



Inferring True Species Richness and Complete Abundance Distribution in Six Reef-fish Communities from Red-sea, Using the Numerical Extrapolation of Incomplete Samplings

Jean Beguinot^{1*}

¹*Department of Biogeosciences, University of Burgundy, Société Histoire Naturelle-Bourgogne Nature, 7 bvd H.P. Schneider, 71200 Le Creusot, France.*

Author's contribution

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

Article Information

DOI: 10.9734/AJEE/2019/V11i330136

Editor(s):

(1). Dr. Ravi Kant Chaturvedi, Centre for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, P. R. China.

Reviewers:

(1). Manoel Fernando Demétrio, Brazil.
(2). R. Uma Maheswari, Arulmigu Palaniandavar Arts College for Women, India.
Complete Peer review History: <http://www.sdiarticle4.com/review-history/53446>

Original Research Article

Received 16 October 2019
Accepted 21 December 2019
Published 30 December 2019

ABSTRACT

Even when ecological communities are incompletely sampled (which is most frequent in practice, at least for species-rich assemblages including many rare species), it remains possible to retrieve much more information than could be expected first, by applying *numerical extrapolation* to incomplete field data. Indeed, recently developed procedures of numerical extrapolation of partial samplings now allow to estimate, with fair accuracy, not only the number of the still unrecorded species but, moreover, the distribution of abundances of each of these unrecorded species, thereby making available the full range of the Species Abundance Distribution, despite dealing with incomplete data only. In turn, this allows to address a series of descriptive and functional aspects of the internal organization of species assemblages, which otherwise would have required disposing of truly exhaustive samplings.

This approach is applied, here, to the previously reported partial samplings of six neighboring reef-fish communities from Tiran Island, Red Sea, with the goal of better understanding their internal organization in relation to their respective environments.

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

In practice, the numerical completion contributes to avoid erroneous interpretations that would likely stem from considering only the incomplete field data. This point is especially relevant when studying reef-associated communities because accurate understanding of their organization will help guiding and refining at best the protective measures required by these particularly vulnerable communities.

Keywords: Coral reef; species diversity; rank abundance distribution; evenness; unevenness; hierarchical structuration; Gulf of Aquaba.

1. INTRODUCTION

Total species richness, taxonomic composition and the hierarchical structuring of species abundances are three key features that appropriately characterize species communities in the wild. Addressing properly these key features obviously requires disposing of exhaustive inventories. Yet, in practice, inventories remaining substantially incomplete are common and even doomed to become still more frequent with the inevitable generalization of “rapid assessments” and “quick surveys”. This is especially true when having to deal with species-rich communities including a lot of rare, hard-to-detect species. Such incomplete inventories prevent deriving reliable inferences and, thus, may often lead to erroneous interpretations regarding the key aspects of species communities evoked above [1–3].

Fortunately, a reliable procedure of *numerical extrapolation* of partial inventories has been developed recently, which can overcome these difficulties and provide least-biased estimations of:

- (i) The number of those species remained undetected and, still further,
- (ii) The respective abundances of each of these undetected species.

Thereby, reliable inferences can finally be derived regarding (i) the *true* (total) species richness and (ii) the *completed* distribution of species abundances, including the set of the still unrecorded species. Only the taxonomic identification of these (rarer) undetected species remains, of course, out of reach.

In turn, once numerically *completed* (and *only* when it is so), the Species Abundance Distribution (“S.A.D.”) can then provide synthetic pieces of information about the process (either deterministic or stochastic) that drive the hierarchical structuring of species abundances within community [4–8]. Accordingly, some light can thus be shed, both qualitatively and

quantitatively, on the biological and ecological determinants of the internal structuration among co-occurring species within community.

Hereafter, I report and discuss the information that has been unveiled thanks to the numerical extrapolation of the partial samplings of six neighboring reef-fish communities located at the mouth of the Gulf of Aquaba, north Red-Sea. Such marine ecosystems, in tropical shallow waters, are of major interest to ecologists and conservationists, as they are considered as embodying remarkably high levels of biological complexity [9–10].

2. MATERIALS AND METHODS

2.1 The Reported Field Data

The present study is based on the survey of six neighboring reef-fish communities, all from Foul Bay of Tiran Island, in the mouth of Gulf of Aquaba, north of Red-Sea. The survey was carried out and reported in detail by Goren and Spanier [11]. The inter-distances between studied communities are comprised within a rather narrow range, from 0.8 km to 8.5 km. All details regarding the precise locations of the six communities (labelled st-4 to st-9 in conformity with the original designations), the sampling procedure and the collected data are provided in the open-access reference [11] and need not being further repeated. Due to substantial incompleteness of the reported samplings (almost unavoidable with relatively species rich communities including a lot of rare species), numerical extrapolations of samplings are required and were implemented accordingly, prior to further analysis. Numerical extrapolation aims at avoiding biased inferences that may likely result from ignoring the rare species remaining unrecorded [12]. Indeed, such rare species may, yet, disproportionately contribute to the functional structuring of communities, as emphasized in [13-23]: “*rare species are critical for bio-assessment*” as quoted in [23]. The number N_0 of collected individuals and the

number R_0 of recorded species in each six studied communities are given in Table 1.

2.2 The Numerical Extrapolation Procedure and its Exploitation

To avoid making seriously biased inferences regarding the main structural descriptors of ecological communities (i.e. total species richness and abundance unevenness), relying upon (sub-) exhaustive inventories is required [24–29]. And, if impossible in practice (when, as here, excessive additional sampling efforts are needed to approach completeness), it is recommended to rely on *numerical extrapolation* of incomplete samplings [29]. As quoted in reference [29]: "*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*".

Beyond estimating the number of unrecorded species, a newly developed extrapolation procedure can provide, in addition, fairly accurate estimates of the *respective abundance of each of these unrecorded species*, as detailed in sections 2.2.1 to 2.2.3. Numerically completed this way (and *only* when it is so [26]), the distribution of species abundances can further reveal some qualitative and quantitative aspects of interest regarding the underlying process that governs the hierarchical structuring of species abundances within community [4,30-33].

2.2.1 Implementation of the procedure of numerical extrapolation

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 of the partially sampled community are computed according to the procedure defined in [34,35] 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 derive the least-biased extrapolation of the "Species Accumulation Curve", which predicts the expected increase in the number of newly 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 allows to *forecast* the expected additional sampling efforts that would be required

to obtain any desirable increment in sampling completeness.

Species Abundance Distribution: as mentioned above, the Species Abundance Distribution ("S.A.D.") is intended to provide the basic data necessary (i) to describe the *pattern* of hierarchical structuration of species abundances within community and (ii) to qualify and quantify the underlying *process* that drives this hierarchical structuration. Yet, to accurately exploit its full potential [36,37], the "S.A.D." requires (i) to be *corrected* for the bias resulting from drawing stochasticity during sampling of finite size and, still more importantly, (ii) to be *completed* by *numerical extrapolation*, to the extent that sampling is suspected to be incomplete, as revealed by the subsistence of singletons. The appropriate procedure of correction and least-biased numerical extrapolation of the as-recorded partial "S.A.D." is described in details in reference [37], briefly summarized in Appendix 2 and concretely exemplified in details in [38]. Classically, the "S.A.D." is graphically presented with the (log-transformed) abundances, a_i , plotted against the rank i of species, the latter being ordered by decreasing values of abundance (with, thus, a_1 and a_{st} respectively standing for the highest and the lowest abundances in a community of S_i co-occurring species).

