

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)

Abstract

The detailed analysis of Species Abundance Distributions ("S.A.D.s") can shed light on how member-species mutually arrange themselves within communities, provided that the complete distribution of species abundances is made available first. In this respect, the numerical extrapolation applied to incomplete "S.A.D." can effectively compensate for incompleteness, when having to deal with substantially incomplete samplings. Indeed, almost as much information as would be obtained from truly complete sampling and truly exhaustive "S.A.D." can finally be released, although the taxonomic identities of unrecorded species remain ignored.

To take full advantage of this new approach, a recently developed procedure allowing the least-biased numerical extrapolation of "S.A.D.s" has been applied to three partially sampled gastropods communities associated to coral-reef in Mannar Gulf (S-E India).

The following main results were derived from the three numerically completed "S.A.D.s": (i) once completed, all three numerically "S.A.D.s" fairly well comply with the log-normal model, suggesting that, in all three communities, the process of hierarchical structuration of species abundances likely involves the combined contributions of many independent factors; (ii) moreover, the same holds true for each of the two involved feeding guilds (primary and secondary consumers) considered separately; (iii) when compared to secondary consumers, the guild of primary consumers has a much lower species richness but, on the other hand, exhibits a more strongly structured distribution of species abundances, which materializes not only in term of the apparent unevenness of species abundances but also after having disentangled the genuine intensity of the underlying process driving the hierarchical structuration.

All the preceding features are shared in common by the three communities studied, in line with the similarity of their environmental contexts. Only minor features can actually slightly distinguish between communities, reflecting the difference in distances between them.

Key-words : coral reef, gastropod, incomplete inventory, ranked abundance distribution, species diversity, evenness, climate change, global warming

1. Introduction

Tropical marine ecosystems in shallow waters are of major concerns to ecologists and conservationists, as they are considered as one of the first ecosystem to suffer significantly from obvious impacts due to the on-going *climate-warming* [1,2]. Moreover, these same ecosystems are of major interest, being considered as embodying remarkably high levels of diversity and biological complexity [1, 3-5]. However, precisely because of their usually high species richness, also including *numerous rare* species, the samplings of these communities often remain substantially incomplete [6]. More generally, sampling incompleteness is doomed to become the unavoidable consequence of the generalization of “rapid assessments” and “quick surveys”, at least when having to deal with species-rich communities, as is usually the case with terrestrial or marine *invertebrates* faunas, especially under tropical climate. In turn, such *partial* inventories prevent a reliable appreciation of the key *descriptive* and *functional* aspects of the internal organization within species communities [7–9]. Hence, the recently acknowledged necessity to implement relevant *numerical extrapolations* applied to both: (i) the as-recorded species accumulation process, beyond the already reached sampling-size and (ii) the as-recorded distribution of species abundances in the studied community.

Indeed, the extrapolated numerical data, regarding all the species that had remained undetected after partial sampling, is absolutely necessary to gain access to the aforementioned descriptive and functional features of the internal organization within species communities.

In particular, once properly numerically *completed* (and *only* when it is so: [6]), the distribution of species abundances can provide synthetic data, both qualitative and quantitative, about the underlying process that drives the hierarchical structuring of species abundances within community [10–14]. Although no deeper details may be extracted from this synthetic overview, it has, yet, the advantage of being straightforward. In particular, this synthetic approach does not involve the long and tedious analytical programs that would be otherwise required to go deeper in the details of the structuring processes. As such, this synthetic approach can serve, at least, as a convenient preliminary investigation.

As complete – or numerically completed – abundance distributions are mandatory for this purpose, **new** procedures for the numerical extrapolation of (i) the Species Accumulation Curve and (ii) the Species Abundance Distribution have recently been developed, aiming at providing reliable, *least-biased* inferences about (i) the number and (ii) the distribution of abundances of all those species that remained undetected after incomplete samplings.

Hereafter, I report and discuss the results derived from the numerical extrapolations applied to the partial inventories of three marine Gastropod communities associated to tropical coral reefs surrounding three small islands within Mannar Gulf Biosphere Reserve (south-eastern coast of India). Partial samplings were carried out by Mohanraj and co-workers [15], while the estimates of the true (total) species richness of each of these three communities were already reported by [16].

In turn, based on the results derived from these numerical extrapolations, I address the following questions:

(i) which kind of process is involved in the hierarchical structuring of species abundances, according to the type of models that fit best to completed Species Abundance Distributions;

(ii) to what extent the intensity of the hierarchical structuring of species abundances actually differs from the resulting apparent degree of abundances unevenness in each case?

(iii) do the different feeding guilds, to which belong the reef associated Gastropods, actually matter (or are indifferent), as regards the hierarchical structuring process (in term of structuring intensity) and the resulting pattern of species abundances (in term of apparent degree of unevenness);

(iv) does any significant difference exist between the three Gastropods communities, as regards abundance structuration and, if any, does these differences can be related to inter-sites distances?

Relevantly addressing all these issues systematically require considering complete Species Abundance Distributions, hence the necessity to *previously extrapolate with minimized bias* the partial samplings that was made available.

2. Materials and Methods

2.1 - field data

The coral reefs surrounding the shores of three small islands, located in “Mannar Gulf Biosphere Reserve” (islands of ‘Hare’, ‘Vaan’ and ‘Koswari’) were sampled for their associated gastropods communities. The recorded data, issued from these partial surveys (on which numerical extrapolations will be based) is reported in details in [15] and the reader is invited to browse this publication for the contextual information regarding the sites and the implemented sampling methods. The number N_0 of collected individuals and the number R_0 of recorded species are recalled here at Table 1, for each investigated site. At all three sites, the numbers f_1 of singletons (species recorded only once) are closed to 20% of the total number R_0 of recorded species, which denotes substantial incompleteness for each of the three samplings, a common situation in practice, as already underlined. As all the species lists reported by the authors include the respective abundances of the recorded species, the corresponding data was appropriate to implement the procedure of numerical extrapolation.

2.2 - the numerical extrapolation procedure and its exploitation

As mentioned above, the Species Abundance Distributions (“S.A.D.”) provides the basic data necessary (i) to describe the *pattern* of structuration of species within community and (ii) to qualify and quantify the underlying *process* that drives this structuration.

Yet, to accurately exploit its full potential [17, 18], the “S.A.D.” requires:

- first, to be *corrected* for bias resulting from drawing stochasticity liable to the finite size of samplings,

- second, and still more importantly, to be *completed* by *numerical extrapolation*, to the extent that sampling is suspected to be incomplete, as revealed by the subsistence of singletons as is the case here.

The appropriate procedure of correction and, subsequently, of least-biased numerical extrapolation of the as-recorded “S.A.D.” is described in details in [18] and briefly

summarized in Appendix 1. In addition, an example of implementation of the procedure is commented in details in [19].

Classically, the “S.A.D.” is graphically presented according to the so-called “Ranked Abundance Distribution” (also known as “Whittaker plot”), according to which the (log-transformed) abundances a_i are plotted against the rank i of species, the species being ordered by decreasing values of abundance (with, thus, a_1 and a_{S_t} respectively standing for the highest and the lowest abundances in a community of S_t species).

2.2.1. – description of the pattern of species abundance structuration

The “S.A.D.” (completed by extrapolation to include undetected species, if any) conveys all the relevant quantitative data relative to the internal structuration of species abundances within community. In particular, it is always advisable to use species-abundance plots to study evenness or unevenness [20]. For commodity, the “S.A.D.” may be synthetically reduced to its two major descriptors: the *total species richness* S_t and the *degree of unevenness*, U , of the abundance distribution. According to the aforementioned, classical mode of representation of “S.A.D.”, it goes natural to quantify the degree of abundance unevenness U as the average slope of the abundance decrease along the whole range of the abundance distribution, as already proposed [21], that is:

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

2.2.2. - qualification of the underlying structuring process: type and intensity

Beyond the mere description of the pattern of structuration, quantified by U , the complete “S.A.D.” can help addressing several important questions regarding the *intensity* and the *type* of the process driving the hierarchical structuration of abundances in the community (i.e. the differential allocations of relative abundances among member-species of the community).

As regards the *kind* of structuring process involved, it is possible to distinguish between two major alternative hypotheses: schematically, the hierarchical structuration of abundances may results either: (i) from the contribution of only one dominant factor or (ii) from the combined contributions of many independent factors acting together. This may be tested by checking the conformity of the corresponding “S.A.D.” to either the *log-series* model or the *log-normal* model, respectively [10, 22-25].

Now, as regards the *intensity* of the structuring process, it is first necessary to remind that the degree of unevenness U is *inadequate* in this respect, due to its additional *mathematical* dependence upon the species richness S_t [26, 27]; see also Appendix 3), which ultimately results in masking the genuine contribution of the structuring process itself [12, 13, 28]. To get rid of this *spurious* dependence, deprived from biological significance, it has been shown appropriate to cancel its influence by comparing the actual “S.A.D.” to the corresponding “broken-stick” distribution [29], computed for the same species richness S_t [19, 30, 31]. Accordingly, the genuine intensity, “ I_{str} ”, of the hierarchical structuring process is relevantly defined through standardizing the unevenness of the actual “S.A.D.” to the unevenness of the corresponding “broken-stick” distribution, computed for the same species richness S_t , that is:

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

with the abundances being classically log-transformed and 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.

Thus defined, freed from the purely mathematical influence of species richness, the index I_{str} accounts for *those biological factors only* that contribute to the hierarchical structuring of abundances within community.

Note at last, that the methodologic approach above may be applied, as well, to any subsets of interest, for example to each of the feeding guilds that co-occur in the community.

3. Results

3.1 Estimated total species richness of each community

The three communities have rather similar estimated total species richness: ‘Koswari’: 53 species, ‘Vaan’: 51 species, ‘Hare’ 49 species (Table 1; [16]).

Table 1 – The number of collected individuals N_0 , the number of recorded species R_0 , the type of nonparametric estimator selected as being the least-biased, 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 . Estimations are computed according to the least-biased procedure: selection key provided in Appendix 2.

parameters of sampling and numerical extrapolation	SITES		
	HARE	VAAN	KOSWARI
nb. collected individuals N_0	310	484	416
nb. recorded species R_0	35	40	38
least-biased estimator	JK-5	JK-5	JK-4
nb. unrecorded species Δ	14	11	15
total species richness S_t	49	51	53
sample completeness R_0/S_t	71%	78%	72%

In this respect, the spatial distance between communities does not seem to have any influence: the (slight) difference of species richness is the same between ‘Hare’ (49 species) and ‘Vaan’ (51 species), separated by > 100 km, than between ‘Koswari’ (53 species) and ‘Vaan’ (51 species), although distant of 3 km only.

