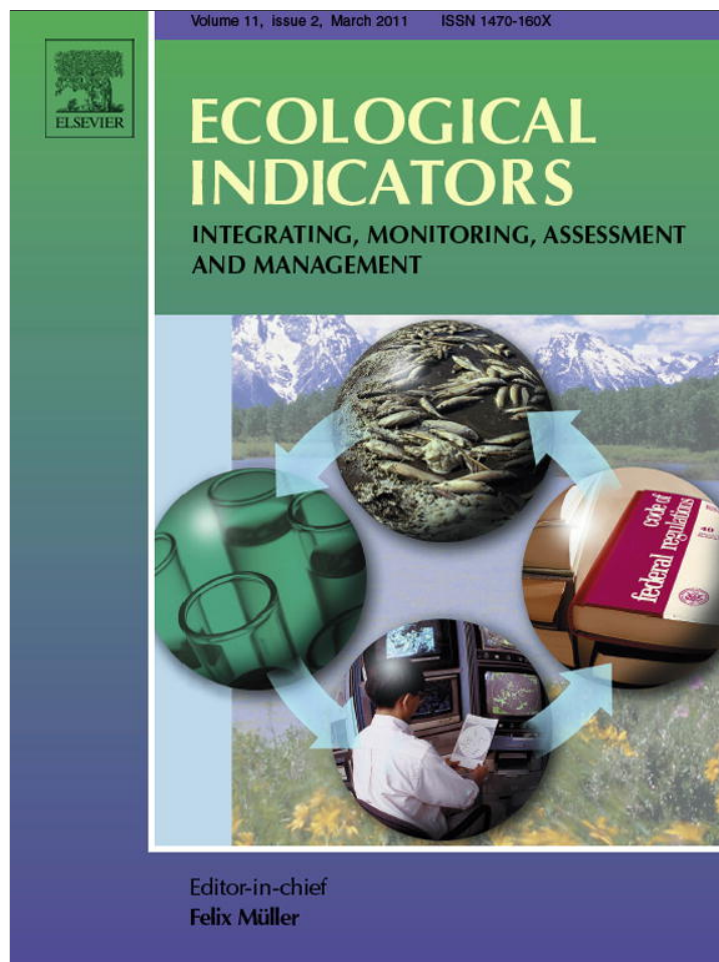


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

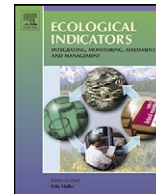
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original article

The relationship between percentage of singletons and sampling effort: A new approach to reduce the bias of richness estimates

Luiz Carlos Serramo Lopez^{a,*}, Maria Paula de Aguiar Fracasso^b, Daniel Oliveira Mesquita^a, Alexandre Ramlo Torre Palma^a, Pablo Riul^c^a Departamento de Sistemática e Ecologia, Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, Cidade Universitária, João Pessoa Paraíba, 58059-900, Brazil^b Departamento de Biologia, Universidade Estadual da Paraíba, Av. das Baraúnas, 351/Campus Universitário, Bodocongó, 58109-753, Campina Grande, Paraíba, Brazil^c Departamento de Engenharia e Meio Ambiente, Centro de Ciências Aplicadas e Educação, Universidade Federal da Paraíba - Campus IV, R: Mangueira s/n, Centro CEP: 58.297-000, Rio Tinto, Paraíba, Brazil

ARTICLE INFO

Article history:

Received 16 February 2011

Received in revised form 1 July 2011

Accepted 10 July 2011

Keywords:

Species richness estimation

Sampling intensity

Singletons

Inventory completeness

Chao1

Jackknife

Bootstrap

Chao2

ACE

ICE

ABSTRACT

Estimate the richness of a community with accuracy despite differences in sampling effort is a key aspect to monitoring high diverse ecosystems. We compiled a worldwide multitaxa database, comprising 185 communities, in order to study the relationship between the percentage of species represented by one individual (singletons) and the intensity of sampling (number of individuals divided by the number of species sampled). The database was used to empirically adjust a correction factor to improve the performance of non-parametrical estimators under conditions of low sampling effort. The correction factor was tested on seven estimators (Chao1, Chao2, Jack1, Jack2, ACE, ICE and Bootstrap). The correction factor was able to reduce the bias of all estimators tested under conditions of undersampling, while converging to the original uncorrected values at higher intensities. Our findings led us to recommend the threshold of 20 individuals/species, or less than 21% of singletons, as a minimum sampling effort to produce reliable richness estimates of high diverse ecosystems using corrected non-parametric estimators. This threshold rise for 50 individuals/species if non-corrected estimators are used which implies in an economy of 60% of sampling effort if the correction factor is used.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Species richness is the most frequently used diversity indicator (Feest et al., 2010), and the demand for more accurate richness estimation grows in parallel with the increased human alteration of our biosphere (Clarke et al., 2011; Gotelli and Colwell, 2001; Magurran, 2004). Although species richness alone does not capture how species interact among themselves and the functional diversity present in the ecosystems (Hooper et al., 2005; Petchey and Gaston, 2002), it still conveys a basic information about biological communities, which can be combined with other parameters to produce high quality indicators (Feest, 2006; Feest et al., 2011).