2.2.2 Abundance unevenness: The pattern of species abundance structuration

Once numerically completed, the "S.A.D." conveys all the relevant quantitative data required to address the internal organization among species within a local community [39]. In turn, the "S.A.D." can be synthetically summarized by two of its major features: the *total species richness* ' S_t ' and the *degree* ' U ' of *unevenness* of the abundance distribution. Indeed, following [40], it is the degree of *unevenness* – rather than *evenness* itself – that should be preferred to address the hierarchical structuring of species abundances in communities. In short, abundance *unevenness* is sensitive to the intensity of interspecific competition within community and, conversely, abundance *evenness* is sensitive to the relaxation of this competitive intensity (with competition understood, here, in its broadest acceptance: see the end of section 2.2.3). Now, according to the mode of representation of "S.A.D.", it goes natural to quantify the degree of

abundance unevenness U as the average slope of the log-transformed abundance decrease, as already proposed by [41], that is:

$$U = [\log(a_1) - \log(a_{S_t})] / (S_t - 1)$$

$$U = [\log(a_1/a_{S_t})] / (S_t - 1) \quad (1)$$

with a_1 and a_{S_t} standing for the highest and the lowest abundances in the studied community.

2.2.3 Abundance unevenness: The underlying process of abundance structuration

Beyond the level of species abundance unevenness U , the underlying *process* of hierarchical structuration of abundances is worth being considered, in terms of (i) the kind of *mechanism* involved and (ii) what determines the *intensity* of this structuring process, from which follows the degree of abundance unevenness.

Very schematically, the kind of *mechanism* driving the hierarchical structuration of abundances may result either (i) from the major contribution of *one dominant* factor or (ii) from the combined contributions of *many mutually independent factors* acting together. This distinction can be tested by checking the conformity of the “S.A.D.” to either the log-series model or the *log-normal* model respectively [4, 42-45].

As regards now the *intensity* of the process of hierarchical structuration, it should be first emphasized that species richness has a *direct*, negative influence on abundance unevenness U , *as a general trend*, a point already highlighted by several authors [46-52]. The likely underlying ecological origin of this overall trend (behind its “mathematical-like” appearance [41,44] is discussed in detail in [47]. Now, each particular community usually deviates more or less – often substantially – from this overall trend. So that it is appropriate to *consider and quantify separately*: (i) on the one hand, the contribution of this overall general trend and (ii) on the other hand, the more or less important deviation from this tendential influence, which specifically singularizes each particular community and is particularly significant ecologically [46-47]. As argued in detail in [47], the direct, negative influence of species richness on abundance unevenness is adequately accounted for by the “broken-stick” theoretical distribution, originally conceptualized by MacArthur [53]. Accordingly, it

looks relevant to standardize the “rough” abundance unevenness U to the corresponding rough abundance unevenness U' of the “broken-stick” distribution, computed for the same species richness [54]. Doing so highlights to what extent the rough abundance unevenness U of any particular community actually deviates from the common overall trend, dictated by the tendential, direct influence of species richness [46-49,55]. Accordingly, a *standardized unevenness index*, “ I_{str} ”, is defined by the ratio U/U' [46,47]:

$$I_{str} = U/U' = [\log(a_1/a_{S_t}) / (S_t - 1)] / [\log(a'_1/a'_{S_t}) / (S_t - 1)]$$

that is:

$$I_{str} = U/U' = \log(a_1/a_{S_t}) / \log(a'_1/a'_{S_t}) \quad (2)$$

with a_1 and a_{S_t} standing for the highest and the lowest abundances in the studied community and a'_1 and a'_{S_t} standing for the highest and the lowest abundances in the corresponding “broken-stick” distribution computed for the same species richness S_t .

Thanks to this standardization, the index I_{str} is made free from the direct influence of species richness and, accordingly, allows for relevant, unbiased comparisons between communities *differing by their species richness* – contrary to the rough abundance, U , which is explicitly sensitive to this influence of species richness. In this respect, the index I_{str} deserves being considered as “genuinely” (idiosyncratically) attached to the corresponding community, *whatever* its particular species richness. Thus, the standardized abundance unevenness I_{str} satisfies the condition emphasized in [50,55]: “to *make sense*, (un)evenness must be independent of species richness”.

Now, from a *functional* point of view, the abundance unevenness U reflects the “mean competitive intensity” in the community (with “competitive intensity” being understood *sensu latissimo*, in its broadest scope, including both biotic and abiotic factors, as detailed in [47]. Accordingly, the standardized structuring index I_{str} reflects the mean competitive intensity once normalized (i.e. compared) to what it is in the broken-stick distribution at the same level of species richness. As the broken-stick model fits fairly well the abundance distribution in most bird communities [4,53], it follows that the mean competitive intensity in a community is equal to I_{str} times that in a *typical bird community having the same species richness*. Thereby, the

standardized structuring index I_{str} offers an *evocative benchmark* to appreciate more concretely the mean competitive intensity within community [47,48]. And, of course, in its *functional* sense, as well as in its *descriptive* acceptance, the index I_{str} allows for relevant, unbiased and meaningful comparisons between communities, regardless of their respective species richness.

3. RESULTS

3.1 Estimated Total Species Richness in Each Community

Fig. 1 provides the numerical extrapolations of each of the six studied reef-fish communities.

As expected from the subsistence in samples of numerous singletons (species recorded only once), numerical extrapolations confirm the incompleteness of all six communities, with completeness levels encompassing the range 59% to 78%, average 69% (Table 1).

Incidentally, note that although sampling completeness $C\%$ increases of course with sampling-size N_0 in any given sampled community, the completeness levels of different communities are by no means necessarily correlated with sampling size N_0 (here completeness $C\%$ is even decreasing with N_0 :

Fig. 2). This could already be expected from the strong difference in the kinetics of approach to sampling completeness between communities: Fig. 1. Thus, as already previously emphasized [56–60], the standardization of sampling-size (or, as well, the implementation of the so-called rarefaction procedure) is in no way a guarantee for reliable comparisons of true (total) species richness between different communities, as is particularly evident in Fig. 1. In fact, sampling completeness level is rather better correlated with the smallest abundance level, a_{st} (Fig. 3, $p = 0.04$), which, indeed, makes sense. And the correlation with a_{st} is still much better when sampling-completeness is considered at a same, arbitrary chosen, sampling-size (Fig. 4, $p = 0.005$). Comparing with Fig. 2, this exemplifies how sampling completeness is far better related to the smallest relative abundance a_{st} in the community than it is to sampling-size.

