Estimation of the number of tree species in French Guiana by extrapolation of permanent plots richness

Eric Marcon, Ariane Mirabel, Jean-François Molino, Daniel Sabatier

Estimation of the number of tree species in French Guiana by extrapolation of permanent plots richness

Eric Marcon[1,4]*

Ariane Mirabel[2,4]

Jean-François Molino[3]

Daniel Sabatier[3]

Abstract

The biodiversity of tropical rainforest is difficult to assess. Yet, its estimation is necessary for conservation purposes, to evaluate our level of knowledge and the risks faced by the forest in relation to global change. Our contribution is to estimate the regional richness of tree species from local but widely spread inventories. We reviewed the methods available, which are nonparametric estimators based on abundance or incidence data, log-series extrapolation and the universal species-area relationship based on maximum entropy. Appropriate methods depend on the scale considered. Harte's self-similarity model is suitable at the regional scale, while the log-series extrapolation is not. GuyaDiv is a network of forest plots installed over the whole territory of French Guiana, where trees over 10 cm DBH are identified. We used its information (1315 species censused in 68 one-hectare plots) to estimate the exponent of the species-area relationship, assuming Arrhenius's power law. We could then extrapolate the number of species from three local, wide inventories (over 2.5 km^2). We evaluated the number of tree species around 2200 over the territory.

Biodiversity, Tropical Forest, Self-Similarity, French Guiana

[1]AgroParisTech, UMR AMAP, CIRAD, CNRS, INRAE, IRD, Univ Montpellier, Montpellier, France.

[2]Department of Biology, Western University, ON, Canada

[3]UMR AMAP, IRD, CIRAD, CNRS, INRAE, Univ Montpellier, Montpellier, France.

[4]UMR EcoFoG, AgroParisTech, CIRAD, CNRS, INRAE, Univ Antilles, Univ Guyane, Kourou, France.

*: eric.marcon@agroparistech.fr,

Contents

1. Introduction

Biodiversity assessment in tropical moist forests is a practical challenge but a major goal considering they are the most diverse terrestrial ecosystems. Estimating the number of tree species is made possible by the long-term effort of sampling resulting in thousands of forest plots organized in various networks (Forest-Plots.net et al., 2021) and a set of methods to apply to their data.

At the local scale, the number of species is related to the sampling effort by species-accumulation curves (Gotelli and Colwell, 2001). The number of sampled species is a matter of well-known statistics based on independent and identically distributed (iid) samples, and estimators of the total number of species of a homogeneous community are available, among which the best known are Chao's (Chao, 1984) and the jack-knife (Burnham and Overton, 1978). These estimators

can be applied to incidence data (i.e. the number of sampled plots that contain a given species) as well as abundance data (the number of sampled individuals of a given species). Yet, these tools fail to estimate regional diversity because increasing the sampled area implies including new, different communities, preventing iid sampling in practice.

Yet, Cazzolla Gatti et al. (2022) successfully applied the incidence-based Chao estimator to 100- by 100-km cells (each cell considered as a plot) covering all forests in the world to assess the number of tree species at the scale of continents. The method requires huge datasets to avoid undersampling and sampling biases.

At very large scales, the unified neutral theory of biodiversity and biogeography (Hubbell, 2001) implies that the distribution of the metacommunity's species abundances is in log-series (Fisher et al., 1943), allowing the extrapolation of the rank-abundance curve of sampled species up to the rarest one, represented by a single individual and counting the number of necessary species. Based on this method, the diversity of tree species has been estimated in Amazonia (ter Steege et al., 2013, 2020) and at the world scale (Slik et al., 2015).

Regional diversity, i.e. at intermediate scales between single communities and the metacommunity, brought less attention. The large and spatially uniform datasets necessary to apply incidence data extrapolation are not easy to gather so alternative methods must be considered: this motivated this study, along with a particular interest for the forest of French Guiana.

The main contribution of this paper is to estimate the number of tree species at the regional scale, in French Guiana (8 million hectares of tropical moist forest with no ecological boundary to distinguish them from the rest of Amazonia) and demonstrate which method is valid to do so. We build on Harte's self-similarity model (Harte et al., 1999) that implies the power-law relationship of Arrhenius (1921) and provides a technique to evaluate its parameters (Harte et al., 1999), previously applied by Krishnamani et al. (2004) in the Western Ghats, India, a 60,000-ha tropical forest with around 1,000 tree species. The current checklist contains close to 1800 tree species (Molino et al., 2022) in French Guiana. Our estimate is around 2200.

We also compare our work to all methods reviewed above and the lesser-known, scale-independent universal species-area relationship based on maximum entropy (Harte et al., 2009). We discuss in depth which method may be applied according to the addressed spatial scale.

2. Methods

2.1 Data
To apply the methods detailed below, a large enough inventory is necessary along with a set of small, widely
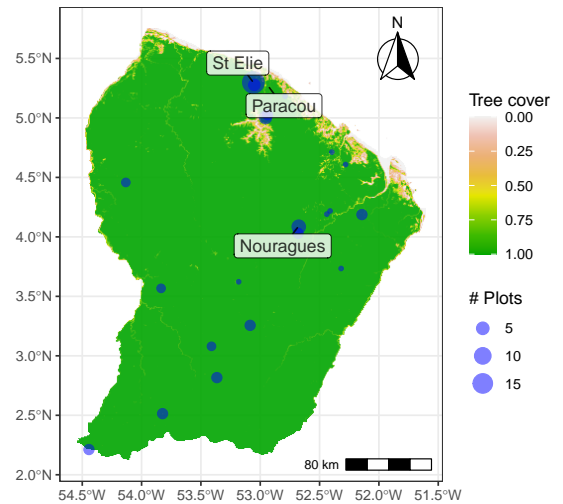

Figure 1. The GuyaDiv network of tropical forest plots, 1-ha plots only. All trees above 10 cm DBH are localized, measured and determined botanically at species level. Each circle represents a location where several forest plots were established: the circle size is proportional to the number of plots. Paracou, Piste de Saint-Elie and Nouragues large inventories locations are shown (Paracou contains no 1-ha GuyaDiv plot). The map background represents the tree cover, as of ESA Land Cover (Zanaga et al., 2021).

spread forest plots. We gathered 3 local, large inventories to account for environmental variability and a network of plot covering the whole region.

Our plot network is GuyaDiv (Engel, 2015). Since the installation of the first plots in 1986, the GuyaDiv network has continuously grown until today. It now consists of 243 plots of various sizes and shapes, distributed in various forest types, in 30 sites across French Guiana. We took into account the 68 one-hectare plots of the network (figure 1). They are located in 21 sites, which provides fairly good coverage of the variability of the forest. They contain 43081 trees among which 415 were removed from the analyses because they could not be assigned to a species or morphospecies.

The Paracou research station (Gourlet-Fleury et al., 2004) contains six 6.25-ha and one 25-ha plots of primary rainforest. Nine 6.25-ha plots were logged between 1986 and 1988 in a forestry experiment that temporarily increased the recruitment of light-demanding species (Mirabel et al., 2021) and the functional diversity (Mirabel et al., 2020).

In a rather conservative approach, we retained only the well-identified trees of the permanent plots (571 species) and added available data from the GuyaDiv network: transects from Molino and Sabatier (2001) and ten 0.49-ha plots around the Guyaflux tower (Bonal et al., 2008) contain 575 species, including 132 new ones. 37 more species at the French Guiana IRD Herbarium (CAY: Gonzalez et al., 2022) were

collected in the area but outside the plots. The total number of species is thus 740 included in a 4.84-km^2 convex envelope.

The Piste de Saint-Elie site has been intensively sampled for 50 years. It encompasses nineteen 1-ha and one half-hectare plots in GuyaDiv and a few small plots added for various studies. Moreover, many herbarium specimen were collected from the site. As a whole, we gathered 763 species in a 3-km^2 area.

Nouragues research station (Bongers et al., 2001) provides 22 hectares of permanent plots. We applied the same protocol, adding 11 Guyadiv plots and herbarium collections up to 850 species in a 2.5-km^2 area.

2.2 Self-similarity

Self-similarity (Harte et al., 1999) is a property based on scale invariance. Consider a species that is present in an area A_0, say French Guiana. The probability to find it in half the whole area, denoted A_1 is a. Then, if it is present in A_1, the probability to find it in turn in half A_1, denoted A_2, is also a, and so on. The probability to find the species in A_n is thus a^n. In other words, the conditional probability to find a species in a sub-area, given that it is present in the area containing it, is constant: it does not depend either on the observation scale nor on the species considered.

The Arrhenius power law (Arrhenius, 1921) both implies and is a consequence of the self-similarity property (Harte et al., 1999). The number of species $S(A)$ observed in an area A is

$$S(A) = cA^z \tag{1}$$

where z is the power parameter and c is the number of species in an area of size 1. Actually, $a = 2^{-z}$. This is a classical relation in macroecology, with long empirical and theoretical support (Williamson et al., 2001; Gárcia Martín and Goldenfeld, 2006).

