

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)

Abstract

Even when it remains substantially incomplete, the partial inventory of a species assemblage can provide much more information than could be expected at first. This can be achieved by applying a rigorous numerical extrapolation procedure that fully extends the incomplete sampling and, thereby, provides reliable estimates regarding not only the number but also the distribution of abundances for all undetected species. As a result, this makes available the full range of the Species Abundance Distribution of the yet partially sampled assemblage and, thus, allows to address a series of interesting issues regarding the process and pattern of the hierarchical structuring of species abundances within the studied assemblage. Moreover, the same kind of numerical extrapolation may be applied separately to each of the subsets, within the whole assemblage, that may have the relevant interest (taxonomic subgroups, feeding guilds, etc...). Thus deconstructing the Species Abundance Distribution can provide further detailed insights into the functional organisation of the studied assemblage.

The mathematical and algorithmic basis for such an extrapolating procedure has been developed recently to be applied to the numerical extension of both the Species Accumulation Curve and the Species Abundance Distribution.

The wide potential interest of this methodological approach, when having to deal with substantially incomplete inventories of species (which is doomed to become inevitable with increasingly species-rich assemblages), is illustrated by a detailed case study of a marine gastropod assemblage on rocky shore under tropical climate.

Keywords: species richness, diversity, rank-abundance, marine snails, trophic rank, incomplete sampling

Introduction

Total species richness, taxonomic composition and hierarchical structuring of species abundances are three key features that appropriately characterise species communities. Unfortunately, partial, incomplete inventories—which are doomed to become even more frequent with the inevitable generalization of “rapid assessments” and “quick surveys” – prevent an in-depth appreciation of each of these key aspects of species communities [1

– 3]. However, a properly implemented procedure of *numerical extrapolation* can provide reliable estimations relative to both the number and the respective abundances of undetected species and, thereby, allows the derivation of reliable inferences as regard (i) the *total* species richness and (ii) the distribution of species abundances *completed* by including the set of still *undetected* species. Only the taxonomic identities of the latter escape, of course, to any attempt of extrapolation. In turn, once numerically *completed* (and *only* when it is so: [4]), the distribution of species abundances can provide some synthetic pieces of information about the process at work (either deterministic or stochastic) that drive the hierarchical structuring of species abundances within the community [5–9]. Although no further mechanistic details may be extracted from this synthetic overview, it has, yet, the advantage of being straightforward, as it does not require the long and tedious analytical approaches that would be required otherwise to go deeper in the details of structuring processes. As such, this synthetic approach can serve as a convenient preliminary approach.

As complete (or completed by numerical extrapolation) abundance distributions are mandatory, a procedure for the numerical extrapolation of the Species Accumulation Curve and the Species Abundance Distribution (the former directly linked to the latter) has recently been developed, aiming at providing reliable, *least-biased* inferences about the number and the respective abundances of undetected species, when having to deal with substantially incomplete inventories.

Tropical marine ecosystems in shallow waters are of major interest to ecologists and conservationists, as they are considered as embodying remarkably high levels of biological complexity among marine communities [10 – 12]. However, and precisely because of their usually high species richness and diversity (i. e. including *numerous rare* species), samplings of these communities often remain substantially incomplete [4]. For all the reasons just mentioned above, such partial inventories thus require implementing reliable numerical extrapolation procedure, so as to release as much information as possible regarding (i) the true species richness of sampled community, (ii) a synthetic overview of the hierarchical structuring of species abundances and (iii) how specific are the respective functional contributions of the different trophic levels involved in the community.

Hereafter, I report and discuss the results from a numerical extrapolation of a partial inventory of an intertidal gastropod community established on a rocky shore located near Rangat, Andaman Islands, India: see JEEVA *et al.* [12].

Materials and Methods

Materials

The rocky shore of Andaman and Nicobar Islands is home to a rich fauna of marine gastropods [13]. Yet, detailed inventories of gastropod communities at the local scale, including species abundances, remain very scarce. A recent report by JEEVA *et al.* [12] opportunely provides such a series of local inventories along the rocky shores of Andaman Islands. But, referring to the high proportion of singletons (species that were detected only once during sampling), these inventories remain substantially incomplete

1 [14 – 16]. Accordingly, implementing *numerical extrapolation* of such inventories is
2 required to uncover the series of useful information that may be derived from complete
3 (here *completed by extrapolation*) Species Abundance Distributions, as mentioned
4 above.

5 Hereafter, I shall focus on the particular community that provides the highest number of
6 recorded species; this community is located near Rangat, in the southern part of Middle
7 Andaman. According to JEEVA *et al.* [12], the sampled area near Rangat is mainly covered
8 by rocky outcrops with boulders and pebbles and the partial inventory was conducted
9 by these authors at the intertidal level.

10 11 ***Numerical extrapolation procedures***

12 * *Total species richness*: the least-biased estimation of the number of still undetected
13 species during partial sampling and the resulting least-biased estimation of the total
14 species richness of the sampled community are derived according to the procedure
15 defined by BÉGUINOT [17, 18] and briefly summarised in Appendix 1, on the basis of the
16 numbers f_x of species observed x -times during partial sampling, as provided in
17 Appendix 3)

18 * *Species Abundance Distribution*: to accurately exploit their full potential, the as-
19 recorded Species Abundance Distributions (“S.A.D.s”) require [19, 20]:

20 - first, to be *corrected* for statistical sampling bias, resulting from the finite size of
21 samplings and,

22 - second, but still more importantly, to be *completed* by numerical extrapolation to
23 the extent that sampling is suspected to be incomplete, as revealed by the subsistence of
24 singletons.

25 The appropriate procedure of *least-biased* numerical extrapolation and correction of as-
26 recorded S.A.D.s, described in details in BÉGUINOT [20] is briefly recalled in Appendix 2

27 After being *corrected* and *extrapolated* accordingly, the S.A.D.

28 - not only provides an overview of both the true species richness of the sampled
29 community and the diversity of the respective abundances of member species

30 - but also, can help addressing several important questions regarding the kind of
31 process driving the hierarchical structuration of the community, including possible
32 specificities according to the different trophic levels involved in the community of
33 interest (Figure 1).

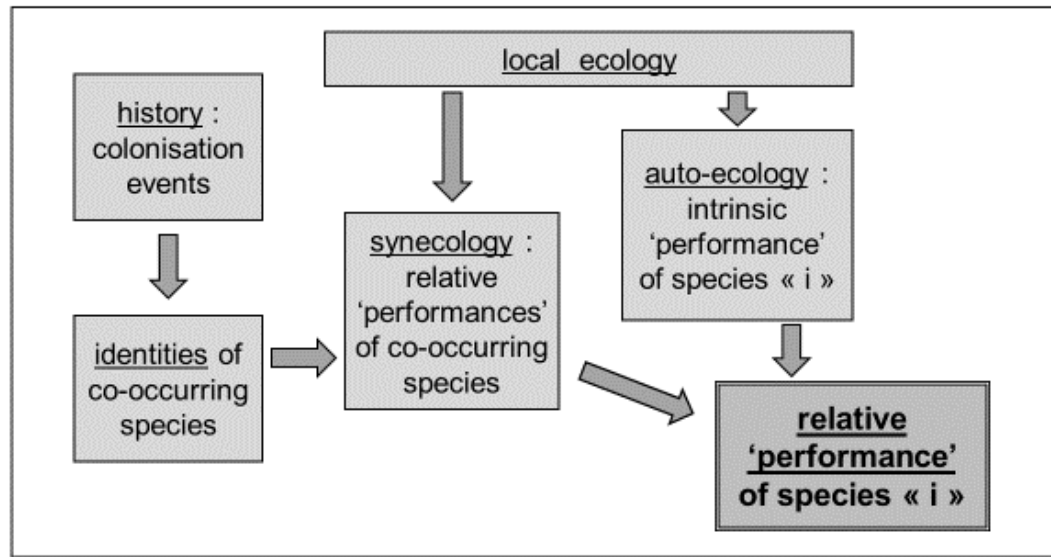


Figure 1 – Schematic sketch showing how the combination of both historical and ecological contexts peculiar to a given community of species drive the relative “performance” - *sensu latissimo* - of each member species “i”, thus generating the *hierarchical structuring of species abundances in the community*.