However, researchers face a trade-off between very complete diversity inventories, which are time and resource consuming, and briefer ones thought to be more imprecise. Longino et al. (2002) and Mao and Colwell (2005) stressed the challenges involved in determining the total richness of a given community, since there is

an overwhelming presence of rare species in mega-diverse ecosystems.

Using non-parametric richness estimators is a potential tool to evaluate the completeness of an inventory (Chao, 1984; Colwell and Coddington, 1994; Smith and van Belle, 1984). Non-parametric estimators are thought to be less dependent on the rate of collection of unseen species discovery or the shape of the assemblage distribution (Chao et al., 2009; Palmer, 1990, 1991; Zelmer and Esch, 1999). However, they demand a minimum sampling effort to produce reliable estimates (Chao et al., 2009; Chiarucci et al., 2003).

Coddington et al. (2009) suggested that many inventories of tropical arthropods suffer from an undersampling bias, strong enough to impair even the use of richness estimators in order to assess the real richness of these assemblages. In a large compilation of tropical arthropod inventories, they also found a significant negative relationship between the percentage of species represented by one individual (singleton) and the sample intensity (abundance divided by richness). Singletons have an intuitive connection with inventory completeness, since we expect that the proportions of singletons should decrease as the sampling effort increases, until we come close to the “real” proportion of singletons present in a

* Corresponding author. Tel.: +55 83 9937 6226.

E-mail address: lcslopez@yahoo.com (L.C.S. Lopez).

community. For instance, Coddington et al. (2009) estimated a true proportion of 4% singletons by lognormal extrapolation from their spider assemblage, which originally presented 29% singletons.

Solutions to obtain more reliable richness estimates are either a dramatic increase in sample effort or the development of better richness estimators (Chiarucci et al., 2003). Here we proposed that it is possible to correct classical non-parametric estimators in order to boost their performance under conditions of undersampling. We empirically derived this correction using the relationship between the intensity of sampling and the proportion of singletons found in a large database of communities we obtained from the literature.

1.1. Deriving the estimator correction

Our correction to improve a non-parametric estimator under low sampling conditions consists in multiplying the original estimate by 1 plus the proportions of singletons in the sample elevated by a constant:

$$S_{estP} = S_{est}(1 + P^z) \quad (1)$$

where S_{estP} is the modified estimate, S_{est} is the original estimate, P is the proportion of singletons (singletons/observed species richness) and z is a constant higher than one. Since the proportions of singletons (P) falls as the sampling effort increases, this basic formula will improve the performance of the estimator under low sampling effort but will converge to the original estimate at high sampling effort conditions.

The constant z in the formula shall mirror the allometric relationship between the proportions of singletons (P) and the intensity of sampling (I) found in natural assemblages. Assuming that:

$$P = I^{-1/z} \quad (2)$$

the value of the constant z can be empirically derived as

$$z = -\frac{\ln I}{\ln P} \quad (3)$$

where $\ln I$ is the natural logarithm of the sampling intensity (the number of individuals divided by the number of species observed in a given sample) and $\ln P$ is the natural logarithm of the proportions of singletons in the same sample. To estimate the value of z we used a large compilation of different communities, varying widely in sampling intensity and taxonomical composition. We found the average value of z in this database of 185 assemblages to be close to 2 (2.06 ± 0.73 SD, $n = 185$), leading us to a general transformation to correct non-parametric estimators under low sampling effort.

$$S_{estP} = S_{est}(1 + P^2) \quad (4)$$

This correction (called P correction) can also be used to species-incidence estimators by substituting the proportion of singletons (species represented by one individual) by the proportions of uniques (species represented in one sample). The transformation will increase the estimate (up to 100%) when the proportion of singletons (or uniques) in the sample is high, but decreases exponentially, converging towards the original estimator value when the proportion is low.

We also used our database of communities to search for trends that could indicate the “real” proportions of singletons in well sampled assemblages and to evaluate the limits that low sample intensities pose to the reliability of non-parametric estimators, with and without the correction we developed. We tested our P correction in the most common used non parametric estimators: Chao1 (Chao, 1984), Chao2 (Chao, 1987), Jackknife1 (Heltshe and Forrester, 1983), Jackknife2 (Burnham and Overton, 1978), ACE, ICE (Chao and Lee, 1992), and Bootstrap (Smith and van Belle, 1984).