Thus, based on the six communities under study, a linear regression of the sampling completeness $C\%$ against the sampling-size N_0 , the total species richness S_t and the species abundance unevenness U satisfies the following equation:

$$C\% = 0.0267 N_0 - 0.576 S_t - 683 U + 133 \quad (3)$$

or, as well:

$$C\% = -0.706 (N_0/S_t) + 355 U + 55.6 \quad (4)$$

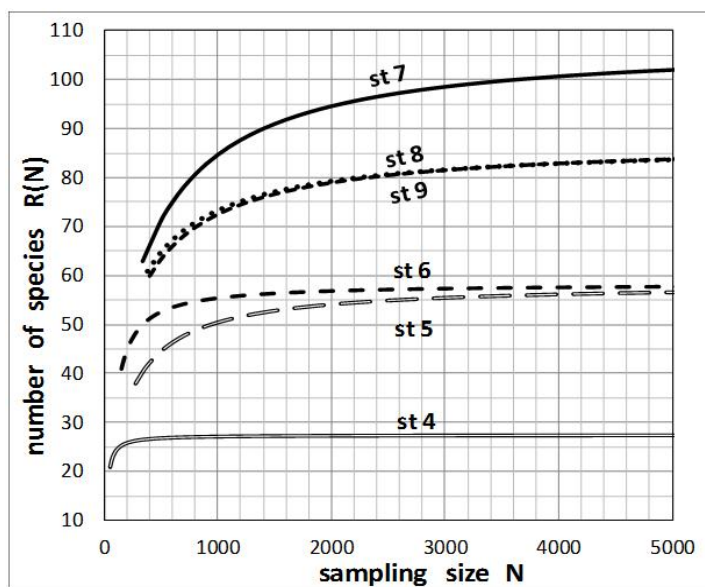


Fig. 1. Numerical extrapolations of the species accumulation curves of the six studied communities. Note, in particular, the strong difference between communities regarding the kinetics of approach to sampling completeness

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

Reef-fish communities	st 4	st 5	st 6	st 7	st 8	st 9
* nb. collected individuals N_0	49	276	147	340	404	385
* nb. recorded species $R_0 = R(N_0)$	21	38	41	63	60	61
* selected least-biased estimator	JK-1	JK-5	JK-2	JK-5	JK-5	JK-5
* number unrecorded species Δ	6	20	17	44	27	26.5
* total species richness S_t	27	58	58	107	87	86.5
* sampling completeness R_0/S_t	78%	65%	71%	59%	69%	71%

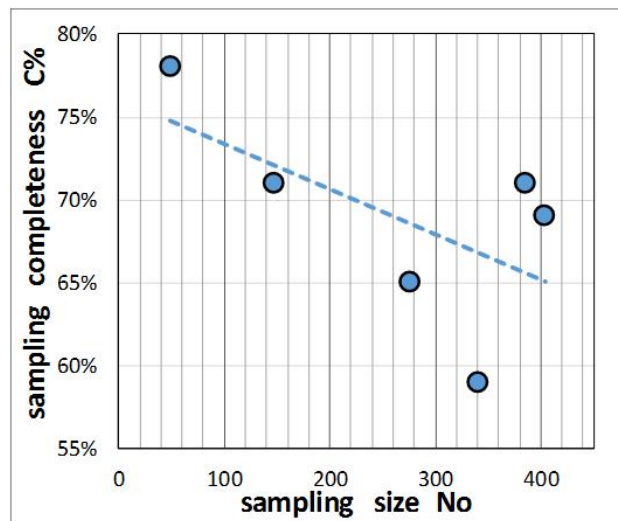


Fig. 2. Lack of positive correlation between sampling completeness C% (from Table 1) and sampling size No, for each of the six studied communities. Here, even a weakly significant a negative correlation ($r = 0.60$, $n = 6$, $p = 0.10$)

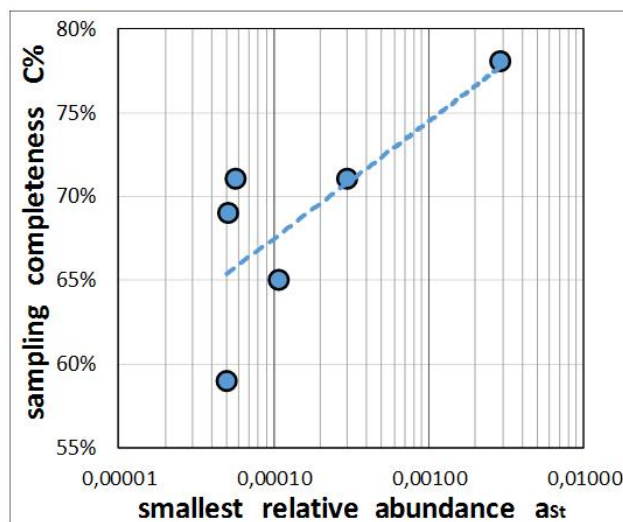


Fig. 3. Sampling completeness C% plotted against the smallest relative abundance, a_{st} , in each of the six studied communities: positive correlation ($r = 0.76$, $n = 6$, $p = 0.04$)

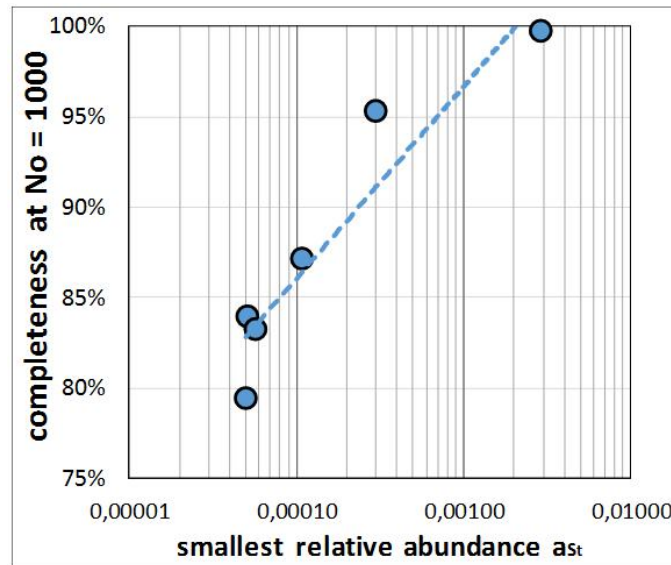


Fig. 4. Sampling completeness C% plotted against the smallest relative abundance, a_{st} , for a same, arbitrary chosen, sampling-size $N_o = 1000$. The positive correlation of sampling completeness with respect to the smallest relative abundance, a_{st} , is, as expected, still stronger ($r = 0.94$, $n = 6$, $p < 0.005$)

Now, $\log(a_{st})$ varies far more than does $\log(a_1)$ (Fig. 12), so that unevenness U (equation (1)) is mainly dependent upon a_{st} and S_t . The sampling completeness is thus mainly driven in terms of N_o , S_t and a_{st} :

$$C\% = 1.46 (N_o/S_t) + 8290 a_{st} + 60.2 \quad (5)$$

All this demonstrate, once again, how *unreliable* can be the “rarefaction” procedure, as a general method aiming at comparing the relative species richness between communities.

3.2 Numerically Completed Species Abundance Distributions

The bias-corrected and numerically extrapolated Species Abundance Distributions (“S.A.D.s”) of the six studied communities (according to the procedure described in [37] and summarized in Appendix 2) are provided in Figs. 5 to 10. The (bias-corrected) abundances of *recorded* species are plotted as grey discs while the *extrapolated* part of the abundance distribution is plotted as a thick double line. A synthetic view is offered at Fig. 11, addressing three of the six reef-fish communities, having respectively the lowest, the highest and an intermediate level of species richness. In particular, it clearly appears that increasing species richness is, for part, accommodated by the enlargement of the overall

range $[a_1 - a_{st}]$ of species abundances, which, in turn, proceeds mainly from the decrease of the minimum abundance level a_{st} (Fig. 12). The rate of this decrease, however, tends to slow down progressively, possibly because approaching some Allee-effect threshold [47].

