




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

## Research Article

Eric Marcon<sup>1,2</sup> , Ariane Mirabel<sup>2,3</sup>, Jean-François Molino<sup>4</sup> and Daniel Sabatier<sup>4</sup>

**Cite this article:** Marcon E, Mirabel A, Molino J-F, and Sabatier D (2024). Estimation of the number of tree species in French Guiana by extrapolation of permanent plots richness. *Journal of Tropical Ecology*. **40**(e11), 1–10. doi: <https://doi.org/10.1017/S0266467424000099>

Received: 8 September 2022  
Revised: 14 March 2024  
Accepted: 17 March 2024

### Keywords:

Biodiversity; species richness; tropical rainforest; biodiversity estimation; hyperdominance; self-similarity

### Corresponding author:

Eric Marcon;  
Email: [eric.marcon@agroparistech.fr](mailto:eric.marcon@agroparistech.fr)

<sup>1</sup>AgroParisTech, AMAP, CIRAD, CNRS, INRAE, IRD, Univ Montpellier, Montpellier, France; <sup>2</sup>UMR EcoFoG, AgroParisTech, CIRAD, CNRS, INRAE, Univ Antilles, Univ Guyane, Kourou, France; <sup>3</sup>Department of Biology, Western University, London, ON, Canada and <sup>4</sup>AMAP, IRD, CIRAD, CNRS, INRAE, Univ Montpellier, Montpellier, France

### 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 occurrence 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<sup>2</sup>). We evaluated the number of tree species around 2200 over the territory.

## 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 (ForestPlots.net *et al.* 2021) and a set of methods to apply to their data.

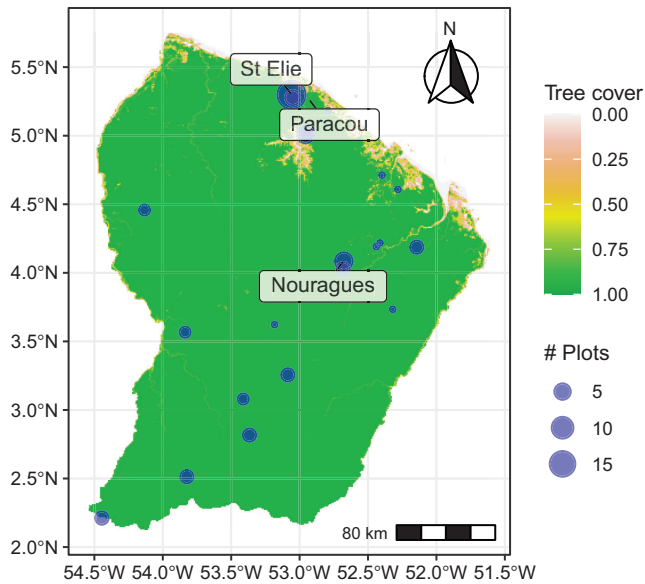
At the local scale, the number of species is related to the sampling effort by species-accumulation curves (Gotelli & Colwell 2001). The number of sampled species is a matter of well-known statistics based on independent and identically distributed (iid) samples, and estimators of the total number of species of a homogeneous community are available, among which the best known are Chao’s (Chao 1984) and the jackknife (Burnham & 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; ter Steege *et al.* 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.* 1999a) that implies the power-law



**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).

relationship of Arrhenius (1921) and provides a technique to evaluate its parameters (Harte *et al.* 1999b), 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.

## Methods

### Data

To apply the methods detailed below, a large enough inventory is necessary along with a set of small, widely 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 & 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<sup>2</sup> 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<sup>2</sup> 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<sup>2</sup> area.

### Self-similarity

Self-similarity (Harte *et al.* 1999a) 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 subarea, given that it is present in the area containing it, is constant: it does not depend either on the observation scale or on the species considered.

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

$$S(A) = cA^z, \quad (2.1)$$

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 (García Martín & Goldenfeld 2006; Williamson *et al.* 2001).

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.* (1999b) 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

$$\chi = 2(S_1 \cap S_2)/(S_1 + S_2), \quad (2.2)$$

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.* 1999b) that can be estimated by the linear model

$$\log(\chi) \sim \log(d). \quad (2.3)$$

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 (2.3) 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  $\chi$  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 2023) v. 4.3.1.

### 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 & Colwell 2001). It is driven by statistical models that address incomplete sampling (Béguinot 2015; Shen *et al.* 2003). 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 & Overton 1978) apply.

The Chao1 estimator is

$$\hat{S}_{\text{Chao1}} = s_{\text{obs}} + \frac{(n-1)f_1^2}{2nf_2}, \quad (2.4)$$

where  $s_{\text{obs}}$  is the number of observed species,  $n$  is the sample size,  $f_1$  and  $f_2$  are 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  $f_1, f_2, \dots, f_k$ , 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 & Overton 1978). For large  $n$ , the jackknife estimator of order 3, used below, is

$$\hat{S}_{\text{Jack3}} = s_{\text{obs}} + 3f_1 - 3f_2 + f_3. \quad (2.5)$$

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 are 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 (Burnham & Overton 1978; Chao 1987).

### 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 & Jost (2015), with the *entropart* package (Marcon & 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.

### Universal species–area relationship

Harte *et al.* (2008) derived a universal species–area relationship based on the maximum entropy theory. 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, which may be a single 1-ha plot or one of our large inventories, the area can be doubled until the target size is reached.

