**

*Plant Ecology & Diversity*, 2018

Vol. 00, Issue 00, 1–xx, http://dx.doi.org/xxxxxxxxxxxxxxxxxxx

Supporting information to the paper

The role of recruitment and dispersal limitation in tree community assembly in Amazonian forests

**APPENDIX S1.** METHODS

***Soil sampling and water table depth***

We collected soil samples at six locations within each plot (0, 50, 100, 150, 200, and 250 m) and four depths (0-5 cm, 5-10 cm, 10-20 cm and 20-30 cm). Samples were dried at room temperature and homogenized within each depth range for physical-chemical analyses, according to standard protocol ([EMBRAPA 1997](#_ENREF_3)). For each plot, soil variables were summarized by averages calculated from across the four depths, including sum of exchangeable bases (Ca+2, K+, Mg+2, Na+), C, N, P, K, organic matter and soil texture.

Between February and May 2012 a piezometer was installed in each plot (depth amplitude 5-16 m) with barometric pressure sensor (Solinst Barologger Edge) and water column sensor (Solinst Levelogger Junior), for continuous measurement in intervals of 30 min, along one year. The data were compensated and scaled ([Solinst 2012](#_ENREF_15)) to express the water table depth.

***Rank Abundance Distributions – RADs***

Many statistical models have been proposed to explain the shape of RADs. The statistical models started with geometric series ([Motomura 1932](#_ENREF_13); [Doi and Mori 2013](#_ENREF_1)). Whittaker highlighted that the RAD of less diverse communities fitted the geometric series ([Whittaker 1965](#_ENREF_19)). He also stressed that log-series ([Fisher *et al.* 1943](#_ENREF_7)) often fits the steep upper and flatter middle parts of RADs, but predicts an excessive number of rare species. Thus, he proposed that lognormal distribution ([Preston 1948](#_ENREF_14)) better explain the relative abundances of species in nature. The Fisher´s log-series predicts that the class with rare species, with only one individual (singletons), is the modal class and includes the most species. The Fisher’s alpha is a parameter associated with this abundance distribution, and independently of the fitness between the observed RAD and the log-series, Fisher’s alpha is broadly used as a diversity measure, because it is not biased by effort sampling ([ter Steege et al. 2003](#_ENREF_16); [ter Steege et al. 2006](#_ENREF_17)). The lognormal distribution probably is the RAD model most applied, and the deviation from this model (lack of fitness to empirical data) have been suggested as indicative of disturbances ([Ugland and Gray 1982](#_ENREF_18)). Some studies found that log-series is a good descriptor of communities in habitats without disturbance, but in a disturbed scenario the lognormal is broken in independent curves, resulting in multimodal abundance distribution ([Ugland and Gray 1982](#_ENREF_18); [Dornelas et al. 2009](#_ENREF_2)). This model was then proposed as a standard method to evaluate the "health" of the ecosystem, in contrast to most of the metrics that require a control reference for comparison with the disturbed community ([Gray 1981](#_ENREF_8); [Kevan *et al.* 1997](#_ENREF_11)). The distributions, parameters (in parenthesis) and formulas was presented below with some comments on best practices for estimation and references, according to McGill (2011) revision.

Log-series (*c*): the log-series is the limit of a negative binomial distribution (Poisson gamma), when it is far from the Poisson distribution with the zero abundances truncated (the probability of sampling abundance 0 is removed by rescaling) ([Fisher et al. 1943](#_ENREF_7)). Fisher’s α is technically not a parameter of the distribution but is instead an implicit function (S= α ln[1+N / α]), where, S is species richness and N is total abundance. The estimation of the single parameter of log-series, c can be done by the numerical solution in the equation:

$$\overbar{n\_{i}}= \frac{c}{-\left(1-c\right)log⁡(1-c)}\*p\left(n\right)=kc^{n}/n$$

where, k = -1/ln(1-*c*).

Lognormal (μ, σ, *cv*): This distribution has two parameters, the mean (μ) and standard deviation (σ) of log-transformed abundances (usually natural). The average is a scale parameter (position in the x-axis abundances), but the format depends on the interaction between the mean and standard deviation. The coefficient of variation (*cv*= σ / μ) serves as a descriptor of shape, *cv* << 1 indicates normality and the skewness increases as the *cv* increases. The estimation is done by logarithmic transformation, followed by the calculation of the mean and standard deviation, and following the equation:

p(n)= exp[-log(n/ μ)²/(2 σ²)]/[n σ(2π)1/2]

ZSM (*m*, θ): The birth-death-immigration equations of unified neutral theory can be solved to produce what is called zero sum multimodal distribution (ZSM) ([Hubbell 2001](#_ENREF_9); [Etienne 2005](#_ENREF_4); [McGill 2011](#_ENREF_12)) despite the assumption of zero sum not fundamental ([Etienne *et al.* 2007](#_ENREF_6)). It has two parameters. The speciation rate (θ) and the percentage of individuals moving from outside the local community (*m*). In practice these values are impossible to measure, then θ and *m* are fitted by maximum likelihood using complex combinatorial methods ([Etienne 2005](#_ENREF_4)).

