Asian Journal of Environment & Ecology
2(1): 1-12, 2017; Article no.AJEE.32040

SCIENCEDOMAIN *international*
www.sciencedomain.org

Least-biased Estimations of True Species Richness of Butterfly Fauna in Sub-urban Sites around Jhansi (India) and the Range of Inter-annual Variation of Species Richness

Jean Béguinot[1*]

[1]*Biogéosciences, Université de Bourgogne, F 21000 – Dijon, France.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/AJEE/2017/32040
Editor(s):
(1) George Tsiamis, Assistant Professor of Environmental Microbiology, Department of Environmental and Natural Resources Management, University of Patras, Agrinio, Greece.
Reviewers:
(1) LeeHsueh Lee, Chung Hua University, Taiwan.
(2) Hamit Ayberk, Istanbul University, Turkey.
(3) G. O. Yager, University of Agriculture, Benue State, Makurdi, Nigeria.
Complete Peer review History: http://www.sciencedomain.org/review-history/18074

Original Research Article

Received 4th February 2017
Accepted 23rd February 2017
Published 6th March 2017

ABSTRACT

As a rule, most biodiversity inventories at local scales remain more or less incomplete, when dealing with relatively speciose taxonomic groups, such as butterfly in tropical regions. Yet, it remains possible to take maximum advantage of partial inventories and to develop reliable predictions by extrapolating the species accumulation curves beyond the already achieved samplings. Besides, due to the wide diversity of available estimators of total species richness, selecting for the less-biased estimator and the associated expression of the species accumulation curve is desirable. Accordingly, the "least-biased extrapolation procedure" is recommended in this respect.

Least-biased extrapolation procedure was applied to nine inventories carried on by Ashok Kumar in (sub-) urban sites in the vicinity of Jhansi (Uttar Pradesh, India), thus providing more accurate evaluations of remnant butterfly species richness in these sites. The range of estimated sampling completeness of inventories was comprised between 65% and 99%, depending on sites and years and the estimated true species richness was comprised between 25 species (along Highway in

**Corresponding author: E-mail: jean-beguinot@orange.fr;*

2010) and 44 species (Jhansi Fort in 2011).
Importantly, the levels of sampling completeness prove to be poorly correlated with sampling size. This highlights the fact that, contrary to still a current opinion, comparisons between levels of species richness may well remain irrelevant, even when made at a same sampling size (for example by using appropriate “rarefaction” procedure).
Four, out of the nine studied inventories, were conducted at two same sites for two successive years (2010-2011) and, thus, provide opportunity to evaluate the range of inter-annual variations of true species richness of butterfly fauna in this sub-urban context. Inter-annual variations within the range 24% to 48% were registered, according to sites.

Keywords: Lepidoptera; diversity; species accumulation curve; estimator; numerical extrapolation; minimum bias.

1. INTRODUCTION

Incomplete inventories of biodiversity are likely doomed to become increasingly frequent, as surveys progressively address new taxonomic groups more difficult to cope with, in particular those groups giving rise to species assemblages with high number of species [1,2,3]. In addition, more commonly investigated taxonomic groups, also, are likely doomed to remain more or less incompletely surveyed at the *local scale*, due to sampling efforts often being far less intensive at these small scales than they usually are through wider areas.

Accordingly, the vast majority of ongoing published inventories are admittedly *more or less incomplete*. This incompleteness may be partially compensated (yet in numerical terms only) by the estimation of the number Δ of “missed” (i.e. unrecorded) species, thereby leading to the evaluation of the total species richness S_t of the sampled assemblage of species. Many different (nonparametric) estimators of the number Δ of “missed” species have been proposed in recent decades (reviewed in [1,2]). As expected, these different types of estimators provide divergent evaluations of Δ, without any consensus having ever been reached in favor of one or the other of those estimators, supposedly being more accurate than the others. And the commonly accepted suggestion to consider all these divergent estimates without being able to choose between them [3] remains frustrating. This, in turn, probably contributes to explain why many partial inventories are still not extrapolated numerically, as would be highly desirable, in order to derive a reliable estimation of the total species richness. Indeed, evaluating the richness of species assemblages is highly desirable, at least in relative if not in absolute terms. Note that even in relative terms, a relevant comparison of species richness between two or several assemblages requires that inventories are actually compared at a *same level of completeness*. A mandatory condition that neither standardised sampling procedures nor rarefaction to a same sampling size may actually secure, contrary to what is still too often asserted in literature (and this, simply because the level of completeness reached at given sample size is tightly dependent on the degree of heterogeneity of species abundances distribution which may usually differs between sampled assemblages).

Now, a rational method of selection of the least-biased estimator (among the most commonly referenced ones) has recently been developed [4,5], enlarging the path initiated by BROSE *et al.* [6]. This newly derived procedure avoids the above mentioned frustration of having to deal with divergent estimates without knowing how to choose the most accurate of them all.

Hereafter, advantage is taken from using this procedure to extrapolate a series of inventories of Butterflies in and around the City of Jhansi (Uttar Pradesh, India) carried out and published by Ashok KUMAR of Lucknow University, making use of the recorded data published by this author [7,8]. Thereby, reliable estimates of the total species richness of each of the nine sites are expected, thus providing a more accurate appreciation of the true local diversity of butterfly fauna. Moreover, reliable predictions of the additional sampling efforts required to improve the completeness of inventories are derived from the least-biased extrapolation of samplings. At last, these extrapolations are also considered to address appropriately several questions of more general interest, in particular the evaluation of the degree of inter-annual variability of true species richness in butterfly faunas.