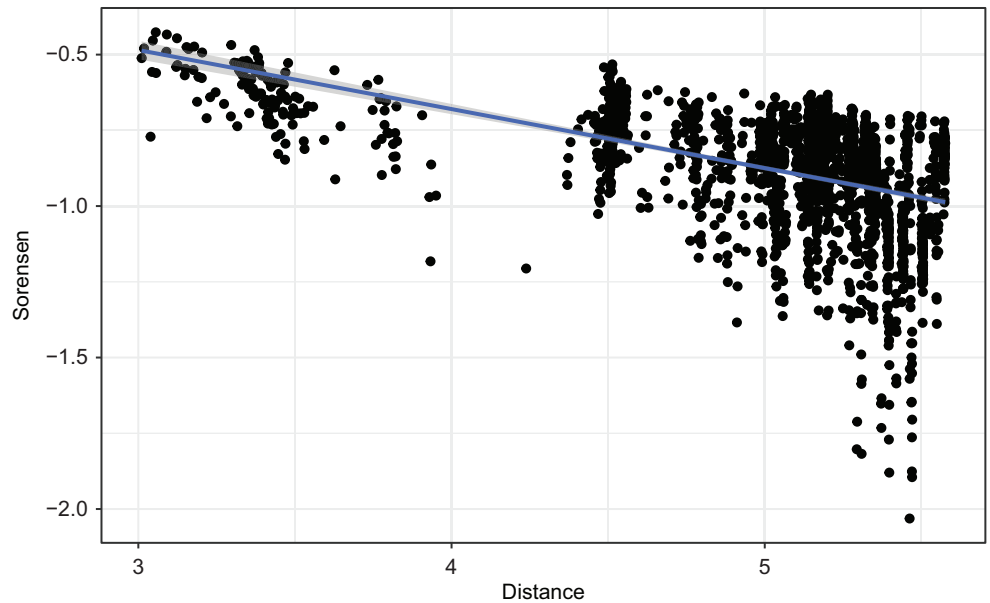
The number of trees per hectare is estimated from the Guyadiv network to obtain a single starting point rather than a different one for each plot. To be consistent with the model, the geometric mean is applied: its logarithm equals the average logarithm of the number of trees in all 1-ha plots.

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.

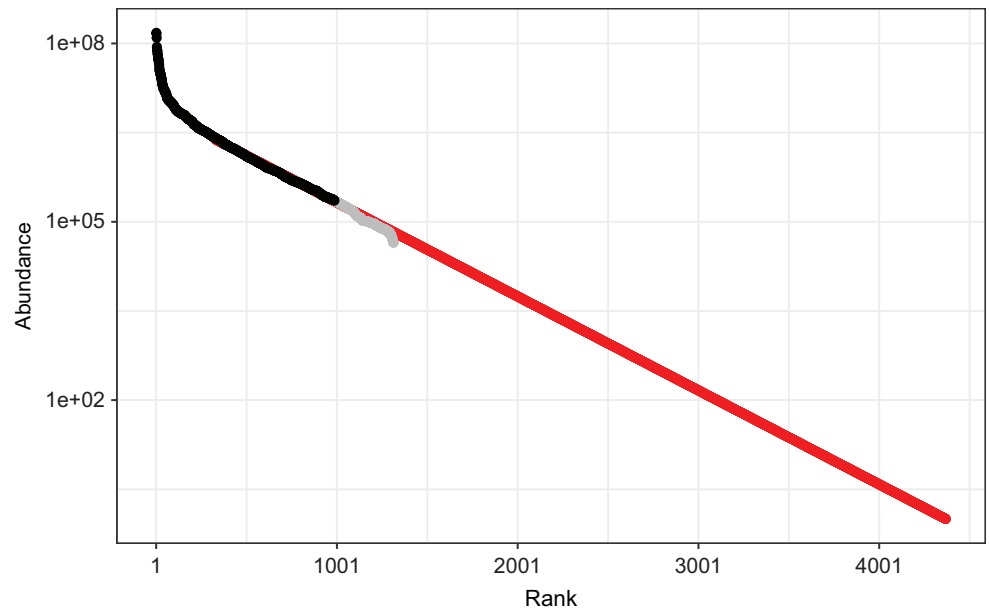
## Results

### 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<sup>2</sup> inventory) are shown, and the



**Figure 2.** Relation between Sørensen's similarity and the distance between pairs of plots. Both axes are in base-10 logarithms, and 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  $-2z$ .



**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) are 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.

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 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.

### 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  $100 \times 100$  km cells is 1643. Its confidence interval is between 1436 and 1564 at the 5% risk level.