More precisely, these questions may relate to:

- the *process of structuration* of a community of species: for example, does only one (or very few) dominant factor is (are) at work to structure the community or, on the contrary, do many independent factors are contributing together; which 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 [5, 21 – 24] ;

- the *degree of structuration* of a community of species, which broadly refers to the level of unevenness between species abundances within the community. This may be appropriately tested by comparing the slope of the corresponding S.A.D. to either the “ideally even” model or the “broken-stick” model. These two models provide two reference levels of structuration, namely the “ideally even” model characterises the *zero* level of structuration while the “broken-stick” model accounts for the degree of structuration that would be obtained by a *random apportionment* of relative abundances among all co-occurring species in the community. Standardising the degree of structuration (the slope of the S.A.D.) to the “broken-stick” model is particularly relevant as this allows to set apart the “mechanistic”, trivial influence of the number of member species to the level of unevenness of species distribution in the community [7, 8, 25]. Thus standardised, the degree of structuration of the community becomes truly independent of its richness in species. This “mechanistic” influence of the level of species richness on the degree of community structuration – that is on the slope of the abundance distribution – is demonstrated in Appendix 3.

Results

Estimation of the total species richness of the community as a whole and separately by kind of feeding guild

Accounting for the values of the numbers f_x of species observed x -times during partial sampling (Figure A1 in Appendix 1), the *least-biased* nonparametric estimator of the number of undetected species during the partial sampling proves to be Jackknife-5 (see the selective key in Appendix 1). The corresponding least-biased estimations of (i) the number Δ of undetected species, (ii) the resulting least-biased estimation S_t of the total species richness of the sampled community and (iii) the level of completeness of the partial sampling Ro/St is provided in Table 1. With 54% level of completeness only, the partial sampling confirms remaining very far from exhaustivity. Separate estimations of Δ and S_t according to feeding guilds – primary and secondary consumers respectively – are subsequently derived. Similar levels of sampling completeness are inferred for both guilds (56% and 53% respectively).

Table 1 – Numerical characteristics of a marine Gastropod community at the intertidal level of a rocky shore at Rangat (Andaman Islands), including: the sampling-size No , the number of detected species Ro ($= R(No)$), the estimated number of undetected species Δ , the resulting evaluation of total species richness S_t and the level of sampling completeness Ro/St .

	No	Ro	Δ	S_t	Ro/St
<u>whole community</u>	157	42	35.7	77.7	54 %
<u>primary consumers</u>	96	17	13.3	30.3	56 %
<u>secondary consumers</u>	61	25	22.4	47.4	53 %
<u>primary / secondary</u>	/	0.68	0.59	0.64	1.06

Due to the relatively low level of achieved sampling completeness, further sampling could be considered of interest. In this perspective, the least-biased extrapolation of the Species Accumulation Curve highlights the expected increase in the number of detected species, $R(N)$, as a function of growing sampling size N , beyond the actually achieved sampling. And, thereby, the additional sampling efforts that would be required to obtain any desirable increase in sampling completeness can be forecasted, as shown in Figure 2.

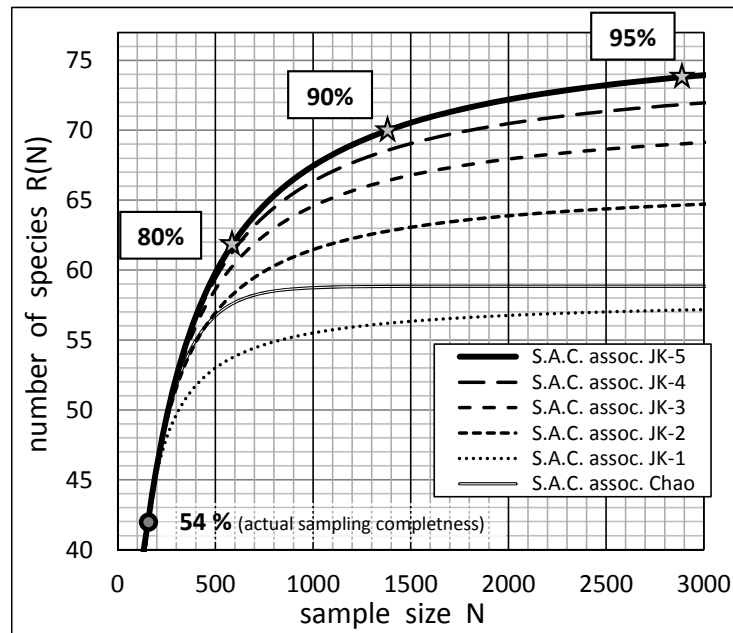


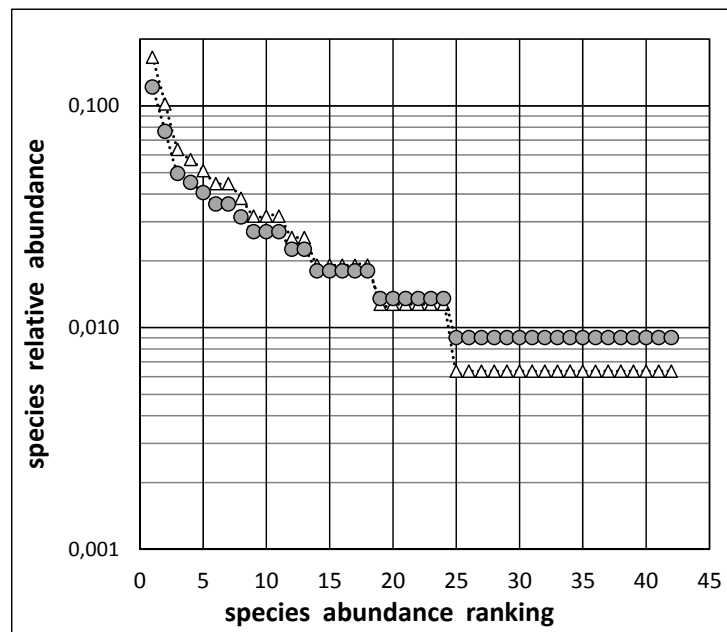
Figure 2 – Extrapolated part of the Species Accumulation Curve (coarse solid line) accounting for the increase of the number of detected species $R(N)$ as a function of growing sample size N beyond the actually achieved sampling ($N_0 = 157$, $R(N_0) = 42$). Here, the selected, least-biased, nonparametric estimator of the number of undetected species is Jackknife-5, leading to a total species richness $St = 78$, with the associated least-biased extrapolation plotted as the coarse solid line. Also plotted are the extrapolations of the S.A.C. associated to non-selected, more biased estimators. In practice, the least-biased extrapolation highlights the expected additional sampling effort required to reach higher levels of sampling completeness (for example, the sample sizes required to reach 80%, 90% and 95% completeness would be around $N = 600$, 1400 , 2900 respectively).

