Shelled-Molluscs Fauna at Abrolhos Bank (Brazil): Assessment of Both Total Species Richness and the Completed Distribution of Species Frequencies by Numerical Extrapolation of a Partial Survey

Jean Béguinot$^{1,2}$*

$^{1}$Société Histoire Naturelle-Bourgogne Nature, 7 bvd H.P. Schneider, 71200 Le Creusot, France. $^{2}$Université de Bourgogne, 6, Boulevard Gabriel, 21000 Dijon, France.

Author's contribution

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

Article Information

DOI: 10.9734/ARRB/2020/v35i1130304

Editor(s):
(1) Dr. Bechan Sharma, University of Allahabad, India.

Reviewer(s):
(1) U. K. Maurya, ICAR- National Bureau of Soil Survey & Land Use Planning (NBSS&LUP), India.
(2) Victor Ribeiro Cedro, Universidade Federal de Alagoas (UFA), Brazil.
Complete Peer review History: http://www.sdiarticle4.com/review-history/62548

Received 22 August 2020
Accepted 28 October 2020
Published 26 November 2020

ABSTRACT

Numerous anthropogenic threats to the exceptionally rich coral-reef ecosystem at Abrolhos Bank (Brazil) arguably require implementing drastic conservation policy and meanwhile, urge for the prior detailed assessment of species richness and the species distribution across the Bank. Due to their unavoidable incompleteness, the already implemented “Rapid Assessment Surveys” at Abrolhos Bank deserve being completed, at least numerically, by implementing an appropriate extrapolation procedure, to avoid serious bias precisely due to ignoring both the number and the frequency distribution of those species still remaining undetected after Rapid Assessment Surveys. Complying with this concern, I report on the results of a numerical extrapolation of a previously achieved partial survey of the soft-bottom shelled-mollusc fauna at Abrolhos Bank. This numerical extrapolation provides least-biased estimates regarding not only the number of species which have remained unrecorded but, moreover, additional information on the respective frequencies of these still unrecorded species. As a result, the extrapolated total species richness at Abrolhos Bank reaches at least 435 shelled-molluscs species (instead of only 293 species actually recorded by the Rapid Assessment Survey), out of which 30 species (instead of 19) are expected to be Brazilian.

*Corresponding author: E-mail: jean-beguinot@orange.fr;
endemics. Accordingly, the soft-bottom shelled-mollusc fauna – an admittedly fairly reliable indicator for the whole marine biodiversity – definitely demonstrate the major biological interest of the whole reef ecosystem at Abrolhos Bank and the imperative necessity of implementing truly efficient conservation programs of this ecosystem.

Keywords: Rapid assessment survey; marine biodiversity; coral reef; bivalve; gastropod; sampling effort.

1. INTRODUCTION

The Abrolhos Bank (Brazil), while being considered as the largest and richest reef complex in the south-west Atlantic [1,2], is yet submitted to steadily increasing detrimental pressures and threats, in particular from overfishing, tourism, nearby mangrove loss and oil drilling programs [1]. Molluscan fauna – by far the most diverse group in marine environment and a fairly good indicator for marine biodiversity as a whole [3] – is expected, accordingly, to be a major contributor to the marine biodiversity at Abrolhos Bank. Yet, much remains to be surveyed in this respect, especially regarding the smallest and the rarest molluscan species.

An extended survey is therefore urgently needed as a prerequisite for the properly informed implementation of appropriate conservation programs. Yet, the exhaustive survey of so rich an ecosystem is clearly far beyond practical reach and, here as elsewhere, only ‘Rapid Assessment Programs’ can reasonably be carried out. Such rapid – but thus partial – assessments are likely being the best compromise in practice, given, on the one hand, the multiplicity of demands of urgent surveys of biodiversity worldwide (especially under tropical climate) and, on the other hand, the limited available resource to be devoted to field investigations in general.

In this perspective, a Rapid Assessment has been carried out and reported by R.S. ABSALAO [4], focused on the soft-bottom shelled-molluscan fauna at Abrolhos Bank. No less than 293 species were recorded, viz: 229 species of Gastropods, 53 species of Bivalves, 6 species of Polyplacophores and 5 species of Scaphopodes.