3.3 Testing for the Type of Process Involved in the Structuring of Species Abundances

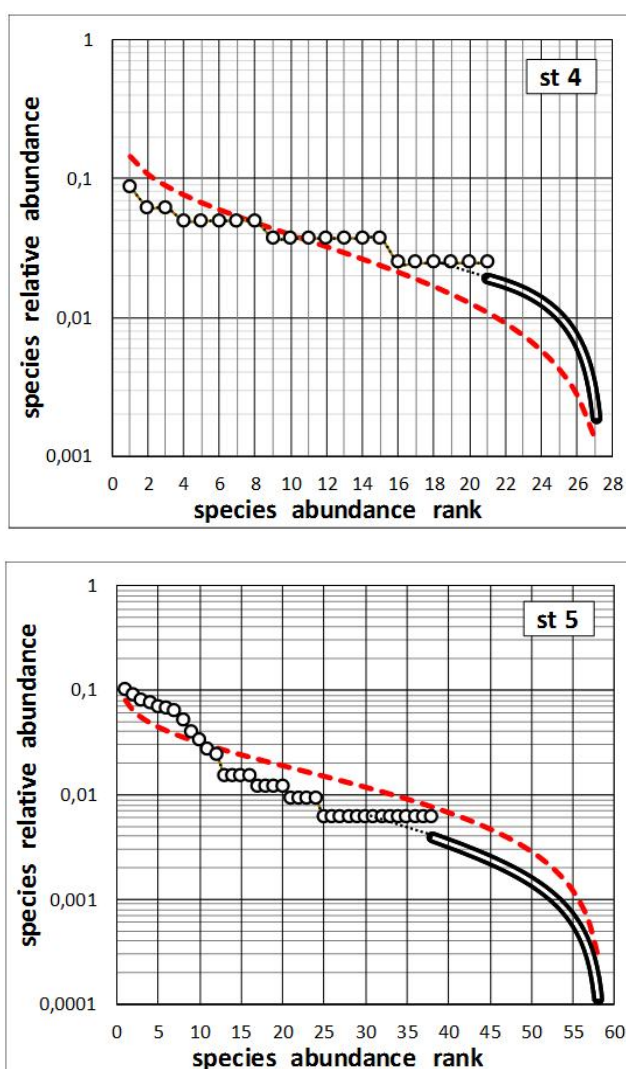
The numerically completed “S.A.D.s” of the six studied communities exhibits a more or less sigmoidal shape (Figs. 5 to 10) which far better fits the typically sigmoidal shape of the “log-normal” model than the J shape of the “log-series” model (the latter not represented but see [38,60-66]). Note, however, that in communities st-5, st-7, st-8, the abundance of the dominant species, is somewhat lower than predicted by the “log-normal” model, which might perhaps be related to some density-dependent predation (or any other density-dependent negative influence) applying to the dominant species.

3.4 The Degree of Unevenness of Species Abundance Distributions and Its Dependence to Species Richness

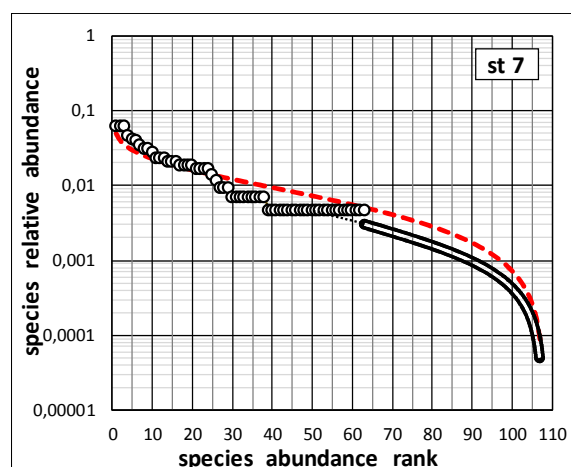
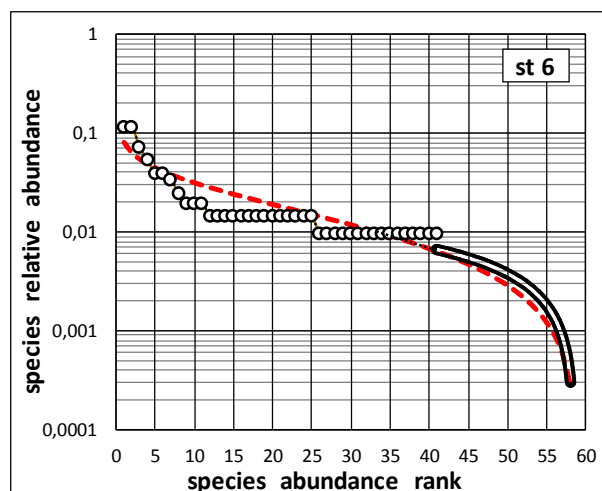
Fig. 13 and Table 2 show the variations of species abundance unevenness U with species

richness among the six reef-fish communities. The decrease of abundance unevenness with growing species richness is highly significant ($r = 0.982$, $n = 6$, $p < 0.001$). As already emphasized, this decrease of fits a common global trend [46-52] which, in turn, highlights a logically expected tendency for increasing species richness to be accommodated, at least for part, by a decreasing – relaxed – competitive intensity among co-occurring species, as previously argued in [47]. As already suggested in [48] and further emphasized in [47], it turns out that the observed global trend for decreasing unevenness U with species richness follows a rate which

corresponds rather well to that of the “broken-stick” model, as already mentioned above. Yet, each particular community, can, of course, more or less deviates from this global trend and, indeed, often does so. The standardized unevenness index I_{str} was precisely defined (section 2.2.3) (i) to highlight and quantify the deviation of the focused community from the global trend of influence of species richness on unevenness and, thereby, (ii) to allow reliable comparisons between communities differing in species richness. The variations of the standardized unevenness I_{str} among the six communities are given in Fig. 14 and Table 2.



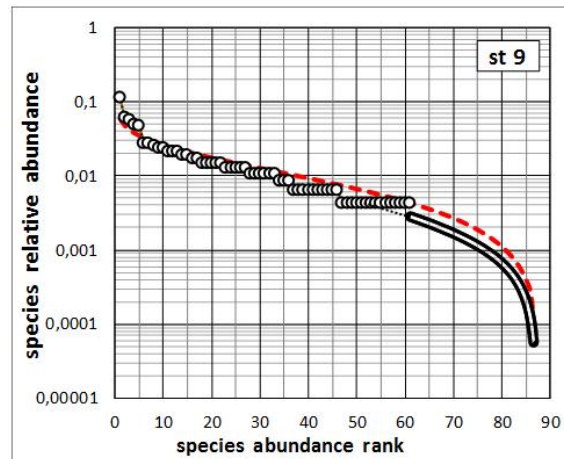
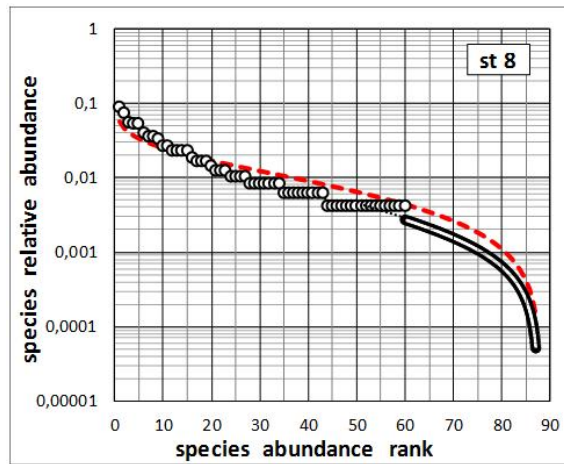
Figs. 5 & 6. The corrected and numerically completed species abundance distributions of communities st 4 and st 5 (white discs: recorded species, double line: numerically extrapolated part of the S.A.D.). For comparison, the “broken-stick” (dashed line)



Figs. 7 & 8. The corrected and numerically completed species abundance distributions of communities st 6 and st 7 (white discs: recorded species, double line: numerically extrapolated part of the S.A.D.). For comparison, the “broken-stick” (dashed line)