2. MATERIALS AND METHODS

Nine inventories of butterfly fauna in Jhansi and the vicinity (Uttar Pradesh, India) have been conducted during years 2010 and 2011 and the results published in detail by A. KUMAR, including the respective abundances of each recorded species [7,8]. Accounting for species abundances is of prime interest in the perspective of the extrapolation of partial samplings, since abundance data provides estimates of the numbers f_1, f_2, f_3,…, f_x, … of those species recorded respectively 1-, 2-, 3-, …, x- times during the realised partial sampling. These numbers are required, in turn, to reliably extrapolate the species accumulation curve, as explained below.

All details relative to the environmental context of each of the nine inventories and the list of species with their respective abundances are provided on-line with free access [7,8] and, accordingly, will not be recalled here. Sampling localities were: University Campus Jhansi (2010 & 2011), Jhansi Fort (2010 & 2011), Parichha Dam (2010), side of Jhansi Gwalior Highway (2010), Medical College Campus (2011), Narayan Bagh (2011), B.I.E.T (Bundelkhand Institute Engeering & Technology) (2011).

2.1 Numerical Extrapolation beyond Achieved Sampling Size

As sampling size increases, the number R of recorded species is monotonically growing, at first rapidly and then less and less quickly. The so-called 'Species Accumulation Curve' R(N) accounts for the growth kinetics of the recorded species richness R with increasing sampling size N (N: typically, the number of observed individuals). The mathematical expression (and thus the details of the shape) of the Species Accumulation Curve are dependent upon both the total species richness of the sampled assemblage of species and the degree of heterogeneity of the species abundance distribution within the sampled assemblage of species. This would apparently make the extrapolation of the Species Accumulation Curve rather difficult to compute, since both preceding factors are unknown *a priori*. Yet, the numbers f_1, f_2, f_3,…, f_x, …, of those species recorded respectively 1-, 2-, 3-, …, x- times during sampling are directly dependent also upon the total species richness and the degree of heterogeneity of the species abundances. This explains why these numbers f_1, f_2, f_3,…, f_x, … may serve as an appropriate numerical basis from which to extrapolate the Species Accumulation Curve, beyond the actual size of the sample under consideration. In particular, the most commonly used estimators of the number of unrecorded species (i.e. 'Chao' and the series of 'Jackknife') are computed from the recorded value of the numbers f_x [1]. In practice, a problem remains however: as already mentioned, each of these different types of estimators provides a substantially distinct estimate and none among these estimators reveals being consistently more appropriate. Accordingly the traditional practice has become to consider together all of them without making any choice [3], an admittedly frustrating situation!

Yet, it has been shown recently that although none of the available estimators consistently remains the more accurate, each of them may, in turn, reveal being the less biased, depending on the value taken by f_1 as compared to the other $f_{x>1}$ [4]. Accordingly, in practice, the most appropriate – i.e. *the least biased* – estimator of the number of unrecorded species may be selected by comparing the value of f_1 to the values of the other $f_{x>1}$ [4,5]. Selecting this way the least-biased type of estimator hereby provides the best possible estimate of the number Δ of "missing" species and, in turn, the best estimate of the total species richness S_t of the partially sampled assemblage. In addition, the less biased expression for the extrapolation of the species accumulation curve is straightforwardly derived.

In practice, Appendix 2 provides (i) the expressions of Δ, S_t and R(N), according to each of the most commonly used types of estimators and (ii) the *key to select* the less biased estimator and, thereby, the less-biased expressions for Δ, S_t and R(N). Also, in order to reduce the influence of drawing stochasticity, which affects the *as-recorded* values of the f_x, it is advisable to regress the as-recorded distribution of the f_x versus x (cf. Appendix 1).

3. RESULTS

3.1 Least-biased Estimations of the Total Species Richness and of the Extra-sampling Effort Required for Improving Sampling Completeness

For the nine inventories of butterfly fauna carried out within and around Jhansi during years 2010

and/or 2011, Table 1 provides: the achieved sample size N_0, the number of recorded species R_0 (= $R(N_0)$), the type of *least-biased* estimator selected according to Appendix 2, the estimated number Δ of missing species, the estimated true (total) species richness S_t and the resulting estimate of sampling completeness. Figs. 1 and 2 provide a convenient graphic overview of the main results.

A few examples of extrapolations of the Species Accumulation Curves are also presented at Fig. 3 (where the extrapolations associated to six different types of estimators are compared for a same inventory) and at Fig. 4.

These extrapolations allow to gauge the predicted extra-sampling effort that would be required to obtain any given increment in recorded species richness. This is, in particular, of practical intest to make a rationnally informed decision, as regards the opportunity (or not) to extend further the inventory. Fig. 3 examplifies the importance of selecting the least-biased extrapolation among the set of possible extrapolations. Not only the predicted number Δ of "missing" species may vary from simple to double, according to the considered extrapolation, but the predicted extra-sampling effort required to reach a given level of completeness may vary in still larger ranges, as examplified in Fig. 3.

3.2 Evolution of the Numbers of Species Recorded 1- 2- 3- 4- 5- times with Increasing Sampling Completeness

The series of the nine inventories of butterfly diversity conducted in and around Jhansi also offers the opportunity to address a rather theoretical, but nevertheless quite an interesting question: how does each of the numbers f_1, f_2, f_3, …, f_x, … of species recorded respectively 1-, 2-, 3-,…, x- times vary with increasing level of sampling completeness. The more straightforward way to cope with the subject would be, of course, to simply monitor progressive sampling (all along its actual progression), thus registering directly the variations of the f_x with increasing sampling size. This, yet, has rather rarely been achieved. Yet, an alternative, indirect, procedure may be envisaged, however, as a possible surrogate. This would consist in taking the opportunity of a series of inventories addressing a similar type of fauna, each of them being conducted at a different level of completeness.

