Original Research article

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

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, the distribution of their respective frequencies. As a result, the extrapolated total species richness at Abrolhos Bank reaches at least 435 species (instead of only 293 actually recorded), out of which 30 species (instead of 19) are expected to be Brazilian 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.

Key-words : 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, although 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 partial – assessments are likely being the best compromise, given, on the one hand, the multiplicity

of demands worldwide (especially under tropical climate) and the limited available resource to be devoted to field investigations on the other hand.
In this perspective, a Rapid Assessment has been carried out and reported by R.S. ABSALAO [4], focused on the soft-bottom shelled-molluscs fauna at Abrolhos Bank. No less than 293 species were recorded: 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 given increment in the actually recorded species richness. This, with the desirable prospect of a future reinforcement of survey operations at Abrolhos Bank.

2. Material 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 urgency of this survey. The corresponding sampling procedure and the collected data are provided in detail in the open-access reference [4] and, accordingly, need not being further repeated here.
Due to the substantial incompleteness of the reported samplings (almost unavoidable, dealing 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_t, 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_x 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 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 computation. In practice, this extrapolation of the Species Accumulation Curve allows to *forecast* and thereby *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 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 equivalent to that of species abundances) is described in details in reference [30], briefly summarized in Appendix 2 and exemplified in details in [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 frequency (with, thus, a_1 and a_{St} respectively standing for the highest and the lowest frequencies in a community of S_t 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.

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 Δ of unrecorded species, the resulting estimate of the "true" total species richness S_t (= R_0 + Δ), 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.

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

With sampling completeness level around two third (Table 1), it could be considered of interest to improve the completeness by further pursuing sampling. Yet, to make a rationnaly 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 bein considered: Figures 1 and 2. As said, this extrapolation allows to predict and thus to anticipate the additional sampling effort that would be required to obtain any desired increment in sampling completeness: Figure 2.

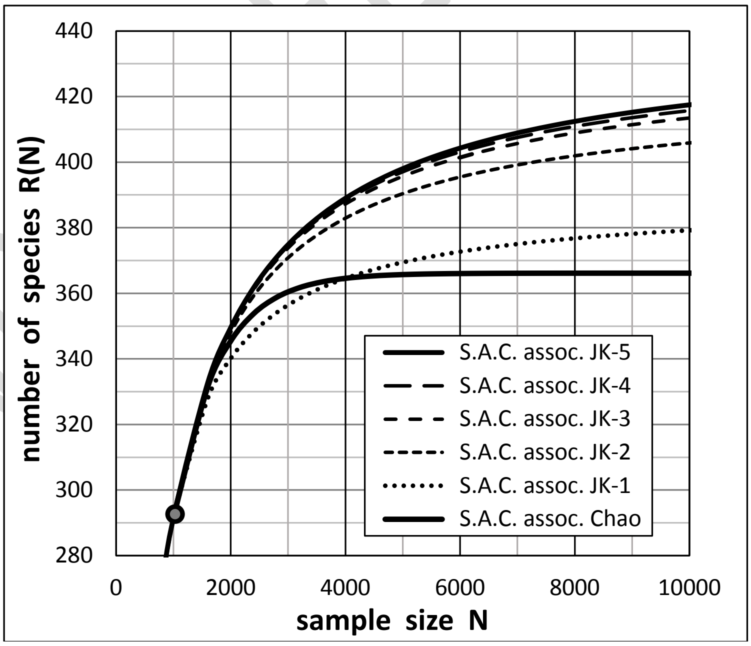

Figure 1 – Extrapolated part of the Species Accumulation Curve accounting for the increase of the number of detected species R(N) as a function of growing sample size N beyond the actually achieved sampling (N_0 = 1019, $R(N_0)$ = 293). Superimposed, here, are the extrapolations computed for six non-parametric 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.