Table 2. A synthetic summary of the main quantitative features of the hierarchical organization of species abundances within community, as derived from numerically completed “S.A.D.s” : (i) the total species richness S_t of the community ; (ii) the relative abundances a_1 and a_{S_t} of the most and least abundant species (species rank 1 and S_t) ; (iii) the same, a'_1 and a'_{S_t} , for the “broken-stick” model, (iv) the mean relative variation of abundance, $\Delta a_i/a_i = (a_i/a_{i+1}) - 1$, between two successive species along the S.A.D., (v) the unevenness of abundances in the community: $U = \log(a_1/a_{S_t})/(S_t-1)$; (vi) the unevenness of abundances in the corresponding “broken-stick” distribution: $U' = \log(a'_1/a'_{S_t})/(S_t-1)$ and, at last, (vii) the standardized unevenness index $I_{str} = U/U'$

community	S_t	a_1	a_{S_t}	a_1/a_{S_t}	a'_1	a'_{S_t}	mean $\Delta a_i/a_i$	U	U'	I_{str}
st 4	27	.0869	.00189	46	.1441	.0014	16%	.0639	.0777	0.822
st 5	58	.1006	.00011	915	.0801	.00030	13%	.0519	.0426	1.217
st 6	58	.1149	.00030	383	.0801	.00030	11%	.0453	.0426	1.063
st 7	107	.0605	.00005	1210	.0491	.00009	7%	.0291	.0260	1.121
st 8	87	.0890	.00005	1780	.0580	.00013	9%	.0376	.0307	1.224
st 9	86.5	.1164	.00006	1940	.0584	.00013	9%	.0387	.0309	1.252



Figs. 9 & 10. The corrected and numerically completed Species Abundance Distributions of communities st 8 and st 9 (white discs: recorded species, double line: numerically extrapolated part of the S.A.D.). For comparison, the “broken-stick” (dashed line)

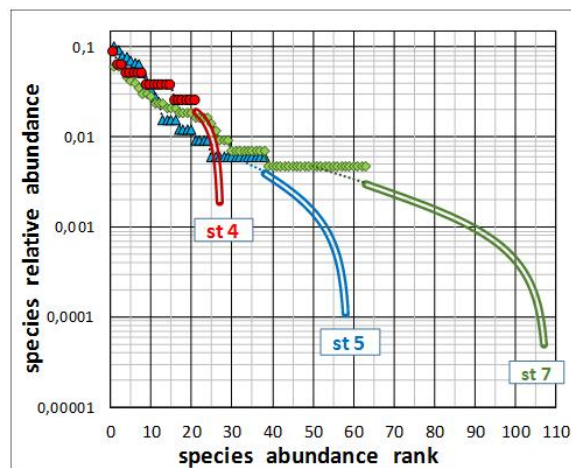


Fig. 11. Comparison of the corrected and numerically completed Species Abundance Distributions of three reef-fish communities of Tiran Island, Red-Sea, having respectively the lowest, intermediate and highest species richness

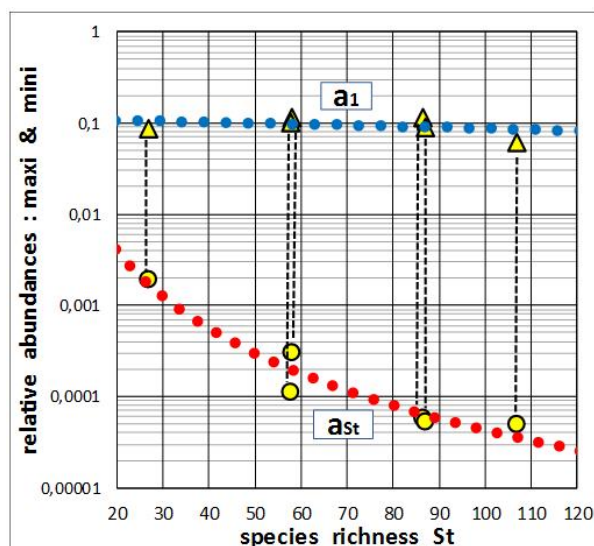


Fig. 12. Variations with species richness of the relative abundances a_1 and a_{St} of the dominant and the rarest species, among six reef-fish communities of Tiran Island, Red-Sea

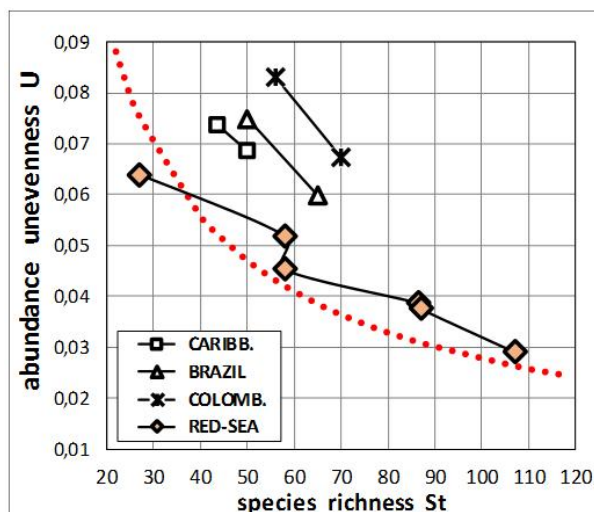


Fig. 13. Variations of the species abundance unevenness U with species richness among six reef-fish communities of Tiran Island, Red-Sea. Also plotted for comparison, the data relative to reef-fish communities from Caribbean [63], Brazil [65] and Colombia [64].
The dotted line corresponds to the “broken-stick” model

3.5 Seeking for the Main Drivers of Difference in Species Composition and Species Richness

The Jaccard index is one of the most commonly used metrics of composition similarity between species communities. Jaccard index is, yet, also sensitive to difference in species richness [67-69]. As argued in [67] and [69], it is relevant, in this respect, to substitute a modified version J_{mod}

$= a/\min\{b, c\}$ to the genuine Jaccard index $J = a/(b + c - a)$, where ‘a’ is the number of shared species and ‘b’, ‘c’ are the species richness of each compared community.

Thus, the *modified* Jaccard index, J_{mod} , accounts more reliably for the degree of similarity of species composition itself, while the genuine Jaccard index provides a compound evaluation of similarity, mixing both the similarity of

composition and the similarity of species richness between compared communities.

Hereafter, are highlighted the respective roles of (i) the inter-distance between communities and (ii) the degree of relative remoteness of the sites sheltering each community.

3.5.1 The role of inter-distance between communities on their degree of dissimilarity

In Table 3 and Figs. 15 and 16, both expressions of the Jaccard index - the genuine J and the modified J_{mod} - are plotted together against the inter-distance between the studied communities taken two by two. The *modified* index J_{mod} , i.e. the relevant measure of composition similarity, decreases slightly, but non-significantly, with inter-distance ($r = 0.22$, $n = 15$, $p = 0.22$). The difference in species richness also decreases slightly and non-significantly with inter-distance (Table 3, $r = 0.25$, $n = 15$, $p = 0.18$). The *genuine* Jaccard index, which accounts for both composition similarity and richness similarity, thus also decreases and now, with a moderate statistical significance: $r = 0.51$, $n = 15$, $p = 0.03$.