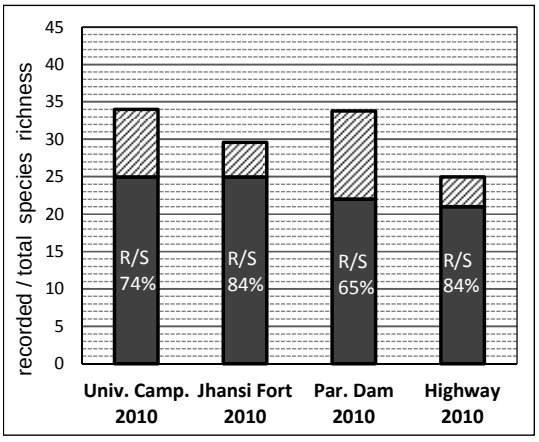


Fig. 1. Actually recorded species richness and extrapolated total species richness for year 2010. Provided are the numbers of species (i) recorded (black) and (ii) still unrecorded (hatched) and the sampling completeness R/S (%)

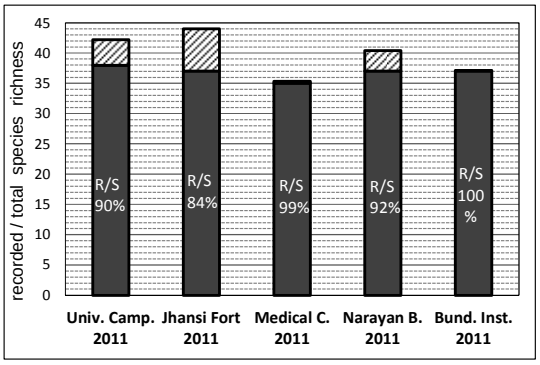


Fig. 2. Actually recorded species richness and extrapolated total species richness for year 2011. Provided are the numbers of species (i) recorded (black) and (ii) still unrecorded (hatched), and the sampling completeness R/S (%).

Here, we dispose, precisely, of such a series with the nine inventories of butterfly fauna conducted in Jhansi. Accordingly, the dependence of each of the f_x upon the level of completeness were computed directly from the regressed values of the f_x (Appendix 1) and the values of R_0, S_t and R_0/S_t given at Table 1. As a result, the numbers f_1, f_2, f_3, f_4, f_5 of species recorded 1-, 2-, 3-, 4-, 5- times are plotted at Fig. 5, against increasing sampling completeness levels. In the range of investigated completeness, the number f_1 of species recorded only once has already enter its phase of continuous decrease, while the numbers f_2, f_3, f_4, f_5 are still in their ascending

stage. For completeness levels up to ≈ 80%, f_1 remains higher than all the other f_x, but, then, falls successively under f_2, f_3, f_4, f_5, as completeness increases further. The same will happen, in turn, to f_2, which, although still being in its ascending phase, is surpassed by f_3 at ≈ 94% completeness. Later on, after sampling has reached exhaustivity, the same process, of course, will happen sequentially for the successive f_x. If the thresholds values of sampling completeness just mentioned above are *case-specific*, the global trends outlined above have general relevance and, indeed, conform to intuitive expectation. In particular, clearly highlighted here is the tight dependence between the level of sampling completeness and

Table 1. The sample size (number of sampled individuals) N_0, the number of recorded species R_0, the selected, *least-biased* type of estimator, the estimated number Δ of missing (unrecorded) species, the resulting estimated total species richness S_t and the resulting sampling completeness R_0/S_t (%), as computed for each of the nine inventories of butterfly fauna at Jhansi, carried out by Kumar [7,8]

	Univ. Camp. 2010	**Jhansi Fort 2010**	**Parich. Dam 2010**	**S.J.G. Highw. 2010**	**Univ. Camp. 2011**	**Jhansi Fort 2011**	**Medic. Coll. 2011**	**Naray. Bagh 2011**	**Bund. Inst. 2011**
Sample size N_0	182	186	125	145	179	173	175	257	164
nb. record. sp. R_0	25	25	22	21	38	37	35	37	37
selec. estimator	***JK.3***	***JK.1***	***JK.3***	***JK.1***	***JK.1***	***JK.1***	***Chao***	***JK.1***	***Chao***
nb. missing sp. Δ	9.0	4.6	11.8	4.0	4.2	7.1	0.3	3.4	0.1
total sp. richn. St	34.0	29.6	33.8	25.0	42.2	44.1	35.3	40.4	37.1
completeness	74%	84%	65%	84%	90%	84%	99%	92%	100%