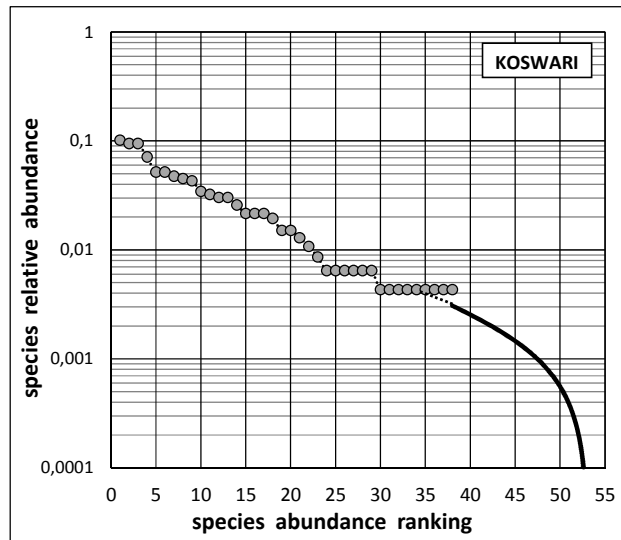
A similar lack of singling out of ‘Hare’ community, as compared to ‘Koswari’ and ‘Vaan’ communities, is reported in an another study of reef-associated gastropod assemblages [32], with respect to the levels of (i) species diversity, (ii) species richness and (iii) evenness of species abundances.

3.2 Species Abundance Distribution for each community, duly corrected and extrapolated

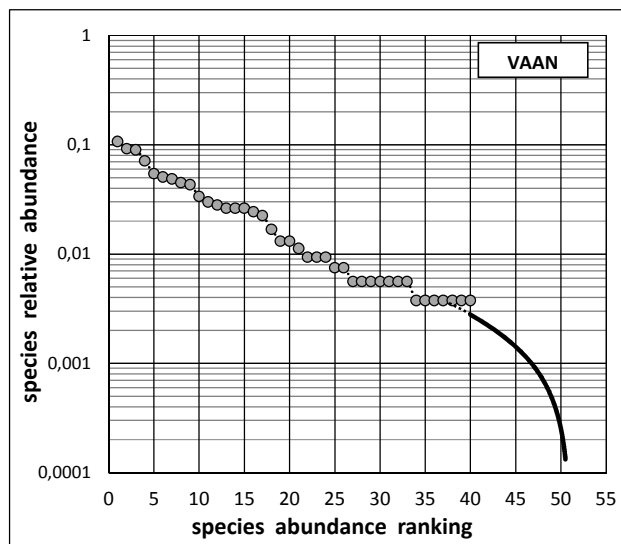
Figures 1, 2, 3, provide the graphical representations of the Species Abundance Distribution (“S.A.D.”) for each of the three communities under study. These “S.A.D.s” have been (i) bias-corrected and then (ii) extrapolated according to equations (A1.1) and (A1.2) respectively: see Appendix 1.

According to the usual convention of graphical representation for S.A.D.s,

214 - along the abscissa, member-species are ranked by decreasing values of their
 215 respective abundances (relative abundance ' a_i ' for the species of rank ' i ');
 216 - along the ordinate, the relative abundances are plotted according to a logarithmic
 217 scale ($\log_{10}$) (although, for specific purpose, an ordinary untransformed scale may be
 218 adopted, as suggested in [29]).
 219 The (bias corrected) abundances of the *recorded* species are plotted as grey circles,
 220 while the *extrapolated* part of the abundance distribution is plotted as a thick solid line.
 221



222
 223
 224 **Figure 1** – The bias-corrected and extrapolated species abundance distribution (“S.A.D.”) for the
 225 reef-associated gastropod community around ‘Koswari’ island: *grey circles*: recorded part of the
 226 “S.A.D.” after correction for bias; *solid line*: least-biased extrapolation of abundance distribution for
 227 the set of unrecorded species due to incomplete sampling (sampling completeness: 72%).
 228



229

Figure 2 – The bias-corrected and extrapolated species abundance distribution (“S.A.D.”) for the reef-associated gastropod community around ‘Vaan’ island: *grey circles*: recorded part of the S.A.D. after correction for bias; *solid line*: least-biased extrapolation for the set of unrecorded species due to incomplete sampling (sampling completeness: 78%).

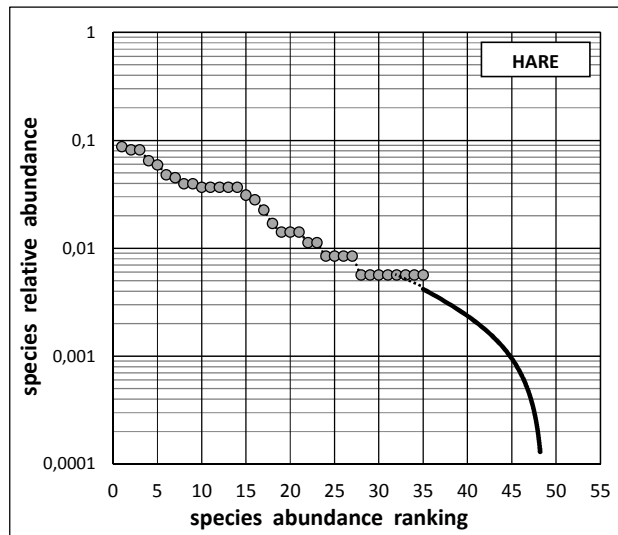


Figure 3 – The bias-corrected and extrapolated species abundance distribution (“S.A.D.”) for the reef-associated gastropod community around ‘Hare’ island: *grey circles*: recorded part of the “S.A.D.” after correction for bias; *solid line*: least-biased extrapolation for the set of unrecorded species due to incomplete sampling (sampling completeness: 71%).

3.3 testing for the kind of process involved in the structuring of species abundances

Figures 4-5, 8-9, 12-13 provide the comparison of the “S.A.D.” of each community to the corresponding “log-series” and “log-normal” models, computed for the *same* species richness respectively. As mentioned in Methods section, these comparisons thereby allow to infer which kind of process is more likely to be involved in the hierarchical structuration of species abundances: schematically, the “log-series” model suggests that *only one* (or very few) *dominant* factor is structuring the distribution of species abundances in the community while “log-normal” model suggests the intermingled contributions of *numerous* and *independent* factors, involved together in the structuration of the community.

For all three communities, the species abundance distributions clearly fit best the “log-normal” model than the “log-series” model.

Now, Figures 6-7, 10-11, 14-15 allow a closer examination of the fitting with the “log-normal” model and, in particular, highlight in all three cases a local departure from the otherwise fairly well fitted “log-normal” model. Specifically, the most abundant species (species of rank $i = 1$ and, to a much lesser extent, of rank $i = 2$) has an abundance level which is only about half the prediction of the “log-normal” model. In all three cases, the gap is statistically highly significant (statistical test based on the Bayesian inference of

the probability distribution of values for the abundance a_1 of the species of rank $i = 1$: $p < 0.001$). The likely origin and ecological meaning of this gap is discussed later.

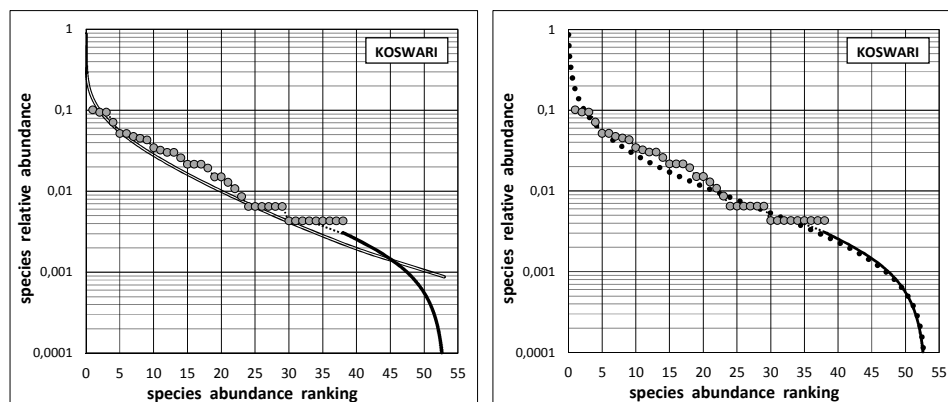


Figure 4 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Koswari’ island (grey circles and solid line) and the corresponding “log-series” model (double line).

Figure 5 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Koswari’ island (grey circles and solid line) and the corresponding “log-normal” model (dotted line).

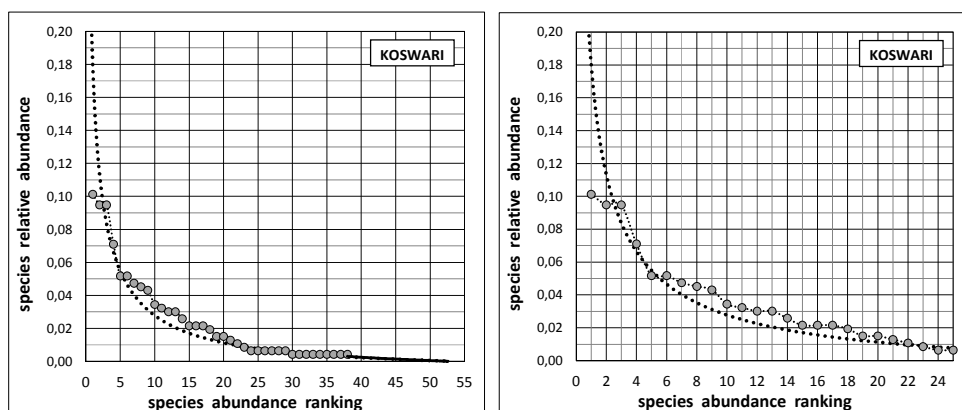


Figure 6 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Koswari’ island (grey circles and solid line) and the corresponding “log-normal” model (dotted line).

Figure 7 – Same as Figure 6 with arithmetical scale for abundances (instead of the traditional logarithmic scale), which offers a more direct visual appreciation of relative abundances, as suggested by [29].

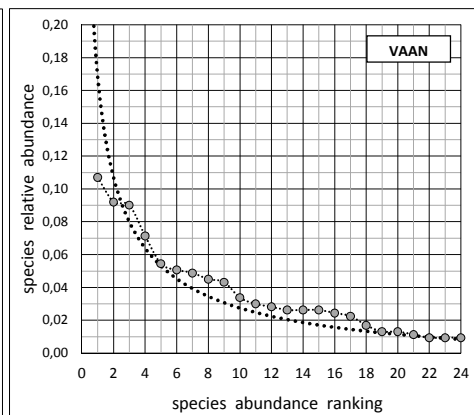
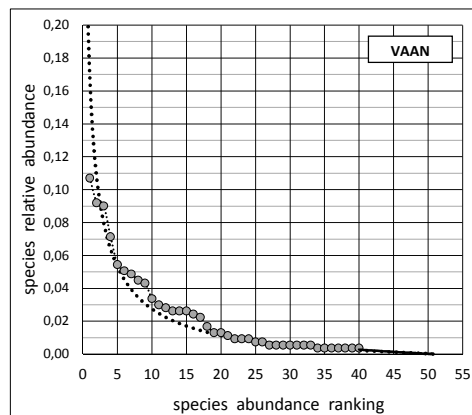
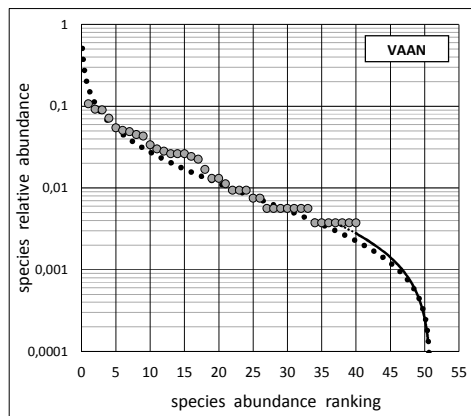
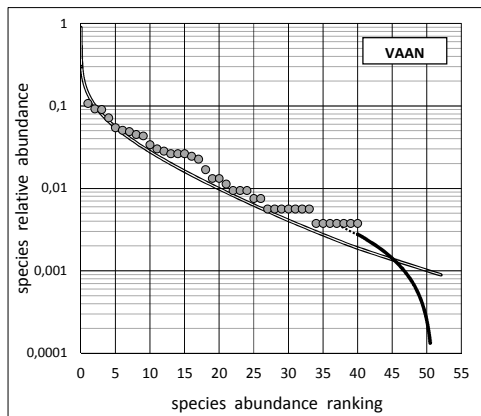


Figure 10 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Vaán’ island (grey circles and solid line) and the corresponding “log-normal” model (dotted line).