3.5.2 The role of location remoteness on the degree of dissimilarity between communities

In this respect, the site sheltering the community st-5 immediately appears singular, being

segregated at the bottom of a small cove, while the five other communities are more readily exposed to the open sea. Accordingly, the distribution of the degrees of similarity of community st-5 with each of the other five communities was compared, statistically, to the distribution of the degrees of similarity between these five communities taken two by two, using t-test. Here statistical significance reveals substantially stronger than when considering the role of inter-distances:

- according to the genuine Jaccard index J: t-test, $n = 15$, $t = 3.02$, $p = 0.011$,
- according to the modified Jaccard index J_{mod} : t-test, $n = 15$, $t = 4.45$, $p = 0.0007$.

4. DISCUSSION

Coral reefs as a whole, and the associated reef fish communities in particular, are considered embodying among the highest levels of diversity and biological complexity on Earth. On the other hand, tropical marine ecosystems, especially those hosted by coral reefs, are under increasing threat, being particularly sensitive to ongoing anthropogenic deteriorations of environment. This, in turn, urges to monitor the progressive alteration of these ecosystems, especially focusing on the on-going reduction in species richness as well as the increased unevenness of species abundances.

Table 3. The inter-distance (kilometers), the genuine Jaccard index of similarity, the modified Jaccard index and the difference in species richness within each 15 couples of communities

Couple of communities	inter-distance (km)	<i>genuine</i> Jaccard	<i>modified</i> Jaccard	difference in sp. richness
st 6 – st 8	0.8	0.295	0.561	29
st 8 – st 9	2.0	0.198	0.333	0.5
st 4 – st 7	2.3	0.135	0.476	80
st 5 – st 8	2.3	0.114	0.263	29
st 6 – st 9	2.3	0.275	0.537	28.5
st 5 – st 6	2.8	0.162	0.289	0
st 5 – st 9	3.5	0.193	0.421	28.5
st 6 – st 7	3.5	0.238	0.488	49
st 7 – st 9	3.7	0.228	0.377	20.5
st 7 – st 8	4.0	0.295	0.467	20
st 4 – st 6	5.8	0.192	0.476	31
st 4 – st 9	5.8	0.139	0.476	59.5
st 5 – st 7	6.2	0.110	0.263	49
st 4 – st 8	6.3	0.157	0.524	60
st 4 – st 5	8.5	0.113	0.286	31

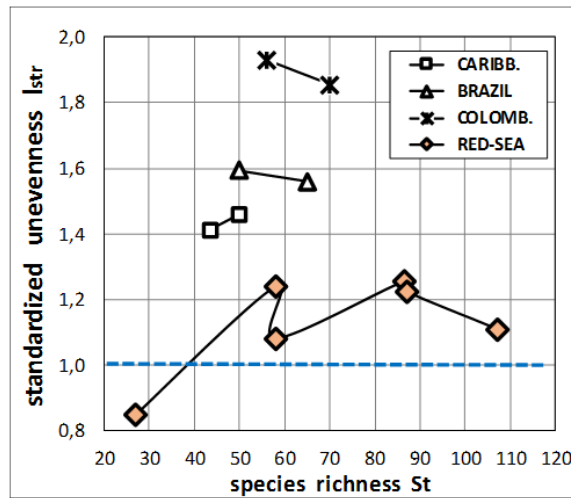
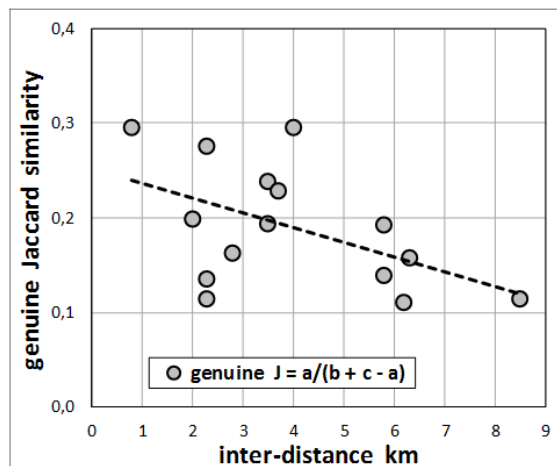
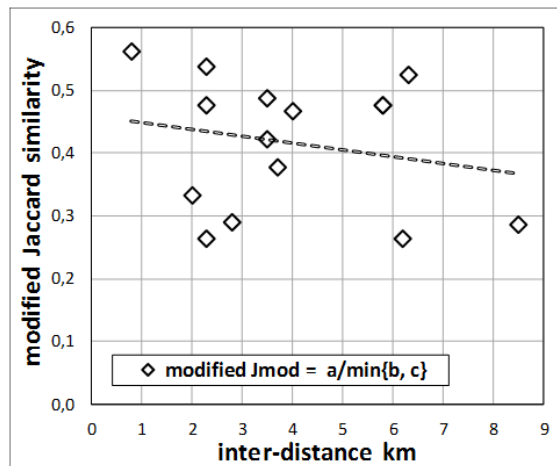


Fig. 14. Variations of the standardized abundance unevenness I_{str} with species richness among six reef-fish communities of Tiran Island, Red-Sea. Also plotted for comparison are the data relative to reef-fish communities from Caribbean [63], Brazil [65] and Colombia [64]



Figs.15, 16. The modified and the genuine Jaccard index of similarity plotted against the inter-distance between the studied communities taken two by two

Properly monitoring the variations in species richness and abundance unevenness yet requires observing some essential methodological precautions, in particular:

- To consider only data issued from (sub-) exhaustive samplings of local reef communities or, failing that, to implement reliable procedure of numerical extrapolation of incomplete samplings;
- To clearly distinguish and separate the observed modifications attributed to detrimental anthropogenic activities from the modifications due to other, *non-anthropogenic* determinants. Accordingly, the proper influences of the latter need to be assessed for their own, so as to be able, then, to disentangle finally, what is the genuine contribution of anthropogenic degradation of the environment from what is the mere consequence of “natural” causes. Improving first our knowledge regarding the influence of different kinds of “natural” factors is thus imperative. Here, for example, assessing the role of the degree of site remoteness on species composition and species richness.

Analysis, in this respect, is favored by the relatively close localizations of the six communities: as all of them are, thus, likely colonized from the same regional pool of species, these six communities are placed in equal terms as regards the recruitment source.

4.1 The True (total) Species Richness of Communities

The reported surveys of each six studied communities prove remaining substantially incomplete and, moreover, incomplete to variable extents (from 59% to 78%: Table 1). Therefore, the species richness of the different communities could not be reliably appreciated, *even in relative terms*. In addition, the procedure of rarefaction (numerical standardization of sampling efforts), usually suggested – too often wrongly [see 56–59] – as a surrogate in such circumstance, would indeed remain inefficient, due to the disparity between the levels of abundance unevenness among communities (Table 2, Fig. 13). More precisely, the degree of sampling completeness depends on three main drivers, (i) the sampling-size of course, but also (ii) the ratio of sampling-size to total species richness and (iii) the species abundance unevenness: the respective influences of each of these three drivers were

assessed by linear regressions: equations (3), (4), (5). Accordingly, no reliable interpretation could be soundly assessed on the uncertain basis of the reported, variously incomplete, samplings. This, once more, emphasized the necessity of implementing numerical extrapolation of such incomplete samplings.

Here, the total, true species richness substantially differs among the six communities – and far more than could have been expected: no less than from 27 up to 107 species (Table 1). With an average estimate of 71 species, the total species richness per community at Tiran Island appreciably exceeds the average figure of 50 species reported for the reef-fish communities of the Atlantic and Pacific coasts of South-America [63-66].