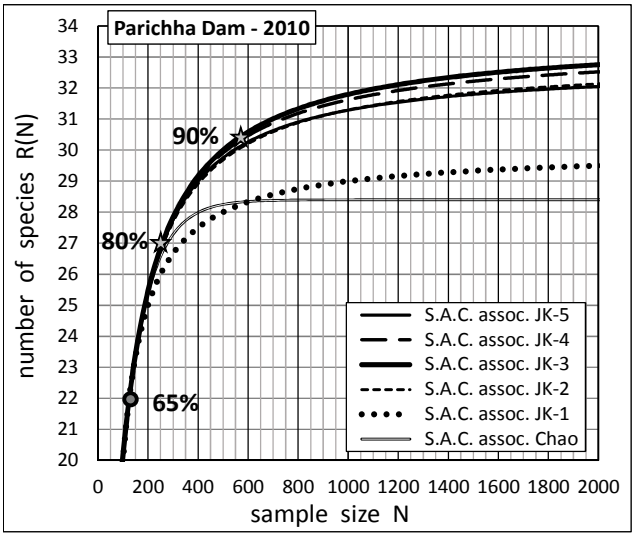


Fig. 3. Extrapolations of the Species Accumulation Curves for the inventory of butterfly fauna at Parichha Dam (2010). The grey point is for the actually performed sampling: N_0 = 125 individuals, $R(N_0)$ = 22 recorded species. The extrapolations respectively associated to each of the six estimators are plotted simultaneously (N.B.: JK-2 turns out to be almost confounded with JK-3). Here, the selected, least-biased extrapolation is according to JK-3. Extrapolations differ markedly according to the type of estimators, as is also the case for the estimates of the number Δ of missing species (from Δ = 6.4 for Chao to Δ = 11.8 for JK-3). Selecting the least-biased extrapolation is therefore very important, not only for a reliable extrapolation of Δ and of the total species richness S_t, but also for a reliable prediction of the extra-sampling effort required to reach a given level of completeness. According to the least-biased extrapolation (JK-3), reaching 80% or 90% completeness would require to increase the sampling size up to N ≈ 260 or N ≈ 570, while, according to the non-selected Chao estimator, the corresponding required sampling sizes would be N ≈ 140 or N ≈ 205

the respective value taken by each of the f_x relatively to the others. This, indeed, is at the very base of the – somewhat fascinating – idea of *being able to estimate the level of sampling completeness from the simple knowledge of the values of the few first f_x*, recorded in a partial sampling. And the commonly used non-parametric estimators of species richness (Chao, Jackknife series, etc.), ultimately find their deep explanation in this relationship between sampling completeness level and the numbers of species still collected at low frequency (once, twice, thrice, …, only).

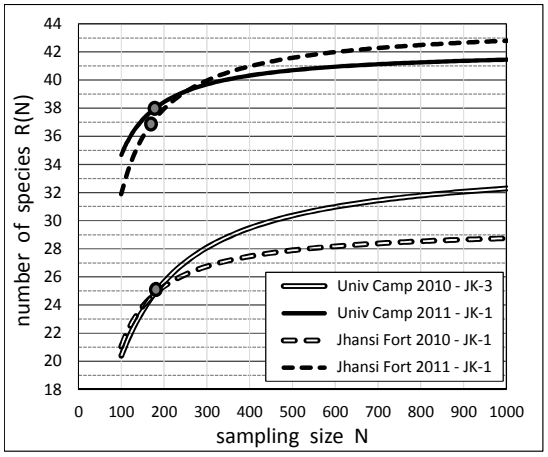


Fig. 4. Least-bias extrapolations of the Species Accumulation Curve for inventories of sites "University Campus" and "Jhansi Fort", during years 2010 and 2011. Actually realised inventories are marked by grey circles. Selected estimators for least-bias extrapolations: Jackknife-3 for University Camp 2010, Jackknife-1 for University Camp 2011 and for Jhansi Fort 2010 & 2011

4. DISCUSSION

4.1 Estimations of the True (Total) Species Richness of the Nine Sampled Assemblages

According to sites locations and years, the *recorded* species richness ranges from 21 to 38 species and the estimated *true* (*total*) species richness ranges from 25 to 44 species (Figs. 1 and 2, Table 1). Thus, most inventories prove being more or less incomplete (as was already expected from the remanence of various numbers of "singletons" among the recorded species), with the levels of completeness varying substantially according to sites and years: Table 1 and Figs. 1 and 2. The *as-recorded* species richness, thus, does not allow any reliable prediction regarding the true (total) richness. And this stands not only in term of absolute value but, as well, in term of relative value, i.e. when trying to compare several samples.

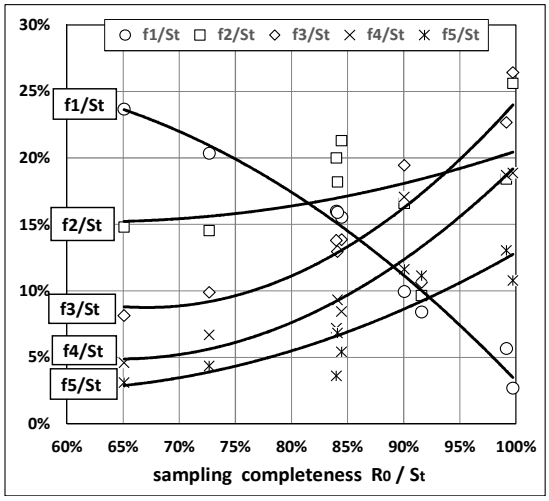


Fig. 5. The numbers f_1, f_2, f_3, f_4, f_5 of species recorded 1-, 2-, 3-, 4-, 5- times according to the estimated level of sampling completeness. As the nine inventories involved in the computation have different species richness S_t, it is the ratio f_x/S_t, rather than f_x itself, which makes sense and is relevantly plotted in this Figure

4.2 Equality of Sampling Sizes Does Not Mean Equality of Sampling Completeness and, thus, Does Not Allows Any Reliable Comparison Between True Species Richness

Moreover, even made at a same sampling size, the comparison between inventories *does not allow* any reliable prediction of total species richness, in general. This is simply because, as a rule, completeness and sampling size are very poorly correlated, as demonstrated at Fig. 6. Thus, in general, the equality of sampling sizes *does not* guarantee the equality of the levels of sampling completeness. As equal levels of sampling completeness are required to make meaningful comparisons between partial inventories, it follows that, in any case, *extrapolation is mandatory*, prior to any further speculation!

And, of course, *least-biased* extrapolations are especially desirable in this perspective.