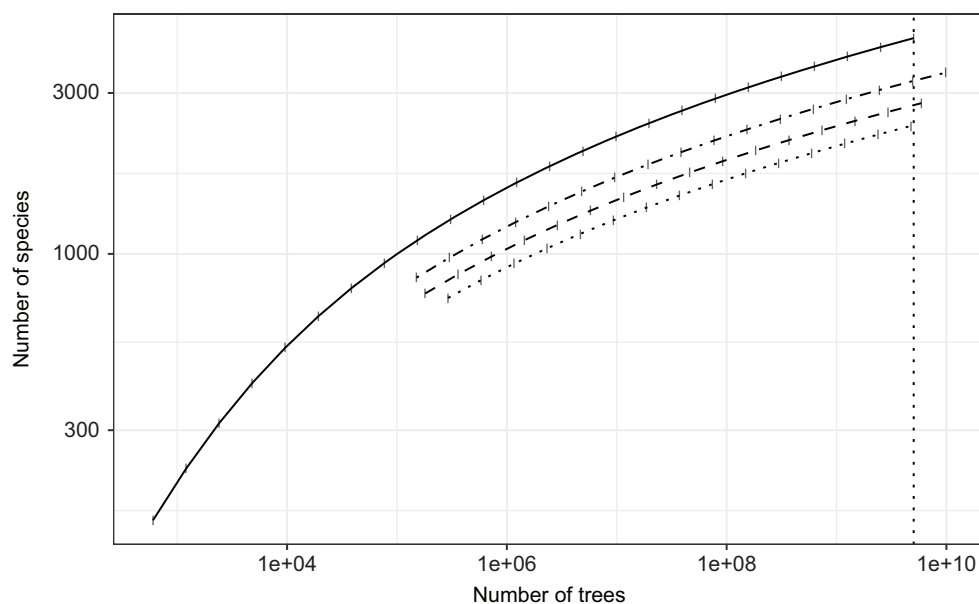
### Log-series extrapolation

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.





**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.

### 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 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 & Burk 1934)  $\frac{1}{\log S} \sim \frac{1}{\log n}$ , where  $S$  is the number of species and  $n$  is 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.

### Summary

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

### Discussion

#### 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 & Colwell 2001). It is driven by

statistical models that address incomplete sampling (Béguinot 2015; Shen *et al.* 2003). Local SACs have 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 (May 1975; Preston 1948, 1962). 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 & 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. (2.1), 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  $S(A) = a \ln A + c_f$  if the number of trees is large and is propor-

**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

tional to the area (Engen 1977). There is no theoretical support to apply Gleason's model at the regional scale (García Martín & 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. It follows a log-series distribution (Hubbell 2001) with Fisher's  $\alpha$  equal to  $\theta$ , known as the fundamental biodiversity number. The log-series does not apply to the local or regional scale: the empirical estimates of Fisher's  $\alpha$  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  $\alpha$  is a constant. At the regional scale, Fisher's  $\alpha$  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 & 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 the 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.

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

The power law is equivalent to self-similarity (Harte *et al.* 1999a), 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 & 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, which 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  $c$ , which 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.

### 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 & 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 × 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.

### 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 S1 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.

### 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 are 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 are 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.

**Supplementary material.** The supplementary material that contains the appendix of this article can be found at <https://doi.org/10.1017/S0266467424000099>

**Financial support.** This work benefited from an 'Investissement d'Avenir' grant managed by the Agence Nationale de la Recherche (LABEX CEBA, ref. ANR-10-LBX-25).

**Competing interests.** The authors declare none.

### References

- Arrhenius O (1921) Species and area. *Journal of Ecology* 9, 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, 1–9.
- Bonal D, Bosc A, Ponton S, Goret J-Y, Burban B, Gross P, Bonnefond J-M, Elbers J, Longdoz B, Epron D, Guehl J-M and Granier A (2008) Impact of severe dry season on net ecosystem exchange in the Neotropical rainforest of French Guiana. *Global Change Biology* 14, 1917–1933.
- Bongers F, Charles-Dominique P, Forget P-M and Théry M (2001) *Nouragues: Dynamics and Plant-Animal Interactions in a Neotropical Rainforest*. Dordrecht: Springer Science & Business Media.
- Burnham KP and Overton WS (1978) Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* 65, 625–633.
- Cazzolla Gatti R, Reich PB, Gamarra JGP, Crowther T, Hui C, Morera A, Bastin J-F, de-Miguel S, Nabuurs G-J, Svenning J-C, Serra-Diaz JM, Merow C, Enquist B, Kamenetsky M, Lee J, Zhu J, Fang J, Jacobs DF, Pijanowski B, Banerjee A, Giaquinto RA, Alberti G, Almeyda Zambrano AM, Alvarez-Davila E, Araujo-Murakami A, Avitabile V, Aymard GA, Balazy R, Baraloto C, Barroso JG, Bastian ML, Birnbaum P, Bitariho R, Bogaert J, Bongers F, Bouriaud O, Brancalion PHS, Brearley FQ, Broadbent EN, Bussotti F, Castro da Silva W, César RG, Češljár G, Chama Moscoso V, Chen HYH, Cienciala E, Clark CJ, Coomes DA, Dayanandan S, Decuyper M, Dee LE, Del Aguila Pasquel J, Derroire G, Djuikouo MNK, Van Do T, Dolezal J, Đorđević ID, Engel J, Fayle TM, Feldpausch TR, Fridman JK, Harris DJ, Hemp A, Hengeveld G, Herault B, Herold M, Ibanez T, Jagodzinski AM, Jaroszewicz B, Jeffery KJ, Johannsen VK, Jucker T, Kangur A, Karminov VN, Kartawinata K, Kennard DK, Kepfer-Rojas S, Keppel G, Khan ML, Khare PK, Kileen TJ, Kim HS, Korjus H, Kumar A, Kumar A, Laermann D, Labrière N, Lang M, Lewis SL, Lukina N, Maitner BS, Malhi Y, Marshall AR, Martynenko OV, Monteagudo Mendoza AL, Ontikov PV, Ortíz-Malavasi E, Pallqui Camacho NC, Paquette A, Park M, Parthasarathy N, Peri PL, Petronelli P, Pfautsch S, Phillips OL, Picard N, Piotto D, Poorter L, Poulsen JR, Pretzsch H,