Figure 11 – Same as Figure 10 with arithmetical scale for abundances (instead of the traditional logarithmic scale), which offers a more direct visual appreciation of relative abundances, as suggested by [29].

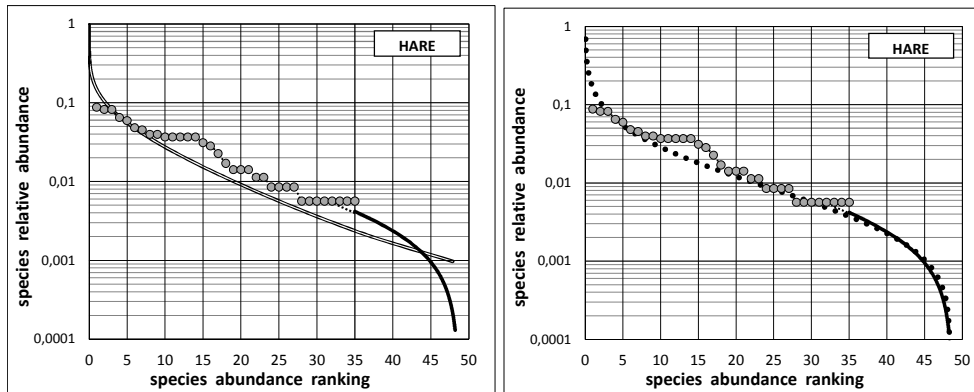


Figure 12 – Comparison between the bias-corrected and extrapolated "S.A.D." for the community at 'Hare' island (grey circles and solid line) and the corresponding "log-series" model (double line).

Figure 13 – Comparison between the bias-corrected and extrapolated "S.A.D." for the community at 'Hare' island (grey circles and solid line) and the corresponding "log-normal" model (dotted line).

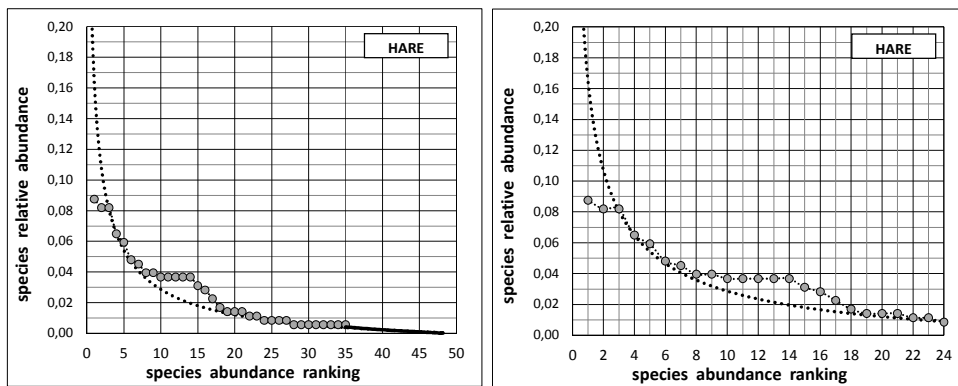


Figure 14 – Comparison between the bias-corrected and extrapolated "S.A.D." for the community at 'Hare' island (grey circles and solid line) and the corresponding "log-normal" model (dotted line).

Figure 15 – Same as Figure 14 with arithmetical scale for abundances (instead of the traditional logarithmic scale), which offers a more direct visual appreciation of relative abundances, as suggested by [29].

3.4 the genuine intensity of the process of hierarchical structuration of abundances

As emphasized in Methods section, the genuine intensity of the process driving the hierarchical structuration of abundances is relevantly appreciated by comparing the "S.A.D." of the studied community to the corresponding "broken-stick" model computed for the same species richness. Figures 16 to 21 provide such comparisons of the "S.A.D." of each community with the corresponding "broken-stick" model, computed for the same species richness.

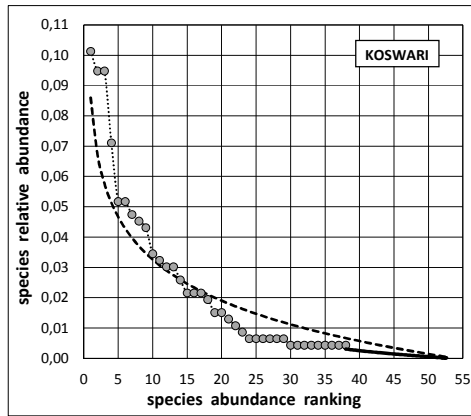
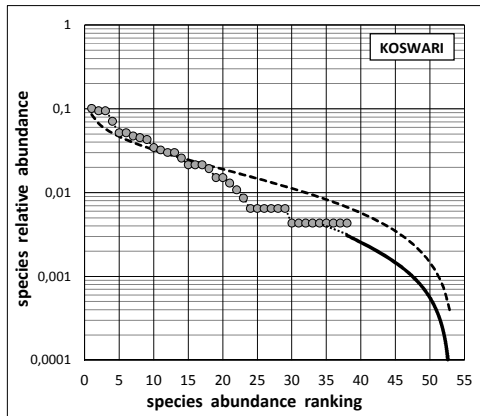


Figure 16 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Koswari’ island (grey circles and solid line) and the corresponding “broken-stick” model (dashed line) – ordinate with *logarithmic* scale.

Figure 17 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Koswari’ island (grey circles and solid line) and the corresponding “broken-stick” model (dashed line) – ordinate with *arithmetic* scale.

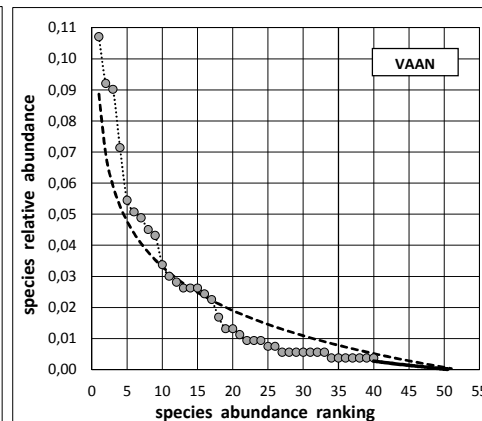
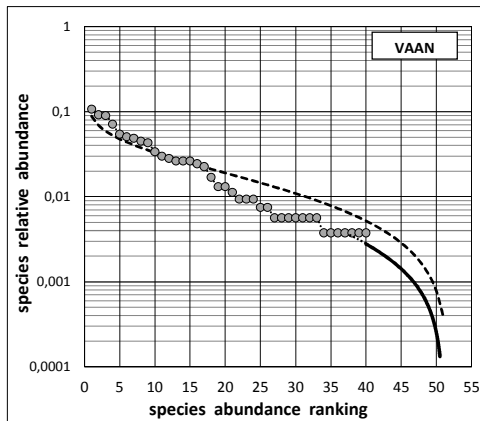


Figure 18 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Vaan’ island (grey circles and solid line) and the corresponding “broken-stick” model (dashed line) – ordinate with *logarithmic* scale.

Figure 19 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Vaan’ island (grey circles and solid line) and the corresponding “broken-stick” model (dashed line) – ordinate with *arithmetic* scale.

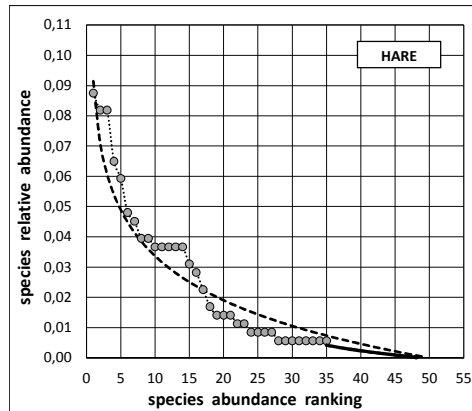
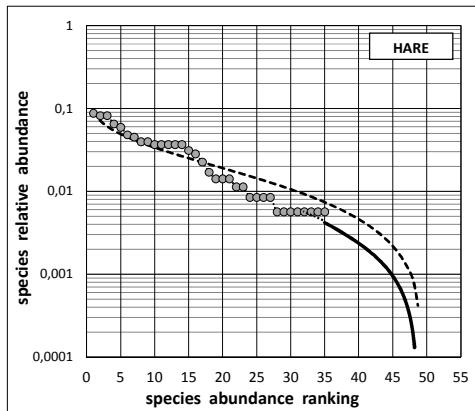


Figure 20 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Hare’ island (grey circles and solid line) and the corresponding “broken-stick” model (dashed line) – ordinate with *logarithmic* scale.

Figure 21 – Comparison between the bias-corrected and extrapolated “S.A.D.” for the community at ‘Hare’ island (grey circles and solid line) and the corresponding “ideally even” model (dotted line) and “broken-stick” model (dashed line) – ordinate with *arithmetic* scale.

In a more synthetic way, Table 3 will provide the evaluations of the genuine intensity, I_{str} , of the process driving the hierarchical structuration of species abundances, for each of the three communities.

3.5 Species richness and the degree of hierarchical structuring of species abundances in each feeding guild considered separately

Two feeding guilds coexist in the three studied **Gastropod** communities: *primary* consumers (“herbivores”) and *secondary* consumers (“carnivores”). The numerical extrapolation procedure, applied above to each community *as a whole*, may be implemented, as well, to each of these two feeding guilds, *separately*.

The respective contributions of each feeding guild – in terms of total species richness and cumulated number of individuals per guild – are derived accordingly (Table 2). In all three communities, the guild of *secondary* consumers is far more species-rich (by twice to almost thrice) than is the guild of *primary* consumers. Regarding the cumulated number of individuals per guild, the trend is less marked because *primary* consumers are more numerous by species, on average, than are *secondary* consumers.

Table 2 – Respective contributions, per *feeding guild* (*primary* and *secondary* consumers), to: (i) the recorded and the total species richness and (ii) the cumulated numbers of individuals.

respective contributions of primary and secondary feeders	SITES					
	HARE		VAAN		KOSWARI	
	<i>primar</i>	<i>secund</i>	<i>primar</i>	<i>secund</i>	<i>primar</i>	<i>secund</i>
nb. recorded species R_0	6	29	7	33	6	32
nb. unrecorded species Δ	2	12	1	10	4	11
total species richness S_t	8	41	8	43	10	43
contrib. to sp. richness %	16 %	84 %	16 %	84 %	19 %	81 %

<i>contrib. to nb. individ. %</i>	28 %	72 %	32 %	68 %	31 %	69 %
-----------------------------------	-------------	-------------	-------------	-------------	-------------	-------------

The Species Abundance Distributions specific to each of the two feeding guilds co-occurring in the community at 'Koswari' (the richest one, taken as an example) are jointly provided in Figure 22.