If z is known, the inventory of a reasonably large area b allows computing $c = S(b)/b^z$. Then, $S(A)$ can be calculated for any value of A.

Harte et al. (1999) showed that under the assumption of self-similarity, z can be inferred from the dissimilarity between small and distant plots of equal size distributed across the area. The Sørensen (1948) similarity between two plots is

$$\chi = 2(S_1 \cap S_2)/(S_1 + S_2) \tag{2}$$

where S_1 (respectively S_2) is the number of species in plot 1 (resp. plot 2) and $S_1 \cap S_2$ is the number of common species.

Applied to plots of the same size separated by distance d, Sørensen's similarity decreases with distance following the relation $\chi \sim d^{-2z}$ (Harte et al., 1999) that can be estimated by the linear model

$$\log(\chi) \sim \log(d). \tag{3}$$

The logarithm of the Sørensen dissimilarity between pairs of plots can be regressed against the logarithm of the distance between the plots: the slope of the regression is $-2z$.

The relation (3) holds at the same scale as the power law, i.e. at the regional scale (Grilli et al., 2012). Krishnamani et al. (2004) estimated $z \approx 0.12$ with a very good fit to the linear model at distances up from 1 km but not below.

The number of plots varies across locations so the estimation of z must be made with care. We sampled one random plot at each location to obtain $21 \times 20/2 = 210$ pairs of plots. We calculated the Sørensen dissimilarity χ and the geographic distance d between each pair of plots. We estimated z as half the coefficient of the distance variable in the linear model $\log(\chi) \sim \log(d)$. We repeated these steps 1000 times to obtain a distribution of estimated z values depending on the plots drawn in each location. z was estimated as the empirical mean of the distribution and its 95% confidence interval was obtained by eliminating the 2.5% extreme values on both tails.

The confidence interval of the estimation of the number of species was assessed by combining the uncertainty in c and A^z. The variance of c was estimated by the empirical variance of the values calculated at Paracou, Piste de Saint-Elie and Nouragues. That of A^z was obtained from the empirical distribution of z. The variance of their product was calculated (the formula and its derivation are in the appendix). Finally, we assumed the normality of the distribution of the product of the estimates to retain an approximate 95% confidence interval of $\pm$ 2 standard deviations.

All analyses were made with R (R Core Team, 2024) v. 4.3.1.

2.3 Nonparametric estimators

At smaller scales, i.e. inside a single community, the relation between area and number of species is described by species accumulation curves (SAC: Gotelli and Colwell, 2001). It is driven by statistical models that address incomplete sampling (Shen et al., 2003; Béguinot, 2015). After replacing the sampled area by the number of individuals it contains, well-known estimators of richness such as Chao's (Chao, 1984) or the jackknife (Burnham and Overton, 1978) apply.

The Chao1 estimator is

$$\hat{S}_{Chao1} = s_{obs} + \frac{(n-1)f_1^2}{2nf_2}, \tag{4}$$

where s_{obs} is the number of observed species, n is the sample size, f_1 and f_2 the number of species observed once and twice. Since n is large, $(n-1)/2n$ can be approximated by $1/2$.

The jackknife estimator depends on the sampling level of the data. The estimator of order k includes $f_1, f_2, \ldots, f_k$, the number of species observed up to k times. Increasing the order implies increasing both

the estimate and its uncertainty: starting from order 1, the order is incremented as long as the new estimator is significantly higher than the previous one (Burnham and Overton, 1978). For large n, the jackknife estimator of order 3, used below, is

$$\hat{S}_{Jack3} = s_{obs} + 3f_1 - 3f_2 + f_3. \tag{5}$$

An alternative, following Cazzolla Gatti et al. (2022), consists of paving the territory with a grid whose size does not change the estimation, say 100 km. In each 100 by 100 km cell of the grid, all available data is aggregated to obtain an incidence dataset. The Chao2 estimator (whose formula is identical to that of Chao1, with n equal to the number of grid cells) is finally applied: it combines the number of species observed in only one or two cells to estimate the number of unobserved species.

The Chao and Jackknife estimators variance can be estimated and a confidence interval is available (Chao, 1987; Burnham and Overton, 1978).

2.4 Log-series extrapolation
Assuming that the plots are samples of a metacommunity that follows a log-series distribution, the rank-abundance curve can be extrapolated following ter Steege et al. (2013).

First, the total number of trees is estimated by extrapolation of the average number of trees per 1-ha plot of the Guyadiv network to the 8 million hectares of the French Guiana forest.

The probability for one of these trees to belong to a given species is obtained by averaging the frequency of the species among plots. Each plot is a sample of a local community whose composition is not completely known: many rare species are not in the sample. The observed frequency of a species in a plot is not the probability of the species in the community: frequencies sum up to 1 while the sum of the actual probabilities of observed species, called the sample coverage (Good, 1953), sums up to 1 minus that of the unobserved species. The actual probabilities of observed species can be estimated following Chao and Jost (2015), with the *entropart* package (Marcon and Hérault, 2015).

The number of trees per species is then obtained by multiplying the total number of trees by the probability of each species. A rank-abundance curve is produced. Its center part is a straight line (see figure 3) that can be extrapolated down to the last species, represented by a single tree. The number of species is finally counted.

Its confidence interval is not available: the extrapolation of the curve is very robust, but the estimation of the total number of trees and of the probabilities of species are sources of uncertainty.

2.5 Universal species-area relationship
Harte et al. (2008) derived a universal species-area relationship based on the maximum entropy theory.
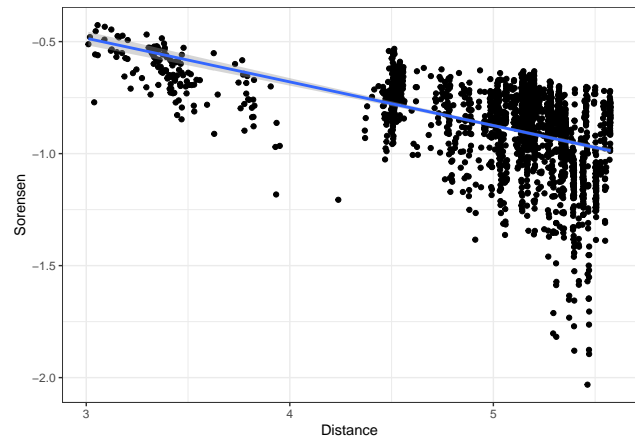

Figure 2. Relation between Sørensen's similarity and the distance between pairs of plots. Both axes are in base-10 logarithms, distances are in meters. Each point is a pair of plots more than 1000 m (10^3) apart, up to 377 km. A linear model is fitted: the slope of the regression is $-2z$.

Assuming only that the area, the total numbers of species and individuals, and the summed metabolic energy rate of all individuals are fixed, many features of the species distribution at any scale can be predicted. Of particular interest is the possibility to derive the number of species in a doubled area from the number of species in a sampled, reference area (Harte et al., 2009; Xu et al., 2012). Starting from a local sample, that may be a single 1-ha plot or one of our large inventories, the area can be doubled until the target size is reached.

The number of trees per hectare is estimated from the Guyadiv network to obtain a single starting point rather than a different one for each plot. To be consistent with the model, the geometric mean is applied: its logarithm equals the average logarithm of the number of trees in all 1-ha plots.

Each step of the estimation consists of doubling the area and calculating the new number of species. This operation is repeated until the target area (8 Mha) is reached, i.e. between 15 times for Paracou (the largest inventory: 484 ha) and 24 times for the 1-ha plots.

3. Results

3.1 Self-similarity
The relation between Sørensen's similarity and distance is presented in figure 2. All pairs of plots more than 1 km apart (the scale of Paracou's 0.625-km^2 inventory) are shown and the regression line of the figure illustrates the relation. Actually, the estimation of z was made as explained in the methods by 1000 random draws of sets of a single random plot per location.

The estimated value of z is 0.104 with a 95% confidence interval between 0.088 and 0.120.

The estimated number of species per square kilometer, c, is respectively 629, 681 and 773 in Paracou, Piste de Saint-Elie and Nouragues. The average value
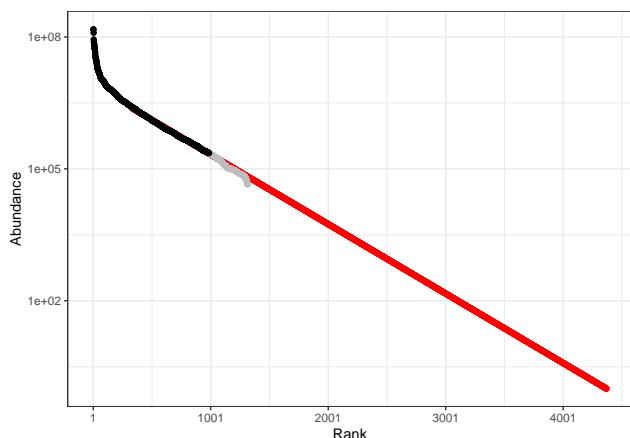
Figure 3. Extrapolation of the rank-abundance curve built from the GuyaDiv plots. Extrapolated abundances (in log scale) of observed species are plotted against the rank of their species. The abundances of unobserved species (the red curve) is extrapolated linearly from the center 50% of the distribution of the observed species. The rarest 25%, ignored for the extrapolation, are plotted as grey points.

is 694.