- Ramírez-Angulo H, Restrepo Correa Z, Rodeghiero M, Rojas González RDP, Rolim SG, Rovero F, Rutishauser E, Saikia P, Salas-Eljatib C, Schepaschenko D, Scherer-Lorenzen M, Šebeň V, Silveira M, Slik F, Sonké B, Souza AF, Stereńczak KJ, Svoboda M, Taedoumg H, Tchebakova N, Terborgh J, Tikhonova E, Torres-Lezama A, van der Plas F, Vásquez R, Viana H, Vibrans AC, Vilanova E, Vos VA, Wang H-F, Westerlund B, White LJ, Wisser SK, Zawila-Niedzwiecki T, Zemagho L, Zhu Z-X, Zo-Bi IC and Liang J (2022) The number of tree species on Earth. *Proceedings of the National Academy of Sciences* **119**, e2115329119.
- Chao A (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**, 265–270.
- Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791.
- Chao A and Jost L (2015) Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution* **6**, 873–882.
- Chao A, Wang Y-T and Jost L (2013) Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution* **4**, 1091–1100.
- Colwell RK and Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **345**, 101–118.
- Condit R, Hubbell SP, Lafrankie JV, Sukumar R, Manokaran N, Foster RB, Ashton PS (1996) Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. *Journal of Ecology* **84**, 549–562.
- Connor EF and McCoy ED (1979) The statistics and biology of the species-area relationship. *The American Naturalist* **113**, 791–833.
- Dengler J (2009) Which function describes the species-area relationship best? A review and empirical evaluation. *Journal of Biogeography* **36**, 728–744.
- Duque A, Muller-Landau HC, Valencia R, Cardenas D, Davies S, De Oliveira A, Pérez AJ, Romero-Saltos H and Vicentini A (2017) Insights into regional patterns of Amazonian forest structure, diversity, and dominance from three large terra-firme forest dynamics plots. *Biodiversity and Conservation* **26**, 669–686.
- Engel J (2015) Plot networks & teams. ATDN tree morphospecies website, <http://atdnmorphospecies.myspecies.info/node/781>. Accessed 5 March 2024.
- Engen S (1977) Exponential and logarithmic species-area curves. *The American Naturalist* **111**, 591–594.
- Fisher RA, Corbet AS and Williams CB (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**, 4258.
- ForestPlots.net, Blundo C, Carilla J, Grau R, Malizia A, Malizia L, Osinaga-Acosta O, Bird M, Bradford M, Catchpole D, Ford A, Graham A, Hilbert D, Kemp J, Laurance S, Laurance W, Ishida FY, Marshall A, Waite C, Woell H, Bastin J-F, Bauters M, Beeckman H, Boeckx P, Bogaert J, De Canniere C, De Haulleville T, Doucet J-L, Hardy O, Hubau W, Kearsley E, Verbeeck H, Vleminckx J, Brewer SW, Alarcón A, Araujo-Murakami A, Arets E, Arroyo L, Chavez E, Fredericksen T, Villaroel RG, Sibauty GG, Killeen T, Licona JC, Lleigue J, Mendoza C, Murakami S, Gutierrez AP, Pardo G, Peña-Claros M, Poorter L, Toledo M, Cayo JV, Viscarra LJ, Vos V, Ahumada J, Almeida E, Almeida J, De Oliveira EA, Da Cruz WA, De Oliveira AA, Carvalho FA, Obermuller FA, Andrade A, Carvalho FA, Vieira SA, Aquino AC, Aragão L, Araújo AC, Assis MA, Gomes JAMA, Baccaro F, De Camargo PB, Barni P, Barroso J, Bernacci LC, Bordin K, De Medeiros MB, Broggio I, Camargo JL, Cardoso D, Carniello MA, Rochelle ALC, Castilho C, Castro AAJF, Castro W, Ribeiro SC, Costa F, De Oliveira RC, Coutinho I, Cunha J, Da Costa L, Da Costa Ferreira L, Da Costa Silva R, Da Graça Zacarias Simbine M, De Andrade Kamimura V, De Lima HC, De Oliveira Melo L, De Queiroz L, De Sousa Lima JR, Do Espírito Santo M, Domingues T, Dos Santos Prestes NC, Carneiro SES, Elias F, Eliseu G, Emilio T, Farrapo CL, Fernandes L, Ferreira G, Ferreira J, Ferreira L, Ferreira S, Simon MF, Freitas MA, García QS, Manzatto AG, Graça P, Guilherme F, Hase E, Higuchi N, Iguatemy M, Barbosa RI, Jaramillo M, Joly C, Klipel J, Do Amaral IL, Levis C, Lima AS, Dan ML, Lopes A, Medeiros H, Magnusson WE, Dos Santos RM, Marimon B, Junior BHM, Grillo RMM, Martinelli L, Reis SM, Medeiros S, Meira-Junior M, Metzker T, Morandi P, Do Nascimento NM, Moura M, Müller SC, Nagy L, Nascimento H, Nascimento M, Lima AN, De Araújo RO, Silva JO, Pansonato M, Sabino GP, De Abreu KMP, Rodrigues PJFP, Piedade M, Rodrigues D, Rodrigues Pinto JR, Quesada C, Ramos