Publication reference
DOI: xxx/xx

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

Eric Marcon[1*]
Ariane Mirabel[2]
Jean-François Molino[3]
Grégoire Vincent[4]
Daniel Sabatier[5]

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. Guyadiv is a network of forest plots installed over the whole territory of French Guiana, where trees over 10 cm DBH are identified. We use its information (1180 species censused in 76 one-hectare plots) to estimate the exponent of the species-area relationship, assuming Arrhenius's power law. We can then extrapolate the number of species from a local, wide inventory (62.5 ha in Paracou research station). We evaluate the number of tree species around 2000 over the territory.

keyword1, keyword2, etc

[1]Department / University
Street address, Zip code, Country.
[2]Department / University
Street address, Zip code, Country.
[3]Department / University
Street address, Zip code, Country.
[4]Department / University
Street address, Zip code, Country.
[5]Department / University
Street address, Zip code, Country.
*: eric.marcon@agroparistech.fr, https://www.company.com

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. In French Guiana, the Guyadiv network consists of close to 250 plots across the whole forest. Based on similar datasets, 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). The methods used in those studies are not appropriate to estimate regional diversity, i.e. at a smaller scale where dispersal limitation is critical. The 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 relatioship 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. We show that the log-series model underlying the work of ter Steege et al. (2013) does not apply at the regional scale.

2. Methods

Self-similarity (Harte et al., 1999) is a property based on scale invariance. Consider a species s 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 h_s. Then, if it is present A_1, the probability to find it in turn in half A_1, denoted A_2, is also h_s, and so on. The probability to find the species in A_n is thus h_s^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, only depends on the relative size of the sub-area (half the parent area here for simplicity): it does not depend on the observation scale.

Arrenhius's power law (Arrhenius, 1921) is a consequence (Harte et al., 1999) of the self-similarity property. 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. This is a classical relation in macroecology, with long empirical and theoretical support (Gárcia Martín and Goldenfeld, 2006).

If z is known, the inventory of a reasonably large area a allows computing $c = a^z/S(a)$. 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 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, Sorensen'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 Sorensen dissimilarity between pair 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. Our data confirm that.

A large enough inventory, provided by a permanent forest facility, is necessary along with a set of small, widely spread forest plots.

The Paracou research station (Gourlet-Fleury et al., 2004) is located at latitude 5°18 N and longitude 52°53 W. It contains six 6.25-ha and one 25-ha plots of primary rainforest summing up to a compact 0.625-ha inventory that can be considered continuous at the scale of French Guiana (80 million hectares).

Decription of Guyadiv here

We take into account the 68 one-hectare plots of the network. They are located in 21 locations that allow a quite good coverage of the variability of the forest in French Guiana (**map here**). 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 Sorensen 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 100 times to obtain a distribution of estimated z values depending on the plots drawn in each location. The empirical mean μ_z and standard devation σ_z of the distribution were calculated. z was estimated as μ_z with a 95% confidence interval equal to $\pm t\sigma_z/\sqrt(100)$.

All analyses were made with R (R Core Team, 2022) v. 4.1.2.

3. Results

The estimated value of z is 0.103 with a 95% confidence interval between 0.087 and 0.123.

The number of species per squared kilometer, c, is 634.

Finally, the estimated number of species is 2031. Taking into account the uncertainty about z, its 95% confidence interval is between 1686 and 2532.

4. Discussion

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) and this paper.

At smaller scales, i.e. inside a single community, the relation between area and number of species is described by species accumulation curves (SAC). It is driven by statistical models that address incomplete sampling. After replacing the sampled area by the number of individuals it contains, well-known estimators of richness such as Chao's or the jackknife apply. Krishnamani et al. (2004) did not have the necessary 100-ha inventory to estimate the number of species at this scale, so they used the self-similarity model to extrapolate small plot data. This lead them to estimate successive scale-dependent values of z with no theoretical support: the model is arguably not valid.