These 293 species represent, however, only an unknown proportion of the richness of the whole soft-bottom molluscan fauna at Abrolhos: a likely substantial underestimate of the actual true species richness, since Rapid Assessment Surveys only lead to more or less incomplete samplings [4]. Accordingly, the present study aims at improving this reported incomplete field data, by implementing a recently developed procedure of least-biased numerical extrapolation of the so-called “Species Accumulation Curve” (which accounts for the kinetics of increase in the number of recorded species during progressive sampling). In turn, this numerical extrapolation helps to address the following issues:

- To provide a least-biased estimate of the number of unrecorded species and, thereby of the true species richness of the molluscan fauna at Abrolhos Bank;
- To provide the estimated complete distribution of species frequencies for the soft-bottom molluscan fauna, that is, including also the estimated frequencies of occurrence of each of the still unrecorded species;
- To furthermore focus on the Brazilian endemic species occurring at Abrolhos Bank, highlighting specifically their estimated number and their estimated respective frequencies of occurrence on the Bank;
- At last, to anticipate the additional sampling effort that would be required to gain any chosen increment in the actually recorded species richness. This, with the desirable prospect of a future reinforcement of survey operations at Abrolhos Bank.

2. MATERIALS AND METHODS

2.1 The Reported Field Data

The present study is based on the survey of the molluscan fauna at 39 sites at Abrolhos Bank, carried out during the expedition devoted to the “Abrolhos Rapid Assessment Survey Program”, and reported by R.S. ABSALAO [4]. As mentioned above, the Abrolhos Bank (centered approximately 18° S) harbors the richest coral-reef ecosystem located off the Brazilian coast, but is, unfortunately, threatened by a series of detrimental anthropogenic pressures [1], hence
the interest and the urgency of performing this survey. The corresponding sampling procedure and the collected data are provided in detail in the open-access reference [4] and, accordingly, needs not being further repeated here.

Due to the substantial incompleteness of the reported samplings (almost unavoidable when having to deal with such species-rich communities, comprising a large part of rare species), the numerical extrapolation of these incomplete samplings is required to get a relevant envision of the field reality.

2.2 The Numerical Extrapolation Procedure and Its Exploitation

Numerical extrapolation of the reported Abrolhos Rapid Assessment Survey was therefore implemented prior to further analysis, thereby aiming at avoiding biased inferences which can likely result from ignoring the substantial set of more or less rare species that remained unrecorded [5]. This is all the more important that rare species (beyond their own intrinsic interest) may disproportionately contribute to the functional structuring of communities, as often emphasized [6-16]: “rare species are critical for bio-assessment” as quoted in [16]. The numbers \( N_0 \) of observed individuals and the numbers \( R_0 \) of recorded species for the whole community and for each of its two main components, Gastropods and Bivalves, are given in Table 1. Numerical extrapolation can thus serve as a convenient palliative to the unavoidable lack of exhaustive inventories, which otherwise would be required [17–22]. As quoted in reference [22], “virtually always, species richness cannot be observed but needs to be estimated because some species may be present but remain undetected. This fact is commonly ignored in ecology and management, although it will bias estimates of species richness and related parameters”.

Furthermore, beyond the first aim of estimating the number of unrecorded species, the newly developed extrapolation procedure can provide, in addition, fairly accurate estimates of the respective frequencies of occurrence of each of these unrecorded species, as detailed in sections below. Numerically completed this way (and only when it is so [19]), the distribution of species frequencies can further reveal some qualitative and quantitative aspects of interest regarding the underlying process that governs the spatial structuration of species occurrences across the studied communities or ecosystem [23-27].

2.2.1 Estimation of the total species richness

The least-biased estimation of the number of still unrecorded species after partial sampling and the resulting estimation of the total species richness, \( S_0 \), of the partially sampled community are computed according to the procedure defined in [28,29] and briefly summarized in Appendix 1, on the basis of the numbers \( f_i \) of species observed \( x \)-times during partial sampling (\( x = 1 \) to 5). The same procedure allows to further derive the least-biased extrapolation of the “Species Accumulation Curve”, which can also predict, in turn, the expected increase in the number of recorded species, \( R(N) \), as a function of the growing sampling size \( N \) (\( N: \) number of currently recorded individuals); see Appendix 1 for the main aspects of the computation procedure. In practice, this extrapolation of the Species Accumulation Curve allows to forecast and, thereby, to anticipate the expected additional sampling efforts that would be required to obtain any desirable increment in sampling completeness.

2.2.2 Extrapolation of the distribution of species frequencies

As is the case for the Distribution of Species Abundance, the Distribution of Species Frequencies requires.