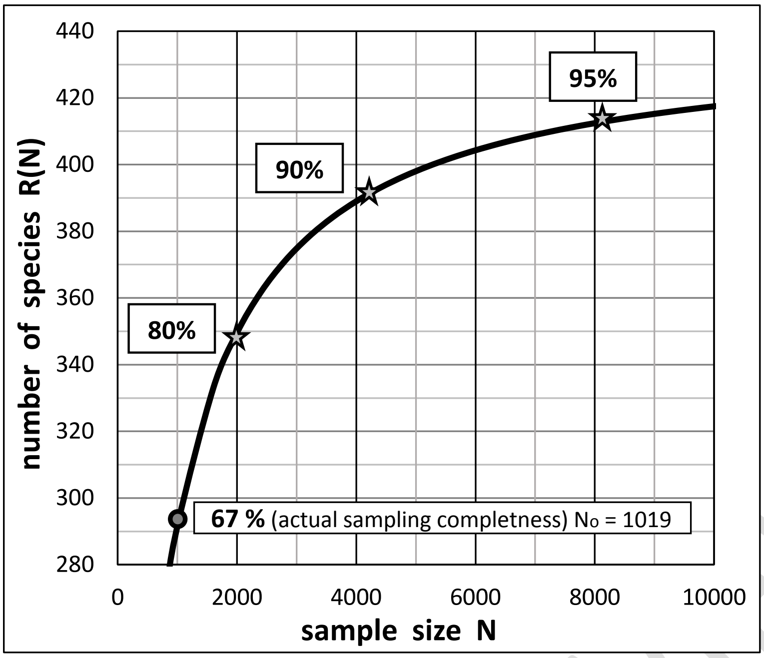

Figure 2 – The least-biased extrapolation of the Species Accumulation Curve (according to the selected estimator, *here*, Jackknife-5) highlighting the expected additional sampling effort, N, required to obtain a given increment in the number R(N) of recorded species: for example, increasing completeness 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 ≈ 2000, ≈ 4200, ≈ 8200, respectively.

3.2 Corrected and extrapolated Distributions of Species Frequencies

The corrected and extrapolated Distributions of Species Frequencies are provided in Figures 3 to 5, for the estimated (i) 435 species of the whole soft-bottom molluscan fauna, (ii) 80 species of Bivalves (Pelecypoda) and (iii) 331 species of Gastropods, respectively. The species are, classically, ranked by decreasing order of frequencies.

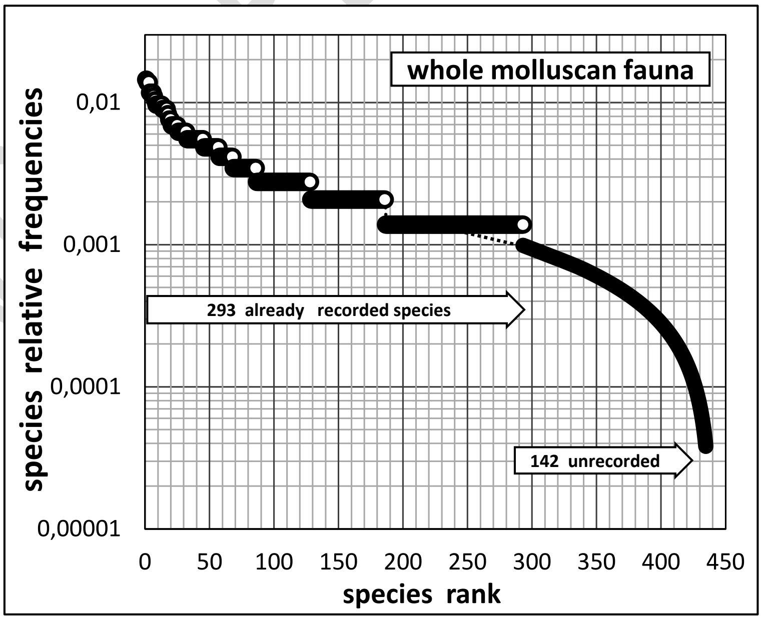

Figure 3 – The numerically completed Distribution of Species Frequencies of occurrence for the *whole* soft-bottom molluscan fauna at Abrolhos Bank: (i) *discs*: the 293 already recorded species, ranks 1 to 293; (ii) *double line*: the estimated 142 species remaining still unrecorded, ranks i = 294 to 435.