4.2 The Numerically Completed Distributions of Species Abundances within Community

Once numerically completed, the Species Abundance Distributions in all six studied communities show rather sigmoidal shapes, which better comply with the *log-normal* model than with the *log-series* model (Figs. 5 to 10), thereby suggesting that in all these communities, the distribution of abundances is mainly driven by the cooperation of numerous mutually independent factors, which, indeed, is the usual case [4,38,42-45,47,60-66,69,70]. In addition, some possible density-dependent predation (or any other negative influence) seems to apply to the subset of dominant species.

Species abundance unevenness consistently decreases with species richness (Table 2, Fig. 13), following a common trend [46-52]. The latter trend resulting, in turn, from a largely shared tendency (among marine as well as terrestrial communities) to partly accommodate larger numbers of co-occurring species by an improved relaxation of the mean competitive intensity at the community scale [47]. A complementary contribution to the accommodation of increasing species richness being the enlargement of the overall range of species abundances, especially by continuously lowering the abundance level of the rarest species (Fig. 12), until reaching the minimal threshold fixed by Allee effects (or the like) [47]. The more or less pronounced deviation of each particular community from the common trend of decreasing unevenness is conveniently quantified by standardizing the recorded unevenness to the corresponding unevenness

level in the “broken-stick” model [47,48], as highlighted in Fig. 14.



Image 1. *Chaetodon paucifasciatus* Ahl, 1923, the ‘Eritrean butterfly-fish’, a species endemic to the Red Sea and the Gulf of Aden, can be found in most reef-fish communities at Tiran Island. © Bernard E. Picton



Image 2. *Pseudobalistes fuscus* (Bloch & Schneider, 1801), the “Yellow-spotted triggerfish”, recorded only in the community st-8 at Tiran Island. © Hectonichus

4.3 Factors of Differentiation of Species Composition and Species Richness among Communities

More or less contrasted ecological conditions are expected to play a significant role on the species richness and the species composition among local communities. Yet, attempts to establish such linkage can reveal more uneasy than expected. For example, in [11], a demonstrative relationship is advocated between (i) the contrasted environments of communities st-5 and st-6 and (ii) the substantial differences in both species richness and abundance unevenness

recorded between these two communities. However, this interpretation, based on the data from the available *incomplete* samplings, is by no means further supported by the numerically completed samplings, according to which the communities st-5 and st-6 are, in fact, very close to each other, as regards both total species richness and abundance unevenness (Table 2) in spite of their contrasted environments.

A more straightforward influence on the degree of differentiation of species composition (less clearly on the degree of differentiation of species richness) is the degree of remoteness of the community location from the open sea, which reveals, here, very significant. This might suggest that, even for fish that are considered highly mobile species, spatial confinement may be more decisive than expected. Inter-distance between communities having only a secondary role, at least on the scale of ten kilometers (Table 3, Figs. 15 & 16), as already pointed out [69].

5. CONCLUSION

Specificities of reef fish communities at Tiran Island (Red Sea) have been highlighted in terms of each of the three main features that relevantly characterize any species community in the wild, namely: (i) the true (total) species richness, (ii) the hierarchical structuring of species abundances and (iii) the taxonomic composition of the species co-occurring in the community. High species richness of reef fish communities in Red Sea has already been emphasized many times, being second only to the reef communities of the celebrated “Coral Triangle”. The origin and peculiarities of the hierarchical organization of species abundances among each of the six studied communities have been highlighted and subsequently discussed. At last, regarding the degree of dissimilarity in species composition among the six communities, the weak influence of moderate inter-distances was again demonstrated, which indeed is not surprising for fishes having substantial dispersal capacities, especially at their larval stage. However, the role of the degree of site remoteness on the distinctiveness in species composition among communities has proved being stronger than might have been expected.

ACKNOWLEDGEMENTS

I acknowledge the fruitful comments of the Editor and two anonymous Reviewers. Moreover, I thank colleagues, M. Goren and E Spanier, for

their reported field data displayed in reference [11].

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Cam E, Nichols JD, Saue JR, Hines, JE. On the estimation of species richness based on the accumulation of previously unrecorded species. *Ecography*. 2002;25: 102-108.
2. 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.
3. Chen Y, Shen TJ. Rarefaction and extrapolation of species richness using an area-based Fisher's logseries. *Ecology and Evolution*. 2017;7:10066-10078.
4. 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. 1975;81-120.
5. 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.
6. Ulrich W, Ollik M, Ugland KI. A meta-analysis of species-abundance distributions. *Oikos*. 2010;119:1149-1155.
7. 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.
8. 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.
9. Wells JW. Coral reefs: 609-632; in J.W. Hedspeth editor, *Treatise on marine ecology and paleoecology*. Geological Society of America. Mem. 1957;67.
10. Glynn PW. High complexity food webs in low-diversity eastern Pacific reef-coral communities. *Ecosystems*. 2004;7:358-367.
11. Goren M. & Spanier E. The communities of benthic fish in Foul Bay (Tiran Island, Red Sea). *Oceanologica Acta*. 1985;8(4): 471-478.
12. Willis A. Rarefaction, alpha diversity and statistics. *bioRxiv*; 2017. DOI: 10.1101/231878
13. Rumohr H, Karakassis I, Jensen JN. Estimating species richness, abundance and diversity with 70 macro-benthic replicates in the Western Baltic Sea. *Marine Ecology Progress Series*. 2001; 214:103-110.
14. Fontaine B, Bouchet P et al. The European union's 2010 target: Putting rare species in focus. *Biodiversity and Conservation*. 2007;139:167-185.
15. 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.
16. Bracken M, Low N. Realistic losses of rare species disproportionately impact higher trophic levels. *Ecology Letters*. 2012;15: 461-467.
17. 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.
18. 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.
19. Ignatiades L, Gotsis-Skretas O. The contribution of rare species to coastal phytoplankton assemblages. *Marine Ecology*. 2014;35:132-145.
20. 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.
21. 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; 283:0084.
DOI: 10.1098/rspb.2016.0084
22. Violle C, Thuillier W, Mouquet N, Munoz F, Kraft NJB, Cadotte MW, Livingstone SW & Moullot D. Functional rarity: the ecology of outliers. *Trends in Ecology*; 2017.
DOI: 10.1016/j.tree.2017.02.002
 23. 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.
 24. 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
 25. 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.
 26. 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.
 27. Connolly SR, Hughes TP & Bellwood DR. A unified model explains commonness and rarity on coral reefs. *Ecology Letters*. 2017; 20:477-486.
 28. Chen Y, Shen TJ. Rarefaction and extrapolation of species richness using an area-based Fisher's logseries. *Ecology and Evolution*. 2017;7:10066-10078.
 29. 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 EURING Technical Meeting and Workshop*, Dunedin, New Zealand; 2007.
 30. 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.
 31. Ulrich W, Ollik M, Ugland KI. A meta-analysis of species-abundance distributions. *Oikos*. 2010;119:1149-1155.
 32. 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.
 33. 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.
 34. 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>
 35. 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 >
 36. Chao A, Hsieh T, Chazdon RL, Colwell RK & Gotelli NJ. Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology*. 2015;96(5):1189-1201.
 37. 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
 38. 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>
 39. Heip CHR, Herman PMJ & Soetaert K. Indices of diversity and evenness. *Océanis*. 1998;24(4):61-87.
 40. Strong WL. Assessing species abundance unevenness within and between plant communities. *Community Ecology*. 2002; 3(2):237-246.
DOI: 10.1556/COMEC.3.2002.2.9
 41. Grzès IM. Ant species richness and evenness increase along a metal pollution gradient in the Boleslaw zinc smelter area. *Pedobiologia*. 2009;53:65-73.