2. Methods

2.1. Database of communities

We expanded the Coddington et al. (2009) original compilation of terrestrial arthropod inventories, adding other taxa (tropical trees, corals and terrestrial vertebrates) to produce a set of 185 datasets where the singletons–richness ratio and the intensity of sampling (abundance/richness) were calculated (see Appendix S1 in Supporting Information for details).

2.2. Testing the efficacy of the P transformation for Chao1 estimator using data from large plots of tropical forests

We used the data from six inventories produced by research teams belonging to The Center for Tropical Forest Science network of large forests plots around the world (Condit et al., 2005; CTFS, 2009). We used data from 3 different continents: Africa (Korup Forest, census 1998 and Egoro Forest, census 2000), Americas (BCI, census 2005 and Luquillo, census 1995) and Asia (Huai Kha Khaeng, census 1999 and Pasoh, census 1995). For each plot, we obtained simulated sets of 100 rarefied sub-samples with increasing average intensities (5, 25, 75 and 100 ind/spp.).

We used the classical Chao1 (Chao, 1984) estimator to perform a series of comparisons between the original Chao1 formula and the P corrected “Chao1P”. We calculated the average Chao1 estimates and our corrected Chao1P for each set of rarefaction simulations at different intensities. Using these estimates, we calculated the bias and precision of these two estimators using the scaled mean error (SME) and the coefficient of variation respectively (Walther and Moore, 2005). Fitting the average estimates from these rarefactions to power curves, we also inferred the minimum sampling intensity necessary to estimate 100% and 95% of the original richness.

2.3. Using the database of communities to test the efficiency of Chao1 estimator

We calculated the percentage difference between Chao1 and Chao1P estimates for the 185 communities in our database. This difference can be derived from the Chao1P formula (Eq. (2)) as follows:

$$\text{Difference}_{\text{Chao1 vs Chao1P}} = \left(\frac{f_1}{S_{obs}} \right)^2 \times 100 \quad (5)$$

To check the validity of our rarefactions, we calculated the mean difference between Chao1 and Chao1P estimates from simulated intensities of 5, 25, 75 and 100 (obtained from the high intensity forest plots) and compared them with the differences obtained from our multi-taxa database with similar non-rarefied intensities.

2.4. Testing the P correction for Chao 2, Jackknife 1, Jackknife 2, ACE, ICE and Bootstrap estimators

To test the performance of the P correction for other non-parametrical estimators, besides the Chao 1, we made simulations using data from BCI tree plot (census 2005) (CTFS, 2009). We created 50 pairs of rarefied samples drawn from BCI data with 5 levels of sampling intensity (5, 25, 50, 75 and 100 ind/spp.). These subsamples were used as an input data for EstimateS 8.2 (Colwell, 2009) to create the original, uncorrected, estimates using Chao2, Jack1, Jack2, ACE, ICE and Bootstrap estimators (100 simulations per sample). Using the formula (1) we transformed the original estimates in their corrected P versions (Chao2P, Jack1P, etc.) and compared the ability of corrected and uncorrected estimators to estimate 100% of BCI dataset richness (299 species) using sub-sets with reduced intensity of sampling (Fig. 1).

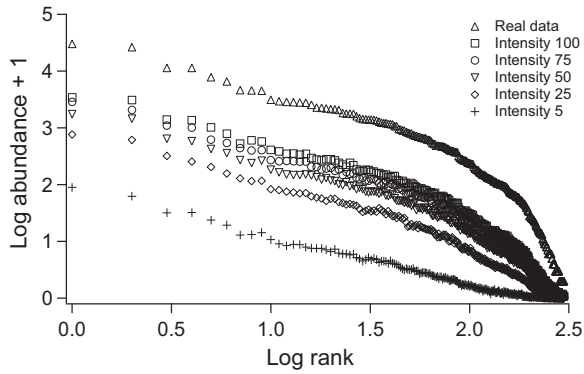


Fig. 1. Abundance (log 10) vs. Rank (log 10) of tree species from BCI utilizing real data (census 2005, real intensity 697 ind/spp.) and the abundance averages from 100 rarefaction simulations under five sampling intensities (100, 75, 50, 25 and 5 ind/spp.).