(i) To be corrected for the bias resulting from drawing stochasticity during samplings of finite sizes and, still more importantly,

(ii) To be completed by numerical extrapolation addressing the set of undetected species (to the extent that sampling is suspected to be incomplete, as revealed by the subsistence of “uniques” [i.e. species only recorded once]). The appropriate procedure of correction and of numerical extrapolation of the distribution of species frequencies (formally treated the same way as the distribution of species abundances) is described in details in reference [30], briefly summarized in Appendix 2 and exemplified in details in reference [31]. Classically, the Species Frequencies Distribution is graphically presented with the (log-transformed) frequencies, \( a_i \), plotted against the rank \( i \) of species, the latter being ordered by decreasing values of frequencies (with, thus, \( a_1 \) and \( a_{\text{rn}} \) respectively standing for the highest and the lowest frequencies in a community of \( S_i \) co-occurring species).
3. RESULTS

3.1 Estimated Total Species Richness for the Molluscan Fauna as a Whole and for Its Two Main Components

Table 1 provides the least-biased estimates of the total species richness of shelled-molluscan fauna inhabiting soft-bottoms at Abrolhos Bank: at first as a whole and subsequently for each of its two main components, Bivalves and Gastropods.

With sampling completeness level around two third (Table 1), it could be considered of interest to improve the sampling completeness by further pursuing investigations on the bank. Yet, to make a rationally based decision as to whether it seems materially possible or not to pursue sampling operation any further, the extrapolation of the species accumulation curve beyond the actual sampling size deserves being considered: Figs. 1 and 2. As said, this extrapolation allows, in turn, to predict and thus to anticipate the additional sampling effort that would be required to obtain any desired increment in sampling completeness: Fig. 2.

3.2 Corrected and Extrapolated Distributions of Species Frequencies

The corrected and extrapolated Distributions of Species Frequencies are provided in Figs. 3 to 5, for the estimated:

Table 1. The number of collected individuals $N_0$, the number of recorded species $R_0$, the type of nonparametric estimator (Jackknife) selected as being the least-biased one, the estimated number $\Delta$ of unrecorded species, the resulting estimate of the “true” total species richness $S_t$ ($= R_0 + \Delta$), the resulting estimated level of sampling completeness $R_0/S_t$.

Estimations are computed according to the least-biased procedure [28], the selection key being provided in Appendix 1

<table>
<thead>
<tr>
<th>Molluscan Fauna Abrolhos</th>
<th>ALL Taxa</th>
<th>Bivalves</th>
<th>Gastropods</th>
<th>Other taxa</th>
</tr>
</thead>
<tbody>
<tr>
<td>nb. collected individuals $N_0$</td>
<td>1019</td>
<td>173</td>
<td>826</td>
<td>20</td>
</tr>
<tr>
<td>nb. recorded species $R_0 = R(N_0)$</td>
<td>293</td>
<td>53</td>
<td>229</td>
<td>11</td>
</tr>
<tr>
<td>selected least-biased estimator</td>
<td>Jackknife-5</td>
<td>Jackknife-3</td>
<td>Jackknife-4</td>
<td>/</td>
</tr>
<tr>
<td>number unrecorded species $\Delta$</td>
<td>142</td>
<td>27</td>
<td>102</td>
<td>13</td>
</tr>
<tr>
<td>estimated total species richness $S_t$</td>
<td>435</td>
<td>80</td>
<td>331</td>
<td>24</td>
</tr>
<tr>
<td>sampling completeness $R_0/S_t$</td>
<td>67 %</td>
<td>66 %</td>
<td>69 %</td>
<td>46 %</td>
</tr>
</tbody>
</table>

Fig. 1. Extrapolated part of the species accumulation Curve accounting for the increase of the number of detected species $R(N)$ as a function of growing sample size $N$ beyond the actually achieved sampling ($N_0 = 1019$, $R(N_0) = 293$). Superimposed, here, are the extrapolations computed for six non-parametric estimators: Chao and the Jackknife series for orders 1 to 5. Substantial differences between these extrapolations highlight the importance of rationally selecting the least-biased extrapolation, corresponding, here, to the nonparametric estimator Jackknife-5.
Fig. 2. The least-biased extrapolation of the species accumulation curve (according to the selected estimator, here, Jackknife-5) highlighting the expected additional sampling effort, $N$, required to obtain a given increment in the number $R(N)$ of recorded species: For example, increasing completeness from the actual level of 67% up to 80%, or 90%, or 95% levels would require increasing sampling efforts from the actual sampling size $N_0 = 1019$ to $N \approx 2000$, $\approx 4200$, $\approx 8200$, respectively.