Figs. 4 and 7 illustrate more directly the pitfalls resulting from a systematic trust in the validity of

comparisons between inventories having same sampling sizes (or with sampling sizes brought back to a same value using the classical procedure of "rarefaction" [1,3]).

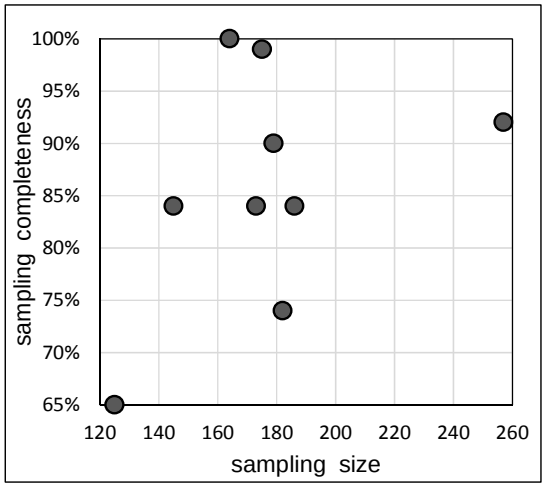


Fig. 6. Very weak correlation between sampling completeness (%) and sampling size: determination coefficient $r^2 = 0.18$; which means that the variations of the level of sampling completeness are explained by sampling size for less than 20%.

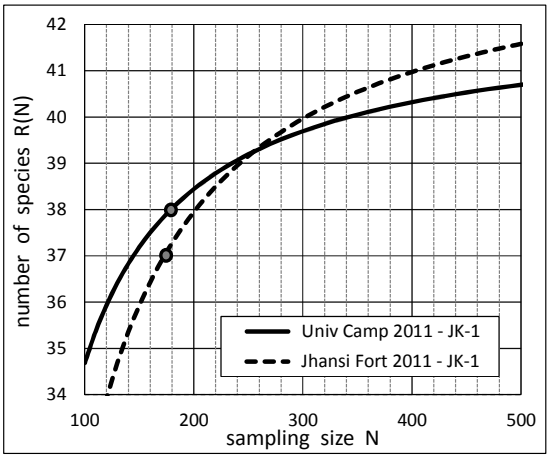


Fig. 7. Enlargement of Fig. 4 showing the intersection and crossing over, at size N ≈ 260, between the two Species Accumulation Curves (for University Campus 2011 and Jhansi Fort 2011). Thus, comparing both inventories at a same size may lead to contradictory conclusions regarding the expected total species richness of each site: for a same size N < 260, University Camp would be expected to be the richest, while for a larger same size, N > 260, it is now Jhansi Fort that would be expected to be the richest.

Yet, referring to the equalisation of sampling sizes, using "rarefaction" procedure still remains, regrettably, a common practice. For example, DE VRIES & WALLA [9] still implement "rarefaction" to compare inventories of butterfly fauna carried on at different height, areas and periods of investigation in an Ecuador tropical forest. And this, although the authors actually recognized that Species Accumulation Curves may well intersect (their Fig. 3). This, indeed, is no lack of rigour from the authors but, as mentioned above, an understandable reluctance to consider extrapolation methods, as long as no reliable procedure was made available to select the appropriate, minimum-bias solution among the wide series of nonparametric estimators of total species richness desribed in the literature. Now that such a selection procedure is made available, this reluctance is no longer justified!

This clearly demonstrate the pitfalls attached to systematically trust in the validity of comparisons between inventories, even having same sampling sizes (or with sampling sizes brought back to a same value by the classical procedure of "rarefaction"): see reference [10] in particular. Only *implementing a reliable extrapolation procedure allows to conclude relevantly* (here, that Jhansi Fort actually has the larger total species richness).

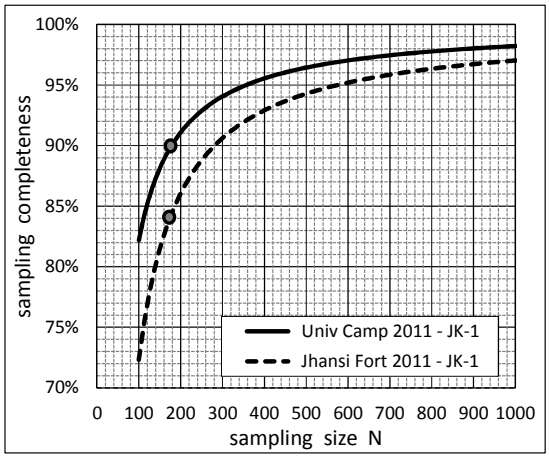


Fig. 8. Compared sampling completeness [$R(N)/S_t$] between the inventories of University Campus 2011 and Jhansi Fort 2011 (extrapolations as for Figs. 4 and 7). Completeness level remains consistently different (higher) for University Campus as compared to Jhansi Fort, and this at any sampling size.

4.3 Variations of True Species Richness between Two Successive Years

Least-biased extrapolations show that the levels of sampling completeness are globally higher in 2011 as compared to 2010 (Figs. 1 and 2 and Table 1). This contributes to the higher levels of *recorded* species richness in 2011 as compared to 2010, but higher *true* (total) species richness in 2011 may also be involved. This, at least, is the case considering the two sites - "University Campus" and "Jhansi Fort" - for which inter-annual comparisons are possible: the estimated total species richness S_t actually reveals higher in 2011 as compared to 2010: Fig. 9.