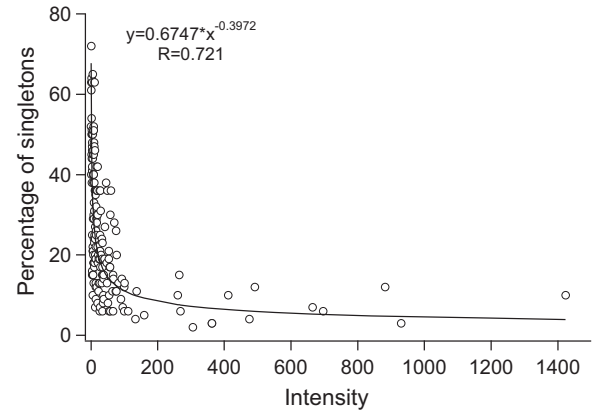


Fig. 2. Scatterplot between percentage sampling intensity (ind/spp.) and percentage of singletons for 185 communities belonging to 4 major taxa (arthropods, corals, trees and vertebrates). The percentage of singletons falls sharply between 0 and 75 ind/spp., but tends to stabilize around 8% singletons above intensity 100.

3. Results

Our expanded database encompasses 185 communities, ranging from 1 to 1423 ind/spp. in intensity and between 2% and 72% singletons. The median intensity was 20.3 and the median percentage of singletons 19.2%. The community samples belonged to four major groups: terrestrial vertebrates ($n = 79$), terrestrial arthropods ($n = 72$), corals ($n = 22$) and trees ($n = 12$) (see [Appendix S1](#)). The correlation between intensity and percentage of singletons was very highly significant (log-transformed power curve, $r = 0.72$; $p < 0.0001$; [Fig. 2](#)). The percentage of singletons tended to decline as sampling intensity increased, with samples of intensity 5 or less

($n = 23$) having an average of 46% singletons ($\pm 3\%$ SE), while communities with a sampling intensity of 100 or more ($n = 20$) had an average of 8% of singletons ($\pm 1\%$ SE) ([Fig. 3](#)).

The Chao1P estimator had less bias and precision compared to Chao1, when tested in rarefied sub-samples from six large plots of trees ([Table 1](#)). At higher intensities, both estimators yield very similar results (0.8% on average bias difference at original intensities). However, at the low intensity of 5 ind/spp., Chao1P outperforms Chao1 by 17.1% in terms of bias, with an average precision loss of 2.8% compared to Chao1 ([Table 1](#)).

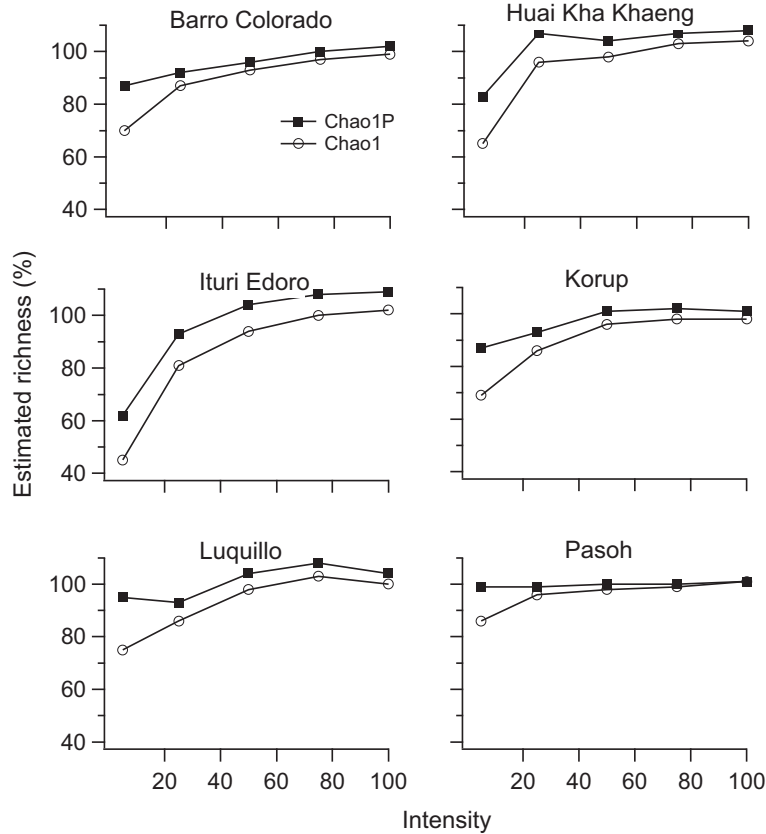


Fig. 3. Average percentage of original richness estimated with uncorrected Chao1 and corrected Chao1P estimators from rarefied simulations with different sampling intensity efforts. At higher intensities, both estimators tend to converge, but Chao1P estimates approaches faster than Chao1 towards 100% of estimation as intensity increased in the six large plots of trees used in the simulations.