E, Ramos R, Rodrigues P, De Sousa TR, Salomão R, Santana F, Scaranello M, Bergamin RS, Schiatti J, Schöngart J, Schwartz G, Silva N, Silveira M, Seixas CS, Simbine M, Souza AC, Souza P, Souza R, Sposito T, Junior ES, Do Vale JD, Vieira ICG, Villela D, Vital M, Xaud H, Zanini K, Zartman CE, Ideris NKH, Metali FBH, Salim KA, Saparudin MS, Serudin RM, Sukri RS, Begne S, Chuyong G, Djuikouo MN, Gonmadje C, Simo-Droissart M, Sonké B, Taedoumg H, Zemagho L, Thomas S, Baya F, Saiz G, Espejo JS, Chen D, Hamilton A, Li Y, Luo T, Niu S, Xu H, Zhou Z, Álvarez-Dávila E, Escobar JCA, Arellano-Peña H, Duarte JC, Calderón J, Bravo LMC, Cuadrado B, Cuadros H, Duque A, Duque LF, Espinosa SM, Franke-Ante R, García H, Gómez A, González-MR, Idárraga-Piedrahíta Á, Jimenez E, Jurado R, Oviedo WL, López-Camacho R, Cruz OAM, Polo IM, Paky E, Pérez K, Pijachi A, Pizano C, Prieto A, Ramos L, Correa ZR, Richardson J, Rodríguez E, Rodríguez MGM, Rudas A, Stevenson P, Chudomelová M, Dancak M, Hédli R, Lhota S, Svatek M, Mukinzi J, Ewango C, Hart T, Yakusu EK, Lisingo J, Makana J-R, Mbayu F, Toirambe B, Mukendi JT, Kivist L, Nebel G, Báez S, Céron C, Griffith DM, Andino JEG, Neill D, Palacios W, Peñuela-Mora MC, Rivas-Torres G, Villa G, Demissie S, Gole T, Gonfa T, Ruokolainen K, Baisie M, Bénédet F, Betian W, Bezard V, Bonal D, Chave J, Droissart V, Gourlet-Fleury S, Hladik A, Labrière N, Naisso P, Réjou-Méchain M, Sist P, Blanc L, Burban B, Derroire G, Dourdain A, Stahl C, Bengone NN, Chezeaux E, Ondo FE, Medjibe V, Mihindou V, White L, Culmsee H, Rangel CD, Horna V, Wittmann F, Adu-Bredu S, Affum-Baffoe K, Foli E, Balinga M, Roopsind A, Singh J, Thomas R, Zagt R, Murthy IK, Kartawinata K, Mirmanto E, Priyadi H, Samsodin I, Sunderland T, Yassir I, Rovero F, Vinceti B, Hérault B, Aiba S-I, Kitayama K, Daniels A, Tuagben D, Woods JT, Fitriadi M, Karolus A, Khoon KL, Majalap N, Maycock C, Nilus R, Tan S, Siteo A, Coronado GI, Ojo L, De Assis R, Poulsen AD, Sheil D, Pezo KA, Verde HB, Moscoso VC, Oroche JCC, Valverde FC, Medina MC, Cardozo ND, De Rutte Corzo J, Del Aguila Pasquel J, Llampazo GF, Freitas L, Cabrera DG, Villacorta RG, Cabrera KG, Soria DG, Saboya LG, Rios JMG, Pizango GH, Coronado EH, Huamantupa-Chuquimaco I, Huasco WH, Aedo YTH, Peña JLM, Mendoza AM, Rodríguez VM, Vargas PN, Ramos SCP, Camacho NP, Cruz AP, Arevalo FR, Huaymacari JR, Rodríguez CR, Paredes MAR, Bayona LR, Del Pilar Rojas Gonzales R, Peña MER, Revilla NS, Shareva YCS, Trujillo RT, Gamarra LV, Martínez RV, Arenas JV, Amani C, Ifo SA, Bocko Y, Boundja P, Ekoungoulou R, Hockemba M, Nzala D, Fofanah A, Taylor D, Bañares-de Dios G, Cayuela L, La Cerda ÍG, Macía M, Stropp J, Playfair M, Wortel V, Gardner T, Muscarella R, Priyadi H, Rutishauser E, Chao K-J, Munishi P, Bánki O, Bongers F, Boot R, Fredriksson G, Reitsma J, Ter Steege H, Van Andel T, Van De Meer P, Van Der Hout P, Van Nieuwstadt M, Van Ulft B, Veenendaal E, Vernimmen R, Zuidema P, Zwerts J, Akite P, Bitariho R, Chapman C, Gerald E, Leal M, Mucunguzi P, Abernethy K, Alexiades M, Baker TR, Banda K, Banin L, Barlow J, Bennett A, Berenguer E, Berry N, Bird NM, Blackburn GA, Brearley F, Brienen R, Burslem D, Carvalho L, Cho P, Coelho F, Collins M, Coomes D, Cuni-Sanchez A, Dargie G, Dexter K, Disney M, Draper F, Duann M, Esquivel-Muelbert A, Ewers R, Fadrique B, Fauset S, Feldpausch TR, França F, Galbraith D, Gilpin M, Gloor E, Grace J, Hamer K, Harris D, Jeffery K, Jucker T, Kalamandeen M, Klitgaard B, Levesley A, Lewis SL, Lindsell J, Lopez-Gonzalez G, Lovett J, Malhi Y, Marthews T, McIntosh E, Melgaço K, Milliken W, Mitchard E, Moonlight P, Moore S, Morel A, Peacock J, Peh KS-H, Pendry C, Pennington RT, De Oliveira Pereira L, Peres C, Phillips OL, Pickavance G, Pugh T, Qie L, Riutta T, Roucoux K, Ryan C, Sarkinen T, Valeria CS, Spracklen D, Stas S, Sullivan M, Swaine M, Talbot J, Taplin J, Van Der Heijden G, Vedovato L, Willcock S, Williams M, Alves L, Loayza PA, Arellano G, Asa C, Ashton P, Asner G, Brncic T, Brown F, Burnham R, Clark C, Comiskey J, Damasco G, Davies S, Di Fiore T, Erwin T, Farfan-Rios W, Hall J, Kenfack D, Lovejoy T, Martin R, Montiel OM, Pipoly J, Pitman N, Poulsen J,