Finally, the estimated number of species is 2234. Taking into account the uncertainty about c and z, its 95% confidence interval is between 1587 and 2882.

3.2 Species accumulation

The observed number of species is 1314 among which 204 and 119 are sampled once and twice. The lower-bound estimation of the number of species by the Chao1 estimator is 1489. The best jackknife estimator (of order 3) is 1677. Its confidence interval is between 1563 and 1791 at the 5% risk level.

The Chao2 estimator applied to the same plots aggregated into 100 x 100 km cells is 1643. Its confidence interval is between 1436 and 1564 at the 5% risk level.

3.3 Log-series extrapolation

The mean number of trees per ha in the Guyadiv 1-ha plots is 627. There are close to 5 billion trees in French Guiana.

Figure 3 is the rank-abundance curve of the species. The most abundant tree species is *Eperua falcata* with around 151 million trees. The log-abundances of the 25 to 75 percentiles of species are linearly related to the rank, allowing the extrapolation of the curve (the red line).

The estimated number of species according to this model is 4368.

3.4 Universal species-area relationship

The method from Harte et al. (2009) is applied to our data. The geometric mean number of trees per hectare estimated from the Guyadiv network is 602 trees/ha.

Initial inventories, e.g. 740 trees species in 484 ha in Paracou and the geometric mean number of species
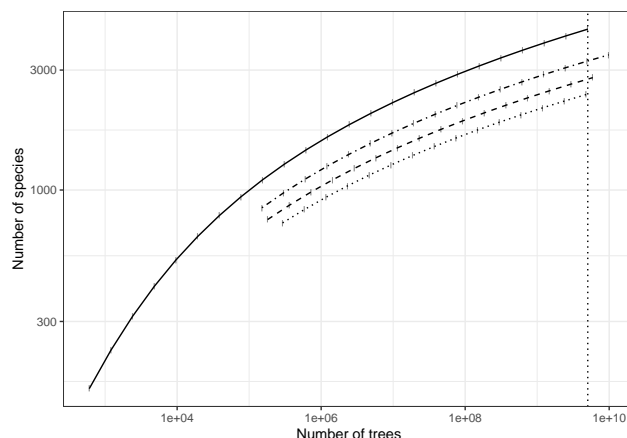

Figure 4. Extrapolation of the initial inventories up to 8 Mha (vertical line, assuming 602 trees/ha). The vertical line corresponds to the area of French Guiana. Curves are the estimated species-area curves from the Guyadiv 1-ha plots (solid line), Nouragues (dot-dashed line), Piste de Saint-Elie (dashed-line) and Paracou (dotted line) starting points. Estimated values are marked as vertical lines on the curves.

in Guyadiv plots, are the starting points of the estimation. Figure 4 shows the species-area curves obtained by successive doubling of the areas.

The curves are almost perfectly fitted by a Michaelis-Menten model, estimated by the linear model (Lineweaver and Burk, 1934) $\frac{1}{\log S} \sim \frac{1}{\log n}$, where S is the number of species and n the number of trees, allowing a very accurate interpolation at any number of trees. The estimated number of species is thus obtained for n equal to 8 Mha times 602 trees per ha:

- From Nouragues: 3238 species.
- From Piste de Saint-Elie: 2739 species.
- From Paracou: 2385 species.

The extrapolation from the average 1-ha plot is 4737 species. Since it is far less reliable than those from the wide inventories, with 7 to 9 more doubling steps, we do not retain it to produce the average estimate of the universal SAR. Finally, we obtain 2787 species.

3.5 Summary

The estimated number of species according to the different methods is summarized in table 1.

4. Discussion

4.1 The species-area relationship varies across scales

We consider three different spatial scales where different models apply.

At the local scale, i.e. inside a single community, the relation between the area and the number of species is described by species accumulation curves (SAC: Gotelli and Colwell, 2001). It is driven by statistical models that address incomplete sampling (Shen et al., 2003; Béguinot, 2015). Local SACs have

Table 1. Estimated number of tree species in French Guiana, according to all methods detailed in the text. Self-similarity is the appropriate method at this scale.

Method	Number of species	Lower bound	Upper bound
Self-similarity	2234	1587	2882
Species accumulation (abundance)	1677	1563	1791
Species accumulation (incidence)	1643	1436	1564
Log-series extrapolation	4368		
Universal species-area relationship	2787	2385	4737

been extensively studied and are out of the scope of this paper, but a few results are important here. The distributions of local, tropical moist forest communities are often approximately log-normal. This has been shown empirically (e.g. Duque et al., 2017) and theoretically (Preston, 1948, 1962; May, 1975). In the framework of the neutral theory (Hubbell, 2001), the local community follows a zero-sum multinomial distribution, derived by Volkov et al. (2003) but challenged empirically by McGill (2003), in favor of the log-normal distribution.

The SAC, plotted as the number of species against the number of individuals in natural scale, is concave downwards since its slope is the probability for the next individual to belong to a new species (Chao et al., 2013; Grabchak et al., 2017), which decreases with the sampling effort. This means that the Arrhenius power law does not apply at the local scale. The power law can be estimated empirically (Condit et al., 1996; Plotkin et al., 2000) but then the value of z depends on the distance, which is in contradiction with the model, which relies on a constant z.

At the regional scale, the mixture of local communities makes a new pattern emerge, namely the power law of Arrhenius (1921). Its origin is empirical, with a lot of support (e.g. Dengler, 2009; Triantis et al., 2012; Williamson et al., 2001). Theoretically, Hubbell (2001) showed that the power law applied to intermediate scales of the neutral theory and Grilli et al. (2012) derived it from a spatially-explicit model only based on the clustering of species. Preston (1962) showed that local, log-normal communities imply the power law at the regional scale. At this scale, the species-area relationship (SAR) properly speaking is not just a matter of accumulation due to sampling (SAC) but the consequence of the inclusion of different communities.

A long empirical controversy (Connor and McCoy, 1979) opposed Arrhenius and Gleason (1922), who argued that the number of species predicted by the power law was far too high and proposed $S(A) = z \ln A + c_g$ rather than the equivalent of eq. (1), i.e. $\ln S(A) = z \ln A + c_a$ (where c_g, c_a and c_f below are constants). Actually, Gleason's model is equivalent to Fisher's, where $S(A) = \alpha \ln A + c_f$ if the number of trees is large

and is proportional to the area (Engen, 1977). There is no theoretical support to apply Gleason's model at the regional scale (Gárcia Martín and Goldenfeld, 2006); in other words, the regional distribution of species is not log-series.

The widest scale is that of the metacommunity, in the sense of the neutral theory. Its follows a log-series distribution (Hubbell, 2001) with Fisher's α equal to θ, known as the fundamental biodiversity number. The log-series does not apply to the local or regional scale: the empirical estimates of Fisher's α at these scales increase with the sampling size (e.g. Condit et al., 1996), which is again in contradiction with the model, which implies that α is a constant. At the regional scale, Fisher's α increases with area because Arrhenius's law, and not Gleason's law, is valid. Yet, our data fit a log-series distribution quite well (figure 3): empirical tests are not efficient to reject the model at the regional scale (Connor and McCoy, 1979). We must rely on theory.

The limits between scales are obviously not sharp. Krishnamani et al. (2004) found that z stabilized when plots more than 1 km apart were considered. We followed them here. Increasing the regional scale makes it converge to the metacommunity. In absence of dispersal limitation, i.e. with migration parameter equal to 1 in the neutral theory, any regional sample would represent the metacommunity and follow a log-series distribution. So the wider the sampled area, the less distinguishable from the metacommunity the data will be, but at the scale of French Guiana, roughly 1% of Amazonia, many less species are present than in a sample of the same size taken across the whole metacommunity, even if we ignore environmental filtering.

4.2 The self-similarity model can be applied at the regional scale

The power law is equivalent to self-similarity (Harte et al., 1999), justifying our preferred method to estimate the richness of the French Guiana forest.

The self-similarity model allows estimating the number of species of tropical forests at a regional scale. It requires a network of plots at a wide range of distances from each other to estimate Arrhenius's power law parameter. It should be completed by a continuous inventory whose size is consistent with

the smallest scale of the power law. These constraints explain why the method has not been widely applied, beyond Krishnamani et al. (2004).