Fig. 3. The numerically completed distribution of species frequencies of occurrence for the whole soft-bottom molluscan fauna at Abrolhos Bank: (i) Discs: the 293 already recorded species, ranks 1 to 293; (ii) Double line: the estimated 142 species remaining still unrecorded, ranks $i = 294$ to 435.
Fig. 4. The numerically completed distribution of species frequencies of occurrence for the Bivalves subset: (i) Discs: the 53 already recorded species, ranks 1 to 53; (ii) Double line: the 27 still unrecorded species, ranks i = 54 to 80

Fig. 5. The numerically completed distribution of species frequencies of occurrence for the Gastropods subset: (i) Discs: the 229 already recorded species, ranks 1 to 229; (ii) Double line: the 102 still unrecorded species, ranks i = 230 to 331

(i) 435 species of the whole soft-bottom molluscan fauna,
(ii) 80 species of Bivalves (Pelecypoda),
(iii) 331 species of Gastropods, respectively.

The species are, classically, ranked by decreasing order of frequencies.

3.3 Brazilian Endemics

Estimation of their total number and the distribution of their respective frequencies of occurrence across Abrolhos Bank.

Among the 293 recorded species, 19 are recognized as Brazilian endemics, all of them belonging to Gastropods [4]. The procedure of least-biased extrapolation was applied to this sub-group of endemic species (as it can be applied to any other kinds of subsets, as argued in [28,32]) leading to an estimated 11 unrecorded endemic species and, accordingly, an estimated total of no less than 30 endemic species for the sampled area of Abrolhos Bank.

The numerically completed distribution of species frequencies, among these 30 endemic species,
is provided in Fig. 7. The comparison with the
distribution of species frequencies within the
whole molluscan fauna (Fig. 3) shows that the
frequencies of these endemic species are
approximately evenly distributed all across the
range of frequencies of the whole molluscan
fauna at Abrolhos Bank.

The sampling completeness for endemic species
turns out to be just slightly lower (63%) than it is

Fig. 6. The least-biased extrapolation of the species accumulation curve (according to the
selected estimator, here also, Jackknife-5) for the subset of Brazilian endemic species,
highlighting the expected additional sampling effort N required to obtain a given increment in
the number R(N) of recorded endemic species: for example, increasing completeness from
63% to 80%, 90%, 94% would require increasing sampling efforts from N₀ = 1019 to N ≈ 2300,
≈ 5200, ≈ 9600, respectively

Fig. 7. The numerically completed distribution of species frequencies of occurrence for the
subset of Brazilian endemic species: (i) Discs: the 19 already recorded species, ranks 1 to 19;
(ii) Double line: the estimated 11 species still remaining unrecorded: ranks i = 20 to 30
for the whole molluscan fauna (67%). As shown above for the whole molluscan fauna, the numerical extrapolation of the Species Accumulation Curve (Fig. 6) allows to forecast the additional sampling effort required to record new endemic species.

4. DISCUSSION

Conservation concerns regarding threatened areas, such as Abrolhos Bank, incite, first, to carry out biodiversity surveys aiming at being both quickly completed and as comprehensive as possible [33].

Unfortunately, simultaneously complying with both these requirements is usually beyond practical and/or economical possibilities, given the ordinarily limited available resource to be devoted. Hence the policy of implementing, as a surrogate, ‘Rapid Biodiversity Assessments’, as those performed at Abrolhos Bank, off the Brazilian coast [4] with, usually, substantial – although unknown – degree of sampling incompleteness, as an unavoidable consequence. In turn, such uncontrolled sampling incompleteness will inevitably lead to unacceptably biased estimates of species richness and regrettably truncated knowledge relative to the distribution of species frequencies.

Hence also, the interest of implementing convenient numerical extrapolations of the actually achieved incomplete samplings, to reach – at least numerically – exhaustive surveys.