- Primack R, Silman M, Steininger M, Swamy V, Terborgh J, Thomas D, Umunay P, Uriarte M, Torre EV, Wang O, Young K, Aymard CGA, Hernández L, Fernández RH, Ramírez-Angulo H, Salcedo P, Sanoja E, Serrano J, Torres-Lezama A, Le TC, Le TT and Tran HD (2021) Taking the pulse of Earth's tropical forests using networks of highly distributed plots. *Biological Conservation* **260**, 108849.
- García Martín H and Goldenfeld N (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**, 10310–10315.
- Gleason HA (1922) On the relation between species and area. *Ecology* **3**, 158–162.
- Gonzalez S, Bilot-Guérin V, Delprete PG, Geniez C, Molino J-F and Smock J-L (2022) L'herbier IRD de Guyane, <https://herbier-guyane.ird.fr/>. Accessed 5 March 2024.
- Good IJ (1953) The population frequency of species and the estimation of population parameters. *Biometrika* **40**, 237–264.
- Gotelli NJ and Colwell RK (2001) Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4**, 379–391.
- Gourlet-Fleury S, Guehl JM and Laroussinie O (2004) *Ecology & Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana*. Paris: Elsevier.
- Grabchak M, Marcon E, Lang G and Zhang Z (2017) The generalized Simpson's entropy is a measure of biodiversity. *Plos One* **12**, e0173305.
- Grilli J, Azaele S, Banavar JR and Maritan A (2012) Spatial aggregation and the species-area relationship across scales. *Journal of Theoretical Biology* **313**, 87–97.
- Guité S, Péliissier R, Brunaux O, Jaouen G and Sabatier D (2015) Geomorphological landscape features explain floristic patterns in French Guiana rainforest. *Biodiversity and Conservation* **24**, 1215–1237.
- Harte J, Kinzig A and Green J (1999a) Self-similarity in the distribution and abundance of species. *Science* **284**, 334–336.
- Harte J, Mccarthy S, Taylor K, Kinzig A and Fischer ML (1999b) Estimating species-area relationships from scale plot to landscape data using species spatial-turnover. *Oikos* **86**, 45–54.
- Harte J, Smith AB and Storch D (2009) Biodiversity scales from plots to biomes with a universal species-area curve. *Ecology Letters* **12**, 789–797.
- Harte J, Zillio T, Conlisk E and Smith AB (2008) Maximum entropy and the state-variable approach to macroecology. *Ecology* **89**, 2700–2711.
- Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, New Jersey: Princeton University Press.
- Jaouen G, Dourdain A and Derroire G (2021) Guyafor Data Dictionary, CIRAD Dataverse. doi: [10.18167/DVN1/B8FHHA](https://doi.org/10.18167/DVN1/B8FHHA).
- Krishnamani R, Kumar A and Harte J (2004) Estimating species richness at large spatial scales using data from small discrete plots. *Ecography* **27**, 637–642.
- Lineweaver H and Burk D (1934) The determination of enzyme dissociation constants. *Journal of the American Chemical Society* **56**, 658–666.
- MacArthur RH and Wilson EO (1967) The theory of island biogeography. In *Monographs in population biology*, Vol. 1, Princeton University Press.
- Marcon E and Hérault B (2015) Entropart, an R package to measure and partition diversity. *Journal of Statistical Software* **67**, 1–26.
- May RM (1975) Patterns of species abundance and diversity. In M. L. Cody & J. M. Diamond, eds, *Ecology and Evolution of Communities*, Cambridge, Massachusetts: Harvard University Press, pp. 81–120.
- McGill BJ (2003) A test of the unified neutral theory of biodiversity. *Nature* **422**(6934), 881–885.
- Mirabel A, Hérault B and Marcon E (2020) Diverging taxonomic and functional trajectories following disturbance in a Neotropical forest. *Science of The Total Environment* **720**, 137397.
- Mirabel A, Marcon E and Hérault B (2021) 30 Years of postdisturbance recruitment in a Neotropical forest. *Ecology and Evolution* **11**, 14448–14458.
- Molino J-F and Sabatier D (2001) Tree diversity in tropical rain forests: a validation of the intermediate disturbance hypothesis. *Science* **294**, 1702–1704.