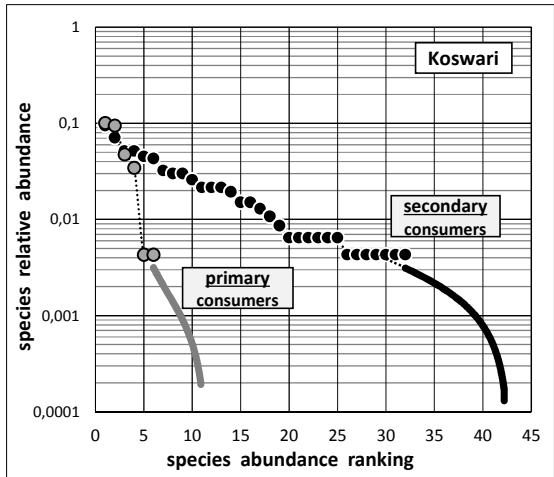


Figure 22 – The bias-corrected and extrapolated “S.A.D.” for the two feeding guilds in the marine gastropod community at ‘Koswari’ island. Grey figures: the *primary* consumers; black figures: the *secondary* consumers. The distribution of species abundances among *primary* consumers is strikingly more uneven (steeper slope of abundance decrease) than it is among *secondary* consumers.

These abundance distributions are drastically different, with a strongly steeper slope of the “S.A.D.” for *primary* consumers than for *secondary* consumers. The distribution of species abundances among *primary* consumers appears, thus, far more uneven than it is among *secondary* consumers. This striking difference between the two feeding guilds still remains obvious when the slope of each species abundance distribution is compared to the slope of the corresponding “broken-stick” model (computed for the same species richness): Figure 23. This means that not only the apparent unevenness (as a *descriptive* parameter), but also the intensity I_{str} of the underlying process driving the hierarchical structuration of abundances (as a *functional* parameter) are stronger for *primary* consumers as compared to *secondary* consumers. $U = 0.272$ and $I_{str} = 1.79$ for *primary* consumers against $U = 0.070$ and $I_{str} = 1.27$ only for *secondary* consumers.

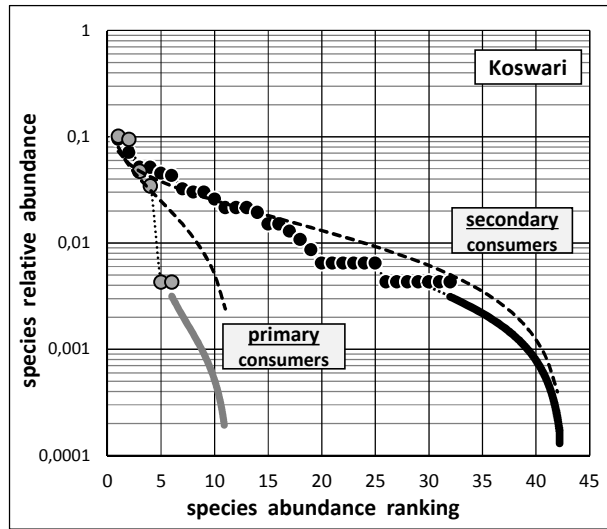


Figure 23 – Same as Figure 22, with the additional superimposition of the corresponding “broken stick” models (dashed lines) for *primary* consumers and for *secondary* consumers respectively. The average slope of the “S.A.D.” of *secondary* consumers is just slightly steeper than the corresponding “broken stick” model, while the average slope of the “S.A.D.” of *primary* consumers is strongly steeper than the corresponding “broken stick” model.

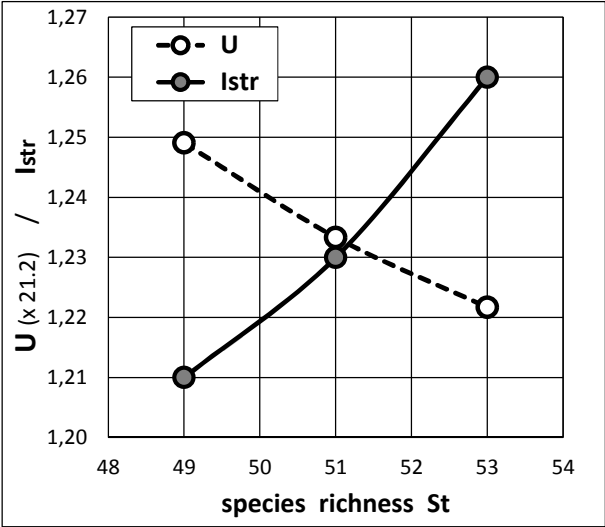
3.6 Overview of the main quantitative features characterizing the three communities

The (extrapolated) *total species richness* on the one hand; the pattern $\bar{U}$ and the intensity I_{str} of the underlying process of *hierarchical structuring of species abundances* (exhaustively unveiled by numerical extrapolation) on the other hand, altogether convey much of the quantitative aspects that characterize the internal structuring of natural communities of species in their particular environment. The corresponding results for each of three communities are summarized in Table 3.

Table 3 – A synthetic summary of the main quantitative features of the species abundances organization within community, as derived from each “S.A.D.”, duly *completed by extrapolation*: (i) the total species richness S_t of the community ; (ii) the relative abundances a_1 and a_{st} of the most and the least abundant species (ranks $i = 1$ and $i = S_t$) ; (iii) the unevenness of abundances in the community: $U = \log(a_1/a_{st})/(S_t - 1)$ and in the corresponding “broken-stick”: $U' = \log(a'_1/a'_{st})/(S_t - 1)$ and (iv) the genuine intensity of the structuring process $I_{str} = U/U'$.

SITES	parameters of the distribution of species abundances						
	S_t	a_1	a_{st}	a_1/a_{st}	U	U'	I_{str}
KOSWARI	53	0.1012	0.000102	992	0.0577	0.0458	1.26
VAAN	51	0.1070	0.000132	811	0.0582	0.0473	1.23
HARE	49	0.0875	0.000130	673	0.0589	0.0487	1.21

411 In turn, the main trends are highlighted in Figure 24 in which both the abundance
 412 unevenness U and the intensity of the structuring process I_{str} are plotted together
 413 against the species richness S_t for each three communities.
 414 The (positive) dependence of the intensity I_{str} of the structuring process upon species
 415 richness is *functionally* relevant, since I_{str} was precisely defined *freed* from any direct
 416 dependence on S_t *a priori*. Of course, three communities only do not suffice to draw any
 417 statistically sound conclusion and to propose any generalization, but the crude
 418 observation yet deserves being acknowledged.
 419 Now, at first sight paradoxically, the level of unevenness is, on the contrary, decreasing
 420 with S_t , which, yet, is nothing else than the consequence of the mathematical, negative
 421 influence of the species richness on the apparent level of unevenness U , as already
 422 argued above. Thus, no functional relevance should be attributed to this decrease of U
 423 with S_t .
 424 From a methodological point of view, this exemplifies the fact that sticking to the first
 425 appearance – that is, here, considering the level of unevenness as the likely faithful
 426 mirror of the strength of hierarchical ranking of species abundances – would lead to
 427 quite an erroneous appreciation. However, relying only on unevenness level still
 428 remains **too often** the usual practice...
 429



430
 431
 432 **Figure 24** – The degree U of unevenness of species abundances (*dashed* line) and the intensity I_{str} of
 433 the underlying structuring process (*solid* line) plotted against the total species richness S_t , for the
 434 three studied marine Gastropod communities. The degree of abundance unevenness decreases with
 435 increasing species richness of communities. Yet, the opposite holds true for the genuine intensity of
 436 the structuring process driving this uneven distribution of species abundances: I_{str} increases with
 437 increasing species richness of communities. The discrepancy results from the negative mathematical
 438 dependence of U upon S_t . Note that for commodity of graphical comparison between U and I_{str} , the
 439 degrees of unevenness are uniformly multiplied by a same factor 21.2.
 440

441 3.7 Three more specific comparisons

Until now, all comparisons above failed to singularize any of the three communities with respect to the others. However, the following, more specifically focused comparisons best succeed in this respect.

3.7.1. the pattern of abundance distribution according to feeding guild

The “S.A.D.” of each feeding guild (primary and secondary feeders) was surveyed as a whole at section 3.5. Yet, further insight may be obtained by considering how each feeding guild is represented among the half most abundant species (i.e. species with ranks $i < S_t/2$) and among the half less abundant species (i.e. species with ranks $i \geq S_t/2$). As shown in Table 4, fairly equal distributions (all rather close to 50%) are recorded for both guilds in communities at ‘Vaani’ and at ‘Koswari’. By contrast, at ‘Hare’, the *primary* consumer species are more than twice as frequent among the most abundant (ranks $i < S_t/2$) than among the least abundant (ranks $i \geq S_t/2$): 69% against 31%. The figure, however, still lacks statistical support, due to the small numbers of species involved.

Table 4 – The numbers and proportions (%) of species in each feeding guild that occur either (i) among the half *more abundant*, that is with rank $< S_t/2$ or (ii) among the half *less abundant*, that is with rank $\geq S_t/2$. According to *primary* consumers and *secondary* consumers).

range of species abundances	SITES					
	HARE		VAANI		KOSWARI	
	primar	secund	primar	secund	primar	secund
half <i>more</i> abundant (rank $< S_t/2$)	5	19	4	21	5	21
half <i>less</i> abundant (rank $\geq S_t/2$)	2.2	22.5	4	22	5.6	21.5
half <i>more</i> abundant (rank $< S_t/2$)	69 %	45 %	50 %	49 %	47 %	49 %
half <i>less</i> abundant (rank $\geq S_t/2$)	31 %	55 %	50 %	51 %	53 %	51 %

3.7.2. the degree of hierarchical dominance of the three most abundant species

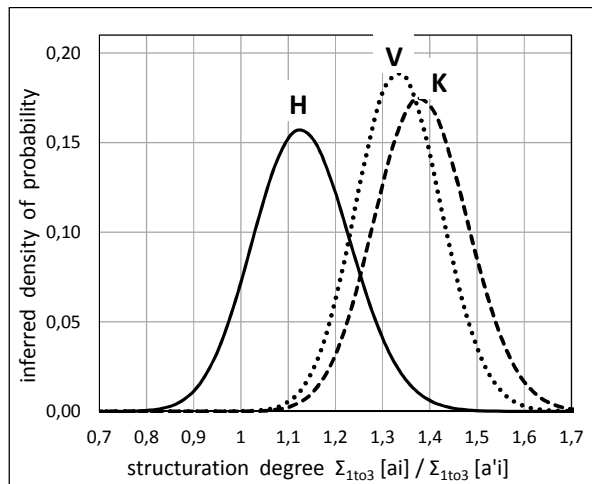


Figure 25 – Bayesian inferences applied to the degree of hierarchical dominance of the three most abundant species (Σa_i for ranks $i = 1$ to 3) after standardization to the broken-stick model ($\Sigma a'_i$ for

ranks $i = 1$ to 3). Thus defined, the degree of structuration is: $\Sigma_{1to3} [a_i] / \Sigma_{1to3} [a'_i]$. Labels H, V, K are for Hare, Vaan, Koswari.