Correction and extrapolation of the Species Abundance Distribution

The as-recorded part of the Species Abundance Distribution (i.e. the part restricted to the set of actually detected species only) needs being corrected, as shown in Figure 3. Corrections, made according to equation (A2.1) in Appendix 2, involve both (i) a *positive* correction due to the multiplicative factor $(1+1/n_i)$ being >1 and (ii) a *negative* correction due to the multiplicative factor $(1-f_1/N_0)/(1+R_0/N_0)$ being < 1 . The positive correction, decreasing progressively with increasing species abundances, outweighs the negative correction only for the lower species abundances, that is along the second part of the recorded Species Abundance Distribution (here for species ranks $i = 19$ to $i = 42$, as highlighted in Figure 3).

The least-biased extrapolation of the Species Abundance Distribution, from rank 43 to rank 78, is computed according to equation (A2.3) in Appendix 2 and is plotted as the solid line in Figure 4. This Figure 4 thus provides the *entire development* of the Species Abundance Distribution, duly corrected for its recorded part and extrapolated with minimized bias for the set of still undetected species. Note that, although most undetected species are expected to be among the less abundant in the community, their ecological importance is no less and, thus, deserve as much attention, as emphasised by several authors [26 – 28].

1 The same procedure of correction and extrapolation is then conducted separately for
2 *primary* and for *secondary* consumers, as shown in Figures 5 and 6. The superposition of
3 the Species Abundance Distributions for each guild in Figure 7 allows an easy, direct
4 comparison of the structuration of species abundances in each guild. An alternative
5 comparative approach is provided in Figures 8 and 9, in the form of a common
6 histogram of species abundances for the two guilds together. In these Figures, the
7 abundances are plotted on an arithmetic scale, rather than the usual log-transformed
8 scale, in order to offer a direct visual appreciation of the respective contributions of the
9 member species of each guild. As shown in Figures 7 and 8, the dominant species in
10 term of number of individuals, are predominantly primary consumers: thus the four
11 numerically dominant species all belong to the herbivorous family Neritidae: *Nerita*
12 *albicilla* Linnaeus 1758, *Nerita polita* Linnaeus 1758, *Nerita insculpta* Récluz 1841,
13 *Nerita costata* Gmelin 1791 (with corrected relative abundances $a_1 = 0.122$, $a_2 = 0.077$,
14 $a_3 = 0.049$, $a_4 = 0.045$, respectively).
15



16
17 **Figure 3** – The as-recorded part of the Species Abundance Distribution (white triangles) and the
18 same after correction (grey discs): see text for further explanation.
19
20

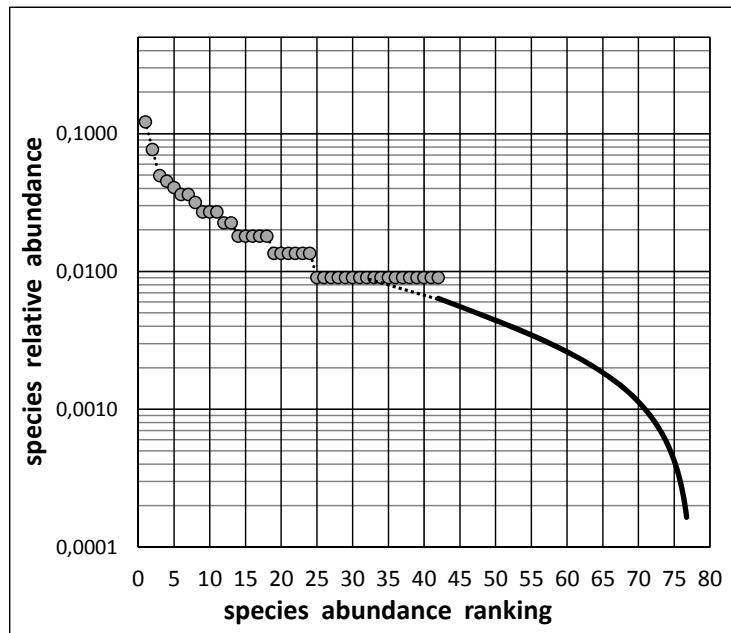


Figure 4 – The completed Species Abundance Distribution, including the correction of the recorded part (involving the 42 detected species: grey discs) and the least-biased extrapolation of the unrecorded part (involving the 36 undetected species: ranks $i = 43$ to 78: solid line).

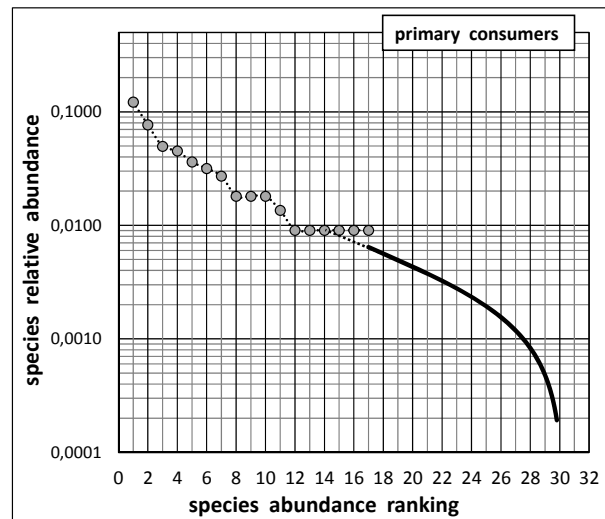


Figure 5 – The completed Species Abundance Distribution (corrected: grey discs and extrapolated: solid line) considering the guild of primary consumers (“herbivores”).

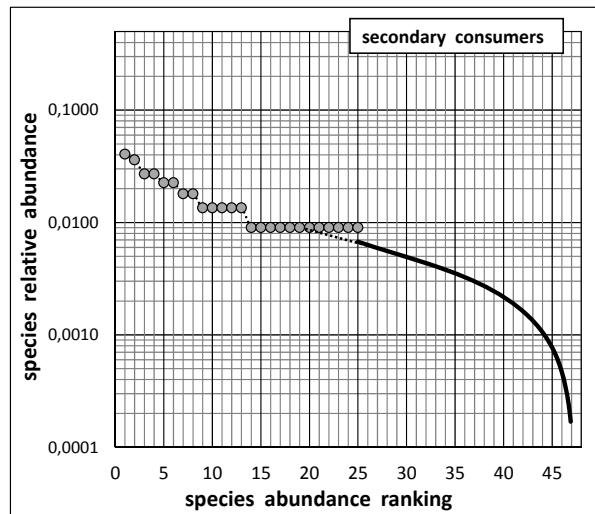


Figure 6 – The completed Species Abundance Distribution (corrected: grey discs and extrapolated: solid line) considering the guild of *secondary consumers* (“carnivores, scavengers,...”).

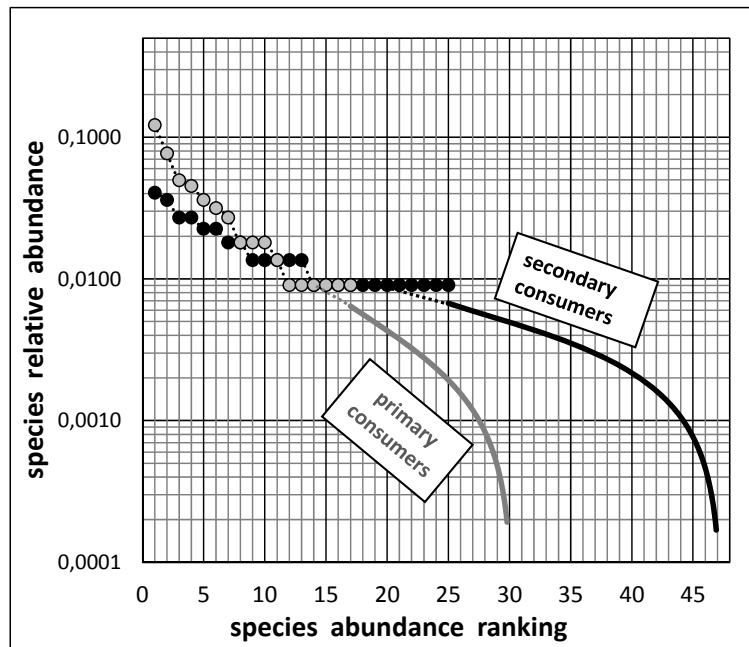


Figure 7 – Superposition of the completed Species Abundance Distributions for primary and secondary consumers (plotted in Figures 5 and 6), allowing direct comparison between both feeding guilds.