- Molino J-F, Sabatier D, Grenand P, Engel J, Frame D, Delprete PG, Fleury M, Odonne G, Davy D, Lucas EJ and Martin CA (2022) An annotated checklist of the tree species of French Guiana, including vernacular nomenclature. *Adansonia* **44**, 345–903.
- Picard N, Karembe M and Birnbaum P (2004) Species-area curve and spatial pattern. *Ecoscience* **11**, 45–54.
- Plotkin JB, Potts MD, Leslie N, Manokaran N, LaFrankie JV and Ashton PS (2000) Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology* **207**, 81–99.
- Preston FW (1948) The commonness, and rarity, of species. *Ecology* **29**, 254–283.
- Preston FW (1962) The canonical distribution of commonness and rarity: part I. *Ecology* **43**, 185–215.
- R Core Team (2023) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Shen T-J, Chao A and Lin C-F (2003) Predicting the number of new species in a further taxonomic sampling. *Ecology* **84**, 798–804.
- Slik JWF, Arroyo-Rodríguez V, Aiba S-I, Alvarez-Loayza P, Alves LF, Ashton PS, Balvanera P, Bastian ML, Bellingham PJ, van den Berg E, Bernacci L, da Conceição Bispo P, Blanc L, Böhning-Gaese K, Boeckx P, Bongers F, Boyle B, Bradford M, Brearley FQ, Breuer-Ndoundou Hockemba M, Bunyavejchewin S, Calderado Leal Matos D, Castillo-Santiago M, Catharino ELM, Chai S-L, Chen Y, Colwell RK, Robin CL, Clark CJ, Clark DB, Clark DA, Culmsee H, Damas K, Dattaraja HS, Dauby G, Davidar P, DeWalt SJ, Doucet J-L, Duque A, Durigan G, Eichhorn KAO, Eisenlohr PV, Eler E, EWango C, Farwig N, Feeley KJ, Ferreira L, Field R, de Oliveira Filho AT, Fletcher C, Forshed O, Franco G, Fredriksson G, Gillespie T, Gillet J-F, Amarnath G, Griffith DM, Grogan J, Gunatilleke N, Harris D, Harrison R, Hector A, Homeier J, Imai N, Itoh A, Jansen PA, Joly CA, de Jong BHJ, Kartawinata K, Kearsley E, Kelly DL, Kenfack D, Kessler M, Kitayama K, Kooyman R, Larney E, Laumonier Y, Laurance S, Laurance WF, Lawes MJ, Amaral IL, Letcher SG, Lindsell J, Lu X, Mansor A, Marjokorpi A, Martin EH, Meilby H, Melo FPL, Metcalfe DJ, Medjibe VP, Metzger JP, Millet J, Mohandass D, Montero JC, de Morisson Valeriano M, Mugerwa B, Nagamasu H, Nilus R, Ochoa-Gaona S, Onrizal, Page N, Parolin P, Parren M, Parthasarathy N, Paudel E, Permana A, Piedade MTF, Pitman NCA, Poorter L, Poulsen AD, Poulsen J, Powers J, Prasad RC, Puyravaud J-P, Razafimahaimodison J-C, Reitsma J, dos Santos JR, Roberto Spironello W, Romero-Saltos H, Rovero F, Rozak AH, Ruokolainen K, Rutishauser E, Saiter F, Saner P, Santos BA, Santos F, Sarker SK, Satdichan M, Schmitt CB, Schöngart J, Schulze M, Suganuma MS, Sheil D, da Silva Pinheiro E, Sist P, Stevart T, Sukumar R, Sun I-F, Sunderland T, Suresh HS, Suzuki E, Tabarelli M, Tang J, Targhetta N, Theilade I, Thomas DW, Tchouto P, Hurtado J, Valencia R, van Valkenburg JLCH, Van Do T, Vasquez R, Verbeeck H, Adekunle V, Vieira SA, Webb CO, Whitfield T, Wich SA, Williams J, Wittmann F, Wöll H, Yang X, Adou Yao CY, Yap SL, Yoneda T, Zahawi RA, Zakaria R, Zang R, de Assis RL, Garcia Luize B and Venticinque EM (2015) An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 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**, 1–34.
- ter Steege H, Pitman NCA, Sabatier D, Baraloto C, Salomão RP, Guevara JE, Phillips OL, Castilho CV, Magnusson WE, Molino J-F, Monteagudo A, Núñez Vargas P, Montero JC, Feldpausch TR, Coronado ENH, Killeen TJ, Mostacedo B, Vasquez R, Assis RL, Terborgh J, Wittmann F, Andrade ACS, Laurance WF, Laurance SGW, Marimon BS, Marimon B-H, Guimarães Vieira IC, Amaral IL, Brienen R, Castellanos H, Cárdenas López D, Duivenvoorden JF, Mogollón HF, de Almeida Matos FD, Dávila N, Garcia-Villacorta R, Stevenson Diaz PR, Costa F, Emilio T, Levis C, Schiatti J, Souza P, Alonso A, Dallmeier F, Montoya AJD, Fernandez Piedade MT, Araujo-Murakami A, Arroyo L, Gribel R, Fine PVA, Peres CA, Toledo M, Aymard C. GA, Baker T, Cerón C, Engel J, Henkel TW, Maas P, Petronelli P, Stropp J, Zartman CE, Daly D, Neill D, Silveira M, Paredes MR, Chave J, de Andrade Lima Filho D, Jørgensen PM, Fuentes A, Schöngart J, Cornejo Valverde F, Di Fiore A, Jimenez EM,