The degree of hierarchical structuring of species abundances at each location was addressed at sections 3.4 and 3.6. Yet, further insight may be obtained by detailing the contribution to hierarchical dominance of the three most abundant species at each location (species of ranks 1, 2, 3). As already underlined, this contribution is best quantified after standardization to the corresponding “broken-stick” model. Accordingly, the ratio $\Sigma_{1to3} [a_i] / \Sigma_{1to3} [a'_i]$ is considered for each of three communities, with a_i standing for the actual “S.A.D.” and a'_i for the corresponding “broken-stick” distribution. Bayesian inference of the distribution of probabilities for the value of this ratio are provided at Figure 25 for each location. The community at ‘Hare’ singularizes as compared to the other two communities.

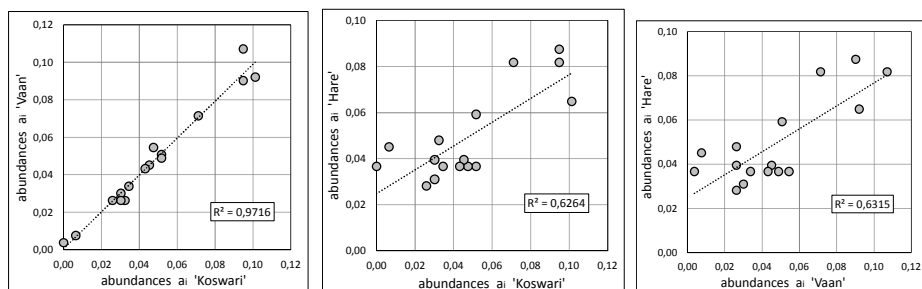
3.7.3. the congruence of taxonomic identity within the set of most abundant species

Beta-diversity, when comparing communities two by two, is a third, potentially distinctive criterion between communities at the three locations. To preserve significance, focus is placed on the set of the 16 most abundant species (Table 5).

Table 5 – Relative abundances of the 16 most abundant species in each of three locations ‘Koswari’ (K), ‘Vaan’ (V) and ‘Hare’ (H).

Species	Sites		
	K	V	H
<i>Trochus niloticus</i> Linné 1767	0,101	0,092	0,065
<i>Trochus radiatus</i> Gmelin 1791	0,095	0,107	0,082
<i>Murex trapa</i> Roding 1798	0,095	0,090	0,088
<i>Murex tribulus</i> Linné 1758	0,071	0,071	0,082
<i>Hemifusus pugilinus</i> (Born1778)	0,052	0,051	0,059
<i>Cypraea moneta</i> Linné 1758	0,052	0,049	0,037
<i>Lambis lambis</i> Linné 1758	0,047	0,054	0,037
<i>Monetaria annulus</i> (Linné 1758)	0,045	0,045	0,040
<i>Cypraea caurica</i> (Linné 1758)	0,043	0,043	0,037
<i>Umbonium</i> sp.	0,034	0,034	0,037
<i>Cypraea teres</i> Gmelin 1791	0,032	0,026	0,048
<i>Cyparea labrolineata</i> Gaskoin 1849	0,030	0,030	0,031
<i>Chicoreus ramosus</i> (Linné 1758)	0,030	0,026	0,040
<i>Vasum ceramicum</i> (Linné 1758)	0,006	0,008	0,045
<i>Nerita polita</i> Linné 1758	0,000	0,004	0,037
<i>Cypraea miliaris</i> Gmelin 1791	0,026	0,026	0,028

The correlation between the relative abundances a_i (ranks $i = 1$ to 16) are plotted in Figures 26 to 28 for the three couples ‘Koswari’-‘Vaan’, ‘Koswari’-‘Hare’ and ‘Vaan’-‘Hare’. While a very strong correlation is observed between communities at ‘Koswari’ and ‘Vaan’ ($r^2 = 0.972$), the correlations between the community at ‘Hare’ and communities at either ‘Koswari’ or ‘Vaan’ are clearly weaker ($r^2 = 0.626$ and $r^2 = 0.632$ respectively).



Figures 26, 27, 28 – Correlations between the relative abundances a_i for 'Koswari'-'Vaani', for 'Koswari'-'Hare' and for 'Vaani'-'Hare'. The sixteen most abundant species are considered.

4. Discussion

Coral reefs in shallow waters are among the most threatened ecosystems, as a consequence of anthropic activities, in particular the on-going climate change [1, 2]. Assessing, at the local scale, the main quantitative descriptors of the animal communities associated to coral-reefs (in particular, the total species richness, the apparent level of unevenness of species abundances and the genuine intensity of the structuring process within communities) is thus a first urgency. Yet, the urgency on the one hand and the completeness of assessments on the other hand are two contradictory requirements, difficult to conciliate in practice. Hence, the usual constraint of having to cope with substantially *incomplete* samplings and, accordingly, to apply least-biased *numerical extrapolations* to these partial samplings. Implementing such a procedure allows, in turn, to infer all necessary quantitative information relative to the set of *unrecorded* species: (i) estimating their *number* and (ii) inferring their *abundance distribution*. Thereby, *total* species richness and the *full-range* Species Abundance Distribution ("S.A.D."), *including the undetected species*, may both be derived, thus providing essential information that, otherwise, would have remain out of reach. The procedures to carry out these extrapolations, with *minimum bias*, have been recently made available [18, 33, 34], so that now, nothing prevents from making the most of "rapid samplings" or "quick assessments".

4.1 Common features to all three gastropod communities

4.1.1 Each community considered as a whole (independently of feeding guilds)

4.1.1.1 Total species richness and abundance distribution

The three studied gastropod communities, associated to coral reefs surrounding the small islands of 'Koswari', 'Vaani' and 'Hare' show similar total species richness (53, 51 and 49 species respectively, Table 1) and also share rather similar "S.A.D.s" features (Figures 1, 2, 3). This, indeed, was expected from the similarity of environments between these three locations [15].

Moreover, the expected additional sampling efforts that would be required to progressively unveil the *taxonomical identities* of the still unrecorded species are provided by the numerically extrapolated *Species Accumulation Curves* of each community, as already reported in [16].

The average total species richness ($\approx$ fifty) found for these *local* communities of marine shelled-Gastropods may be put in the *regional* context of the same fauna for Mannar Gulf considered as a whole. The figures for marine Gastropods in Mannar Gulf are comprised between 260 species [35] and 450 species [15]. That is, between 240 and 410 shelled-species, after setting aside the non-shelled species (which are less than 10%, according to [36]). Thus, the proportion of this regional richness actually present at the local scale of the studied coral reefs is likely comprised between 50/410 and 50/240, i.e. between 12% and 21% of the regional stock. This is rather in line with the range of values reported in [37] for the ratios of species richness between the local and the corresponding regional scale, in tropical benthic ecosystems.

4.1.1.2 Inferred origin of the hierarchical structuration of species abundances

For all three communities, the abundance distributions ("S.A.D.s") reveal clearly more consistent with the "log-normal" model than with the "log-series" model (Figures 4-5, 8-9, 12-13). This, in fact, seems to be a rather frequent trend, likely being common to most wild communities of species approaching or having reached some kind of dynamic equilibrium [22, 24, 25]. Note that the affiliation to "log-normal" rather than to "log-series" models would have been less obvious if only non-extrapolated "S.A.D.s" were available (see Figures 4, 8, 12). This is in line with the opinion according to which incomplete distributions, lacking the set of rarest, unrecorded species may often lead – erroneously – to a better fitting to the "log-series" model [24]. Thus, relevantly addressing the kind of process driving the hierarchical structuration of species abundances in communities imperatively requires (here also and once again) considering the *full range* of "S.A.D.s" and, thus, implementing *numerical extrapolation* as far as necessary.

This consistency of all three "S.A.D.s" with the "log-normal" model suggests a process of hierarchical structuration of species abundances likely involving the *combined contributions of many independent factors*, rather than the sole contribution only one (or very few) dominant factor [10, 22 – 25].

Now, a more detailed examination (see Figures 6-7, 10-11, 14-15) shows, however, that in all three communities, the most abundant species (species of rank $i = 1$ and, to a much lesser extent, of rank $i = 2$) depart significantly from the "log-normal" model, with an abundance a_1 of the most abundant species being only $\approx$ half the expected value according to the "log-normal" model for $i = 1$. This gap between observation and expectation – statistically highly significant ($p < 0.001$) – may result from a variety of possible causes [38, see also 22]:

- from a "*saturation*" effect: while, according strictly to the "log-normal" model, the most abundant species is intended to profit by the freely cumulative contributions of all positively influencing factors, this cumulative process may, yet, tend to saturate by itself;

- from the communities still not being sufficiently close to their potential dynamic equilibrium that would permit to the most competitive species to reach its asymptotic level of abundance [38];

- from some *positively density-dependent detrimental factor*, becoming predominant when species relative abundance approaches 0.10; this detrimental factor may, in turn, result either:

- from *intra-specific competition* for resource, [39] or

580 - from *inter-specific* interactions, such as predation (which, often, is positively
581 dependent on prey density).

582 Whatever may be the nature of the cause involved, the latter is likely being non-taxon
583 specific since the identity of the most abundant species differs in each of the three
584 studied community: *Trochus niloticus* at 'Koswari', *Trochus radiatus* at 'Vaan', *Murex*
585 *trapa* at 'Hare'.

586 At last, it should be noted, by contrast, that such a (punctual) departure from the "log-
587 normal" model is not reported for the intertidal gastropod fauna along rocky shore at
588 Middle Andaman Island, which strictly follows the "log-normal" model, even at the level
589 of the most abundant species [19].
590

591 4.1.1.3 The intensity of the hierarchical structuration of species abundances

592 The hierarchical structuration of species abundances – in short, graphically, the slope of
593 the "S.A.D." – may be considered two ways: (i) either purely descriptively as a *pattern* or
594 (ii) more deeply analyzed, by considering the underlying *process* driving the differential
595 allocations of abundances among co-occurring species. Figures 1 to 3 and Figures 16 to
596 21 respectively address these two focus. As regards the *pattern*, the average steepness
597 of the "S.A.D." slope (directly readable from Figures 1 to 3 and quantified by unevenness
598 level $U (= \log(a_1/a_{St}) / (S_t - 1)$, in Table 3) is relatively similar in all three communities.
599 Accordingly, the apparent degree of unevenness of species abundances is not distinctive
600 among the three communities.

601 Yet, as argued in Methods section, the degree of unevenness of species abundances in a
602 community does not reliably reflect the intensity of the hierarchical structuring of
603 species abundances, because unevenness is also mathematically sensitive to the total
604 species richness of the community.

605 However, a relevant appreciation of the intensity of the hierarchical structuring process,
606 freed from the (spurious) influence of species richness, was obtained:

607 - *graphically*, by plotting simultaneously the Species Abundance Distribution under
608 study and the corresponding "broken-stick" model, computed for the same level of
609 species richness (Figures 16 to 21),

610 - *quantitatively*, by standardizing the average slope of the Species Abundance
611 Distribution under study to the average slope of the corresponding "broken-stick"
612 model, thus leading to the definition (equation (1)) of the genuine intensity I_{str} of the
613 process driving the hierarchical structuring of species abundances (Table 3). Thereby
614 leaving aside the purely mathematical influence of species richness on unevenness, the
615 index I_{str} reliably reflects the specifically *biological* contribution made to the apparent
616 level of unevenness.

617 Moreover, this mathematical independence between I_{str} and species richness confers a
618 *true biological significance* to any empirical correlation that would happen to be
619 recognized between I_{str} and species richness. Hence, the interest of plotting I_{str} against
620 S_t , as proposed in Figure 24. In addition, this representation has the advantage of
621 providing a synthetic overview of the main results derived from this new kind of
622 methodological approach.