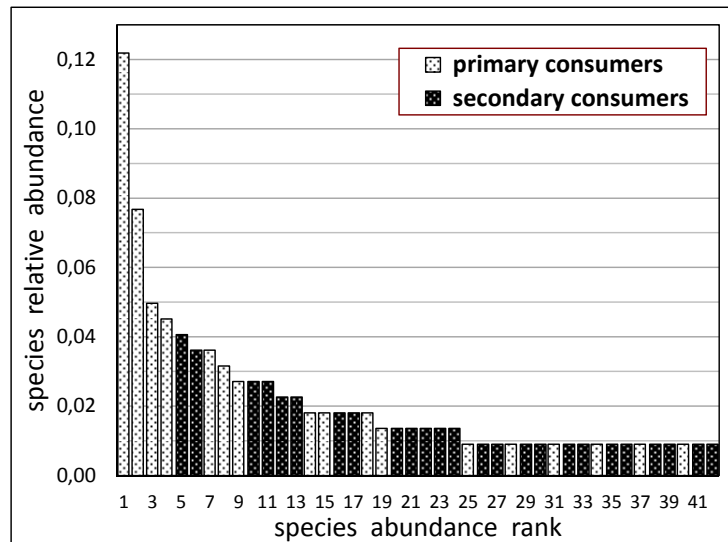


Figure 8 – The Species Abundance Distribution for the 42 *already detected* species, highlighting the feeding mode of each species. Note the arithmetic (instead of log-transformed) scale for abundances. Directly derived from Figures 5 & 6.

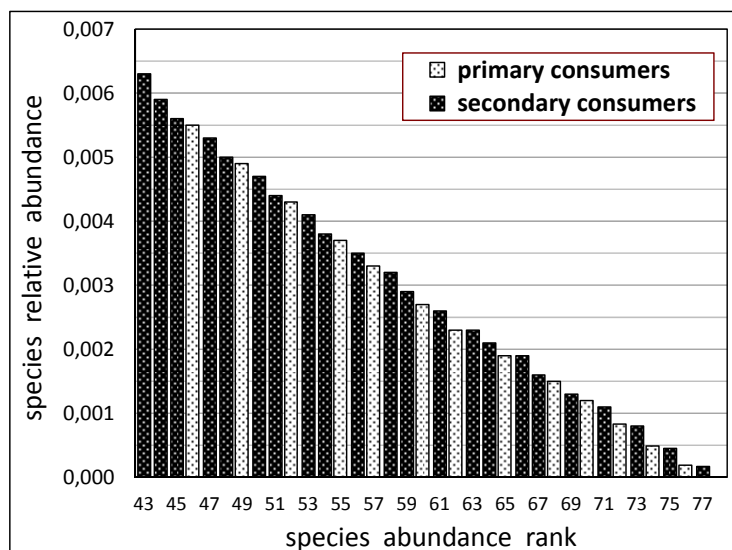


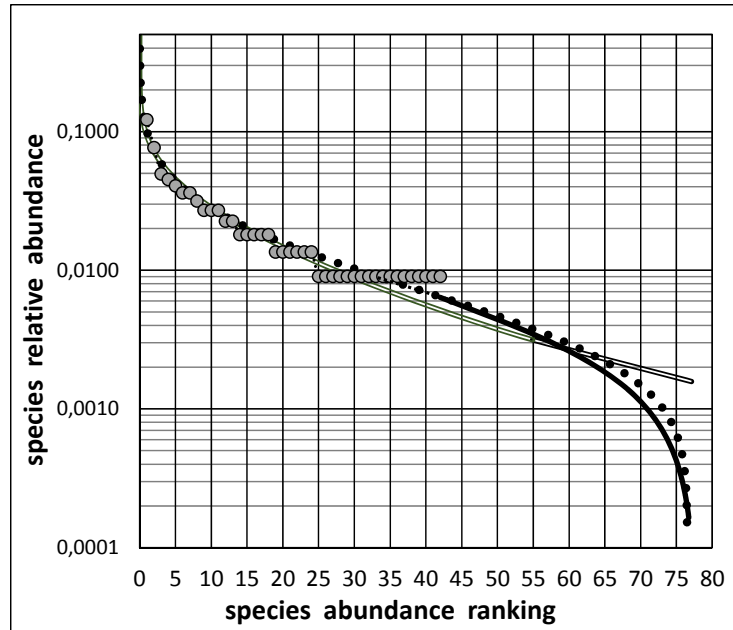
Figure 9 – The Species Abundance Distribution *extrapolated* for the 35 *still undetected* species, highlighting the feeding mode of each species. Note the arithmetic (instead of log-transformed) scale for abundances. Directly derived from Figures 5 & 6.

Quantitative characterisation of the hierarchical structuring of species abundances in the studied community

** the type of structuring process*

In order to investigate which kind of structuring process is at work in the studied community, two models of abundance distribution – the “log-normal” distribution and the “log-series” distribution – were tentatively fitted to the completed Species Abundance Distribution provided at Figure 4. As shown in Figures 10 and 11, a fairly good fit is obtained with the *log-normal* model while accordance with *log-series* is comparatively unsatisfactory. The same holds true for both feeding guilds, primary and secondary consumers considered separately: Figures 12, 13, 14, 15.

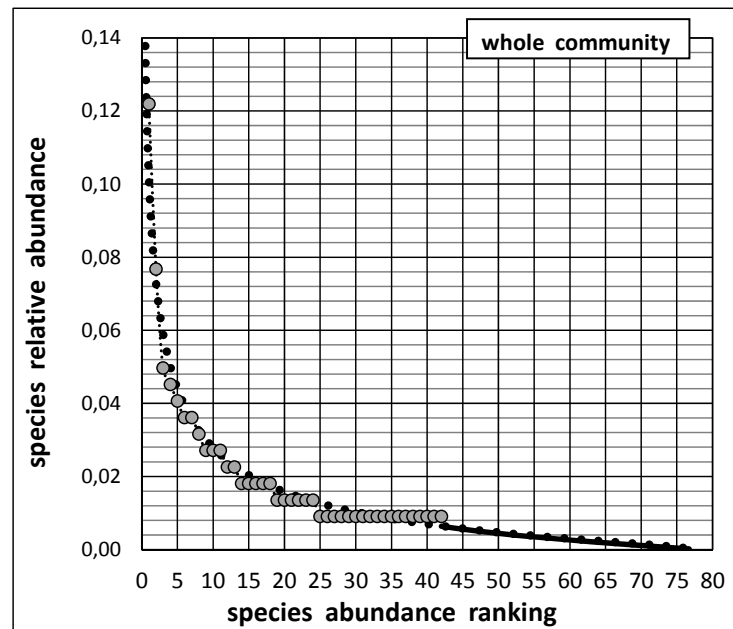
1



2

3 **Figure 10** – Two classical models: “log-normal” (coarse dotted line) and “log-series” (double line)
 4 fitted to the Species Abundance Distribution of the whole community (corrected and extrapolated
 5 cf. Figure 4). The best fit is with "log-normal" distribution.