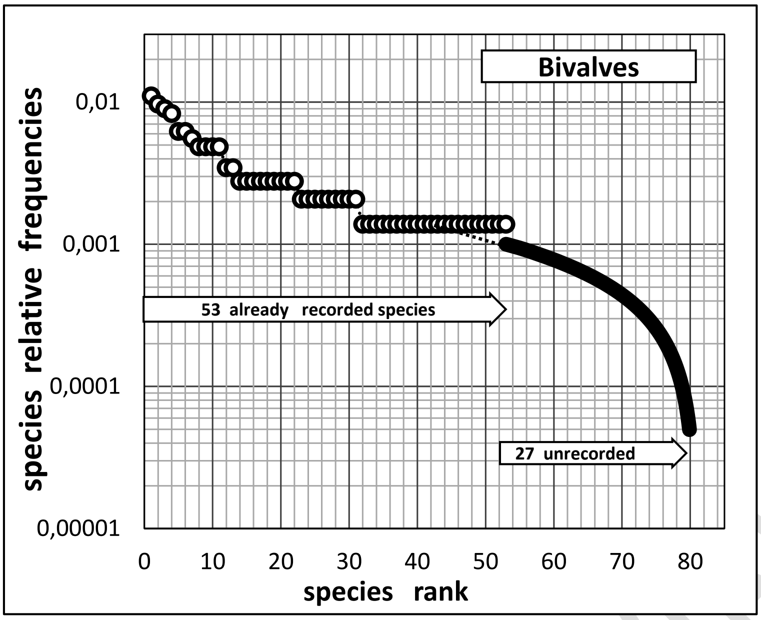

Figure 4 – The numerically completed Distribution of Species Frequencies of occurrence for the *Bivalves* subset: (i) *discs*: the 53 already recorded species, ranks 1 to 53; (ii) *double line*: the 27 still unrecorded species, ranks i = 54 to 80.

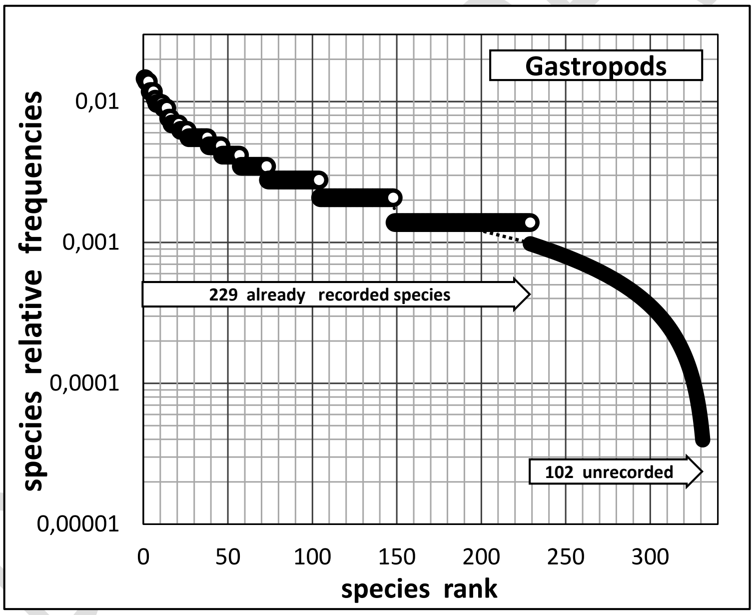

Figure 5 – The numerically completed Distribution of Species Frequencies of occurrence for the *Gastropods* subset: (i) *discs*: the 229 already recorded species, ranks 1 to 229; (ii) *double line*: the 102 still unrecorded species, ranks i = 230 to 331.

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.

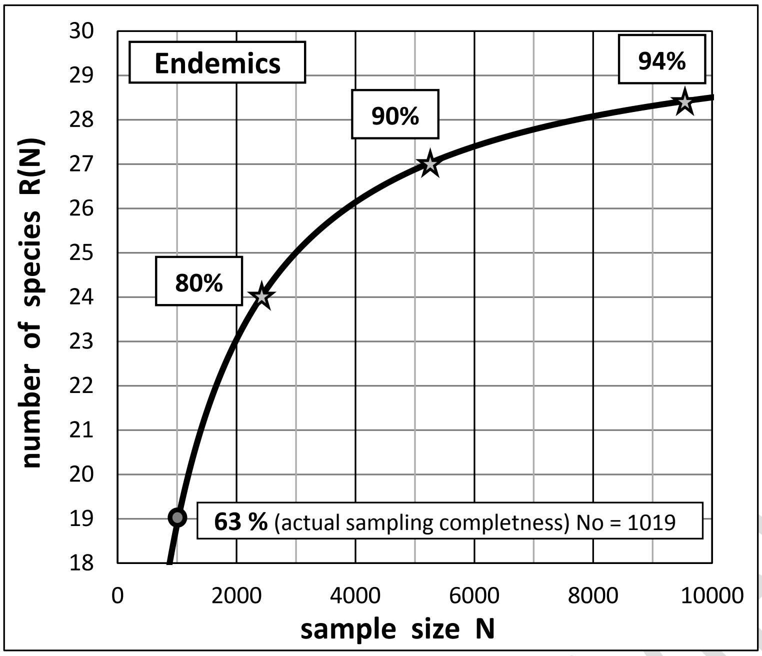

Figure 6 – The least-biased extrapolation of the Species Accumulation Curve (according to the selected estimator, *here also*, Jackknife-5) for the subset of Brazilian endemic species, highlighting the expected additional sampling effort N required to obtain a given increment in the number R(N) of recorded *endemic* species: for example, increasing completeness from 63% to 80%, 90%, 94% would require increasing sampling efforts from N_0 = 1019 to N ≈ 2300, ≈ 5200, ≈ 9600, respectively.