As shown in figure 2, the fit of the linear model is not perfect. The theory does not address habitat variation, that is well-described in French Guiana (Guitet et al., 2015). The dissimilarity between plot pairs is thus explained by distance and habitat dissimilarity, with the latter ignored in the model. Yet, the estimation of z is quite robust because the GuyaDiv network covers a wide range of habitats, allowing to cancel out local variability. Adding more plots or describing a few more species in the previous plots may not change z significantly since it is obtained from the dissimilarity between plots. Its value 0.104 is in line with that of Krishnamani et al. (2004) in another tropical forest: it is very small compared to the classical 0.25 of Arrhenius (1921) or 0.263 of Preston (1962). This was discussed by MacArthur and Wilson (1967), chapter 2. The power law applies to embedded scales of the same ecosystem here, in contrast to the usual sets of isolated islands providing the data (Triantis et al., 2012): in our case, the number of species increases less with the area, leading to smaller z values.

The critical aspect of the estimation is the accuracy of the starting point of the extrapolation, that mainly depends on the representativeness of the local inventories. Again, the self-similarity model assumes that c, the number of species per square kilometer, is the same everywhere. Local, observed values must be understood as variations around the real c, that should be estimated by replicating inventories across the whole region. This is of course restricted by the huge resources needed to settle a single one: three replicates are an exceptional amount of data. Paracou, Piste de Saint-Elie and Nouragues represent quite well the variability of local richness of the forest of French Guiana. We made a strict selection of the data to count the numbers of species, which are thus lower bounds. Ongoing efforts of botanists may increase a bit the value of c, implying a proportional increase in the estimation of the number of species.

4.3 Chao2's estimator is a valid alternative

Nonparametric estimators of richness are widely used to estimate the asymptotic richness of a community, because they are designed to estimate the number of unobserved species due to incomplete sampling (Colwell and Coddington, 1994). Yet, their underlying assumptions are limited: they do not depend on any distribution model or scale of observation. The only constraint is independent and identically distributed (iid) sampling, even though at the local scale spatial aggregation is often neglected (Picard et al., 2004).

The asymptotic estimation based on the Chao1 or jackknife estimator is less than 1700 species, i.e. less than the total number of known species (Molino et al., 2022). As already underlined by ter Steege et al. (2013), nonparametric asymptotic estimation of richness is not appropriate at large scales because of severe undersampling: many local communities are just not included in the data. Yet, increasing the sampling effort would not be enough: mixing local samples (the 1-ha plots) to mimic an iid sampling of a whole region is clearly not a valid approximation, because each plot has its own distribution.

Cazzolla Gatti et al. (2022) applied a similar method on a large-scale grid (100 x 100 km cells) where species occurrences were reported in each cell. Considering each cell as a plot, the Chao2 estimator (Chao, 1987) allows estimating the total richness. The practical advantage of this approach is the opportunity to combine several sources of occurrence data to improve the sampling coverage. Theoretically, it is far more robust than the mixture of abundance data: the local distribution of each plot is cancelled out by its transformation into incidence data. An appropriate spatial distribution of sampling plots, covering all habitats or at least a regular grid in absence of more detailed knowledge, can be seen as a valid sampling. When applied to our data, aggregated in 100-km square cells, the estimation is similar to that obtained directly from the abundance data of the plots because of undersampling but the method must not be rejected.

4.4 Log series extrapolations are not valid at the regional scale

At the scale of the metacommunity, defined as of the neutral model of biogeography, the species distribution is in log-series (Hubbell, 2001; Volkov et al., 2003). ter Steege et al. (2013) fitted a log-series to data provided by a network of plots to estimate the number of species in Amazonia. We applied the same method to our data. Its estimation is well over 4000 species in French Guiana: a very unlikely result according to the current expert knowledge and the recent checklist (Molino et al., 2022). The regional species pool does not follow a log-series distribution because of dispersal limitation (Grilli et al., 2012). In other words, the regional community is not a sample of the metacommunity: many of the metacommunity's species are not present. As a consequence, the log-series estimation of the richness of a regional species pool leads to severe overestimation. For the same reasons, hyperdominance is less pronounced: 4% of the species contain half the trees (figure 5 in the appendix), compared to 1.4% in Amazonia as a whole (ter Steege et al., 2013).

The universal species-area relationship (Harte et al., 2008) allowed the extrapolation of observed richness up to the 8 million hectares of French Guiana. The number of species estimated from Paracou, Piste de Saint-Elie and Nouragues starting points (their number of species and area) is on average 2787, and over 4500 when extrapolating from an average Guyadiv 1-ha plot. Again, this model implies a log-series distribution as it integrates as few assumptions as possible (Harte et al., 2008). On the log-log representation of figure 4, the species-area relationships are never linear, as predicted by the

power law at the regional scale. The arguments for overestimation are the same as those against the extrapolation of the log-series at the regional scale.

4.5 The number of species is around 2200

Finally, our estimations of the number of tree species in the 8-million-hectare forest of French Guiana is close to 2200, with a quite wide confidence interval due to the variability in the estimation of both the number of trees in a square kilometer and the power-law parameter. Their distribution is highly unequal: 90 species (4%) contain half the trees.

A recent work (Molino et al., 2022) lists nearly 1800 species of indigenous trees in French Guiana, based on herbarium collections on the one hand, and on data from the GuyaDiv and GuyaFor plot networks (Engel, 2015) on the other. However, this checklist is only a state of the art of our knowledge of the tree flora. Even in the most intensively explored areas, botanists conducting botanical inventories have identified a number of entities that are morphologically distinct from all known species in French Guiana, and which they therefore consider to be still unnamed species. They gave them provisional names (e.g. *Pouteria* sp. A), until more information is available to either recognize species known in other parts of the world, or to describe them and give them a valid name according to the Code of Nomenclature. The GuyaDiv and GuyaFor databases together currently list more than 300 of these unnamed species, but Molino et al. (2022) selected only 143 of them for their checklist, the ones that were best characterized and best illustrated by good quality herbarium specimens. Although it cannot be excluded that some of the other 150-200 unnamed species are in fact simple morphological variants of already described species, they believe that most of them represent distinct species. In other words, the number of known species in French Guiana (named and unnamed) is probably already close to 2000. Furthermore, the available data is very unevenly distributed across the territory. The south and especially the north-west of French Guiana are poorly explored botanically (few inventory plots, relatively few herbarium specimens), while their floras are significantly different from the better inventoried northern and central zones. It is thus very likely that the exploration of these little-known areas will add new species to the list. Therefore, the estimate of 2234 spp. given here seems quite plausible, given the state of our knowledge.

A improvement perspective is the aggregation of all sources of localized data at the scale of French Guiana, including all Guyadiv plots, the Guyafor network (Jaouen et al., 2021), herbarium collections and various scientific projects, to proceed with an incidence-based estimation of richness following Cazzolla Gatti et al. (2022). A main issue of this approach is the standardization of the taxonomy, already well advanced by Molino et al. (2022), so it may be feasible in a near future.
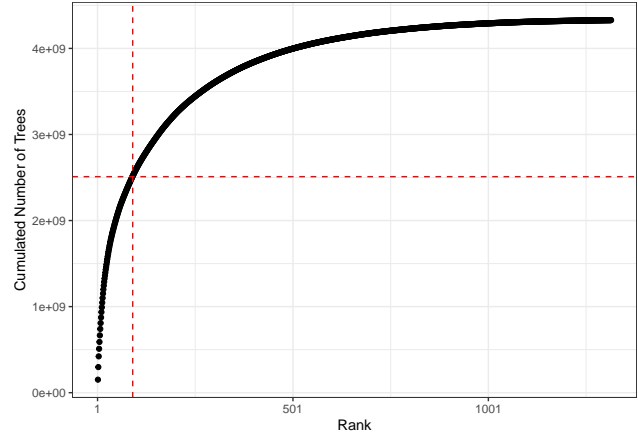

Figure 5. Accumulation of the number of individuals from the most abundant to the rarest species. The horizontal line corresponds to half the individuals. The vertical line allows reading the corresponding rank of the species. Unobserved species are not represented on the curve.

5. Appendix

5.1 Variance of the product of two independent random variables

Let X and Y two random variables, here the estimators of c and A^z.

The variance of their product XY is

$$
\begin{aligned}
\mathrm{Var}(XY) &= \mathbb{E}(X^2Y^2) - \mathbb{E}^2(XY) \\
&= \mathbb{E}(X^2)\mathbb{E}(Y^2) + \mathrm{Cov}(X^2,Y^2) \\
&\quad - [\mathbb{E}(X)\mathbb{E}(Y) + \mathrm{Cov}(X,Y)]^2.
\end{aligned} \tag{6}
$$

If X and Y are independent (this applies to c and A^z), covariances are 0 and the variance reduces to

$$
\begin{aligned}
\mathrm{Var}(XY) &= \mathbb{E}(X^2)\mathbb{E}(Y^2) - [\mathbb{E}(X)\mathbb{E}(Y)]^2 \\
&= [\mathrm{Var}(X) + \mathbb{E}(X)^2][\mathrm{Var}(Y) + \mathbb{E}(Y)^2] \\
&\quad - [\mathbb{E}(X)\mathbb{E}(Y)]^2 \\
&= \mathrm{Var}(X)\mathrm{Var}(Y) \\
&\quad + \mathbb{E}(X)^2\mathrm{Var}(Y) + \mathbb{E}(Y)^2\mathrm{Var}(X).
\end{aligned} \tag{7}
$$

5.2 Hyperdominance

Hyperdominance is a characteristic of many distributions of species. Figure 5 shows the accumulation of individuals from the most abundant to the rarest species.