6



7

8 **Figure 11** – The “log-normal” model (coarse dotted line) fitted to the Species Abundance Distribution
 9 (as in Figure 2). Note arithmetic scale for abundances (instead of log-transformed scale), which
 10 allows a more direct visual appreciation of relative abundances, as suggested by MACARTHUR [29]
 11

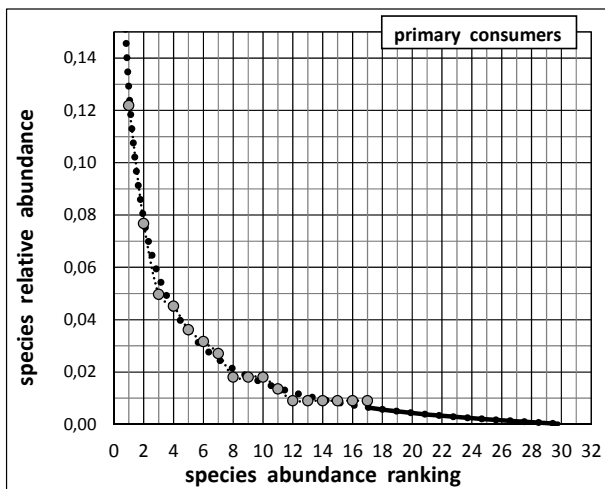
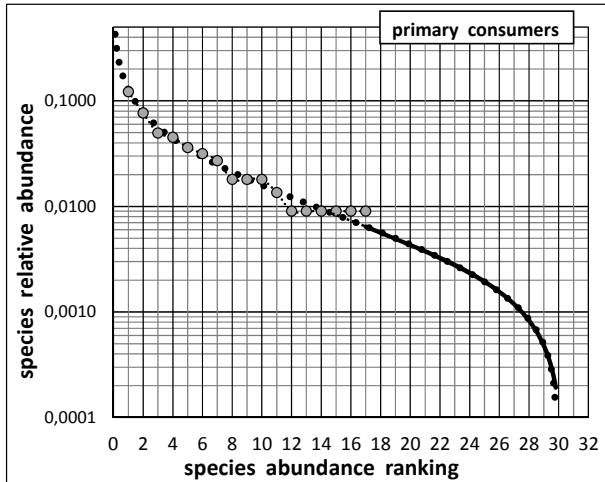
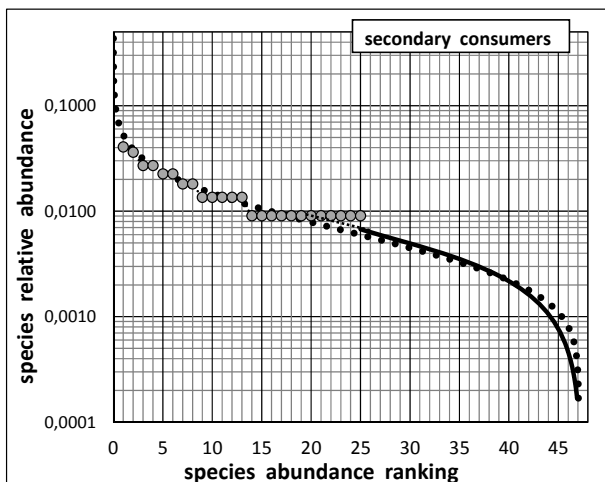


Figure 12 – The “log-normal” model (coarse dotted line) fitted to the Species Abundance Distribution for the guild of *primary consumers*.

Figure 13 – Same as Figure 12 but with arithmetic scale for abundances (instead of log-transformed scale), which allows a more direct visual appreciation of relative abundances, as suggested by MACARTHUR [29].



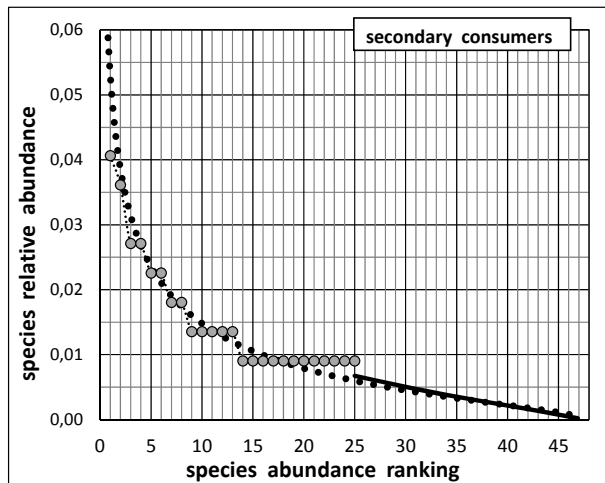


Figure 14 – The “log-normal” model (coarse dotted line) fitted to the Species Abundance Distribution for the guild of *secondary consumers*.

Figure 15 – Same as Figure 14 but with arithmetic scale for abundances (instead of log-transformed scale), which allows a more direct visual appreciation of relative abundances, as suggested by MACARTHUR [29].

** the intensity of the structuring process*

The shape of the Species Abundance Distribution characterizes in details the intensity of hierarchical structuring of species abundances in the community.

As emphasised in the Methods section, the slope of the Species Abundance Distribution should preferably be standardised to the slope of the corresponding “broken-stick” distribution in order to unveil the *genuine* intensity of the structuring process, thereby leaving aside the trivial contribution of the level of species richness. Accordingly, in Figures 16, 17, 18, the completed (corrected and extrapolated) Species Abundance Distribution is plotted together with the corresponding “broken-stick” distribution. Comparing the slopes on the same graph thus provides a reliable appreciation of the degree of hierarchical structuring of species abundances.

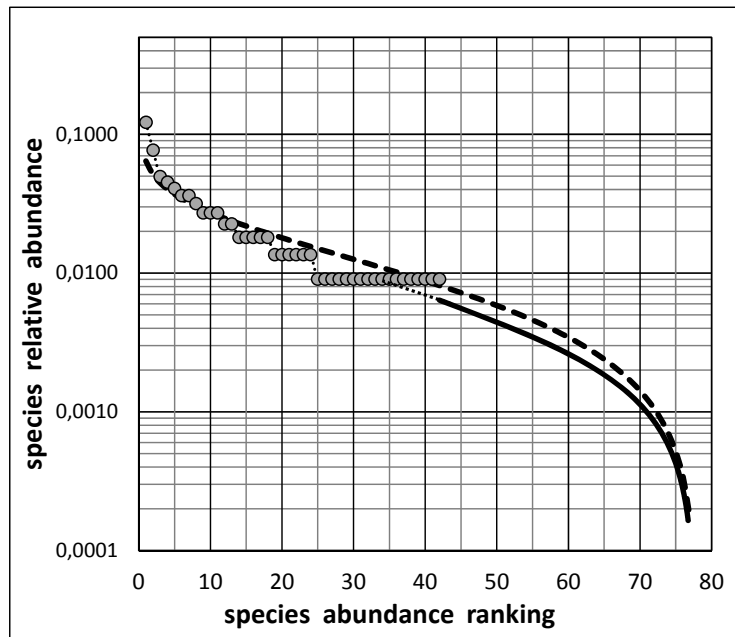
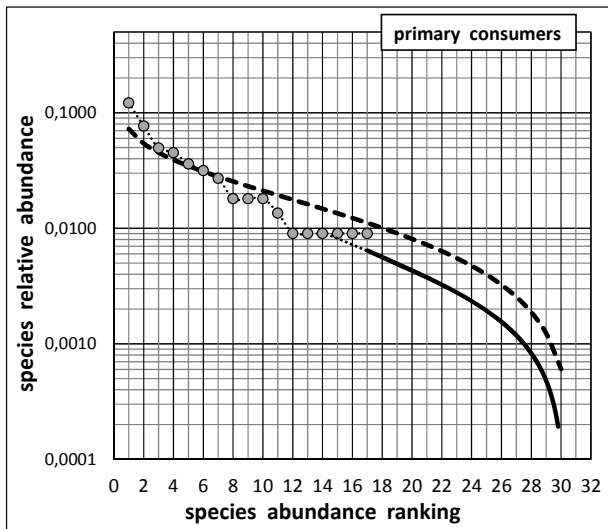