Table 1

Average bias and loss of precision percent values (100×) for estimates of richness found with the uncorrected Chao1 and the corrected Chao1P using 100 rarefied simulations with different sampling intensities (INT) drawn from six large plots of tropical trees (standard error between parentheses). At lower intensities Chao1P showed a good trade-off between bias reduction and loss of precision compared to Chao1.

	INT 5	INT 25	INT 50	INT 75	INT 100	Real
Bias						
Average Bias Chao1	-32 (±6)	-11 (±2)	-4 (±1)	0 (±1)	1 (±1)	7 (±2)
Average Bias Chao1P	-14 (±5)	-4 (±2)	1 (±1)	4 (±2)	4 (±1)	7 (±2)
Bias Chao1P–Bias Chao1	-17 (±1)	-8 (±1)	-5 (±1)	-4 (±1)	-4 (±1)	-1 (±0)
Precision						
Avg Precision Loss (A.P.L.) Chao1	19 (±4)	10 (±3)	10 (±3)	8 (±2)	7 (±1)	N.A.
Avg Precision Loss (A.P.L.) Chao1P	22 (±4)	12 (±3)	11 (±3)	8 (±2)	7 (±1)	N.A.
A.P.L. Chao1P–A.P.L. Chao1	-3 (±0)	-2 (±0)	-1 (±0)	0 (±0)	0 (±0)	N.A.

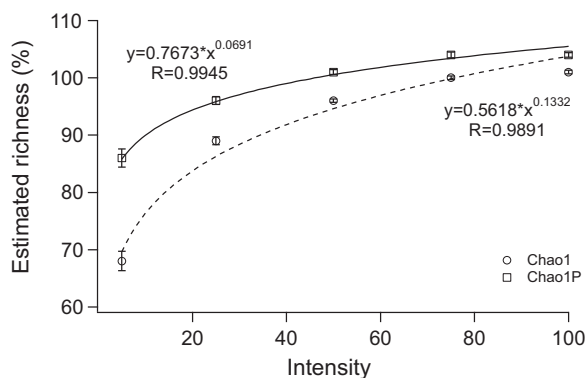


Fig. 4. Mean richness (SE bars) estimated by Chao1 and Chao1P estimators obtained from rarefactions of six large plots of tropical trees. The fitted power curves were used to calculate the minimum intensity necessary to estimate 95% and 100% of original richness. Chao1P crosses these thresholds (95% and 100%) with less sampling effort than Chao1.

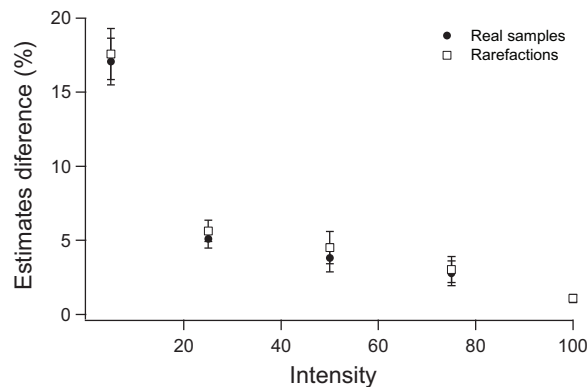


Fig. 5. Average differences (SE bars) between Chao1 and Chao1P estimates for rarefactions extracted from high intensity samples compared to differences obtained from real samples with original low intensities. The pattern of improved estimation by Chao1P at low intensities followed by convergence at higher intensities is very similar between rarefaction simulations and real data.

The estimated richness increased with rarefaction intensities in a pattern that fits very well to the power curves for both Chao1P and Chao1 ($R^2 = 0.99$ for Chao1P and 0.98 for Chao1) (Fig. 4). These curves predict that, on average, Chao1P will estimate 100% of original richness at an intensity of 52.0 ind/spp. (± 9.4 SE), while Chao1 will reach 100% at an intensity of 78.7 ind/spp. (± 5.5 SE) (Fig. 4). If we use 95% of original richness instead of 100% as a good approximation of original richness (as proposed by Chao et al. (2009)), the thresholds change to 20.7 ind/spp. (± 5.9 SE) for Chao1P and 50.0 ind/spp. (± 9.4 SE) for Chao1.

The difference between the two estimators obtained from the rarefaction simulations (from tree plots with intensity 261 or more) showed good agreement with the average difference obtained from samples of taxa that had low intensity (Fig. 5). For example, on rarefactions to an intensity of 5 the Chao1P estimates were, on average, 17.1% (± 1 SE) higher than Chao1 while for the 57 communities (29 of arthropods and 28 of vertebrates) in our database with intensities ranging between zero and 10 (midpoint intensity 5) the average difference was 17.2% (± 2 SE) (Fig. 5).