623 The intensity I_{str} of the structuring process slightly differs according to communities,
624 ranging from 1.21 to 1.26, (Table 3). Interestingly, here, I_{str} increases with the total
625 species richness S_t of communities. And, as just mentioned, this empirically assessed

dependence between I_{str} and S_t should be considered as biologically relevant. This, indeed, is an original and important finding, contrasting with preceding reported results obtained for a series of tropical frog communities in N-W Ghats, where, on the contrary, I_{str} consistently remains independent from S_t [30].

4.1.2 Feeding guilds considered separately within each community

Feeding guilds constitute major functional sub-units within species communities. As such, they deserve separate consideration, as regards both their own species richness and their own distribution of species abundances. In the studied gastropod communities, two feeding guilds are co-occurring: *primary* consumers (“herbivores”) and *secondary* consumers (“carnivores”). The “S.A.D.s” specific to each feeding guild were therefore assessed separately and the corresponding numerical extrapolation were implemented for each specific “S.A.D.” as well: Figure 22.

4.1.2.1 Contrasted contributions of primary and secondary consumers to species richness and to the numbers of individuals

The respective contributions of *primary* and *secondary* consumers regarding the number of species and the number of individuals are given in Table 2. Relatively similar figures were obtained for the three studied communities, with a markedly reduced contribution of *primary* consumers (“herbivores”) as compared to *secondary* consumers (“carnivores”), for both the species richness and the cumulated numbers of individuals. *Primary* consumers contribute to an overall 15% - 20% of species richness and 28% - 32% of the cumulated number of individuals for each three communities. Accordingly, in these communities associated to coral reefs, *secondary* consumers make the bulk of the gastropod fauna in terms of both species richness and number of individuals. Interestingly, this is in relative contrast with the corresponding figures obtained with the intertidal gastropod fauna sampled along rocky shore at Middle Andaman Island [19], where *primary* consumers, contribute for no less than 39% of total species richness, thus approaching the contribution of *secondary* consumers. The contrast is still stronger, regarding the cumulated number of individuals among *primary* consumers, which reaches 60% at Andaman rocky shore and, thus, exceeds the figure for *secondary* consumers. This should arguably be related to the difference in algal cover between coral reefs and rocky shores. As a rule, algal cover is significantly lower on healthy coral reefs [40 – 43] than on degraded reefs [1] or rocky shores [44]. Thus, healthy coral reefs – such as those studied here – are expected to be less appealing to *primary* consumers than are the degraded coral reefs or the rocky shores, all of them supporting higher algal covers. A similar conclusion was drawn from the observation of healthy coral reefs along Kenyan coast, where most reef-associated Gastropods turn out to be *secondary* consumers: the proportion of predatory gastropods is comprised between 70% and 99% (average 86%), depending on reefs types and location [45].

4.1.2.2 Contrasted degree of hierarchical structuration of species abundances between primary and secondary consumers

Both the apparent *unevenness* U of species abundances (the average slopes of “S.A.D.s” at Figure 22) and the intensity I_{str} of the underlying structuring *process* (the slopes of “S.A.D.s” compared with those of the “broken-stick” distribution at Figure 23) are markedly stronger for *primary* consumers than for *secondary* consumers: $U = 0.300$ and

$I_{str} = 1.92$ for *primary* consumers, against $U = 0.068$ and $I_{str} = 1.27$ only for *secondary* consumers. The still more striking gap regarding apparent unevenness U (a 4.4 times higher unevenness for *primary* consumers) results, as expected, from the additional influence on unevenness of the smaller species richness of *primary* consumers.

Interestingly, the same tendency is highlighted for the intertidal gastropod fauna of rocky shore in Andaman Islands [19] and for the gastropod fauna associated to coral-reefs in Fiji Islands [31]. In fact, the trend for higher unevenness in the guild of *primary* consumers might well be general for most marine and terrestrial ecosystems, as argued on both theoretical and empirical basis in [46].

Moreover, the much stronger value of the intensity I_{str} of the structuring process for *primary* consumers suggests that the trend for higher unevenness of abundances among *primary* consumers does not result only from the influence of their reduced species richness. In fact, there is also a significant contribution of some *purely biological cause(s)* to the much stronger unevenness of species abundances within the guild of *primary* consumers. Indeed, this was a question which was still remaining unanswered after the publication of reference [46], as the author does not consider the genuine intensity of the structuring process, nor even the apparent evenness pattern, but the “Hill number”, which, in fact, combines apparent evenness and species richness and thus quantifies the diversity rather than evenness itself. The present results thus shed significant new light on this issue.

4.2 Minor features singling out the community at ‘Hare’ as compared to ‘Koswari’ and ‘Vaan’

Until now, the three gastropods communities were hardly found differing from each other, which is not surprising, considering their very similar environmental conditions. Yet, addressing a series of more subtle factors allows to slightly singularize the community at ‘Hare’ from the other two communities, at ‘Koswari’ and ‘Vaan’.

First, as shown in Table 4, the species belonging to the guild of *primary* consumers appear fairly equally distributed all along the range of species abundances, as regards the communities at ‘Vaan’ and ‘Koswari’. By contrast, at ‘Hare’, the *primary* consumer species are more than twice more frequent in the upper half than in the lower half of the range of species abundances.

Second, as shown in Figure 25, the degree of hierarchical dominance of the three most abundant species differs according to sites; with a slightly, but yet significantly, lower dominance value at ‘Hare’ as compared to both ‘Vaan’ and ‘Koswari’.

Third, as shown in Figures 26 to 28, there is a trend for higher congruence in the relative abundances, considered species by species, when comparing ‘Vaan’ and ‘Koswari’ than when comparing ‘Hare’ to either ‘Vaan’ or ‘Koswari’.

Thus, for each of the three criteria above, it is always the community at ‘Hare’ which singularizes from the other two communities, located at ‘Vaan’ and ‘Koswari’. This is suggestive, as ‘Vaan’ and ‘Koswari’ are only a few kilometers apart while ‘Hare’ is distant from about 150 kilometers from the other two localities. The results above thus anecdotally highlight the (limited) extent to which distance may contribute to slightly affect the features of communities which, on the other hand, are placed in likely very similar environments and profit by the same regional stock of species. A point, however discussed in more details by [47].

Besides, it is worth noting that Mohanraj and co-workers [15], although relying only on the partial samplings of the gastropod assemblages of each three islands, had already concluded also for a slightly greater taxonomical proximity between the assemblages of “Vaan” and “Koswari”, as compared to “Hare”. A pattern that they had already hypothesized as the consequence of the closer spatial proximity of “Vaan” and “Koswari”, as compared to “Hare”.

4.3 Numerical extrapolation: a useful complement to partial species inventories

Considering in details the data implicitly conveyed by Species Abundance Distributions can provide an astonishing diversity of information relative to the internal structuration of species communities, having both theoretical and practical interest. Yet, taking such full advantage from the treatment of Species Abundance Distributions requires dealing with the *exhaustive range* of the distributions of abundances, which in practice is rarely possible, especially when having to cope with species-rich communities, as are, for example, most invertebrate assemblages. Hence the necessity of implementing a reliable procedure of *numerical extrapolation* to extend the still incomplete Species Abundance Distributions [48], as was proceeded here.

This is all the more important that rare species, that often escape recording in practice, may yet disproportionately contribute to the functional structuring of communities in the wild [49]. Accordingly, numerical extrapolation remains therefore the *only way to account for their existence and role*. Moreover – as frustrating as it may seem – the lack of taxonomic identification of these rare species (remaining undetected in the field but highlighted by numerical extrapolation) has more limited consequences that could have been thought at first. Indeed, the specific taxonomic identification turns out to be of less importance, as soon as one is aware of the generally great variability of taxonomic compositions that can equally well answer a same type of functionality [50].

4.4 Final consideration regarding the methodological approach to the internal organization within species communities

For a local assemblage of species to be relevantly regarded – and fruitfully understood – as a *community* of species, it is necessary to go beyond the purely descriptive pattern of community organization, trying to address the meaningful underlying process driving this organization [38]. In other words, it is necessary to go beyond the empirical level of *unevenness* (as the apparent pattern of community organization), aiming at characterizing the genuine *intensity of the underlying structuring process* driving the differential allocation of relative abundances among co-occurring species. This distinction between the immediately apparent pattern and the underlying process makes, here, specific sense and, even, has essential implications, because it happens that – for fundamental reason – the pattern of the as-observed unevenness cannot faithfully mirror the intensity of the structuring process itself. This is because the level of unevenness is also influenced, purely mathematically [26, 27], by the species richness of the community. An important point indeed, already argued previously [12, 13, 26, 27, 28] but, yet, still remaining too frequently ignored in common practice. This systematic – and biologically irrelevant – influence, should thus be *set aside* if the true biological meaning of species abundance unevenness is to be relevantly understood.

The (“biologically spurious”) mathematical influence of species richness on abundance unevenness may be evidenced quantitatively by considering how unevenness would vary with species richness while the nature of the structuring process is constrained to remain unchanged. The “broken-stick” distribution [29] provides such an opportunity, as it involves an unchanged process of random allocation of relative abundances among species, whatever species richness can be. So that, the “broken-stick” distribution is uniquely dependent upon (and solely parametrized by) the species richness only [10]. The “broken-stick” distribution may thus serve as a relevant “null model” against which to compare empirical Species Abundance Distributions.

This has been fully taken into account, here, as a *major methodological improvement* aiming at a better understanding of the drivers of the internal structuring within species communities.

5. Conclusions

After being numerically completed, using a recently developed least-biased extrapolation procedure, the Species Abundance Distributions of the three reef-associated gastropod communities reveal remarkably similar to each other, although no specific connection exists between these three communities, one of them being even located 150 km apart from the other two. In particular, the Species Abundance Distributions of the three communities all comply fairly well with the *log-normal* model, suggesting that, in each case, the process of hierarchical structuration of species abundances likely involves the combined contributions of many independent factors. The same trend to similarity holds true for all other major characteristic of each of the three Species Abundance Distributions.

Only a fine, in-depth analysis proves able to highlight subtle distinctions between the three communities and, interestingly, precisely for the site (Hare) located 150 km apart from the others.

In addition to the information specific to the studied Gastropod communities associated to coral-reefs in Mannar Gulf, some new results, of more general interest, were highlighted, in particular regarding the two co-occurring feeding guilds.

Thus, as compared to *secondary* consumers (“carnivores”), the guild of *primary* consumers (“herbivores”) is clearly distinguished by a comparatively lower species richness and a strongly higher degree of abundance unevenness (as already emphasized recently. But, moreover, the intensity of the process of hierarchical structuring of species abundance *itself* was also demonstrated as being markedly stronger within the guild of *primary* consumers. To my knowledge, this last point is reported here for the first time.

Given the growing threat posed by global warming to the integrity of coral reefs [1, 2, 51], it features essential to multiply detailed and *exhaustive* inventories – such as those conducted here with the support of numerical extrapolation – in order to accumulate sufficient contemporary data to serve as reference for subsequent monitoring of the future evolution of coral reefs ecosystems.

Appendix 1

Formatted: Font color: Red

Comment [AP1]: Delete these refrees intoe conclusion

Formatted: Font color: Red

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 [18].