While numerical extrapolations remain, of course, silent as regards the identities of the unrecorded species they have, yet, the major practical advantage of providing – with minimum time and resource expenditure – the reliable estimates of both the number of unrecorded species and the distribution of their respective frequencies of occurrence. And the “silence” of numerical extrapolation, as regards the taxonomic identities of unrecorded species, has finally limited inconvenience, since unrecorded species – most of them being rare – often remain taxonomically ill-defined presently, especially among invertebrates [3] and, accordingly, would simply be anonymously accounted as “morphospecies” or “RTU” (i.e. Recognizable Taxonomic Units), if any further additional sampling efforts could have been conducted [3].

As a whole, numerical extrapolations – at least when conducted in compliance with sound and accurate procedures [28–30] – can therefore provide very valuable quantitative information regarding both the number and the more or less uneven frequencies and spatial distribution of species across the studied ecosystem – and this, in an incomparably shorter time and at very significantly lowered cost.

Among the various threats to Abrolhos Bank already highlighted in Introduction section, the prospect of large oil drilling operations should be emphasized, as likely being particularly harmful to local marine biodiversity and becoming more prominent at present. This more specifically justifies the urgency of having undertaken the numerical extrapolations of the previously reported Rapid Assessment (yet waiting for the desirable future achievement of some more complete samplings).

As expected, the molluscan species richness at Abrolhos Bank proves being substantially larger than suggested by Rapid Assessment, reaching an estimated level of 435 species, out of which 142 species were thus remained unrecorded (Table 1, Figs. 1, 2). The corresponding degree of sampling incompleteness typically falls in the usual range for marine biodiversity rapid assessments in tropical areas [34]. Now, here, this estimated species richness of 435 units far exceeds what is usually reported in tropical coral reefs: at the local scale, rarely more than one hundred species, even often less are usually reported [31,35,36]. Indeed, this high level of biodiversity at Abrolhos Bank may also be partly due, here, to the comprehensive sampling of the smaller-size species as well (i.e. including species less than 10 mm long, a range of size which still remains often neglected during most surveys).

Among these 435 shelled-mollusc species, an estimated 30 species are Brazilian endemics, out of which 11 were let unrecorded (Fig. 7). These 11 unrecorded endemic species, thus remain to be detected, which would yet require substantial supplementary sampling effort to be consented in the future (as anticipated by considering Fig. 6).

The numerically completed distribution of species frequencies over the Bank (Fig. 3) – including the estimated distribution of frequencies of each of the 142 unrecorded species – shows a sigmoidal shape, reminiscent of a log-normal distribution, as such suggesting the involvement of many independent factors governing together the distribution of species frequencies. The same
holds true for each of the two main components of the shelled-molluscan fauna: Bivalves (80 species out of which 27 remain to be recorded) and Gastropods (331 species out of which 102 remain to be recorded). Indeed, there again, sigmoidal shapes characterize the respective distributions of frequencies for each of these two classes, considered separately (Figs. 4 and 5). More precisely, the distributions of species frequencies – for the whole molluscan fauna as well as for the Bivalves and the Gastropods subsets separately – comply fairly well with the “broken-stick” model [37] (parametrized with the corresponding levels of estimated species richness): Figs. 8, 9, 10. This fair compliance with the “broken-stick” model – characterized by a typically random-generation process – still further highlights the stochastic-like consequences of the extraordinarily complex network of both historical and environmental factors that jointly contribute to the respective frequencies and spatial distributions of the more than 400 co-occurring species on Abrolhos Bank.

Fig. 8. The distribution of species frequencies for the whole soft-bottom Molluscan fauna at Abrolhos Bank, compared to the corresponding “broken-stick” model (dashed line) computed at the same level of species richness

Fig. 9. The distribution of species frequencies for the Bivalves fauna at Abrolhos Bank, compared to the corresponding “broken-stick” model (dashed line) computed at the same level of species richness
At last, the two minor components of the shelled-molluscan fauna, namely Polyplacophora and Scaphopoda, contribute together for an estimated 24 species only, out of which 11 were recorded and 13 still remain unrecorded (Table 1).

As a whole, the respective contributions of the four classes (Gastropoda, Pelecypoda (Bivalves), Polyplacophora and Scaphopoda) at Abrolhos Bank fairly mirror what is recognized worldwide for shelled-molluscs, in terms of relative proportions of species richness in each class [38]. More specifically, the approximately four to one ratio between the respective numbers of species of Gastropods and Bivalves, as reported from the richest molluscan hotspot of Koumac, New-Caledonia [3], is similarly highlighted on Abrolhos Bank.