The P transformation improved the performance of the other 6 non-parametrical estimators (Chao2, Jack1, Jack2, ACE, ICE and Bootstrap) in similar way it did for Chao1. The transformed estimators produced estimates that were more close to BCI real richness compared to their untransformed versions under simulated conditions of low sampling effort (9% less biased, in average, compared to the uncorrected formulas at the intensity of 5 ind/spp.) and converge to the untransformed values as sampling effort increases (Table 2). The ICE corrected estimator (“ICEP”) showed the best overall performance, in these simulations, estimating, in average, 83% of BCI real richness at intensity 5 ind/spp. compared to 67% made using its uncorrected version (Fig. 6).

4. Discussion

Our expanded dataset confirmed the trend found by Coddington et al. (2009) for their arthropod database: the percentage of singletons tends to decrease with an increase in the sampling intensity in a very consistent way. At lower intensities, one needs to increase the intensity of sampling by five-fold in order to halve the frequency of singletons. However, at higher intensities (roughly, above intensity 100), the frequency of singletons tends to stabilize around 8% (Fig. 2). Given that we have a phylogenetically diverse group of assemblages present in our database (corals, arthropods, vertebrates and trees), we assume that this pattern is a general one among communities. Consequently, the value of 8% (± 4 SD) is

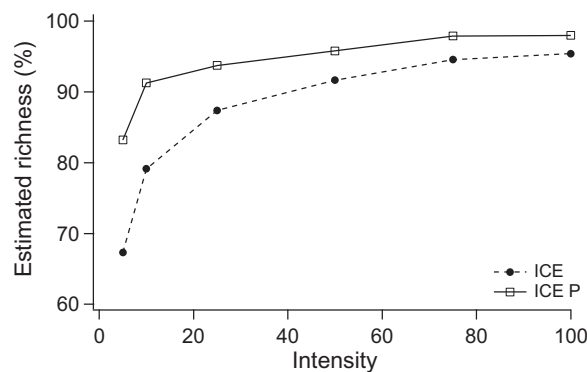


Fig. 6. Comparative performance between corrected (ICEP) and uncorrected (ICE) version of the ICE non-parametrical richness estimator trying to predict the richness of BCI 50 ha plot (census 2005, richness = 299, intensity = 697 ind/spp.). The corrected estimator produced less biased estimates and make better predictions with less sampling effort.

Table 2
Biases of six non-parametric richness estimators (ACE, ICE, Chao2, Jackknife 1, Jackknife 2 and Bootstrap), using their original uncorrected formulas and their *P* corrected versions, while trying to estimate BCI 50 ha plot richness (299 tree species and intensity sample of 697 individuals/species) using subsamples with low intensity (5 ind/spp.) and high intensity (100 ind/spp.). The *P* corrected versions of the estimators produce less biased estimates compared to their uncorrected versions.

	Low intensity uncorrected bias	Low intensity corrected bias	High intensity uncorrected bias	High intensity corrected bias
ACE	34%	26%	4%	3%
ICE	26%	14%	1%	0%
Chao2	44%	35%	9%	7%
Jackknife 1	41%	34%	3%	3%
Jackknife 2	41%	31%	3%	2%
Bootstrap	46%	37%	7%	5%
Average (\pm SD)	39% (\pm 7%)	30% (\pm 9%)	5% (\pm 1%)	3% (\pm 1%)

probably close to the percentage of singletons expected from most natural communities after severe undersampling bias is removed.

Since the proportion of singletons has a robust statistical relationship with the degree of undersampling, it can be used to adjust the results from non-parametrical estimators. Our transformation, empirically derived from this relationship, was able to reduce the bias from all the non-parametric estimators tested compared to their untransformed versions at low sampling intensities, while both versions converged to very similar values at high intensities.

The accuracy of an estimator is a compromise between the variation among estimations (precision) and the distance between the estimated richness and the real richness (bias) (Brose et al., 2003; Walther and Moore, 2005). For example, at lower intensities (intensity 5), the corrected Chao1P showed a reasonable trade-off, losing on average 3% precision, but gaining 17% in bias reduction compared to Chao1. A mean of 17% less bias was found in both the rarefaction simulations, drawn from high intensity samples, and from non-rarified lower intensity samples (Fig. 5). This agreement between rarefied sub-samples from large tree plots and other multi-taxa data suggests that the simulations were able to reproduce realistic patterns in low intensity samples of natural situations.