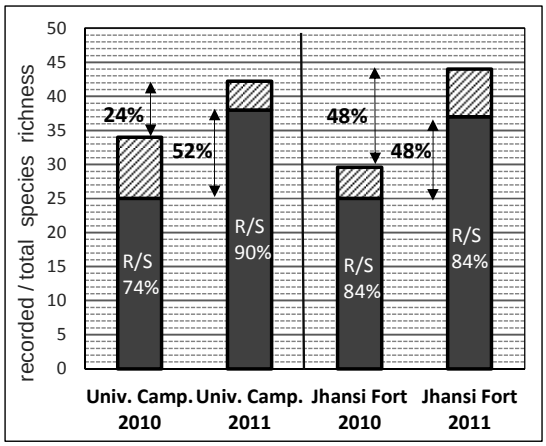


Fig. 9. Recorded species richness and extrapolated total species richness, *comparing* "University Campus" *to* "Jhansi Fort" and year 2011 *to* year 2010. Provided are the numbers of species recorded (black) and unrecorded (hatched), the sampling completeness R/S (%) and the increment (%) of recorded and total species richness when comparing year 2011 to year 2010

More precisely:

* at "Jhansi Fort", the relative increment in total species richness in 2011 reaches 48% (incidentally, this is the same relative increment as for recorded species richness, simply due to the purely coincidental equality of sampling completeness for years 2010 and 2011 [as a rule, sampling completeness would more or less differs substantially between inventories at different dates or different locations]).

* at "University Campus", the estimated increment of total species richness in 2011 is 24%, while the increment of recorded species is far larger: 52%, due to 2011 inventory being far more complete - 90% as compared to 74% in 2010. Accordingly, the 52% increment of recorded species cumulates both the true increase of total species richness and the consequence of more complete sampling in 2011.

Thus, in conclusion, year 2011 actually shows an appreciable enrichment in butterfly true diversity (24% or 48%, as compared to 2010), at least for the two investigated sites: Fig. 9.

While *monthly or seasonal* variations of species richness along one year have been studied and reported rather often, *yearly* (inter-annual) variations of species richness have rarely been addressed otherwise.

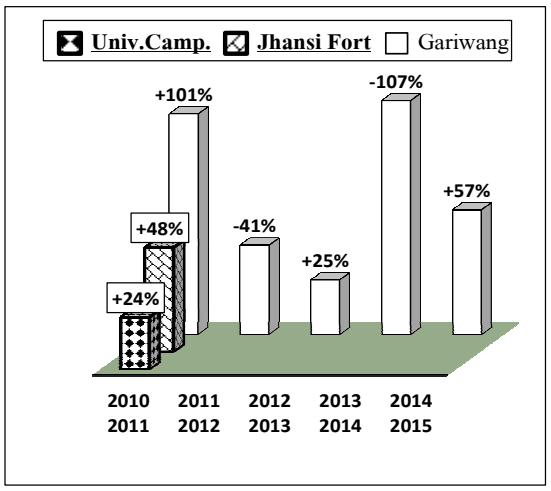


Fig. 10. Variation of estimated true species richness at University Campus and Jhansi Fort between 2010 and 2011 and, for comparison, the inter-annual variations of estimated true species richness at mount Gariwang-San (S-Korea), extrapolated (BÉGUINOT *unpublished*) from field data by LEE *et al.*: [11]. [Note the alternating sign of the estimated true species richness along successive years at Mount Gariwang-San]

Yet, a continuous five years-long study of the variations of butterfly species richness at Mount Gariwang-San (S-Korea) has been reported [11]. After having subjected the crude, as-recorded data to least-biased extrapolation (BÉGUINOT *unpublished data*), the inter-annual variations of estimated true richness were quantified as +101%, –41%, +25%, –107% and +57%, successively, during the period 2010 to 2015:

Fig. 10 above. On the other hand, a much more limited range of inter-annual variations (5% to 15% variations, yet based on crude, as-recorded richness only) is reported in [9] for the butterfly fauna of an undisturbed tropical forest in Ecuador.

The yearly variations of the butterfly fauna around Jhansi, between 2010 and 2011, thus fall in an intermediate range.

5. CONCLUSION

Incomplete inventories of local biodiversity, which are the ordinary rule in practice, at least for speciose taxonomic groups, may yet provide *much more information* than the crude recorded data would let suppose. Releasing this additional information requires, however, that inventories include also the respective abundances of the recorded species. Under this condition, extrapolating the Species Accumulation Curve, beyond the actually realised inventory, may easily be considered. Reliable extrapolation, however, is conditioned by the rational selection, for each inventory, of the *least-biased* estimator of total species richness, among the series of estimators made available in the literature. This selection may now be implemented using the procedure described in [4] and summarised in practice at Appendix 2.

In turn, such reliable extrapolations may allow to address a series of issues that could not have been answered properly otherwise, as shown above with a few examples.

ACKNOWLEDGEMENTS

Ashok KUMAR of Lucknow University (Uttar Pradesh, India) is gratefully acknowledged for the achievement and detailed publication [7, 8] of all nine inventories involved in the present analysis. The author also thanks four anonymous reviewers for their appreciations and useful comments on the original manuscript