- Peñuela-Mora MC, Phillips JF, Rivas G, van Andel TR, von Hildebrand P, Hoffman B, Zent EL, Malhi Y, Prieto A, Rudas A, Ruschell AR, Silva N, Vos V, Zent S, Oliveira AA, Schutz AC, Gonzales T, Trindade Nascimento M, Ramirez-Angulo H, Sierra R, Tirado M, Umaña Medina MN, van der Heijden G, Vela CIA, Vilanova Torre E, Vriesendorp C, Wang O, Young KR, Baider C, Balslev H, Ferreira C, Mesones I, Torres-Lezama A, Urrego Giraldo LE, Zagt R, Alexiades MN, Hernandez L, Huamantupa-Chuquimaco I, Milliken W, Palacios Cuenca W, Pauletto D, Valderrama Sandoval E, Valenzuela Gamarra L, Dexter KG, Feeley KJ, Lopez-Gonzalez G and Silman MR (2013) Hyperdominance in the amazonian tree flora. *Science* 342, 1243092.
- ter Steege H, Prado PI, de Lima RAF, Pos E, de Souza Coelho L, de Andrade Lima Filho D, Salomão RP, Amaral IL, de Almeida Matos FD, Castilho CV, Phillips OL, Guevara JE, de Jesus Veiga Carim M, Cárdenas López D, Magnusson WE, Wittmann F, Martins MP, Sabatier D, Irumé MV, da Silva Guimarães JR, Molino J-F, Bánki OS, Piedade MTF, Pitman NCA, Ramos JF, Monteagudo Mendoza A, Venticinqué EM, Luize BG, Núñez Vargas P, Silva TSF, de Leão Novo EMM, Reis NFC, Terborgh J, Manzatto AG, Casula KR, Honorio Coronado EN, Montero JC, Duque A, Costa FRC, Castaño Arboleda N, Schöngart J, Zartman CE, Killeen TJ, Marimon BS, Marimon-Junior BH, Vasquez R, Mostacedo B, Demarchi LO, Feldpausch TR, Engel J, Petronelli P, Baraloto C, Assis RL, Castellanos H, Simon MF, de Medeiros MB, Quaresma A, Laurance SGW, Rincón LM, Andrade A, Sousa TR, Camargo JL, Schiatti J, Laurance WF, de Queiroz HL, Nascimento HEM, Lopes MA, de Sousa Farias E, Magalhães JLL, Brienen R, Aymard CGA, Revilla JDC, Vieira ICG, Cintra BBL, Stevenson PR, Feitosa YO, Duivenvoorden JF, Mogollón HF, Araujo-Murakami A, Ferreira LV, Lozada JR, Comiskey JA, de Toledo JJ, Damasco G, Dávila N, Lopes A, García-Villacorta R, Draper F, Vicentini A, Cornejo Valverde F, Lloyd J, Gomes VHF, Neill D, Alonso A, Dallmeier F, de Souza FC, Gribel R, Arroyo L, Carvalho FA, de Aguiar DPP, do Amaral DD, Pansonato MP, Feeley KJ, Berenguer E, Fine PVA, Guedes MC, Barlow J, Ferreira J, Villa B, Peñuela Mora MC, Jimenez EM, Licona JC, Cerón C, Thomas R, Maas P, Silveira M, Henkel TW, Stropp J, Paredes MR, Dexter KG, Daly D, Baker TR, Huamantupa-Chuquimaco I, Milliken W, Pennington T, Tello JS, Pena JLM, Peres CA, Klitgaard B, Fuentes A, Silman MR, Di Fiore A, von Hildebrand P, Chave J, van Andel TR, Hilário RR, Phillips JF, Rivas-Torres G, Noronha JC, Prieto A, Gonzales T, de Sá Carpanedo R, Gonzales GPG, Gómez RZ, de Jesus Rodrigues D, Zent EL, Ruschel AR, Vos VA, Fonty É, Junqueira AB, Doza HPD, Hoffman B, Zent S, Barbosa EM, Malhi Y, de Matos Bonates LC, de Andrade Miranda IP, Silva N, Barbosa FR, Vela CIA, Pinto LFM, Rudas A, Albuquerque BW, Umaña MN, Carrero Márquez YA, van der Heijden G, Young KR, Tirado M, Correa DF, Sierra R, Costa JBP, Rocha M, Vilanova Torre E, Wang O, Oliveira AA, Kalamandeen M, Vriesendorp C, Ramirez-Angulo H, Holmgren M, Nascimento MT, Galbraith D, Flores BM, Scudeller VV, Cano A, Ahuite Reategui MA, Mesones I, Baider C, Mendoza C, Zagt R, Urrego Giraldo LE, Ferreira C, Villarreal D, Linares-Palomino A, Farfan-Rios W, Farfan-Rios W, Casas LF, Cárdenas S, Balslev H, Torres-Lezama A, Alexiades MN, Garcia-Cabrera K, Valenzuela Gamarra L, Valderrama Sandoval EH, Ramirez Arevalo F, Hernandez L, Sampaio AF, Pansini S, Palacios Cuenca W, de Oliveira EA, Pauletto D, Levesley A, Melgaço K and Pickavance G (2020) Biased-corrected richness estimates for the Amazonian tree flora. *Scientific Reports* 10, 10130.
- Triantis KA, Guilhaumon F and Whittaker RJ (2012) The island species-area relationship: biology and statistics. *Journal of Biogeography* 39, 215–231.
- Volkov I, Banavar JR, Hubbell SP and Maritan A (2003) Neutral theory and relative species abundance in ecology. *Nature* 424, 1035–1037.
- Williamson M, Gaston KJ and Lonsdale WM (2001) The species-area relationship does not have an asymptote! *Journal of Biogeography* 28, 827–830.
- Xu H, Liu S, Li Y, Zang R and He F (2012) Assessing non-parametric and area-based methods for estimating regional species richness. *Journal of Vegetation Science* 23, 1006–1012.
- Zanaga D, Van De Kerchove R, De Keersmaecker W, Souverijns N, Brockmann C, Quast R, Wevers J, Grosu A, Paccini A, Vergnaud S, Cartus O, Santoro M, Fritz S, Georgieva I, Lesiv M, Carter S, Herold M, Li L, Tsendbazar N-E, Ramoimo F and Arino O (2021) ESA WorldCover 10 m 2020 V100, Zenodo. doi: [10.5281/ZENODO.5571936](https://doi.org/10.5281/ZENODO.5571936).