42. Loreau M. Species abundance patterns and the structure of ground-beetle communities. *Ann. Zool. Fennici.* 1992;28: 49-56.
43. Magurran AE, Henderson PA. Explaining the excess of rare species in natural species abundance distributions. *Nature.* 2003;422:714-716.
44. Connolly SR, Hughes TP, Bellwood DR & Karlson RH - Community structure of corals and reef fishes at multiple scales. *Science.* 2005;309:1363-1365.
45. Ulrich W, Soliveres S, Thomas AD, Dougill AJ, Maestre FT - Environmental correlates of species rank-abundance distributions in global drylands. *Europe PMC Funders Group.* 2016;20:56-64.
46. Béguinot J. The hierarchical structuring of species abundances within communities: Disentangling the intensity of the underlying structuring process behind the apparent unevenness pattern. *Advances in Research.* 2018;16(1):1-12. DOI: 10.9734/AIR/2018/43918
47. Béguinot J. Disentangling and quantifying the functional determinants of species abundance unevenness in ecological communities. *Advances in Research.* 2019;19(1):1-14. DOI: 10.9734/AIR/2019/v19i130114
48. DeBenedictis PA. On the correlations between certain diversity indices. *The American Naturalist.* 1973;107: 295-302.
49. Stirling G, Wilsey B. Empirical relationships between species richness, evenness and proportional diversity. *The American Naturalist.* 2001;158(3):286-299.
50. Smith B, Wilson JB. A consumer's guide to evenness indices. *Oikos.* 1996;76:70-82.
51. Loiseau N, Gaertner JC. Indices for assessing coral reef fish biodiversity: The need for a change in habits. *Ecology and Evolution.* 2015;5(18):4018-4027.
52. Buzas MA, Hayek LAC. On richness and evenness within and between communities. *Paleobiology.* 2005;31(2): 199-220.
53. MacArthur RH. On the relative abundance of bird species. *Proceedings of the National Academy of Sciences U.S.A.* 1957;43:293-295.
54. Wilson JB. Would we recognize a Broken-Stick community if we found one? *Oikos.* 1993;67(1):181-183.
55. Su Q. A relationship between species richness and evenness that depends on specific relative abundance distribution. *Peer J.* 2018;6:e4951.
56. Lande R, DeVries PJ, Walla TR When species accumulation curves intersect: implications for ranking diversity using small samples. *Oikos.* 2000;89(3):601-605.
57. Béguinot J. 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. *Asian Journal of Environment & Ecology.* 2017;2(1):1-12. DOI:10.9734/AJEE/2017/32040; <hal-01700717>.
58. Béguinot J. Inter-annual variations of true species richness in a subtropical butterfly assemblage: An estimation based on least-biased extrapolations of species accumulation curves. *Asian Journal of Biology.* 2017;2(4):1-16. DOI: 10.9734/AJOB/2017/33876; <hal-01700765>
59. Béguinot J. Extrapolation of total species richness from incomplete inventories: Application to the Gastropod fauna associated to coral reefs in 'Mannar Gulf Biosphere Reserve', India. *Asian Journal of Environment and Ecology.* 2017;4(3):1-14. DOI: 109734/AJEE/2017/36831; <hal-01807046>.
60. 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
61. 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
62. Béguinot J. Inferring total species richness and the exhaustive hierarchical structuring of species abundances in tropical Sea Stars communities (Asteroidea), using numerical extrapolation of partial

- inventories. Asian Journal of Environment and Ecology. 2018;8(2):1-25.
DOI: 10.9734/AJEE/2018/46272
63. Béguinot J. Comparing the complete hierarchical structuration of species abundances in reef fish communities according to coral morphology, using the numerical extrapolation of only incomplete inventories. Asian Journal of Environment and Ecology. 2018;8(1):1-20.
DOI: 10.9734/AJEE/2018/45402
64. Béguinot J. Influence of Coral Architecture on Species Richness and the Hierarchical Structuration of Species Abundances in Reef Fish Communities: A Case Study in the Eastern Tropical Pacific. Asian Journal of Environment & Ecology. 2019;8(3):1-21.
DOI: 10.9734/AJEE/2018/V8I330075
65. Béguinot J. Influence of fishing activity on the total species richness and the abundance unevenness in reef fish communities: A case study in a Brazilian tropical coral complex. International Journal of Environment and Climate Change. 2019;9(1):58-76.
DOI: 10.9734/IJECC/2019/v9i130097
66. Béguinot J. Influence of Coral-reef complexity on Species Richness and the Hierarchical Structuration of Species Abundances in Reef Fish Communities: A Case Study in south-east Brazil. Asian Journal of Environment & Ecology. 2019; 9(3):1-20.
DOI: 10.9734/AJEE/2019/v9i330098
67. Baselga A. Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography. 2010; 19:134-143.
68. Rattray A, Andrello M, Asnaghi V, Bevilacqua S et al. Geographic distance, water circulation and environmental condition shape the biodiversity of Mediterranean rocky coast. Marine Ecology Progress Series. 2016;553:1-11.
69. Béguinot J. Influence of environmental heterogeneity on the species composition, species richness and species abundances unevenness in reef-associated *Conus* communities (Neogastropoda) from Papua New-Guinea. Asian Journal of Environment & Ecology. 2019;10(3):1-21.
DOI: 10.9734/AJEE/2019/v10i330116
70. Béguinot J. Variations in total species richness and the unevenness of species abundance distribution between two distant *Conus* communities (Neogastropoda): A case study in Mannar Gulf (India). Asian Journal of Environment & Ecology. 2019;9(4):1-18.
DOI: 10.9734/AJEE/2019/v9i430102
71. 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>
72. 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>
73. 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.
74. 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.). Oxford University Press, Oxford. 2010;345.
75. Gotelli NJ, Chao A. Measuring and estimating species richness, species diversity, biotic similarity from sampling data. In: Levin S.A. (ed.) Encyclopedia of Biodiversity. 2013; second edition. Waltham, MA: Academic Press. 2013;5: 195-211.
76. Brose U, Martinez ND, Williams RJ. Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. Ecology. 2003;84(9): 2364-2377.

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 [34,71,72]:

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

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 [34]:

* for f_1 up to $f_2 \rightarrow R_1(N) = (R(N_0) + f_1) - f_1 \cdot N_0/N$

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

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

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

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

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:

* $f_1 \leq f_2 \rightarrow \Delta_{j1} = f_1; R_1(N)$

* $f_2 < f_1 \leq 2f_2 - f_3 \rightarrow \Delta_{j2} = 2f_1 - f_2; R_2(N)$

* $2f_2 - f_3 < f_1 \leq 3f_2 - 3f_3 + f_4 \rightarrow \Delta_{j3} = 3f_1 - 3f_2 + f_3; R_3(N)$

* $3f_2 - 3f_3 + f_4 < f_1 \leq 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{j4} = 4f_1 - 6f_2 + 4f_3 - f_4; R_4(N)$

* $f_1 > 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{j5} = 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5; R_5(N)$

N.B. 1: As indicated above (and demonstrated in details in [34]), 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 *under*-estimates of the true number of missing species [26,28, 73-75]. Also, this shows that the approach initially proposed by [76] – 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 $0.6 \times f_2$ (that is when sampling completeness closely approaches exhaustivity), then Chao estimator may alternatively be selected: see reference [35].

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 [37].

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:

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

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 $i > R_0$):

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

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.

© 2019 Beguinet; 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/53446>