The numerically completed distribution of species frequencies among these 30 endemic species is provided in Figure 7. The comparison with the distribution of species frequencies within the whole molluscan fauna (Figure 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.

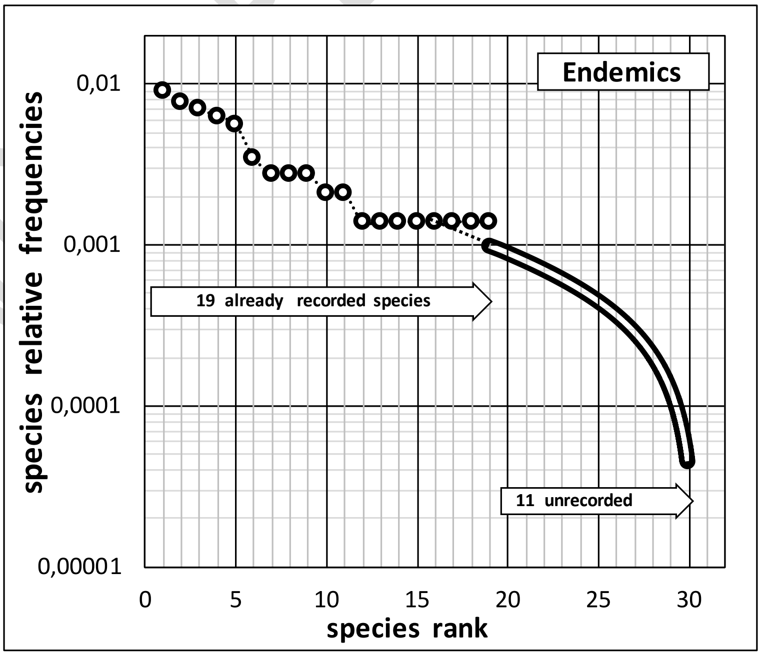

Figure 7 – The numerically completed Distribution of Species Frequencies of occurrence for the subset of Brazilian endemic species: (i) *discs*: the 19 already recorded species, ranks 1 to 19; (ii) *double line*: the estimated 11 species still remaining unrecorded: ranks i = 20 to 30.

The sampling completeness for endemic species turns out to be just slightly lower (63%) than for the whole molluscan fauna (67%). As shown above for the whole molluscan fauna, the numerical extrapolation of the Species Accumulation Curve (Figure 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 (i) *quickly* completed and (ii) 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, as an unavoidable consequence, usually substantial – although unknown – degree of sampling incompleteness. The latter potentially leading to unacceptably biased evaluations of species richness and regrettably truncated knowledge relative to the distribution of species frequencies.

Hence also, in turn, 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 for, as "morphospecies" or "RTU" (recognizable taxonomic units) if any further additional sampling efforts could have been conducted [3].

As a whole, numerical extrapolations 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, which has become more prominent recently. 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, Figures 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 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-molluscan species, an estimated 30 species are Brazilian endemics, out of which 11 were let unrecorded (Figure 7). These 11 unrecorded endemic species, thus remain to be detected, which would yet require substantial supplementary sampling effort (anticipated by considering Figure 6) to be consented in the future.

The numerically completed distribution of species frequencies over the Bank (Figure 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): sigmoidal shapes also characterize the respective distributions of the species frequencies pertaining to each of these two classes considered separately (Figures 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): Figures 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.