Figure 16 – The Species Abundance Distribution of the *whole community* plotted together with the corresponding “broken-stick” distribution (i.e. computed for the same species richness)



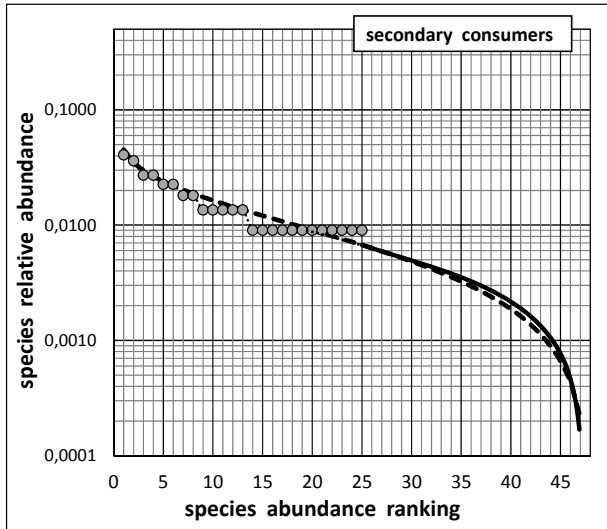


Figure 17 – The Species Abundance Distribution for the guild of *primary consumers* plotted together with the corresponding “broken-stick” distribution.

Figure 18 – The Species Abundance Distribution for the guild of *secondary consumers* plotted together with the corresponding “broken-stick” distribution.

In a more concise, but reductionist approach, the *average* slope of the Species Abundance Distribution provides a convenient appreciation of the degree of hierarchical structuring. Accordingly, a “structuring index” (i) highlighting the degree of structuring and (ii) conventionally standardised to the “broken-stick” distribution, can be defined as the ratio between the average slope of the actual abundance distribution and the average slope of the “broken-stick” model, with abundances being classically plotted log-transformed. Thus defined, the structuring index is equal to:

$$I_{str} = \log(a_1/a_{st}) / \log(a'_1/a'_{st})$$

where a_1 and a_{st} stand for the highest and the lowest abundances in the studied assemblage and a'_1 and a'_{st} stand for the highest and the lowest abundances in the corresponding “broken-stick” distribution having the same species richness S_t . Results are given in Table 2.

Table 2 – The degree of hierarchical structuration (unevenness) of species abundances, quantified as the average slope of the Species Abundance Distribution standardised to the average slope of the corresponding “broken-stick” distribution (see Figures 16, 17, 18).

	Sp. Abund. Distr.		“broken-stick”		structuring index $\log[a_1/a_{st}] / \log[a'_1/a'_{st}]$
	a_1	a_{st}	a'_1	a'_{st}	
<u>whole community</u>	0.12200	0.00016	0.06400	0.00017	1.12
<u>primary consumers</u>	0.12200	0.00019	0.07250	0.00060	1.35
<u>secondary consumers</u>	0.08500	0.00035	0.04510	0.00022	1.03

* *body-size versus relative abundance relationships*

Figures 19, 20, 21 highlight the relationships between the shell-size – as a surrogate to body-size – and the corresponding relative abundance of each actually sampled species

(of course, numerical extrapolation provides no information on body-size of undetected species).

Primary and secondary consumers differ regarding their respective ranges of shell-size (Figures 20 and 21). Considering an arbitrarily fixed threshold-size of 40 mm:

- only one species (out of 17) has a common-size in excess of 40 mm among *primary* consumers ;

- eleven species (out of 25) have a common-size in excess of 40 mm among *secondary* consumers.

The difference in proportions (6% against 44%) is statistically significant: χ^2 with Yates correction = 6.8, $p < 0.01$. The largest species among secondary consumers are: *Harpa major* Röding 1798 (≈ 95 mm), *Vasum turbinellus* (Linnaeus 1758) (≈ 85 mm), *Chicoreus brunneus* (Link 1807) (≈ 75 mm), *Semicassis bisulcata* (Schubert & Wagner 1829) (≈ 55 mm), *Bufonaria echinata* (Link 1807) (≈ 50 mm), *Latirolagena smaragdulus* (Linnaeus 1758) (≈ 50 mm), *Gemmula vagata* (E.A. Smith 1895) (≈ 50 mm), *Bursa granularis* (Röding 1798) (≈ 48 mm), *Tylothais virgata* (Dillwyn 1817) (≈ 45 mm), *Polinia undosa* (Linnaeus 1758) (≈ 45 mm), *Polinices mammilla* (Linnaeus 1758) (≈ 45 mm). The largest species among primary consumers is: *Scutellastra flexuosa* (Quoy & Gaimard 1834) (≈ 70 mm).

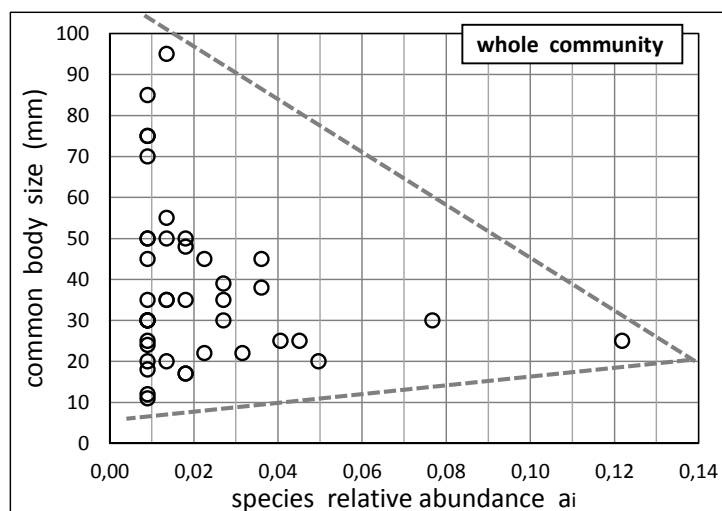
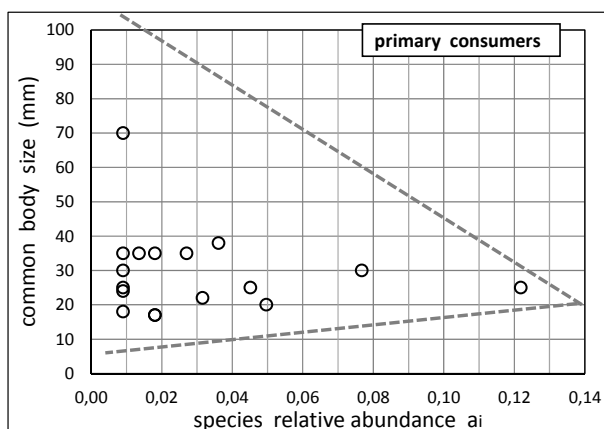


Figure 19 – Body-size versus relative abundance relationship for the 42 sampled species in the *whole* community. Relative abundances corrected according to equation (A2.1).