5. CONCLUSION

With almost three hundred recorded species, the Rapid Assessment of the shelled-molluscan fauna inhabiting soft-bottoms at Abrolhos Bank, Brazil (as reported by R.S. ABALOA [4]) already pointed for the remarkable richness of this threatened reef-associated ecosystem. In turn, the least-biased numerical extrapolation of this rapid – and therefore fairly incomplete – assessment leads to a still 50% higher figure, with, finally, an estimated figure of 435 species of shelled-molluscan fauna at Abrolhos Bank, out of which no less than 30 species feature as Brazilian endemics. And, more or less similar proportions are expected for the other major groups of marine invertebrates and vertebrates occurring at Abrolhos Bank, since shelled-molluscs, as a group, are considered an especially appropriate “indicator” for the whole marine animal biodiversity [3].

All this, indeed, truly represents an invaluable biological richness attributed to Brazil, therefore requiring an effective protection policy. It is not the place, here, to further specify the details of the conservative measures to be taken in this regard: first of all, our purpose was to factually highlight why such an exceptionally rich ecosystem as Abrolhos Bank actually deserves to be preserved. And, by the way, to provide a factual and compelling account of precisely what needs now an urgent protection.

In turn, this clearly emphasizes the responsibility of Brazilian higher decision-makers, in the face of the steadily increasing threats to this exceptionally interesting reef-associated ecosystem.

On a more global point of view, the present study makes a modest but valuable additional contribution to the overall effort that is now required to compensate for the usual incompleteness of marine biodiversity surveys in the tropics [34].
ACKNOWLEDGEMENTS

I acknowledge the fruitful comments of two anonymous Reviewers.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES


19. Rajakaruna H, Drake DAR, Chan FT, Bailey SA. Optimizing performance of


37. MacArthur RH. On the relative abundance of bird species. Proceedings of the


39. 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>

40. 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>


APPENDIX 1

Bias-reduced extrapolation of the Species Accumulation Curve and associated 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 N₀ (with sampling effort N₀ 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 (N₀) species among which f₁, f₂, f₃, f₄, f₅, 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) [see [28,39,40]]:

\[ \frac{\partial^x R(N)}{\partial N^x} = (-1)^{(x)} \frac{f_x(N)}{C_{N,x}} = (-1)^{(x)} (x! / N^x) \frac{f_x(N)}{(N)} \ (\approx \text{as } N >> x) \]  \ (A1.1)

Compliance with the mathematical constraint (equation (A1.1)) warrants reduced-bias expression for the extrapolation of the Species Accumulation Curves R(N) (i.e. for N > N₀). Below are provided, accordingly, the polynomial solutions Rₓ (N) that respectively satisfy the mathematical constraint (A1.1), considering increasing orders x of derivation \[ \frac{\partial^x R(N)}{\partial N^x} \]. Each solution Rₓ (N) is appropriate for a given range of values of f₁ compared to the other numbers fₓ, according to [28]:

* For f₁ up to f₂ \( R₁(N) = (R(N₀) + f₁) - f₁ N₀ / N \)

* For larger f₁ up to 2f₂ - f₁ \( R₂(N) = (R(N₀) + 2f₁ - f₂) - (3f₁ - 2f₂) N₀ / N - (f₂ - f₁) N₀^2 / N^2 \)

* For larger f₁ up to 3f₂ - 3f₁ + f₂ \( R₃(N) = (R(N₀) + 3f₁ - 3f₂ + f₃) - (6f₁ - 8f₂ + 3f₃) N₀ / N - (4f₁ + 7f₂ - 3f₃) N₀^2 / N^2 - (f₁ - 2f₂ + f₃) N₀^3 / N^3 \)

* For larger f₁ up to 4f₂ - 6f₁ + 4f₄ - f₅ \( R₄(N) = (R(N₀) + 4f₁ - 6f₂ + 4f₄ - f₅) - (10f₁ - 20f₂ + 15f₃ - 4f₄) N₀ / N - (-10f₁ + 25f₂ - 21f₃ + 6f₄) N₀^2 / N^2 - (5f₁ - 14f₂ + 13f₃ - 4f₄) N₀^3 / N^3 - (f₁ + 3f₂ - 3f₃ + f₄) N₀^4 / N^4 \)
* For \( f_1 \) larger than \( 4f_2 - 6f_3 + 4f_4 - f_5 \) → \( R_8 (N) = (R(N_0) + 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5) - \)
\((15f_1 - 40f_2 + 45f_3 - 24f_4 + 5f_5)N_0/N - (-20f_1 + 65f_2 - 81f_3 + 46f_4 - 10f_5)N_0^2/N^2 - \)
\((15f_1 - 54f_2 + 73f_3 - 44f_4 + 10f_5)N_0^3/N^3 - (-6f_1 + 23f_2 - 33f_3 + 21f_4 - 5f_5)N_0^4/N^4 - \)
\((f_1 - 4f_2 + 6f_3 - 4f_4 + f_5)N_0^5/N^5 \)