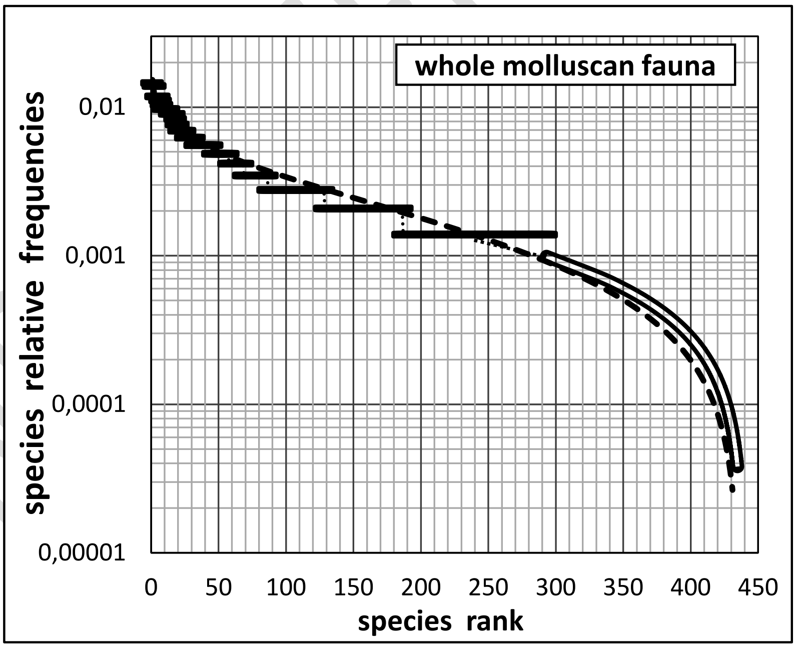

Figure 8 – The Distribution of Species Frequencies for the *whole* soft-bottom molluscan fauna at Abrolhos Bank, compared to the corresponding "broken-stick" model (dashed line) computed at the same level of species richness.

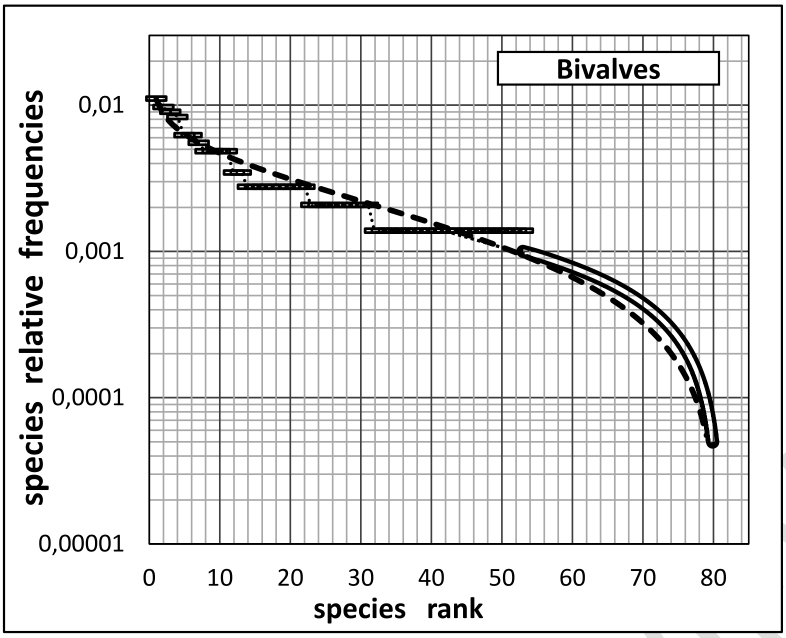

Figure 9 – The Distribution of Species Frequencies for the *Bivalves* fauna at Abrolhos Bank, compared to the corresponding "broken-stick" model (dashed line) computed at the same level of species richness.

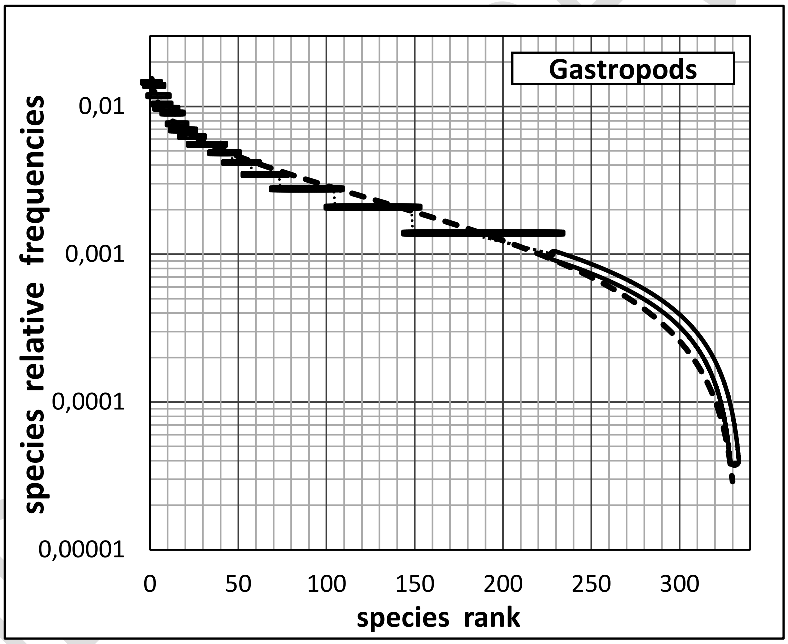

Figure 10 – The Distribution of Species Frequencies for the *Gastropods* fauna at Abrolhos Bank, compared to the corresponding "broken-stick" model (dashed line) computed at the same level of species richness.