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

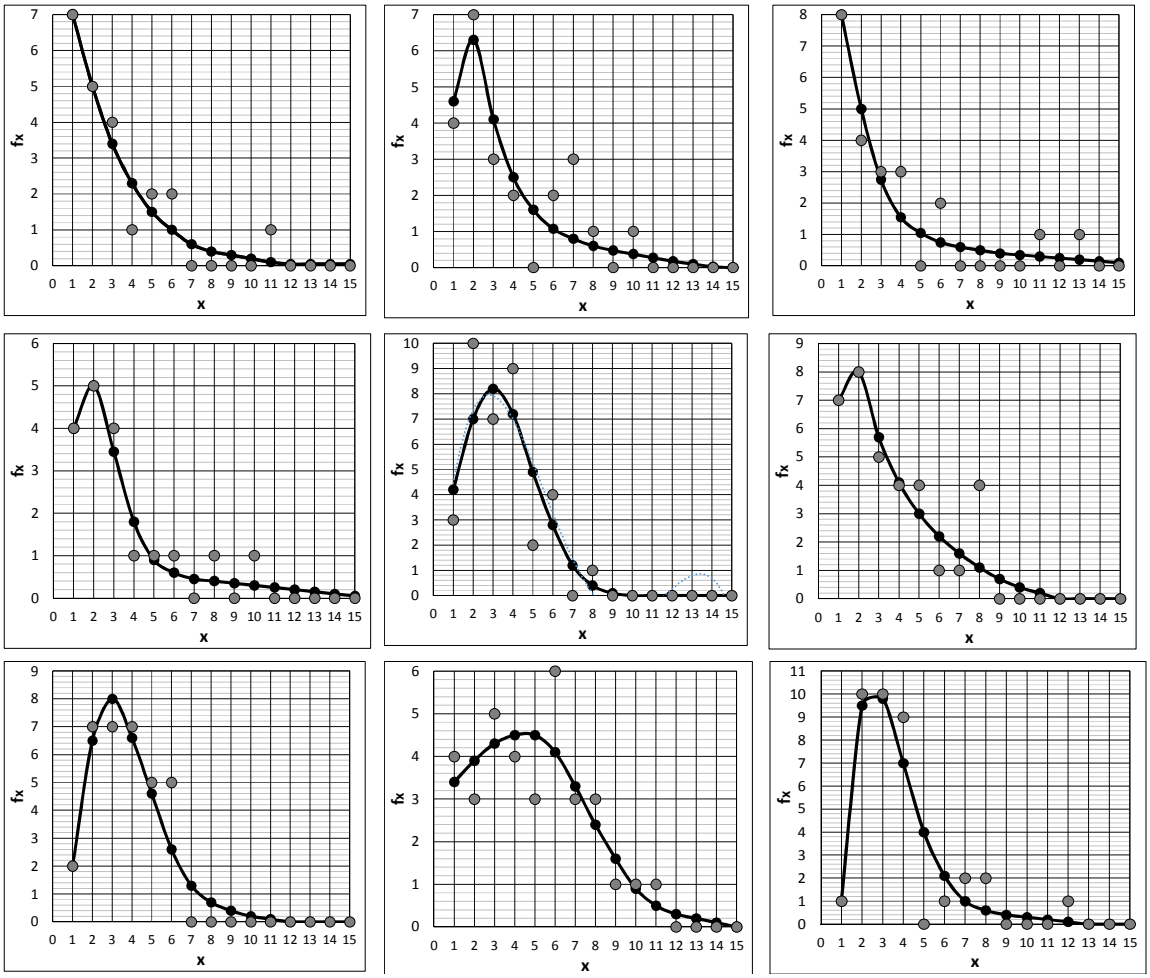
1. Gotelli NJ, Chao A. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In: Levin SA (ed.) Encyclopedia of Biodiversity, Academic Press, second edition. 2013;5:195-211.
2. Béguinot J. Extrapolation of the species accumulation curve for incomplete species samplings: A new nonparametric approach to estimate the degree of sample completeness and decide when to stop sampling. Annual Research & Review in Biology. 2015;8(5):1-9.
DOI: 10.9734/ARRB/2015/22351; <hal-01238720>
3. Colwell RK, Coddington JA. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society London B. 1994;345:101-118.
4. Béguinot J. Theoretical derivation of a bias-reduced expression for the extrapolation of the Species Accumulation Curve and the associated estimation of total species richness. Advances in Research. 2016;7(3):1-16.
DOI: 10.9734/AIR/2016/26387; <hal-01367803>
5. Béguinot J. Extrapolation of the Species Accumulation Curve associated to “Chao” estimator of the number of unrecorded species: A mathematically consistent derivation. Annual Research & Review in Biology. 2016;1(4):1-19.
DOI: 10.9734/ARRB/2016/30522
6. Brose U, Martinez ND, Williams RJ. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. Ecology. 2003;84(9): 2364-2377.
7. Kumar A. A report on the Butterflies in Jhansi (U.P.) India. Journal Applied & Natural Science. 2012;4(1):51-55.
8. Kumar A. Butterfly abundance and species diversity in some urban habitats International Journal of Advanced Research. 2014;2(6):367-374.
9. De Vries PJ, Walla TR. Species diversity and community structure in neotropical fruit-feeding butterflies. Biological Journal of the Linnean Society. 2001;74:1-15.
10. Lande R, DeVries PJ, Walla TR. When species accumulation curves intersect: implications for ranking diversity using small samples. OIKOS. 2000;89(3):601-605.
11. Lee CM, Kim S-S, Kwon T-S. Butterfly fauna in Mount Gariwang-san, Korea.

Journal of Asia-Pacific Biodiversity. 2016;9:198-204.