The number of trees and the probabilities of observed species were calculated for the log-series extrapolation (see the methods in the main text). The total number of species is estimated according to the self-similarity model.

Only 90 species, i.e. 4% of their estimated number, contain half the number of trees.

References

Arrhenius, O. (1921). Species and area. *Journal of Ecology 9*(1), 95–99.

Béguinot, J. (2015). Extrapolation of the species accumulation curve for incomplete species samplings: A new nonparametric approach to estimate the degree of sample completeness and decide when to stop sampling. *Annual Research & Review in Biology 8*(5), 1–9.

Bonal, D., A. Bosc, S. Ponton, J.-Y. Goret, B. Burban, P. Gross, J.-M. Bonnefond, J. Elbers, B. Longdoz, D. Epron, J.-M. Guehl, and A. Granier (2008). Impact of severe dry season on net ecosystem exchange in the Neotropical rainforest of French Guiana. *Global Change Biology 14*(8), 1917–1933.

Bongers, F., P. Charles-Dominique, P.-M. Forget, and M. Théry (2001). *Nouragues: Dynamics and Plant-Animal Interactions in a Neotropical Rainforest*, Volume 80. Springer Science & Business Media.

Burnham, K. P. and W. S. Overton (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika 65*(3), 625–633.

Cazzolla Gatti, R., P. B. Reich, J. G. P. Gamarra, T. Crowther, C. Hui, A. Morera, J.-F. Bastin, S. de-Miguel, G.-J. Nabuurs, J.-C. Svenning, J. M. Serra-Diaz, C. Merow, B. Enquist, M. Kamenetsky, J. Lee, J. Zhu, J. Fang, D. F. Jacobs, B. Pijanowski, A. Banerjee, R. A. Giaquinto, G. Alberti, A. M. Almeyda Zambrano, E. Alvarez-Davila, A. Araujo-Murakami, V. Avitabile, G. A. Aymard, R. Balazy, C. Baraloto, J. G. Barroso, M. L. Bastian, P. Birnbaum, R. Bitariho, J. Bogaert, F. Bongers, O. Bouriaud, P. H. S. Brancalion, F. Q. Brearley, E. N. Broadbent, F. Bussotti, W. Castro da Silva, R. G. César, G. Češljar, V. Chama Moscoso, H. Y. H. Chen, E. Cienciala, C. J. Clark, D. A. Coomes, S. Dayanandan, M. Decuyper, L. E. Dee, J. Del Aguila Pasquel, G. Derroire, M. N. K. Djuik-ouo, T. Van Do, J. Dolezal, I. Đ. Đorđević, J. Engel, T. M. Fayle, T. R. Feldpausch, J. K. Fridman, D. J. Harris, A. Hemp, G. Hengeveld, B. Herault, M. Herold, T. Ibanez, A. M. Jagodzinski, B. Jaroszewicz, K. J. Jeffery, V. K. Johannsen, T. Jucker, A. Kangur, V. N. Karminov, K. Kartawinata, D. K. Kennard, S. Kepfer-Rojas, G. Keppel, M. L. Khan, P. K. Khare, T. J. Kileen, H. S. Kim, H. Korjus, A. Kumar, A. Kumar, D. Laarmann, N. Labrière, M. Lang, S. L. Lewis, N. Lukina, B. S. Maitner, Y. Malhi, A. R. Marshall, O. V. Martynenko, A. L. Monteagudo Mendoza, P. V. Ontikov, E. Ortiz-Malavasi, N. C. Pallqui Camacho, A. Paquette, M. Park, N. Parthasarathy, P. L. Peri, P. Petronelli, S. Pfautsch, O. L. Phillips, N. Picard, D. Piotto, L. Poorter, J. R. Poulsen, H. Pretzsch, H. Ramírez-Angulo, Z. Restrepo Correa, M. Rodeghiero, R. D. P. Rojas Gonzáles, S. G. Rolim, F. Rovero, E. Rutishauser, P. Saikia, C. Salas-Eljatib, D. Schepaschenko, M. Scherer-Lorenzen, V. Šebeň, M. Silveira, F. Slik, B. Sonké, A. F. Souza, K. J. Stereńczak, M. Svoboda, H. Taedoumg, N. Tchebakova, J. Terborgh, E. Tikhonova, A. Torres-Lezama, F. van der Plas, R. Vásquez, H. Viana, A. C. Vibrans, E. Vilanova, V. A. Vos, H.-F. Wang, B. Westerlund, L. J. T. White, S. K. Wiser, T. Zawiła-Niedźwiecki, L. Zemagho, Z.-X. Zhu, I. C. Zo-Bi, and J. Liang (2022). The number of tree species on Earth. *Proceedings of the National Academy of Sciences 119*(6), e2115329119.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics 11*(4), 265–270.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics 43*(4), 783–791.

Chao, A. and L. Jost (2015). Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution 6*(8), 873–882.

Chao, A., Y.-T. Wang, and L. Jost (2013). Entropy and the species accumulation curve: A novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution 4*(11), 1091–1100.

Colwell, R. K. and J. A. Coddington (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 345*(1311), 101–118.

Condit, R., S. P. Hubbell, J. V. Lafrankie, R. Sukumar, N. Manokaran, R. B. Foster, and P. S. Ashton (1996). Species-area and species-individual relationships for tropical trees: A comparison of three 50-ha plots. *Journal of Ecology 84*(4), 549–562.

Connor, E. F. and E. D. McCoy (1979). The statistics and biology of the species-area relationship. *The American Naturalist 113*(6), 791–833.

Dengler, J. (2009). Which function describes the species-area relationship best? A review and empirical evaluation. *Journal of Biogeography 36*(4), 728–744.

Duque, A., H. C. Muller-Landau, R. Valencia, D. Cardenas, S. Davies, A. De Oliveira, Á. J. Pérez, H. Romero-Saltos, and A. Vicentini (2017). Insights into regional patterns of Amazonian forest structure, diversity, and dominance from three large terra-firme forest dynamics plots. *Biodiversity and Conservation 26*(3), 669–686.

Engel, J. (2015). Plot networks & teams. http://atdnmorphospecies.myspecies.info/node/781.

Engen, S. (1977). Exponential and logarithmic species-area curves. *The American Naturalist 111*(979), 591–594.

Fisher, R. A., A. S. Corbet, and C. B. Williams (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology 12*, 42–58.

ForestPlots.net, C. Blundo, J. Carilla, R. Grau, A. Malizia, L. Malizia, O. Osinaga-Acosta, M. Bird, M. Bradford, D. Catchpole, A. Ford, A. Graham, D. Hilbert, J. Kemp, S. Laurance, W. Laurance, F. Y. Ishida, A. Marshall, C. Waite, H. Woell, J.-F. Bastin, M. Bauters, H. Beeckman, P. Boeckx, J. Bogaert, C. De Canniere, T. De Haulleville, J.-L. Doucet, O. Hardy, W. Hubau, E. Kearsley, H. Verbeeck, J. Vleminckx, S. W. Brewer, A. Alarcón, A. Araujo-Murakami, E. Arets, L. Arroyo, E. Chavez, T. Fredericksen, R. G. Villaroel, G. G. Sibauty, T. Killeen, J. C. Licona, J. Lleigue, C. Mendoza, S. Murakami, A. P. Gutierrez, G. Pardo, M. Peña-Claros, L. Poorter, M. Toledo, J. V. Cayo, L. J. Viscarra, V. Vos, J. Ahumada, E. Almeida, J. Almeida, E. A. De Oliveira, W. A. Da Cruz, A. A. De Oliveira, F. A. Carvalho, F. A. Obermuller, A. Andrade, F. A. Carvalho, S. A. Vieira, A. C. Aquino, L. Aragão, A. C. Araújo, M. A. Assis, J. A. M. A. Gomes, F. Baccaro, P. B. De Camargo, P. Barni, J. Barroso, L. C. Bernacci, K. Bordin, M. B. De Medeiros, I. Broggio, J. L. Camargo, D. Cardoso, M. A. Carniello, A. L. C. Rochelle, C. Castilho, A. A. J. F. Castro, W. Castro, S. C. Ribeiro, F. Costa, R. C. De Oliveira, I. Coutinho, J. Cunha, L. Da Costa, L. Da Costa Ferreira, R. Da Costa Silva, M. Da Graça Zacarias Simbine, V. De Andrade Kamimura, H. C. De Lima, L. De Oliveira Melo, L. De Queiroz, J. R. De Sousa Lima, M. Do Espírito Santo, T. Domingues, N. C. Dos Santos Prestes, S. E. S. Carneiro, F. Elias, G. Eliseu, T. Emilio, C. L. Farrapo, L. Fernandes, G. Ferreira, J. Ferreira, L. Ferreira, S. Ferreira, M. F. Simon, M. A. Freitas, Q. S. García, A. G. Manzatto, P. Graça, F. Guilherme, E. Hase, N. Higuchi, M. Iguatemy, R. I. Barbosa, M. Jaramillo, C. Joly, J. Klipel, I. L. Do Amaral, C. Levis, A. S. Lima, M. L. Dan, A. Lopes, H. Madeiros, W. E. Magnusson, R. M. Dos Santos, B. Marimon, B. H. M. Junior, R. M. M. Grillo, L. Martinelli, S. M. Reis, S. Medeiros, M. Meira-Junior, T. Metzker, P. Morandi, N. M. Do Nascimento, M. Moura, S. C. Müller, L. Nagy, H. Nascimento, M. Nascimento, A. N. Lima, R. O. De Araújo, J. O. Silva, M. Pansonato, G. P. Sabino, K. M. P. De Abreu, P. J. F. P. Rodrigues, M. Piedade, D. Rodrigues, J. R. Rodrigues Pinto, C. Quesada, E. Ramos, R. Ramos, P. Rodrigues, T. R. De Sousa, R. Salomão, F. Santana, M. Scaranello, R. S. Bergamin, J. Schietti,