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

* \( f_1 \leq f_2 \) → \( \Delta_{J1} = f_1; R_1 (N) \)

* \( f_2 < f_1 \leq 2f_2 - f_3 \) → \( \Delta_{J2} = 2f_1 - f_2; R_2 (N) \)

* \( 2f_2 - f_3 < f_1 \leq 3f_2 - 3f_3 + f_4 \) → \( \Delta_{J3} = 3f_1 - 3f_2 + f_3; R_3 (N) \)

* \( 3f_2 - 3f_3 + f_4 < f_1 \leq 4f_2 - 6f_3 + 4f_4 - f_5 \) → \( \Delta_{J4} = 4f_1 - 6f_2 + 4f_3 - f_4; R_4 (N) \)

* \( f_1 > 4f_2 - 6f_3 + 4f_4 - f_5 \) → \( \Delta_{J5} = 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5; R_5 (N) \).

**N.B. 1:** As indicated above (and demonstrated in details in [28]), this series of inequalities define the ranges that are best appropriate, respectively, to the use of each of the five estimators, JK-1 to JK-5. That is the respective ranges within which each estimator will benefit of minimal bias for the predicted number of missing species.

Besides, it is easy to verify that another consequence of these preferred ranges is that the selected estimator will always provide the highest estimate, as compared to the other estimators. Interestingly, this mathematical consequence, of general relevance, is in line with the already admitted opinion that all non-parametric estimators provide more or less pronounced under-estimates of the true number of missing species [19,21,34,41-44]. Also, this shows that the approach initially proposed in [45] – which has regretfully suffered from its somewhat difficult implementation in practice – might be advantageously reconsidered, now, in light of the very simple selection key above, of far much easier practical use, namely: the best estimate of the number of species remaining unrecorded after incomplete sampling is provided by the non-parametric estimator providing the highest value, among the Chao and the series of Jackknife estimators. And this is no longer a likely admissible point of view as suggested in [19,21,34,41-44] but, now, a rationally established affirmation.

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

**N. B. 3:** For \( f_1 \) falling beneath 0.6 \( x \) \( f_2 \) (that is when sampling completeness closely approaches exhaustivity), then Chao estimator may alternatively be selected: see reference [29].

**APPENDIX 2**

**Correction and extrapolation (when required) of the as-recorded S.A.D.**

N.B: details regarding the derivation of the following expressions are provided in [30].

Correction for bias of the recorded part of the S.A.D.

The bias-corrected expression of the true abundance, \( \bar{a}_i \), of species of rank ‘i’ in the S.A.D. is given by:

Extrapolation of the recorded part of the S.A.D.

Accounting for the complementary abundance distribution of the set of unrecorded species.
The following expression stands for the estimated abundance, \( a_i \), of the unrecorded species of rank \( i \) (thus for \( i > R_0 \)):

\[
a_i = \frac{2}{N_i} \left( 1 + \frac{R(N_i)}{N_i} \right) \left( 1 - \left[ \frac{\partial R(N)}{\partial N} \right]_{N_i} \right) \quad (A2.2)
\]

which, in practice, comes down to:

\[
a_i \approx \frac{2}{N_i} \left( 1 + \frac{R(N_i)}{N_i} \right),
\]

as \( f_i(N) \) already becomes quite negligible as compared to \( N \) for the extrapolated part.

This equation provides the extrapolated distribution of the species abundances \( a_i \) (for \( i > R(N_0) \)) as a function of the least-biased expression for the extrapolation of the species accumulation curve \( R(N) \) (for \( N > N_0 \)), \( i \) being equal to \( R(N) \). The key to select the least-biased expression of \( R(N) \) is provided at Appendix 1.

© 2020 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://www.sdiarticle4.com/review-history/62548