At last, the two minor components of the shelled-mollusc 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, 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. ABSALAO) already pointed for the remarkable richness of this threatened reef-associated ecosystem. In turn, the least-biased *numerical extrapolation* of this Rapid – and therefore 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*. 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].

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 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) [see [28, 39, 40]:

$$\partial^x R_{(N)}/\partial N^x = (-1)^{(x-1)} f_{x(N)}/C_{N,x} \approx (-1)^{(x-1)} (x!/N^x) f_{x(N)} \quad (\approx \text{ as } N >> x) \quad \text{(A1.1)}$$

Compliance with the mathematical constraint (equation (A.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_x (N) that respectively satisfy the mathematical constraint (A1.1), considering increasing orders x of derivation $\partial^x R_{(N)}/\partial N^x$. Each solution R_x (N) is appropriate for a given range of values of f_1 compared to the other numbers f_x, according to [28]:

* *for* f_1 *up to* f_2 → $R_1(N) = (R(N_0) + f_1) - f_1.N_0/N$

* *for larger* f_1 *up to* $2f_2 - f_3$ → $R_2(N) = (R(N_0) + 2f_1 - f_2) - (3f_1 - 2f_2).N_0/N - (f_2 - f_1).N_0^2/N^2$

* *for larger* f_1 *up to* $3f_2 - 3f_3 + f_4$ → $R_3(N) = (R(N_0) + 3f_1 - 3f_2 + f_3) - (6f_1 - 8f_2 + 3f_3).N_0/N - (-4f_1 + 7f_2 - 3f_3).N_0^2/N^2 - (f_1 - 2f_2 + f_3).N_0^3/N^3$

* *for larger* f_1 *up to* $4f_2 - 6f_3 + 4f_4 - f_5$ → $R_4(N) = (R(N_0) + 4f_1 - 6f_2 + 4f_3 - f_4) - (10f_1 - 20f_2 + 15f_3 - 4f_4).N_0/N - (-10f_1 + 25f_2 - 21f_3 + 6f_4).N_0^2/N^2 - (5f_1 - 14f_2 + 13f_3 - 4f_4).N_0^3/N^3 - (-f_1 + 3f_2 - 3f_3 + f_4).N_0^4/N^4$

* *for* f_1 *larger than* $4f_2 - 6f_3 + 4f_4 - f_5$ → $R_5(N) = (R(N_0) + 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5) - (15f_1 - 40f_2 + 45f_3 - 24f_4 + 5f_5).N_0/N - (-20f_1 + 65f_2 - 81f_3 + 46f_4 - 10f_5).N_0^2/N^2 - (15f_1 - 54f_2 + 73f_3 - 44f_4 + 10f_5).N_0^3/N^3 - (-6f_1 + 23f_2 - 33f_3 + 21f_4 - 5f_5).N_0^4/N^4 - (f_1 - 4f_2 + 6f_3 - 4f_4 + f_5).N_0^5/N^5$

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:

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

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

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

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

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

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 regrettably 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*.

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.

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].

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

The bias-corrected expression of the true abundance, $\tilde{a}_i$, of species of rank 'i' in the S.A.D. is given by:

$$\tilde{a}_i = p_i.(1+1/n_i)/(1+R_0/N_0).(1-f_1/N_0) \qquad (A2.1)$$

where N_0 is the actually achieved sample size, R_0 (=$R(N_0)$) the number of recorded species, among which a number f_1 are singletons (species recorded only once), n_i is the number of recorded individuals of species 'i', so that $p_i = n_i/N_0$ is the recorded frequency of occurrence of species 'i', in the sample. The crude recorded part of the "S.A.D." – expressed in terms of the series of as-recorded frequencies $p_i = n_i/N_0$ – should then be replaced by the corresponding series of expected true abundances, $\tilde{a}_i$, according to equation (A2.1).

2) 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 $i > R_0$):

$$a_i = (2/N_i)/(1+ R(N_i)/N_i).(1- [\partial R(N)/\partial N]_{Ni}) \qquad (A2.2)$$

which, in practice, comes down to: $a_i \approx (2/N_i)/(1+ R(N_i)/N_i)$, as $f_1(N)$ 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 $R(N_i)$. The key to select the least-biased expression of R(N) is provided at Appendix 1.