J. Schöngart, G. Schwartz, N. Silva, M. Silveira, C. S. Seixas, M. Simbine, A. C. Souza, P. Souza, R. Souza, T. Sposito, E. S. Junior, J. D. Do Vale, I. C. G. Vieira, D. Villela, M. Vital, H. Xaud, K. Zanini, C. E. Zartman, N. K. H. Ideris, F. B. H. Metali, K. A. Salim, M. S. Saparudin, R. M. Serudin, R. S. Sukri, S. Begne, G. Chuyong, M. N. Djuikouo, C. Gonmadje, M. Simo-Droissart, B. Sonké, H. Taedoumg, L. Zemagho, S. Thomas, F. Baya, G. Saiz, J. S. Espejo, D. Chen, A. Hamilton, Y. Li, T. Luo, S. Niu, H. Xu, Z. Zhou, E. Álvarez-Dávila, J. C. A. Escobar, H. Arellano-Peña, J. C. Duarte, J. Calderón, L. M. C. Bravo, B. Cuadrado, H. Cuadros, A. Duque, L. F. Duque, S. M. Espinosa, R. Franke-Ante, H. García, A. Gómez, R. González-M., Á. Idárraga-Piedrahíta, E. Jimenez, R. Jurado, W. L. Oviedo, R. López-Camacho, O. A. M. Cruz, I. M. Polo, E. Paky, K. Pérez, A. Pijachi, C. Pizano, A. Prieto, L. Ramos, Z. R. Correa, J. Richardson, E. Rodríguez, G. M. Rodriguez M., A. Rudas, P. Stevenson, M. Chudomelová, M. Dancak, R. Hédl, S. Lhota, M. Svatek, J. Mukinzi, C. Ewango, T. Hart, E. K. Yakusu, J. Lisingo, J.-R. Makana, F. Mbayu, B. Toirambe, J. T. Mukendi, L. Kvist, G. Nebel, S. Báez, C. Céron, D. M. Griffith, J. E. G. Andino, D. Neill, W. Palacios, M. C. Peñuela-Mora, G. Rivas-Torres, G. Villa, S. Demissie, T. Gole, T. Gonfa, K. Ruokolainen, M. Baisie, F. Bénédet, W. Betian, V. Bezard, D. Bonal, J. Chave, V. Droissart, S. Gourlet-Fleury, A. Hladik, N. Labrière, P. Naisso, M. Réjou-Méchain, P. Sist, L. Blanc, B. Burban, G. Derroire, A. Dourdain, C. Stahl, N. N. Bengone, E. Chezeaux, F. E. Ondo, V. Medjibe, V. Mihindou, L. White, H. Culmsee, C. D. Rangel, V. Horna, F. Wittmann, S. Adu-Bredu, K. Affum-Baffoe, E. Foli, M. Balinga, A. Roopsind, J. Singh, R. Thomas, R. Zagt, I. K. Murthy, K. Kartawinata, E. Mirmanto, H. Priyadi, I. Samsoedin, T. Sunderland, I. Yassir, F. Rovero, B. Vinceti, B. Hérault, S.-I. Aiba, K. Kitayama, A. Daniels, D. Tuagben, J. T. Woods, M. Fitriadi, A. Karolus, K. L. Khoon, N. Majalap, C. Maycock, R. Nilus, S. Tan, A. Sitoe, I. Coronado G., L. Ojo, R. De Assis, A. D. Poulsen, D. Sheil, K. A. Pezo, H. B. Verde, V. C. Moscoso, J. C. C. Oroche, F. C. Valverde, M. C. Medina, N. D. Cardozo, J. De Rutte Corzo, J. Del Aguila Pasquel, G. F. Llampazo, L. Freitas, D. G. Cabrera, R. G. Villacorta, K. G. Cabrera, D. G. Soria, L. G. Saboya, J. M. G. Rios, G. H. Pizango, E. H. Coronado, I. Huamantupa-Chuquimaco, W. H. Huasco, Y. T. H. Aedo, J. L. M. Peña, A. M. Mendoza, V. M. Rodriguez, P. N. Vargas, S. C. P. Ramos, N. P. Camacho, A. P. Cruz, F. R. Arevalo, J. R. Huaymacari, C. R. Rodriguez, M. A. R. Paredes, L. R. Bayona, R. Del Pilar Rojas Gonzales, M. E. R. Peña, N. S. Revilla, Y. C. S. Shareva, R. T. Trujillo, L. V. Gamarra, R. V. Martinez, J. V. Arenas, C. Amani, S. A. Ifo, Y. Bocko,

P. Boundja, R. Ekoungoulou, M. Hockemba, D. Nzala, A. Fofanah, D. Taylor, G. Bañares-de Dios, L. Cayuela, Í. G.-d. La Cerda, M. Macía, J. Stropp, M. Playfair, V. Wortel, T. Gardner, R. Muscarella, H. Priyadi, E. Rutishauser, K.-J. Chao, P. Munishi, O. Bánki, F. Bongers, R. Boot, G. Fredriksson, J. Reitsma, H. Ter Steege, T. Van Andel, P. Van De Meer, P. Van Der Hout, M. Van Nieuwstadt, B. Van Ulft, E. Veenendaal, R. Vernimmen, P. Zuidema, J. Zwerts, P. Akite, R. Bitariho, C. Chapman, E. Gerald, M. Leal, P. Mucunguzi, K. Abernethy, M. Alexiades, T. R. Baker, K. Banda, L. Banin, J. Barlow, A. Bennett, E. Berenguer, N. Berry, N. M. Bird, G. A. Blackburn, F. Brearley, R. Brienen, D. Burslem, L. Carvalho, P. Cho, F. Coelho, M. Collins, D. Coomes, A. Cuni-Sanchez, G. Dargie, K. Dexter, M. Disney, F. Draper, M. Duan, A. Esquivel-Muelbert, R. Ewers, B. Fadrique, S. Fauset, T. R. Feldpausch, F. França, D. Galbraith, M. Gilpin, E. Gloor, J. Grace, K. Hamer, D. Harris, K. Jeffery, T. Jucker, M. Kalamandeen, B. Klitgaard, A. Levesley, S. L. Lewis, J. Lindsell, G. Lopez-Gonzalez, J. Lovett, Y. Malhi, T. Marthews, E. McIntosh, K. Melgaço, W. Milliken, E. Mitchard, P. Moonlight, S. Moore, A. Morel, J. Peacock, K. S.-H. Peh, C. Pendry, R. T. Pennington, L. De Oliveira Pereira, C. Peres, O. L. Phillips, G. Pickavance, T. Pugh, L. Qie, T. Riutta, K. Roucoux, C. Ryan, T. Sarkinen, C. S. Valeria, D. Spracklen, S. Stas, M. Sullivan, M. Swaine, J. Talbot, J. Taplin, G. Van Der Heijden, L. Vedovato, S. Willcock, M. Williams, L. Alves, P. A. Loayza, G. Arellano, C. Asa, P. Ashton, G. Asner, T. Brncic, F. Brown, R. Burnham, C. Clark, J. Comiskey, G. Damasco, S. Davies, T. Di Fiore, T. Erwin, W. Farfan-Rios, J. Hall, D. Kenfack, T. Lovejoy, R. Martin, O. M. Montiel, J. Pipoly, N. Pitman, J. Poulsen, R. Primack, M. Silman, M. Steininger, V. Swamy, J. Terborgh, D. Thomas, P. Umunay, M. Uriarte, E. V. Torre, O. Wang, K. Young, G. A. Aymard C., L. Hernández, R. H. Fernández, H. Ramírez-Angulo, P. Salcedo, E. Sanoja, J. Serrano, A. Torres-Lezama, T. C. Le, T. T. Le, and H. D. Tran (2021, August). Taking the pulse of Earth's tropical forests using networks of highly distributed plots. *Biological Conservation 260*, 108849.