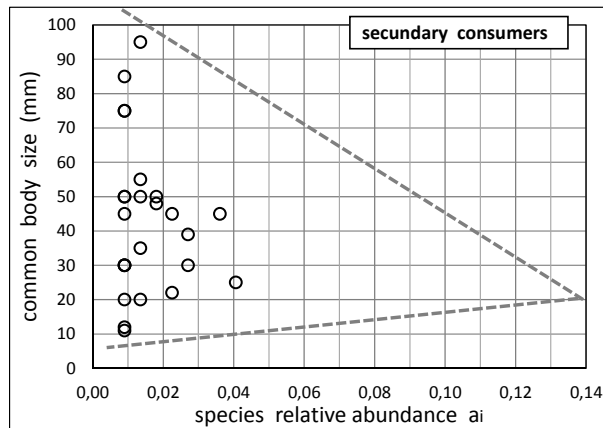


Figure 20 – Body-size versus relative abundance relationship for the 17 sampled species members of the guild of *primary consumers*. Relative abundances corrected according to equation (A2.1).

Figure 21 – Body-size versus relative abundance relationship for the 25 sampled species members of the guild of *secondary consumers*. Relative abundances corrected according to equation (A2.1).

Discussion

A thorough approach to the species diversity and the hierarchical structuring of abundances in species assemblages would arguably require (quasi-) exhaustive samplings. Yet, incomplete samplings usually become unavoidable practice as soon as species-rich assemblages are addressed, especially invertebrate's communities. Hopefully, the implementation of appropriate extrapolating methods can "force" incomplete samplings and partial inventories *to reveal much more information* than one might think *a priori*. Thus, proper numerical extrapolations of both the Species Accumulation Curve and the Species Abundance Distribution provide an unexpected set of additional information relative to all those species that *remained undetected* after partial samplings. This, in turn, allows to tackle the main issues relative to the evaluation of species richness and the internal organisation within partially sampled communities – all subjects that would normally require, first, a complete assessment of the species list.

A thorough analysis of the intertidal gastropod community on the rocky shore at Rangat location (Andaman Islands) was conducted accordingly, in compliance with this methodological approach.

Total species richness estimates and the forecasted additional sampling efforts required to improve sampling completeness

At first, *Jackknife at order 5* reveals being the *least-biased* estimator among the series of classical nonparametric estimators of the number of undetected species (according to the selective key in Appendix 1). This selection of Jackknife-5 proves here satisfying not only for the community as a whole but also for each of the feeding guilds considered separately. The true species richness estimates (Table 1) – 77 species for the whole community, among which 30 primary-consumers and 47 secondary-consumers – substantially exceeds the recorded numbers (42, 17, 25, respectively). Thus, the sampling-completeness levels reached during the inventory of the whole community and the sub-inventories of each feeding guilds all hardly exceed 50%. This justifies, *a posteriori*, the need of implementing an accurate numerical extrapolation of the partial

1 inventory of this marine gastropod community. As a comparison, the total species
2 richness of marine gastropods assemblages on coral reefs around each of three small
3 islands in Mannar Gulf (India) was estimated from 49 to 53 species only [30].

4 Although these extrapolations will provide a lot of additional interesting information
5 regarding the set of still unrecorded species (as described below), further field
6 investigation effort, aiming at increasing the completeness of inventory, might possibly
7 be considered. If so, a reliable forecast of the additional sampling effort required to meet
8 a given target gain in sampling-completeness would be useful for the optimal planning
9 of the resources and efforts to be organised on site. Considering the least-biased
10 extrapolation of the Species Accumulation Curve associated to the least-biased
11 estimator (here Jackknife-5) answers appropriately this demands, as shown in Figure 2.
12 Clearly, substantial further improvement of sampling completeness would rapidly
13 require strong additional investment of resource on site. Thus, increasing completeness
14 from the actual 54% level to 80% or 90% would require multiplying the actual sample-
15 size by *four* or by *ten* respectively.

16 Being able to reliably estimate the level of such additional efforts is of obvious prime
17 interest to rationally decide whether to continue sampling operation any further or to
18 rely only on the actual inventory adequately extended by numerical extrapolation.

20 ***Correction and extrapolation of the Species Abundance Distribution***

21 The as-recorded distribution needs correction and extrapolation because it is both (i)
22 slightly *biased*, due in particular to sampling stochasticity and (ii) most importantly,
23 *incomplete* to the extent of the proportion of those species of the sampled community
24 that remained undetected during sampling.

25 The *corrections*, computed according to equation (A2.1), involve (i) a negative
26 contribution (multiplying factor $(1-f_1/N_0)/(1+R_0/N_0)$ which is < 1) in relation to the
27 degree of sampling incompleteness and (ii) a positive contribution (multiplying factor
28 $(1+1/n_i)$ which is >1) that compensate for statistical bias during sampling [20]. These
29 combined corrections result in a slight reduction of higher abundances and a slight
30 increase of lower abundances, as shown in Figure 3.

31 In turn, the numerical, *least-biased extrapolation* of the distribution of abundances leads
32 to the complete development of the Species Abundance Distribution, as would be
33 expected from an exhaustive inventory of the whole set of member species in the
34 community: Figure 4 for the whole community and Figures 5, 6, 7, 8, 9 for each feeding
35 guild. Indeed, dealing with a *full range* Species Abundance Distribution (using
36 extrapolation as far as necessary) is essential, not only to deliver a full description of the
37 *pattern* of abundances but, also, to question the kind of *process* actually involved in the
38 hierarchical structuring of abundance distribution. This is achieved, in particular, by
39 comparing the studied Species Abundance Distribution to different theoretical models,
40 each of which being considered representative of a particular kind of process at work in
41 the structuring of species abundances in the community [5]. However, it turns out that
42 such comparisons, when conducted with incomplete – non extrapolated – distributions
43 often leads to erroneous diagnostics, as in shown in the present work (Figure 10) and
44 already emphasised by several authors [4, 6, 23, 31, 32]. Hence the requirement to
45 conduct reliable comparisons using *full range* Species Abundance Distribution only.

Beyond considering globally the Species Abundance Distribution of the whole community, it seems worth considering also, separately, the distributions peculiar to each feeding guild, here *primary consumers* and *secondary consumers*. Indeed, deconstructing the global abundance distribution according to the member subsets of the whole community can lead to more focused and detailed analysis, as indicated by [22, 32].

Inferring the type of process driving the hierarchical structuring of species abundances

Considered in its full range, the Species Abundance Distribution of the whole community is clearly fitted best by the “log-normal” distribution than by the “log-series” distribution (Figures 10 & 11). This suggests, accordingly, that the *process of structuration* of the whole community is likely driven by the combined contributions of many independent factors, rather than by only one (or very few) dominant factors [5, 21, 22 – 24]. Moreover, the same still holds true for the structuring process at work in each feeding guild: both *primary consumers* (Figures 12 & 13) and *secondary consumers* (Figures 14 & 15) are likely structured by the combined influences of multiple independent factors, probably rather different in each guild.

Quantifying the degree of hierarchical structuration of species abundances

Here also, considering the full range of the Species Abundance Distribution is necessary not only to duly include the substantial set of still undetected species but, also, to allow the comparison with the corresponding “broken-stick” reference: Figures 16, 17, 18. The degree of hierarchical structuring of species abundances in the whole community, measured with the “structuring index” (Table 2), is slightly higher than the “broken-stick” reference ($I_{str} = 1.12$). Besides, the two feeding guilds differ from each other in this respect, with *secondary consumers* being close to the reference ($I_{str} = 1.03$) while *primary consumers* are more strongly structured ($I_{str} = 1.35$).