***Metacommunity Model***

The metacommunity model was assembled by a set of D neutral communities, each of size $C\_{i}$, i in {1, ..., D}, connected through migration. The metacommunity has a total size of $J\_{M}=\sum\_{i=1}^{D}C\_{i}$. The relative species abundances in the regional pool are summarized by the vector $X≡x\_{j}$, j in {1, …, S}, where $x\_{j}$ is the regional relative abundance of species j. In sample i, the abundance of species j is denoted by $n\_{ij}$, and the community matrix is denoted by $N≡\{n\_{ij}\}$, i in {1, …, D}, j in {1, …, S}. Thus, each of the D communities may be defined as a sample of the metacommunity ([Etienne and Alonso 2005](#_ENREF_5)), limited by both dispersal and by the habitat characteristics of the site ([Jabot et al. 2008](#_ENREF_10)). The metacommunity is not assumed to be itself neutral, whereas we used the full regional species abundance distribution X instead of θ to describe the metacommunity, as it is a sum of local communities with potentially different environmental filters ([Jabot et al. 2008](#_ENREF_10)). The probability that an individual establishes in community i having been produced outside of this community is defined as mi, and write as $M≡\{m\_{i}\}$, i in {1,…,D}.

**References**

Doi H, Mori T. 2013. The discovery of species–abundance distribution in an ecological community. Oikos. 122(2):179-182.

Dornelas M, Moonen AC, Magurran AE, Bàrberi P. 2009. Species abundance distributions reveal environmental heterogeneity in modified landscapes. J Appl Ecol. 46(3):666-672.

EMBRAPA. 1997. Manual de métodos de análise de solo. Rio de Janeiro: Embrapa Solos.

Etienne RS. 2005. A new sampling formula for neutral biodiversity. Ecol Lett. 8(3):253-260.

Etienne RS, Alonso D. 2005. A dispersal‐limited sampling theory for species and alleles. Ecol Lett. 8(11):1147-1156.

Etienne RS, Alonso D, McKane AJ. 2007. The zero-sum assumption in neutral biodiversity theory. J Theor Biol. 248(3):522-536.

Fisher RA, Corbet AS, Williams CB. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. The Journal of Animal Ecology. 12(1):42-58.

Gray JS. 1981. Detecting pollution induced changes in communities using the log-normal distribution of individuals among species. Mar Pollut Bull. 12(5):173-176.

Hubbell SP. 2001. The unified neutral theory of biodiversity and biogeography. Princeton, New Jersey Princeton University Press.

Jabot F, Etienne RS, Chave J. 2008. Reconciling neutral community models and environmental filtering: theory and an empirical test. Oikos. 117(9):1308-1320.

Kevan PG, Greco CF, Belaoussoff S. 1997. Log-normality of biodiversity and abundance in diagnosis and measuring of ecosystemic health: pesticide stress on pollinators on blueberry heaths. J Appl Ecol.1122-1136.

McGill BJ. 2011. Species abundance distributions. In: Magurran AE, McGill BJ, editors. Biological diversity: frontiers in measurement and assessment. New York: Oxford University Press; p. 105-122.

Motomura I. 1932. On the statistical treatment of communities. Zool Mag. 44:379-383.

Preston FW. 1948. The Commonness, And Rarity, of Species. Ecology. 29(3):254-283.

Solinst. 2012. Levelogger Software 4.0.3.

ter Steege H, Pitman N, Sabatier D, Castellanos H, Van Der Hout P, Daly DC, Silveira M, Phillips O, Vasquez R, Van Andel T et al. 2003. A spatial model of tree α-diversity and tree density for the Amazon. Biodivers Conserv. 12(11):2255-2277.

ter Steege H, Pitman NCA, Phillips OL, Chave J, Sabatier D, Duque A, Molino JF, Prévost MF, Spichiger R, Castellanos H et al. 2006. Continental-scale patterns of canopy tree composition and function across Amazonia. Nature. 443(7110):444-447.

Ugland KI, Gray JS. 1982. Lognormal distributions and the concept of community equilibrium. Oikos.171-178.

Whittaker RH. 1965. Dominance and Diversity in Land Plant Communities: Numerical relations of species express the importance of competition in community function and evolution. Science. 147(3655):250-260.