Gárcia Martín, H. and N. Goldenfeld (2006). On the origin and robustness of power-law species-area relationships in ecology. *Proceedings of the National Academy of Sciences of the United States of America 103*(27), 10310–10315.

Gleason, H. A. (1922). On the relation between species and area. *Ecology 3*(2), 158–162.

Gonzalez, S., V. Bilot-Guérin, P. G. Delprete, C. Geniez, J.-F. Molino, and J.-L. Smock (2022).

L'herbier IRD de Guyane.   https://herbier-guyane.ird.fr/.

Good, I. J. (1953). The population frequency of species and the estimation of population parameters. *Biometrika 40*(3/4), 237–264.

Gotelli, N. J. and R. K. Colwell (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters 4*(4), 379–391.

Gourlet-Fleury, S., J. M. Guehl, and O. Laroussinie (2004). *Ecology & Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana.* Paris: Elsevier.

Grabchak, M., E. Marcon, G. Lang, and Z. Zhang (2017). The generalized Simpson's entropy is a measure of biodiversity. *Plos One 12*(3), e0173305.

Grilli, J., S. Azaele, J. R. Banavar, and A. Maritan (2012). Spatial aggregation and the species-area relationship across scales. *Journal of Theoretical Biology 313*, 87–97.

Guitet, S., R. Pélissier, O. Brunaux, G. Jaouen, and D. Sabatier (2015). Geomorphological landscape features explain floristic patterns in French Guiana rainforest. *Biodiversity and Conservation 24*(5), 1215–1237.

Harte, J., A. Kinzig, and J. Green (1999). Self-similarity in the distribution and abundance of species. *Science 284*(5412), 334–336.

Harte, J., S. Mccarthy, K. Taylor, A. Kinzig, and M. L. Fischer (1999). Estimating species-area relationships from scale plot to landscape data using species spatial-turnover. *Oikos 86*(1), 45–54.

Harte, J., A. B. Smith, and D. Storch (2009). Biodiversity scales from plots to biomes with a universal species-area curve. *Ecology Letters 12*(8), 789–797.

Harte, J., T. Zillio, E. Conlisk, and A. B. Smith (2008). Maximum entropy and the state-variable approach to macroecology. *Ecology 89*(10), 2700–2711.

Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography.* Princeton University Press.

Jaouen, G., A. Dourdain, and G. Derroire (2021). Guyafor Data Dictionary.

Krishnamani, R., A. Kumar, and J. Harte (2004). Estimating species richness at large spatial scales using data from small discrete plots. *Ecography 27*(5), 637–642.

Lineweaver, H. and D. Burk (1934). The determination of enzyme dissociation constants. *Journal of the American Chemical Society 56*(3), 658–666.

MacArthur, R. H. and E. O. Wilson (1967). The theory of island biogeography. In *Monographs in Population Biology*, Volume 1. Princeton University Press.

Marcon, E. and B. Hérault (2015). Entropart, an R package to measure and partition diversity. *Journal of Statistical Software 67*(8), 1–26.

May, R. M. (1975). Patterns of species abundance and diversity. In M. L. Cody and J. M. Diamond (Eds.), *Ecology and Evolution of Communities*, pp. 81–120. Harvard University Press.

McGill, B. J. (2003). A test of the unified neutral theory of biodiversity. *Nature 422*(6934), 881–885.

Mirabel, A., B. Hérault, and E. Marcon (2020). Diverging taxonomic and functional trajectories following disturbance in a Neotropical forest. *Science of The Total Environment 720*, 137397.

Mirabel, A., E. Marcon, and B. Hérault (2021). 30 Years of postdisturbance recruitment in a Neotropical forest. *Ecology and Evolution 11*(21), 14448–14458.

Molino, J.-F. and D. Sabatier (2001). Tree diversity in tropical rain forests: A validation of the intermediate disturbance hypothesis. *Science 294*(5547), 1702–1704.

Molino, J.-F., D. Sabatier, Grenand, Pierre, J. Engel, Frame, Dawn, Delprete, Piero G., Fleury, Marie, Odonne, Guillaume, Davy, Damien, Lucas, Eve J., and Martin, Claire A. (2022). An annotated checklist of the tree species of French Guiana, including vernacular nomenclature. *Adansonia 44*(26), 345–903.

Picard, N., M. Karembe, and P. Birnbaum (2004). Species-area curve and spatial pattern. *Ecoscience 11*(1), 45–54.

Plotkin, J. B., M. D. Potts, N. Leslie, N. Manokaran, J. V. LaFrankie, and P. S. Ashton (2000). Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology 207*(1), 81–99.

Preston, F. W. (1948). The commonness, and rarity, of species. *Ecology 29*(3), 254–283.

Preston, F. W. (1962). The canonical distribution of commonness and rarity: Part I. *Ecology 43*(2), 185–215.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Shen, T.-J., A. Chao, and C.-F. Lin (2003). Predicting the number of new species in a further taxonomic sampling. *Ecology 84*(3), 798–804.

Slik, J. W. F., V. Arroyo-Rodríguez, S.-I. Aiba, P. Alvarez-Loayza, L. F. Alves, P. S. Ashton, P. Balvanera, M. L. Bastian, P. J. Bellingham, E. van den Berg, L. Bernacci, P. da Conceição Bispo, L. Blanc, K. Böhning-Gaese, P. Boeckx, F. Bongers, B. Boyle, M. Bradford, F. Q. Brearley, M. Breuer-Ndoundou Hockemba, S. Bunyavejchewin, D. Calderado Leal Matos, M. Castillo-Santiago, E. L. M. Catharino, S.-L. Chai, Y. Chen, R. K. Colwell, C. L. Robin, C. J. Clark, D. B. Clark, D. A. Clark, H. Culmsee, K. Damas, H. S. Dattaraja, G. Dauby, P. Davidar, S. J. DeWalt, J.-L. Doucet, A. Duque, G. Durigan, K. A. O. Eichhorn, P. V. Eisenlohr, E. Eler, C. Ewango, N. Farwig, K. J. Feeley, L. Ferreira, R. Field, A. T. de Oliveira Filho, C. Fletcher, O. Forshed, G. Franco, G. Fredriksson, T. Gillespie, J.-F. Gillet, G. Amarnath, D. M. Griffith, J. Grogan, N. Gunatilleke, D. Harris, R. Harrison, A. Hector, J. Homeier, N. Imai, A. Itoh, P. A. Jansen, C. A. Joly, B. H. J. de Jong, K. Kartawinata, E. Kearsley, D. L. Kelly, D. Kenfack, M. Kessler, K. Kitayama, R. Kooyman, E. Larney, Y. Laumonier, S. Laurance, W. F. Laurance, M. J. Lawes, I. L. Amaral, S. G. Letcher, J. Lindsell, X. Lu, A. Mansor, A. Marjokorpi, E. H. Martin, H. Meilby, F. P. L. Melo, D. J. Metcalfe, V. P. Medjibe, J. P. Metzger, J. Millet, D. Mohandass, J. C. Montero, M. de Morisson Valeriano, B. Mugerwa, H. Nagamasu, R. Nilus, S. Ochoa-Gaona, Onrizal, N. Page, P. Parolin, M. Parren, N. Parthasarathy, E. Paudel, A. Permana, M. T. F. Piedade, N. C. A. Pitman, L. Poorter, A. D. Poulsen, J. Poulsen, J. Powers, R. C. Prasad, J.-P. Puyravaud, J.-C. Razafimahaimodison, J. Reitsma, J. R. dos Santos, W. Roberto Spironello, H. Romero-Saltos, F. Rovero, A. H. Rozak, K. Ruokolainen, E. Rutishauser, F. Saiter, P. Saner, B. A. Santos, F. Santos, S. K. Sarker, M. Satdichanh, C. B. Schmitt, J. Schöngart, M. Schulze, M. S. Suganuma, D. Sheil, E. da Silva Pinheiro, P. Sist, T. Stevart, R. Sukumar, I.-F. Sun, T. Sunderland, H. S. Suresh, E. Suzuki, M. Tabarelli, J. Tang, N. Targhetta, I. Theilade, D. W. Thomas, P. Tchouto, J. Hurtado, R. Valencia, J. L. C. H. van Valkenburg, T. Van Do, R. Vasquez, H. Verbeeck, V. Adekunle, S. A. Vieira, C. O. Webb, T. Whitfeld, S. A. Wich, J. Williams, F. Wittmann, H. Wöll, X. Yang, C. Y. Adou Yao, S. L. Yap, T. Yoneda, R. A. Zahawi, R. Zakaria, R. Zang, R. L. de Assis, B. Garcia Luize, and E. M. Venticinque (2015). An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences of the United States of America 112*(24), 7472–7477.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. *Biologiske Skrifter 5*(4), 1–34.