References

1. Moura RL. Brazilian reefs as priority areas for biodiversity conservation in the Atlantic Ocean. Proccedings 9th International Coral Reef Symposium, Bali, Indonesia. 2000; 2: 917-920.

2. Francini-Filho RB, Coni EOC, Meirelles PM, Amado-Filho GM, Thompson FL, Pereira-Filho GH et al. Dynamics of coral reef benthic assemblages of the Abrolhos Bank, Eastern Brazil: inferences on natural and anthropogenic drivers. PLoS ONE. 2013; 8(1): e54260.

3. Bouchet P, Lozouet P, MaestratiP & Heros V. Assessing the magnitude of species richness in tropical marine environments: exceptionally high numbers of molluscs at a New Caledonia site. Biological Journal of the Linnean Society. 2002; 75: 421-436.

4. Absalao R. Soft-bottom molluscs of the Abrolhos Bank. In Dutra GF, Allen GR, Werner T & McKenna S. editors (2005) A Rapid Marine Biodiversity Assessment of the Abrolhos Bank, Bahia, Brazil (chapter 6), RAP Bulletin of Biological Assessment 38. Conservation International, Washington DC, USA.

5. Willis A. Rarefaction, alpha diversity and statistics. bioRxiv. 2017: doi 10.1101/231878.

6. Rumohr H, Karakassis I & Jensen JN. Estimating species richness, abundance and diversity with 70 macrobenthic replicates in the Western Baltic Sea. Marine Ecology Progress Series. 2001 ; 214 : 103-110.

7. Fontaine B, Bouchet P. et al. The European union's 2010 target: putting rare species in focus. Biodiversity and Conservation. 2007; 139: 167-185.

8. Flöder S, Jaschinski S, Wells G & Burns CW. Dominance and compensatory growth in phytoplankton communities under salinity stress. Journal of Experimental Marine Biology and Ecology. 2010; 395: 223-231.

9. Bracken M & Low N. Realistic losses of rare species disproportionately impact higher trophic levels. Ecology Letters. 2012; 15: 461-467.

10. Mouillot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelin-Vivien M, Kulbicki M, Lavergne S, Lavorel S, Mouquet N, Paine CET, Renaud J & Thuiller W. Rare species support vulnerable functions in high-diversity ecosystems. PLoS Biol. 2013; 11(5): e1001569.

11. Jain M, Flynn DFB, Prager CM, Hart GM, DeVan CM, Ahrestani FS, Palmer MI, Bunker DE, Knops JHM, Jouseau CF & Naeem S. The importance of rare species: a trait-based assessment of rare species contribution to functional diversity and possible ecosystem function in tall-grass prairies. Ecology and Evolution. 2014; 4(1): 104-112.

12. Ignatiades L & Gotsis-Skretas O. The contribution of rare species to coastal phytoplankton assemblages. Marine Ecology. 2014; 35: 132-145.

13. Low-Decarie E, Kolber M, Homme P, Lofano A, Dumbrell A, Gonzalez A & Bell G. Community rescue in experimental communities. Proceedings of the National Academy of Sciences USA. 2015; 112(46): 14307-14312.

14. Leitao RP, Zuanon J, Villéger S, Williams SE, Baraloto C, Fortunel C, Mendonça FP & Mouillot D. Rare species contribute disproportionately to the functional structure of species assemblages. Proceedings of The Royal Society B. 2016; 283: e0084 ; doi: 10.1098/rspb.2016.0084

15. Violle C, Thuillier W, Mouquet N, Munoz F, Kraft NJB, Cadotte MW, Livingstone SW & Mouillot D. Functional rarity: the ecology of outliers. Trends in Ecology. 2017; doi 10.1016/j.tree.2017.02.002

16. Cao Y, Williams DD & Williams NE. How important are rare species in aquatic community ecology and bioassessment? Limnology and Oceanography. 1998; 43(7): 1403-1409.

17. Bellier E, Grotan V, Engen S, Schartau AK, Diserud OH & Finstad AG. Combining counts and incidence data: an efficient approach for estimating the log-normal species abundance distribution and diversity indices. Oecologia. 2012; doi 10.1007/s00442-012-2311-2.

18. Cam E, Nichols JD, Sauer JR & Hines JE. On the estimation of species richness based on the accumulation of previously unrecorded species. Ecography. 2002; 25: 102-108.

19. Rajakaruna H, Drake DAR, Chan FT & Bailey SA. Optimizing performance of nonparametric species richness estimators under constrained sampling. Ecology and Evolution. 2016; 6: 7311-7322.