Notice that the improvement provided by Chao1P applies not only to the average values but also to the 95% boundary, which can be used to produce less conservative richness estimates. For example, for the BCI dataset rarefied at an intensity of 5 ind/spp., Chao1P improved both the average estimate (24% less bias) and the upper 95% estimate (26% less bias) compared to untransformed Chao1.

The rarefactions, using data from six large plots of trees, also allows us to predict the minimum intensity necessary for Chao1 and Chao1P to make estimates close to 100% of original species richness. According to these simulations, it would be necessary to sample, on average, 51% more individuals to be able to make an accurate estimation using Chao1 (minimum intensity 78.7 ind/spp.) compared to Chao1P (minimum intensity 52.0 ind/spp.). A difference of this magnitude can represent a great economy of time and resources while estimating the total richness of very diverse communities.

If one uses 95% of the total richness estimated as a more tenable goal (Chao et al., 2009), the difference in sampling effort between Chao1P and Chao1 becomes even larger, since our simulations predict that one would need to sample, on average, 142% more individuals using Chao1 (minimum intensity 50.0) than for Chao1P (minimum intensity 20.7) to estimate 95% of a total sample richness. Since we found an increase of only 2% on Chao1P precision loss compared to Chao1 at intensity 25 (close to the threshold of 20.7 for Chao1P for 95% estimation), the trade-off between loss of precision and gain in economy of sampling effort in order to estimate 95% of total richness appears to be extremely positive.

The other non-parametric estimators tested (Chao2, Jack1, Jack2, ACE, ICE and Bootstrap) presented the same pattern found with Chao1 (Fig. 6). The *P* corrected versions of each estimator tested produced less biased estimates at low sampling intensities

compared to their original formulas while the corrected values converge with the original ones as the intensity of sampling increases. The corrected version of ICE (ICEP), for example, was able to estimate 95% of BCI plot richness using sub-samples with 50 ind/spp. of intensity while the uncorrected version of ICE only achieved the same feat at intensity 100 ind/spp. (Fig. 6), this difference represents an economy of 50% in terms of sampling effort.

Consequently, our findings strongly indicate that our correction for non-parametric estimators (Eq. (3)) produce less biased results and should be used to estimate the richness in ecological studies that are trying to remove the effects of undersampling. An alternative option is to parametrically extrapolate the number of species to a given area or number of individuals (Melo et al., 2007; Reichert et al., 2010). However, if such information (the total community area, or the final number of individuals expected to be sampled) is not available, a non-parametric estimation using the correction present in Eq. (3) is the best option.

We also demonstrated that the intensity of sampling (the number of individuals sampled divided by the number of species) and the proportion of singletons (the number of species represented by one individual divided by the total number of species) can be used to indirectly access the accuracy of richness estimates. Since these two parameters can be easily determined at each stage of a real sampling program they can provide useful guidelines for planning and evaluating biodiversity surveys. In our multi-taxa database, for example, 74% of the inventories are below the average intensity threshold necessary to estimate at least 95% of the total richness using Chao1, and 50% did not reach the same kind of threshold for Chao1P. These numbers give support to Coddington et al.'s (2009) arguments that we need greater investment in biodiversity inventories in order to get a realistic picture of the true richness of highly diverse ecosystems.

5. Conclusions

- Our results indicate that ecological surveys that present more than 8% of singletons, or less than 100 individuals/species of sampling intensity, probably are suffering from some degree of undersampling and could be improved either by an increase of sampling effort or by using richness estimators.
- Our simulations and database analysis led us to recommend the threshold of 20 individuals/species, or less than 21% of singletons, as a minimum sampling effort to produce reliable richness estimates (at least 95% of richness estimated) using corrected non-parametric estimators.
- The threshold rise for 50 individuals/species, or less than 14% of singletons, if non-corrected estimators are used, which implies in an economy of 60% of sampling effort due to the correction factor.

Acknowledgments

We thank Nicholas Gotelli, Adriano S. Melo, Carlos Eduardo Grelle and two anonymous referees for insightful comments on the

subject. This work is supported by research fellowship from CNPq to DOM and a post doc fellowship from CNPq/FAPESQ to MPAF.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolind.2011.07.012.