At the scale of the metacommunity, defined in the neutral model of biogeography, the species distribution is in log-series. 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 in appendix 5.2. Its

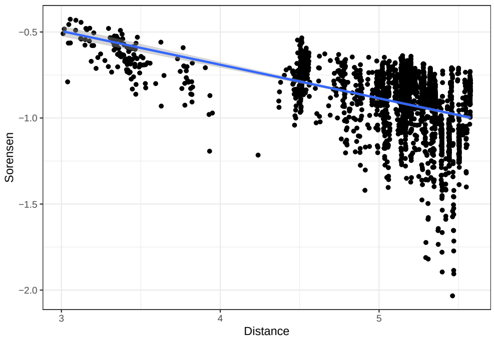

Figure 1. Relation between Sorensen's similarity and the distance between pairs of plots. Both axes are in base-10 logarithms, distances are in meters. Each point is a pair of plots more than 1km apart. A linear model is fitted: the slope of the regression is $-2z$.

estimation is close to 4000 species in French Guiana: a very unlikely result according to the current expert knowledge and the recent checklist (**Reference**). The regional species pool does not follow a log-series distribution because of dispersal limitation. In other words, the regional community is not a sample of the metacommunity: many of the metacommunity's species are not present.

The estimated number of tree species in the 8-million-hectare forest of French Guiana is close to 2000. A lot of approximations were made to obtain this result: - z fit - Paracou is marginal - Guyadiv is not perfect - Species delimitation

Yet it is a very likely estimation according to the current knowledge > JF.

5. Appendix

5.1 Similarity distance decay

The relation between Sorensen's similarity and distance is shown in figure 1. All pairs of plots more than 1 km apart are shown. The estimation of z is not made this way because some locations contain more plots than others so their weight is increased. The technique used in the text of the paper consists of drawing a random plot in each location to estimate z, and repeat this process a large number of times to estimate the expectation of z.

5.2 Log-series estimation of the number of species

Assuming that the plots are samples of a metacommunity that follows a log-series distribution, the rank-abundance curve can be extrapolated (figure 2) following ter Steege et al. (2013).

The estimated number of species according to this model is 3951. This is undoubtedly a severe overestimation, see the discussion section of the paper.

5.3 Hyperdominance

Hyperdominance is a characteristic of many distributions of species. Figure 3 shows the accumulation

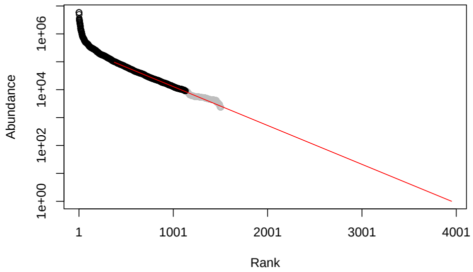

Figure 2. Extrapolation of the rank-abundance curve built from the Guyadiv plots.

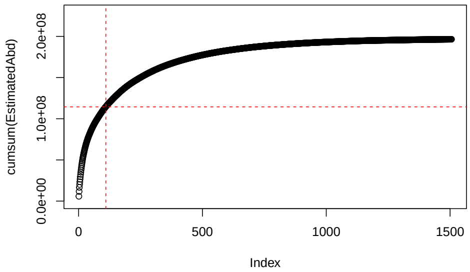

Figure 3. Accumulation of the number of individuals from the most abundant to the rarest species. The horizontal line corresponds to half the individuals. The vertical lines allows reading the corresponding rank of the species.

of individuals from the most adundant to the rarest species.

110 species, i.e. 5% of their estimated number, contain half the number of trees.

References

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

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.

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.

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*(0), 87–97.

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.

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.

R Core Team (2022). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

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. d. A. 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. d. A. 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 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, December). Biased-corrected richness estimates for the Amazonian tree flora. *Scientific Reports 10*(1), 10130.