20. Connolly SR, Hughes TP & Bellwood DR. A unified model explains commonness and rarity on coral reefs. Ecology Letters. 2017; 20: 477-486.

21. Chen Y & Shen TJ. Rarefaction and extrapolation of species richness using an area-based Fisher's logseries. Ecology and Evolution. 2017; 7: 10066-10078.

22. Kery M & Royle JA. Inference about species richness and community structure using species-specific occupancy models in the National Swiss Breeding Bird survey MUB. Proceedings of the 2007 EURING Technical Meeting and Workshop, Dunedin, New Zealand.

23. May RM. Patterns of species abundance and diversity. In Cody M.L. & Diamond J.M. Ecology and Evolution of Communities. 1975; The Belknap Press of Harvard University: 81-120.

24. McGill BJ, Etienne RS, Gray JS et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology Letters. 2007; 10: 995-1015.

25. Ulrich W, Ollik M & Ugland KI. A meta-analysis of species-abundance distributions. Oikos. 2010; 119: 1149-1155.

26. Komonen A & Elo M. Ecological response hides behind the species abundance distribution: community response to low-intensity disturbance in managed grasslands. Ecology and Evolution. 2017; 7: 8558-8566.

27. Wang X, Ellwood F, AI D, Zhang R & Wang G. Species abundance distributions as a proxy for the niche-neutrality continuum. Journal of Plant Ecology. 2017; rtx 013.

28. 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>

29. 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; <hal 01477263 >

30. Béguinot J. How to extrapolate Species Abundance Distributions with minimum bias when dealing with incomplete species inventories. Advances in Research. 2018; 13(4): 1-24 ; doi: 10.9734/AIR/2018/39002.

31. Béguinot J. Numerical extrapolation of the species abundance distribution unveils the true species richness and the hierarchical structuring of a partially sampled marine

gastropod community in the Andaman Islands (India). Asian Journal of Environment and Ecology. 2018; 6(4): 1 – 23 ; doi: 10.9734/AJEE/2018/41293 <hal-01807454>

32. Béguinot J. Basic theoretical arguments advocating Jackknife-2 as usually being the most appropriate nonparametric estimator of total species richness. Annual Research & Review in Biology. 2016; 10(1): 1-12 ; doi: 10.9734/ARRB/2016/25104> ; <hal-01300828>

33. Mazzei EF, Bertoncini AA, Pinheiro HT, Machado LF, Vilar CC et al. Newly discovered reefs in the southern Abrolhos Bank, Brazil: anthropogenic impacts and urgent conservation needs. Marine Pollution Bulletin. 2016; doi 10.1016/j.marpolbul.2016.08.059

34. Menegotto A & Rangel TF. Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. Nature communications. 2018; 9: 4713 ; doi: 10.1038/s41467-018-07217-7

35. Béguinot J. The full hierarchical structuration of species abundances reliably inferred from the numerical extrapolation of still partial samplings: a case study with marine snail communities in Mannar Gulf (India). Asian Journal of environment and Ecology. 2018; 7(3): 1-27 ; doi: 109734/AJEE/2018/36831. <hal- 01963913 >

36. Béguinot J. Analyzing the role of environmental stresses on species richness and the process of hierarchical structuring of species abundances in marine Gastropods communities at Suva (Fiji Islands). International Journal of Environment and Climate Change. 2018; 8(3): 200-233 ; doi: 109734/IJECC/2018/44913. <hal- 01963939 >

37. MacArthur RH. On the relative abundance of bird species. Proceedings of the National Academy of Sciences U.S.A. 1957; 43: 293-295.

38. MolluscaBase, www.molluscabase.org (as of May 2018).

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>

41. O'Hara RB. Species richness estimators: how many species can dance on the head of a pin? Journal of Animal Ecology. 2005; 74: 375-386.

42. Gotelli NJ & Colwell RK. Estimating species richness. pp. 39-54 in: Biological Diversity: Frontiers In Measurement And Assessment. A.E. Magurran and B.J. McGill (eds.), 2010, Oxford University Press, Oxford. 345 pp.

43. Gotelli NJ & Chao A. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In: Levin S.A. (ed.) Encyclopedia of Biodiversity, second edition 2013, volume 5, 195-211; Waltham, MA: Academic Press.

44. Bevilacqua S, Ugland KI, Plicanti A, Scuderi D & Terlizzi A. An approach based on the total-species accumulation curve and higher taxon richness to estimate realistic upper limits in regional species richness. Ecology and Evolution. 2018; 8: 405-415.

45. Brose U, Martinez ND & Williams RJ. (2003). Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. Ecology, 84(9), 2364-2377.