ter Steege, H., N. C. A. Pitman, D. Sabatier, C. Baraloto, R. P. Salomão, J. E. Guevara, O. L. Phillips, C. V. Castilho, W. E. Magnusson, J.-F. Molino, A. Monteagudo, P. Núñez Vargas, J. C. Montero, T. R. Feldpausch, E. N. H. Coronado, T. J. Killeen, B. Mostacedo, R. Vasquez, R. L. Assis, J. Terborgh, F. Wittmann, A. C. S. Andrade, W. F. Laurance, S. G. W. Laurance, B. S. Marimon, B.-H. Marimon, I. C. Guimarães Vieira, I. L. Amaral, R. Brienen, H. Castellanos, D. Cárdenas López, J. F. Duivenvoorden, H. F. Mogollón, F. D. de Almeida Matos, N. Dávila, R. García-Villacorta, P. R. Stevenson Diaz, F. Costa, T. Emilio, C. Levis, J. Schietti, P. Souza, A. Alonso, F. Dallmeier, A. J. D. Montoya, M. T. Fernandez Piedade, A. Araujo-Murakami, L. Arroyo, R. Gribel, P. V. A. Fine, C. A. Peres, M. Toledo, G. A. Aymard, T. Baker, C. Cerón, J. Engel, T. W. Henkel, P. Maas, P. Petronelli, J. Stropp, C. E. Zartman, D. Daly, D. Neill, M. Silveira, M. R. Paredes, J. Chave, D. de Andrade Lima Filho, P. M. Jørgensen, A. Fuentes, J. Schöngart, F. Cornejo Valverde, A. Di Fiore, E. M. Jimenez, M. C. Peñuela-Mora, J. F. Phillips, G. Rivas, T. R. van Andel, P. von Hildebrand, B. Hoffman, E. L. Zent, Y. Malhi, A. Prieto, A. Rudas, A. R. Ruschell, N. Silva, V. Vos, S. Zent, A. A. Oliveira, A. C. Schutz, T. Gonzales, M. Trindade Nascimento, H. Ramirez-Angulo, R. Sierra, M. Tirado, M. N. Umaña Medina, G. van der Heijden, C. I. A. Vela, E. Vilanova Torre, C. Vriesendorp, O. Wang, K. R. Young, C. Baider, H. Balslev, C. Ferreira, I. Mesones, A. Torres-Lezama, L. E. Urrego Giraldo, R. Zagt, M. N. Alexiades, L. Hernandez, I. Huamantupa-Chuquimaco, W. Milliken, W. Palacios Cuenca, D. Pauletto, E. Valderrama Sandoval, L. Valenzuela Gamarra, K. G. Dexter, K. J. Feeley, G. Lopez-Gonzalez, and M. R. Silman (2013). Hyperdominance in the amazonian tree flora. *Science 342*(6156), 1243092.

ter Steege, H., P. I. Prado, R. A. F. de de Lima, E. Pos, L. de Souza Coelho, D. de Andrade Lima Filho, R. P. Salomão, I. L. Amaral, F. D. de Almeida Matos, C. V. Castilho, O. L. Phillips, J. E. Guevara, M. de Jesus Veiga Carim, D. Cárdenas López, W. E. Magnusson, F. Wittmann, M. P. Martins, D. Sabatier, M. V. Irume, J. R. da Silva Guimarães, J.-F. Molino, O. S. Bánki, M. T. F. Piedade, N. C. A. Pitman, J. F. Ramos, A. Monteagudo Mendoza, E. M. Venticinque, B. G. Luize, P. Núñez Vargas, T. S. F. Silva, E. M. M. de Leão Novo, N. F. C. Reis, J. Terborgh, A. G. Manzatto, K. R. Casula, E. N. Honorio Coronado, J. C. Montero, A. Duque, F. R. C. Costa, N. Castaño Arboleda, J. Schöngart, C. E. Zartman, T. J. Killeen, B. S. Marimon, B. H. Marimon-Junior, R. Vasquez, B. Mostacedo, L. O. Demarchi, T. R. Feldpausch, J. Engel, P. Petronelli, C. Baraloto, R. L. Assis, H. Castellanos, M. F. Simon, M. B. de Medeiros, A. Quaresma, S. G. W. Laurance, L. M. Rincón, A. Andrade, T. R. Sousa, J. L. Camargo, J. Schietti, W. F. Laurance, H. L. de Queiroz, H. E. M. Nascimento, M. A. Lopes, E. de Sousa Farias, J. L. L. Magalhães, R. Brienen, G. A. Aymard C., J. D. C. Revilla, I. C. G. Vieira, B. B. L. Cintra, P. R. Stevenson, Y. O. Feitosa, J. F. Duivenvoorden, H. F. Mogollón, A. Araujo-Murakami, L. V. Ferreira, J. R. Lozada, J. A. Comiskey, J. J. de Toledo, G. Damasco, N. Dávila, A. Lopes, R. García-Villacorta, F. Draper, A. Vicentini, F. Cornejo Valverde, J. Lloyd, V. H. F. Gomes, D. Neill, A. Alonso, F. Dallmeier, F. C. de Souza, R. Gribel, L. Arroyo, F. A. Carvalho, D. P. P. de Aguiar, D. D. do Amaral, M. P. Pansonato, K. J. Feeley, E. Berenguer, P. V. A. Fine, M. C. Guedes, J. Barlow, J. Ferreira, B. Villa, M. C. Peñuela Mora, E. M. Jimenez, J. C. Licona, C. Cerón, R. Thomas, P. Maas, M. Silveira, T. W. Henkel, J. Stropp, M. R. Paredes, K. G. Dexter, D. Daly, T. R. Baker, I. Huamantupa-Chuquimaco, W. Milliken, T. Pennington, J. S. Tello, J. L. M. Pena, C. A. Peres, B. Klitgaard, A. Fuentes, M. R. Silman, A. Di Fiore, P. von Hildebrand, J. Chave, T. R. van Andel, R. R. Hilário, J. F. Phillips, G. Rivas-Torres, J. C. Noronha, A. Prieto, T. Gonzales, R. de Sá Carpanedo, G. P. G. Gonzales, R. Z. Gómez, D. de Jesus Rodrigues, E. L. Zent, A. R. Ruschel, V. A. Vos, É. Fonty, A. B. Junqueira, H. P. D. Doza, B. Hoffman, S. Zent, E. M. Barbosa, Y. Malhi, L. C. de Matos Bonates, I. P. de Andrade Miranda, N. Silva, F. R. Barbosa, C. I. A. Vela, L. F. M. Pinto, A. Rudas, B. W. Albuquerque, M. N. Umaña, Y. A. Carrero Márquez, G. van der Heijden, K. R. Young, M. Tirado, D. F. Correa, R. Sierra, J. B. P. Costa, M. Rocha, E. Vilanova Torre, O. Wang, A. A. Oliveira, M. Kalamandeen, C. Vriesendorp, H. Ramirez-Angulo, M. Holmgren, M. T. Nascimento, D. Galbraith, B. M. Flores, V. V. Scudeller, A. Cano, M. A. Ahuite Reategui, I. Mesones, C. Baider, C. Mendoza, R. Zagt, L. E. Urrego Giraldo, C. Ferreira, D. Villarroel, R. Linares-Palomino, W. Farfan-Rios, W. Farfan-Rios, L. F. Casas, S. Cárdenas, H. Balslev, A. Torres-Lezama, M. N. Alexiades, K. Garcia-Cabrera, L. Valenzuela Gamarra, E. H. Valderrama Sandoval, F. Ramirez Arevalo, L. Hernandez, A. F. Sampaio, S. Pansini, W. Palacios Cuenca, E. A. de Oliveira, D. Pauletto, A. Levesley, K. Melgaço, and G. Pickavance (2020). Biased-corrected richness estimates for the Amazonian tree flora. *Scientific Reports 10*(1), 10130.

Triantis, K. A., F. Guilhaumon, and R. J. Whittaker (2012). The island species-area relationship: Biology and statistics. *Journal of Biogeography 39*(2), 215–231.

Volkov, I., J. R. Banavar, S. P. Hubbell, and A. Maritan (2003). Neutral theory and relative species abundance in ecology. *Nature 424*(6952), 1035–1037.

Williamson, M., K. J. Gaston, and W. M. Lonsdale

(2001). The species-area relationship does not have an asymptote! *Journal of Biogeography 28*(7), 827–830.

Xu, H., S. Liu, Y. Li, R. Zang, and F. He (2012). Assessing non-parametric and area-based methods for estimating regional species richness. *Journal of Vegetation Science 23*(6), 1006–1012.

Zanaga, D., R. Van De Kerchove, W. De Keersmaecker, N. Souverijns, C. Brockmann, R. Quast, J. Wevers, A. Grosu, A. Paccini, S. Vergnaud, O. Cartus, M. Santoro, S. Fritz, I. Georgieva, M. Lesiv, S. Carter, M. Herold, L. Li, N.-E. Tsendbazar, F. Ramoino, and O. Arino (2021, October). ESA WorldCover 10 m 2020 v100.