12. Béguinot J. An algebraic derivation of Chao's estimator of the number of species in a community highlights the condition allowing Chao to deliver centered estimates. ISRN Ecology; 2014. Article ID 847328. DOI:10.1155/2014/847328; <hal-01101415>
13. Béguinot J. When reasonably stop sampling? How to estimate the gain in newly recorded species according to the degree of supplementary sampling effort. Annual Research & Review in Biology. 2015;7(5):300-308. DOI: 10.9734/ARRB/2015/18809; <hal-01228695>
14. Béguinot J. Extrapolation of the species accumulation curve associated to "Chao" estimator of the number of unrecorded species: A mathematically consistent derivation. Annual Research & Review in Biology. 2016;11(4):1-19. DOI: 10.9734/ARRB/2016/30522

APPENDICES

Appendix 1 - Regressions on the distributions of recorded f_x to reduce the consequences of drawing stochasticity



Figs. A1.1 to A1.9 – The recorded values of the numbers f_x of species recorded x-times (grey discs) and the regressed values of f_x (black discs) so as to reduce the consequence of stochastic dispersion. Successively from left to right and from top to bottom : University Campus 2010, Jhansi Fort 2010, Parichha Dam 2010, side of J.G. highway 2010, University Campus 2011, Jhansi Fort 2011, Medical College 2011, Narayan Bagh 2011, BIET 2011

Appendix 2 - Bias-reduced extrapolation of the Species Accumulation Curve and associated bias-reduced estimation of the number of missing species, based on the recorded numbers of species occurring 1 to 5 times

Consider the survey of an assemblage of species of size N_0 (with sampling effort N_0 typically identified either to the number of recorded individuals or to the number of sampled sites, according to the inventory being in terms of either species abundances or species incidences), including $R(N_0)$ species among which f_1, f_2, f_3, f_4, f_5, of them are recorded 1, 2, 3, 4, 5 times respectively. The following procedure, designed to select the less-biased solution, results from a general mathematical relationship that constrains the theoretical expression of *any* theoretical Species Accumulation Curves R(N) [4, 12, 13, 14]:

$$\partial^x R_{(N)}/\partial N^x \;=\; (-1)^{(x-1)}\, f_{x(N)}/C_{N,\,x} \;\approx\; (-1)^{(x-1)}\,(x!/N^x)\, f_{x(N)} \quad (\approx \text{as } N >> x) \qquad [A.1]$$

Compliance with the mathematical constraint [1] warrants *reduced-bias* expression for the extrapolation of the Species Accumulation Curves R(N) (i.e. for $N > N_0$). Below are provided, accordingly, the polynomial solutions $R_{\mathbf{x}}(N)$ that respectively satisfy the mathematical constraint [A.1], considering increasing orders x of derivation $\partial^x R_{(N)}/\partial N^x$. Each solution $R_{\mathbf{x}}(N)$ is appropriate for a given range of values of f_1 compared to the other numbers f_x. According to [4] :

* *for* f_1 *up to* f_2 → $R_{\mathbf{1}}(N) = (R(N_0) + f_1) - f_1.N_0/N$

* *for* f_1 *up to* $2f_2 - f_3$ → $R_{\mathbf{2}}(N) = (R(N_0) + 2f_1 - f_2) - (3f_1 - 2f_2).N_0/N - (f_2 - f_1).N_0^2/N^2$

* *for* f_1 *up to* $3f_2 - 3f_3 + f_4$ → $R_{\mathbf{3}}(N) = (R(N_0) + 3f_1 - 3f_2 + f_3) - (6f_1 - 8f_2 + 3f_3).N_0/N - (-4f_1 + 7f_2 - 3f_3).N_0^2/N^2 - (f_1 - 2f_2 + f_3).N_0^3/N^3$

* *for* f_1 *up to* $4f_2 - 6f_3 + 4f_4 - f_5$ → $R_{\mathbf{4}}(N) = (R(N_0) + 4f_1 - 6f_2 + 4f_3 - f_4) - (10f_1 - 20f_2 + 15f_3 - 4f_4).N_0/N - (-10f_1 + 25f_2 - 21f_3 + 6f_4).N_0^2/N^2 - (5f_1 - 14f_2 + 13f_3 - 4f_4).N_0^3/N^3 - (-f_1 + 3f_2 - 3f_3 + f_4).N_0^4/N^4$

* *for* f_1 *larger than* $4f_2 - 6f_3 + 4f_4 - f_5$ → $R_{\mathbf{5}}(N) = (R(N_0) + 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5) - (15f_1 - 40f_2 + 45f_3 - 24f_4 + 5f_5).N_0/N - (-20f_1 + 65f_2 - 81f_3 + 46f_4 - 10f_5).N_0^2/N^2 - (15f_1 - 54f_2 + 73f_3 - 44f_4 + 10f_5).N_0^3/N^3 - (-6f_1 + 23f_2 - 33f_3 + 21f_4 - 5f_5).N_0^4/N^4 - (f_1 - 4f_2 + 6f_3 - 4f_4 + f_5).N_0^5/N^5$

The associated non-parametric estimators of the number Δ_J of missing species in the sample [with $\Delta_J = R(N=\infty) - R(N_0)$] are derived immediately:

* *for* f_1 *up to* f_2 → $\Delta_{J1} = f_1$

* *for* f_1 *up to* $2f_2 - f_3$ → $\Delta_{J2} = 2f_1 - f_2$

* *for* f_1 *up to* $3f_2 - 3f_3 + f_4$ → $\Delta_{J3} = 3f_1 - 3f_2 + f_3$

* *for* f_1 *up to* $4f_2 - 6f_3 + 4f_4 - f_5$ → $\Delta_{J4} = 4f_1 - 6f_2 + 4f_3 - f_4$

* *for* f_1 *larger than* $4f_2 - 6f_3 + 4f_4 - f_5$ → $\Delta_{J5} = 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5$

N.B. 1: for f_1 falling beneath 0.6 x f_2 (that is when sampling completeness closely approaches exhaustivity), then Chao estimator may be selected: see reference [14].

N.B. 2: in order to reduce the influence of drawing stochasticity on the values of the f_x, the as-recorded distribution of the f_x should preferably be smoothened: this may be obtained either by rarefaction processing or by regression of the as-recorded distribution of the f_x versus x.

© 2017 Béguinot; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/18074