen.



**Supplementary Figure S4.** Bi-plot for NMDS analysis with 28 regions and 513 genera found in these regions. Ordination rotated -64 degrees for comparison with Fig. S1.