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:

$$\tilde{a}_i = p_i \cdot (1 + 1/n_i) / (1 + R_0/N_0) \cdot (1 - f_1/N_0) \quad (A1.1)$$

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 (A1.1).

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 $i > R_0$):

$$a_i = (2/N_i) / (1 + R(N_i)/N_i) \cdot (1 - [\partial R(N)/\partial N]_{N_i}) \quad (A1.2)$$

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 2.

Appendix 2

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

Consider the survey of an assemblage of species of size N_0 (with sampling effort N_0 typically identified either to the number of recorded individuals or to the number of sampled sites, according to the inventory being in terms of either species abundances or species incidences), including $R(N_0)$ species among which f_1, f_2, f_3, f_4, f_5 , of them are recorded 1, 2, 3, 4, 5 times respectively. The following procedure, designed to select the less-biased solution, results from a general mathematical relationship that constrains the theoretical expression of *any* theoretical Species Accumulation Curves $R(N)$ [33, 52, 53]:

$$\partial^x R(N) / \partial N^x = (-1)^{(x-1)} f_{x(N)} / C_{N,x} \approx (-1)^{(x-1)} (x! / N^x) f_{x(N)} \quad (\approx \text{as } N \gg x) \quad (A2.1)$$

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 (A2.1), 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 [33]:

856 * for f_1 up to $f_2 \rightarrow R_1(N) = (R(N_0) + f_1) - f_1.N_0/N$
857
858 * for larger f_1 up to $2f_2 - f_3 \rightarrow R_2(N) = (R(N_0) + 2f_1 - f_2) - (3f_1 - 2f_2).N_0/N -$
859 $(f_2 - f_1).N_0^2/N^2$
860
861 * for larger f_1 up to $3f_2 - 3f_3 + f_4 \rightarrow R_3(N) = (R(N_0) + 3f_1 - 3f_2 + f_3) - (6f_1 - 8f_2 +$
862 $3f_3).N_0/N -$
863 $(-4f_1 + 7f_2 - 3f_3).N_0^2/N^2 - (f_1 - 2f_2 + f_3).N_0^3/N^3$
864
865 * for larger f_1 up to $4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow R_4(N) = (R(N_0) + 4f_1 - 6f_2 + 4f_3 - f_4) -$
866 $(10f_1 - 20f_2 + 15f_3 - 4f_4).N_0/N - (-10f_1 + 25f_2 - 21f_3 + 6f_4).N_0^2/N^2 -$
867 $(5f_1 - 14f_2 + 13f_3 - 4f_4).N_0^3/N^3 - (-f_1 + 3f_2 - 3f_3 + f_4).N_0^4/N^4$
868
869 * for f_1 larger than $4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow R_5(N) = (R(N_0) + 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5)$
870 $- (15f_1 - 40f_2 + 45f_3 - 24f_4 + 5f_5).N_0/N - (-20f_1 + 65f_2 - 81f_3 + 46f_4 - 10f_5).N_0^2/N^2 -$
871 $(15f_1 - 54f_2 + 73f_3 - 44f_4 + 10f_5).N_0^3/N^3 - (-6f_1 + 23f_2 - 33f_3 + 21f_4 - 5f_5).N_0^4/N^4 -$
872 $(f_1 - 4f_2 + 6f_3 - 4f_4 + f_5).N_0^5/N^5$
873

874 The associated non-parametric estimators of the number Δ_j of missing species in the
875 sample [with $\Delta_j = R(N=\infty) - R(N_0)$] are derived immediately:
876
877

878 * $f_1 \leq f_2 \rightarrow \Delta_{j1} = f_1 ; R_1(N)$
879
880 * $f_2 < f_1 \leq 2f_2 - f_3 \rightarrow \Delta_{j2} = 2f_1 - f_2 ; R_2(N)$
881
882 * $2f_2 - f_3 < f_1 \leq 3f_2 - 3f_3 + f_4 \rightarrow \Delta_{j3} = 3f_1 - 3f_2 + f_3 ; R_3(N)$
883
884 * $3f_2 - 3f_3 + f_4 < f_1 \leq 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{j4} = 4f_1 - 6f_2 + 4f_3 - f_4 ; R_4(N)$
885
886 * $f_1 > 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{j5} = 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5 ; R_5(N)$
887

888 **N.B. 1:** As indicated above (and demonstrated in details in [33], this series of
889 inequalities define the ranges that are best appropriate, respectively, to the use of each
890 of the five estimators, JK-1 to JK-5. That is the respective ranges within which each
891 estimator will benefit of minimal bias for the predicted number of missing species.
892 Besides, it is easy to verify that another consequence of these preferred ranges is that
893 the selected estimator will *always* provide the highest estimate, as compared to the
894 other estimators. Interestingly, this mathematical consequence, of general relevance, is
895 in line with the already admitted opinion that all non-parametric estimators provide
896 *under-estimates* of the true number of missing species [8, 9, 54 – 56]. Also, this shows
897 that the approach initially proposed in [57] – which has regrettably suffered from its
898 somewhat difficult implementation in practice – might be advantageously reconsidered,
899 now, in light of the very simple selection key above, of *far much easier practical use*.
900

901 **N.B. 2:** In order to reduce the influence of drawing stochasticity on the values of the f_x ,
902 the as-recorded distribution of the f_x should preferably be smoothened: this may be
903 obtained either by rarefaction processing or by regression of the as-recorded
904 distribution of the f_x versus x .
905

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 [33].

Appendix 3

The trivial (“mechanistic”) contribution of the level of species richness to the degree of structuring of species abundances

All things equal otherwise, the larger the species richness, the weaker is the slope of the Species Abundance Distribution.

This can be easily exemplified and quantified, on a theoretical basis, by considering a theoretically constant structuring process - such as the random distribution of the relative abundances that characterises the “broken-stick” distribution model. By applying this model successively to a series of communities with increasing species richness, a steadily decrease of the slope of abundance distributions is highlighted:

Figure A3

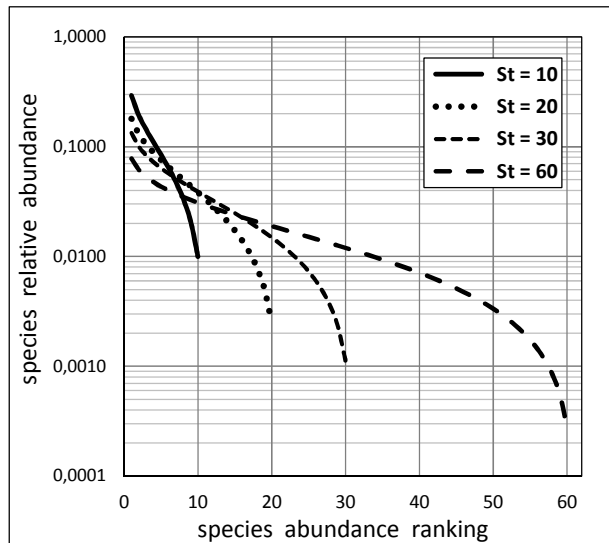


Figure A3 – The “broken-stick” distribution model applied to species communities with increasing species richness $S_t = 10, 20, 30, 60$. Although the theoretical structuring process involved in the “broken-stick” model remains unchanged (random apportionment of relative abundances among member species), the slope of the species abundance distribution strongly depends upon (and monotonously decreases with) the level of species richness S_t .

COMPETING INTERESTS

Author has declared that no competing interests exist.

ACKNOWLEDGEMENTS

Three anonymous reviewers are gratefully acknowledged for their comments and appreciations on the original manuscript

References

937
938 | 1.—Graham NAJ, Wilson SK, Jennings S, Polunin NVC, Bijoux JP & Robinson J.— Dynamic
939 fragility of oceanic coral reef ecosystems. Proceedings of the National Academy of
940 Sciences USA. 2006; 103(22): 8425-8429.
941
942 | 2.—Ziegler M, Quéré G, Ghiglione J-F, Iwankow G, Barbe V, Boissin E, Wincker P, Planes S
943 & Voolstra CR.— Status of coral reefs of Upolu (Independent State of Samoa) in the
944 South West Pacific and recommendations to promote resilience and recovery of coastal
945 ecosystems. Marine Pollution Bulletin. 2018; 129: 392-398.
946
947 | 3.—Wells JW.— Coral reefs: 609-632; in J.W. Hedspeth editor, Treatise on marine ecology
948 and paleoecology. Geological Society of America. 1957; Mem. 67.
949
950 | 4. —Glynn PW.— High complexity food webs in low-diversity eastern Pacific reef-coral
951 communities. Ecosystems. 2004; 7: 358-367.
952
953 | 5.—Jeeva C, Mohan PM, Sabith DB, Ubare VV, Muruganantham M & Kumari RK.—
954 Distribution of Gastropods in the intertidal environment of south, middle and north
955 Andaman Islands, India. Open Journal of Marine Science. 2018; 8: 173-195.
956
957 | 6.—Connolly SR, Hughes TP & Bellwood DR.— A unified model explains commonness and
958 rarity on coral reefs. Ecology Letters. 2017; 20: 477-486.
959
960 | 7. —Cam E, Nichols JD, Sauer JR & Hines JE.— On the estimation of species richness based
961 on the accumulation of previously unrecorded species. Ecography. 2002; 25: 102-108.
962
963 | 8. —Rajakaruna H, Drake DAR, Chan FT & Bailey SA.— Optimizing performance of
964 nonparametric species richness estimators under constrained sampling. Ecology and
965 Evolution. 2016; 6: 7311-7322.
966
967 | 9. —Chen Y & Shen TJ.— Rarefaction and extrapolation of species richness using an area-
968 based Fisher's logseries. Ecology and Evolution. 2017; 7: 10066-10078.
969
970 | 10.—May RM.— Patterns of species abundance and diversity. In Cody M.L. & Diamond
971 J.M. 1975 Ecology and Evolution of Communities. The Belknap Press of Harvard
972 University: 81-120.
973
974 | 11.— McGill BJ, Etienne RS, Gray JS et al. - Species abundance distributions: moving
975 beyond single prediction theories to integration within an ecological framework.
976 Ecology Letters. 2007; 10: 995-1015.
977
978 | 12.—Ulrich W, Ollik M & Ugland KI.— A meta-analysis of species-abundance distributions.
979 Oikos. 2010; 119:1149-1155.
980