References

- Brose, U., Martinez, N.D., Williams, R.J., 2003. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology* 84 (9), 2364–2377.
- Burnham, K.P., Overton, W.S., 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* 65, 623–633.
- Chao, A., 1984. Nonparametric-estimation of the number of classes in a population. *Scandinavian Journal of Statistics* 11 (4), 265–270.
- Chao, A., 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43, 783–791.
- Chao, A., Colwell, R.K., Lin, C.W., Gotelli, N.J., 2009. Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90 (4), 1125–1133.
- Chao, A., Lee, S.M., 1992. Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, 7210–7217.
- Chiarucci, A., Enright, N.J., Perry, G.L.W., Miller, B.P., Lamont, B.B., 2003. Performance of nonparametric species richness estimators in a high diversity plant community. *Diversity and Distributions* 9 (4), 283–295.
- Clarke, K., Lewis, M., Ostendorf, B., 2011. Additive partitioning of rarefaction curves: removing the influence of sampling on species-diversity in vegetation surveys. *Ecological Indicators* 11 (1), 132–139.
- Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M., Hormiga, G., 2009. Under-sampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology* 78 (3), 573–584.
- Colwell, R.K., 2009. EstimateS: statistical estimation of species richness and shared species from samples. Version 8.2. User's Guide and application published at: <http://purl.oclc.org/estimates>.
- Colwell, R.K., Coddington, J.A., 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences* 345, 101–118.
- Condit, R.G., Ashton, M.S., Balslev, H., Brokaw, N.V.L., Bunyavejchewin, S., Chuyong, G.B., Co, L., Dattaraja, H.S., Davies, S.J., Esufali, S., Ewango, C.E.N., Foster, R.B., Gunatilleke, N., Gunatilleke, S., Hernandez, C., Hubbell, S.P., John, R., Kenfack, D., Kiratiprayoon, S., Hall, P., Hart, T.H., Itoh, A., LaFrankie, J.V., Liengola, I., Lagunzad, D., Lao, S., Losos, E.C., Magard, E., Makana, J.R., Manokaran, N., Navarrete, H., Mohammed Nur, S., Okhubo, T., Perez, R., Samper, C., Hua Seng, L.H., Sukumar, R., Svenning, J.C., Tan, S., Thomas, D.W., Thompson, J.D., Vallejo, M.I., Villa Muñoz, G., Valencia, R., Yamakura, T., Zimmerman, J.K., 2005. Tropical tree α -diversity: results from a worldwide network of large plots. *Biologiske Skrifter* 55, 565–582.
- CTFS, 2009. Center for Tropical Forest Science. [WWW document]. <http://www.ctfs.si.edu/>. Accessed: May 25, 2010.
- Feest, A., 2006. Establishing baseline indices for the quality of the biodiversity of restored habitats using a standardized sampling process. *Restoration Ecology* 14 (1), 112–122.
- Feest, A., Aldred, T.D., Jedamzik, K., 2010. Biodiversity quality: a paradigm for biodiversity. *Ecological Indicators* 10 (6), 1077–1082.
- Feest, A., van Swaay, C., Aldred, T.D., Jedamzik, K., 2011. The biodiversity quality of butterfly sites: a metadata assessment. *Ecological Indicators* 11 (2), 669–675.
- Gotelli, N.J., Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4 (4), 379–391.
- Heltshe, J.F., Forrester, N.E., 1983. Estimating species richness using the Jackknife procedure. *Biometrics* 3, 11–91.
- Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A.J., Vandermeer, J., Wardle, D.A., 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs* 75 (1), 3–35.
- Longino, J.T., Coddington, J., Colwell, R.K., 2002. The ant fauna of a tropical rain forest: Estimating species richness three different ways. *Ecology* 83 (3), 689–702.
- Magurran, A.E., 2004. *Measuring Biological Diversity*. Blackwell Publishing, Oxford, UK.
- Mao, C.X., Colwell, R.K., 2005. Estimation of species richness: mixture models, the role of rare species, and inferential challenges. *Ecology* 86 (5), 1143–1153.
- Melo, A.S., Bini, L.M., Thomaz, S.M., 2007. Assessment of methods to estimate aquatic macrophyte species richness in extrapolated sample sizes. *Aquatic Botany* 86 (4), 377–384.
- Palmer, M.W., 1990. The estimation of species richness by extrapolation. *Ecology* 71, 1195–1198.
- Palmer, M.W., 1991. Estimating species richness: the second order jackknife reconsidered. *Ecology* 72, 1512–1513.
- Petchey, O.L., Gaston, K.J., 2002. Functional diversity (FD), species richness and community composition. *Ecology Letters* 5 (3), 402–411.
- Reichert, K., Uglund, K.I., Bartsch, I., Hortal, J., Bremner, J., Kraberg, A., 2010. Species richness estimation: estimator performance and the influence of rare species. *Limnology and Oceanography: Methods* 8, 294–303.
- Smith, E.P., van Belle, G., 1984. Non-parametric estimation of species richness. *Biometrics* 40, 119–129.
- Walther, B.A., Moore, J.L., 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28 (6), 815–829.
- Zelmer, D.A., Esch, G.W., 1999. Robust estimation of parasite component community richness. *Journal of Parasitology* 85, 592–594.