Formatted: Font color: Red

981 | 13.—Komonen A & Elo M.— Ecological response hides behind the species abundance
982 | distribution: community response to low-intensity disturbance in managed grasslands.
983 | Ecology and Evolution. 2017; 7: 8558-8566.
984 |
985 | 14.—Wang X, Ellwood F, Ai D, Zhang R & Wang G.— Species abundance distributions as a
986 | proxy for the niche-neutrality continuum. Journal of Plant Ecology. 2017; rtx 013.
987 |
988 | 15. Mohanraj J, Johnson JA, Ranjan R, Johnson L, Pandi U & Shunmugaraj T.— Coral reef
989 | associated gastropods in Tuticorin coast of Gulf of Mannar Biosphere Reserve, India.
990 | Indian Journal of Science and Technology. 2010; 3(2): 204-206.
991 |
992 | 16.—Béguinot J.— Extrapolation of total species richness from incomplete inventories:
993 | application to the Gastropod fauna associated to coral reefs in 'Mannar Gulf Biosphere
994 | Reserve', India. Asian Journal of Environment and Ecology. 2017;— 4(3): 1-14. doi:
995 | 109734/AJEE/2017/36831.
996 |
997 | 17.—Chao A, Hsieh T, Chazdon RL, Colwell RK & Gotelli NJ.— Unveiling the species-rank
998 | abundance distribution by generalizing the Good-Turing sample coverage theory.
999 | Ecology. 2015; 96(5): 1189-1201.
1000 |
1001 | 18.—Béguinot J.— How to extrapolate species abundance distributions with minimum
1002 | bias when dealing with incomplete species inventories. Advances in Research. 2018;
1003 | 13(4): 1-24. doi: 10.9734/AIR/2018/39002.
1004 |
1005 | 19.—Béguinot J.— Numerical extrapolation of the species abundance distribution unveils
1006 | the true species richness and the hierarchical structuring of a partially sampled marine
1007 | gastropod community in the Andaman Islands (India). Asian Journal of Environment
1008 | and Ecology. 2018; 6(4): 1 – 23. doi: 10.9734/AJEE/2018/41293 <hal-01807454>
1009 |
1010 | 20.—Heip CHR, Herman PMJ & Soetaert K.— Indices of diversity and evenness. Océanis.
1011 | 1998; 24(4): 61-87.
1012 |
1013 | 21.—Grzès IM.— Ant species richness and evenness increase along a metal pollution
1014 | gradient in the Boleslaw zinc smelter area. Pedobiologia. 2009; 53: 65-73.
1015 |
1016 | 22. Loreau M.— Species abundance patterns and the structure of ground-beetle
1017 | communities. Ann. Zool. Fennici. 1992; 28: 49-56.
1018 |
1019 | 23.—Magurran AE & Henderson PA —Explaining the excess of rare species in natural
1020 | species abundance distributions. Nature. 2003; 422: 714-716.
1021 |
1022 | 24. Connolly SR, Hughes TP, Bellwood DR & Karlson RH - Community structure of
1023 | corals and reef fishes at multiple scales. Science. 2005; 309: 1363-1365.
1024 |

Formatted: Justified

- 1025 | 25.—Ulrich W, Soliveres S, Thomas AD, Dougill AJ & Maestre FT.— Environmental
1026 | correlates of species rank-abundance distributions in global drylands. Europe PMC
1027 | Funders Group. 2016; 20: 56-64.
1028 |
1029 | 26. Smith B & Wilson JB.—A consumer's guide to evenness indices. Oikos. 1996;76: 70-
1030 | 82.
1031 |
1032 | 27.—Loiseau N & Gaertner JC.— Indices for assessing coral reef fish biodiversity: the need
1033 | for a change in habits. Ecology and Evolution. 2015; 5(18): 4018-4027.
1034 |
1035 | 28.—MacDonald ZG, Nielsen SE & Acorn JH.—Negative relationships between species
1036 | richness and evenness render common diversity indices inadequate for assessing long-
1037 | term trends in butterfly diversity. Biodiversity Conservation. 2017; 26: 617-629.
1038 |
1039 | 29.—MacArthur RH.—On the relative abundance of bird species. Proceedings of the
1040 | National Academy of Sciences U.S.A. 1957; 43: 293-295.
1041 |
1042 | 30.—Béguinot J.—Estimating true species richness and the degree of hierarchical
1043 | structuring of species abundances in eight frog communities from the north-western
1044 | Ghats of India. —International Journal of Environment and Climate Change. 2018; 8(2):
1045 | 118-137. doi: 10.9734/IJECC/2018/42067
1046 |
1047 | 31. Béguinot J - Analyzing the role of environmental stresses on both species richness
1048 | and the hierarchical structuring of species abundances in marine Gastropods
1049 | communities at Suva (Fiji Islands). in preparation in press or in preparation???: if in
1050 | preparation should not be cited
1051 |
1052 | 32.—Mohanraj J, Chelladurai G, Balakrishnan S & Kuamr IV --Gulf of Mannar island coral
1053 | reef associated gastropods assemblage: distribution and diversity pattern. Journal of
1054 | Coastal life Medicine. 2015; 3(9): 691-695.
1055 |
1056 | 33.—Béguinot J - Theoretical derivation of a bias-reduced expression for the
1057 | extrapolation of the Species Accumulation Curve and the associated estimation of total
1058 | species richness. Advances in Research. 2016; 7(3): 1-16. doi:
1059 | 10.9734/AIR/2016/26387 ; <hal-01367803>
1060 |
1061 | 34.—Béguinot J - Extrapolation of the Species Accumulation Curve associated to “Chao”
1062 | estimator of the number of unrecorded species: a mathematically consistent derivation.
1063 | Annual Research & Review in Biology. 2016; 11 (4): 1-19 doi:
1064 | 10.9734/ARRB/2016/30522; <hal 01477263 >
1065 |
1066 | 35.—Mohanraj T, Jagadis I & Sivanesh H - Association of molluscan fauna with the coral
1067 | reefs of Gulf of Mannar. European Journal of Experimental Biology. 2011; 1(3): 44-48.
1068 |
1069 | 36. Wägle H & Klusmann-Kolb A --Opisthobranchia (Mollusca, Gastropoda) – more
1070 | than just slimy slugs. Shell reduction and its implications on defense and foraging.

Formatted: Font color: Red

Formatted: Font color: Red

1071 | ~~Frontiers in Zoology. 2005; 2(3), doi:10.1186/1742-9994-2-3~~~~doi:10.1186/1742-9994-~~
1072 | ~~2-3~~
1073 |
1074 | 37.—Witman JD, Etter RJ & Smith F - The relationship between regional and local species
1075 | diversity in marine benthic communities: a global perspective. Proceedings of the
1076 | National Academy of Sciences USA. 2004;—101(44): 15664-15669
1077 |
1078 | ~~38.—Connell JH —Some mechanisms producing structure in natural communities: a~~
1079 | ~~model and evidence from field experiments. in Cody M.L. & Diamond J.M. 1975 Ecology~~
1080 | ~~and Evolution of Communities. 1975; The Belknap Press of Harvard University: 1975;~~
1081 | ~~pp460-490.~~
1082 |
1083 | 39.—Skilleter GA & Underwood AJ - Intra- and inter-specific competition for food in
1084 | infaunal coral reef gastropods. Journal of Experimental Marine Biology and Ecology.
1085 | 1993; 173(1): 29-55.
1086 |
1087 | 40.—Villanueva RD, Edwards AJ & Bell JD - Enhancement of grazing Gastropod
1088 | populations as a coral reef restoration tool: predator effects and related applied
1089 | implications. Restoration Ecology. 2010; 18(6): 803-809.
1090 |
1091 | 41.—Rasher DB, Hoey AS & Hay ME - Consumer diversity interacts with prey defenses to
1092 | drive ecosystem function. Ecology. 2013; 94(6): 1347-1358.
1093 |
1094 | 42.—Mailloux NA - Increasing coral cover and decreasing coral species diversity with
1095 | decreasing light availability ta depth: a study of the future effects of sea level rise on
1096 | fringing reef ecosystems. Physis. 2014; 16: 48-54.
1097 |
1098 | 43.—Rempel H - Control of algae on coral reefs by large herbivorous fishes. Physis. 2014;
1099 | 16: 55-65.
1100 |
1101 | 44.—Martinez E, Sanchez J, Alba A & Vazquez AA - Changes in structure and composition
1102 | of two communities of rocky shores molluscs exposed to different human impact in
1103 | Playa Jibacoa. Journal of Ecosystem and Ecography. 2015; 5(2): 1-6.
1104 |
1105 | 45.—McClanahan TR— Kenyan coral reef-associated gastropod fauna: a comparison
1106 | between protected and unprotected reefs. Marine Ecology Progress Series. 1989; 53:
1107 | 11-20.
1108 |
1109 | 46.—Garcia-Callejas D— On the variability of Species Abundance Distributions with
1110 | trophic guild and community structure. bioRxiv. 2018; volume and Pages?? doi:
1111 | 10.1101/289348
1112 |
1113 | ~~47. -Diamond JM— Assembly of species communities. In Cody M.L. & Diamond J.M. 1975~~
1114 | ~~Ecology and Evolution of Communities. The Belknap Press of Harvard University. 1975;~~
1115 | ~~pp: 81-120.~~
1116 |

Formatted: Pattern: Clear

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red

- 1117 48. Kery M & Royle JA – Inference about species richness and community structure
 1118 using species-specific occupancy models in the National Swiss Breeding Bird survey
 1119 MUB. Proceedings of the 2007 EURING Technical Meeting and Workshop, Dunedin, New
 1120 Zealand. 2007.
- 1121
- 1122 49.—Leitao RP, Zuanon J, Villéger S, Williams SE, Baraloto C, Fortunel C, Mendonça FP &
 1123 Mouillot D – Rare species contribute disproportionately to the functional structure of
 1124 species assemblages. Proceedings of The Royal Society B. 2016; 283: 2016 0084 ; doi:
 1125 10.1098/rspb.2016.0084
- 1126
- 1127 50.—Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DL, Wegener Parfrey L, Farjalla
 1128 VF & Doebeli M – High taxonomic variability despite stable functional structure across
 1129 microbial communities. Nature Ecology and Evolution. 2016; 1: 0015 ; doi:
 1130 10.1038/s41559-016-0015
- 1131
- 1132 51.—Bellwood DR, Hughes TP, Folke C & Nyström M - Confronting the coral crisis.
 1133 Nature. 2004; 429: 827-830.
- 1134
- 1135 52.—Béguinot J - An algebraic derivation of Chao's estimator of the number of species in
 1136 a community highlights the condition allowing Chao to deliver centered estimates. ISRN
 1137 Ecology. 2014; vol. 2014 :—article ID 847328,— doi:10.1155/2014/847328 ; <hal-
 1138 01101415>
- 1139
- 1140 53.—Béguinot J - When reasonably stop sampling? How to estimate the gain in newly
 1141 recorded species according to the degree of supplementary sampling effort. Annual
 1142 Research & Review in Biology. 2015; 7(5): 300-308. doi : 10.9734/ARRB/2015/18809 ;
 1143 <hal-01228695>
- 1144
- 1145 54.—O'Hara RB - Species richness estimators: how many species can dance on the head
 1146 of a pin? Journal of Animal Ecology. 2005; 74: 375-386.
- 1147
- 1148 55.—Gotelli NJ & Colwell RK —Estimating species richness. pp. 39-54 in: Biological
 1149 Diversity: Frontiers In Measurement And Assessment. A.E. Magurran and B.J. McGill
 1150 (eds.). ~~2010;~~ Oxford University Press, Oxford. 2010; 345 p.
- 1151
- 1152 56.—Gotelli NJ & Chao A —Measuring and Estimating Species Richness, Species Diversity,
 1153 and Biotic Similarity from Sampling Data. In: Levin S.A. (ed.) Encyclopedia of
 1154 Biodiversity. Waltham, MA: Academic Press; second edition. 2013;—second edition,
 1155 volume 5; pp. 195-211. Waltham, MA: Academic Press.
- 1156
- 1157 57.—Brose U, Martinez ND & Williams RJ —Estimating species richness: sensitivity to
 1158 sample coverage and insensitivity to spatial patterns. Ecology. 2003; 84(9): 2364-2377.

Formatted: Font color: Red

Formatted: Font color: Red

Formatted: Font color: Red