Dependence between body-size and relative abundance

Globally, shell-size (taken as a surrogate to body-size) and relative abundance are fairly independent among the 42 recorded species and the same hold true for each feeding guild considered separately (Figures 19, 20, 21). Yet, the maximum shell-size reached a given level of abundance is clearly decreasing with increasing relative abundance, in the whole community as in each feeding guild. This last pattern is not surprising and may possibly have wide validity [33]; in particular, the same trend was observed in land snail communities as well (BÉGUINOT, *unpublished results*).

Primary and secondary consumers differ however regarding the range of shell-size (Figures 20 and 21), with a significantly larger proportion of species having shell-size above 40 mm among secondary consumers: 44% against 5% (χ^2 with Yates correction = 6.8, $p < 0.01$).

Conclusion

When dealing with substantially incomplete species inventories, the numerical extrapolations of the Species Accumulation Curve and of the Species Abundance Distribution offer remarkable opportunities to unveil an unexpectedly rich sum of information relative to the set of undetected species. In turn, thanks to the resulting access to the full range of the Species Abundance Distribution, interesting additional information may be derived, regarding the process and pattern of the hierarchical organisation within the partially sampled community. Still, further investigations may be conducted by deconstructing the Species Abundances Distribution in its main constitutive subsets, such as feeding guilds, which co-exist and interact within the whole community. This is made possible by the additive properties of least-biased numerical extrapolation [17, 18]. The numerical extrapolations applied to the partial inventory of a marine gastropod community in Andaman Islands concretely demonstrate the wide range of ecological questions that may be addressed and successfully answered, even though just over half of the true species richness of the community had been actually recorded only. In short, this clearly highlights the potential interest of numerical extrapolations of incomplete samplings and partial inventories, doomed to become increasingly frequent, due to the generalised practice of quick assessments of biodiversity, especially for invertebrate faunas under tropical climates.

Appendix 1

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 $R(N)$: see [17, 34, 35]:

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

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 [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 [17]:

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

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

$$* \text{ for } f_1 \text{ up to } 3f_2 - 3f_3 + f_4 \rightarrow R_3(N) = (R(N_0) + 3f_1 - 3f_2 + f_3) - (6f_1 - 8f_2 + 3f_3).N_0/N \\ - (-4f_1 + 7f_2 - 3f_3).N_0^2/N^2 - (f_1 - 2f_2 + f_3).N_0^3/N^3$$

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

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

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

$$* 0.6 f_2 < f_1 \leq f_2 \rightarrow \Delta_{j1} = f_1 ; R_1(N)$$

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

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

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

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

N.B. 1: As indicated above (and demonstrated in details in BÉGUINOT [17], 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 from 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 [2, 3, 15, 16, 36]. Also, this shows that the approach initially proposed by BROSE *et al.* [37] – 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 be selected: see reference [18].

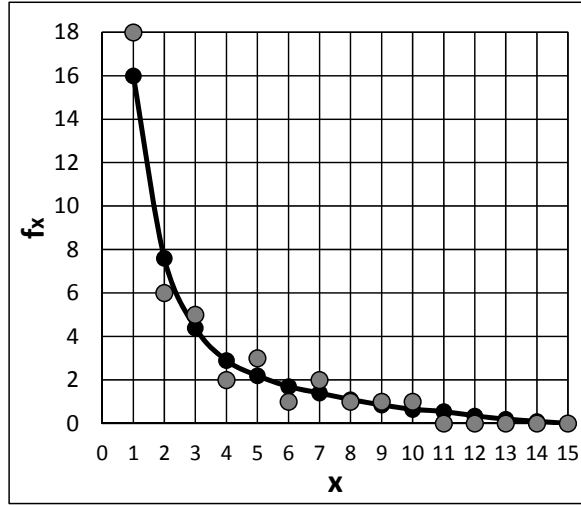


Figure A1 – The recorded values of the numbers f_x of species recorded x -times (grey discs) and the regressed values of f_x (black discs) derived to reduce the consequence of stochastic dispersion.

Appendix 2

Correction and extrapolation of the as-recorded Species Abundance Distribution (S.A.D.)

N.B.: details regarding the derivation of the following expressions are provided in BÉGUINOT (2018).

1) Correction for a 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:

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

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 $i > R_0$):

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

which, in practice, comes down to:

$$a_i \approx (2/N_i) / (1 + R(N_i)/N_i) \quad (A2.3)$$

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.

Appendix 3

The trivial 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

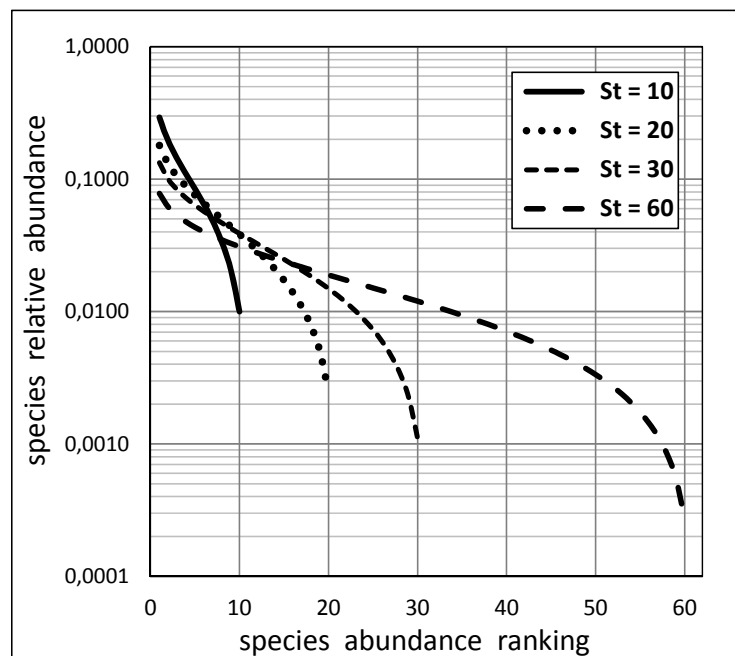


Figure A3 – The “broken-stick” distribution model applied to species communities with increasing species richness $S_t = 10, 20, 30, 60$. Although the theoretical structuring process involved in the “broken-stick” model remains unchanged (random apportionment of relative abundances among member species),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

References

1. 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.
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 T-J Rarefaction and extrapolation of species richness using an area-based Fisher’s logseries. *Ecology and Evolution*. 2017; 7: 10066-10078.

- 1
- 2 4. Connolly SR, Hughes TP & Bellwood DR A unified model explains commonness and
- 3 rarity on coral reefs. *Ecology Letters*. 2017; 20: 477-486.
- 4
- 5 5. May RM Patterns of species abundance and diversity. In Cody M.L. & Diamond J.M.
- 6 1975 *Ecology and Evolution of Communities*. The Belknap Press of Harvard University:
- 7 81-120.
- 8
- 9 6. McGill BJ, Etienne RS, Gray JS et al. Species abundance distributions: moving beyond
- 10 single prediction theories to integration within an ecological framework. *Ecology*
- 11 *Letters*. 2007; 10: 995-1015.
- 12
- 13 7. Ulrich W, Ollik M & Ugland KI A meta-analysis of species-abundance distributions.
- 14 *Oikos*. 2010; 119:1149-1155.
- 15
- 16 8. Komonen A & Elo M Ecological response hides behind the species abundance
- 17 distribution: community response to low-intensity disturbance in managed grasslands.
- 18 *Ecology and Evolution*. 2017; 7: 8558-8566.
- 19
- 20 9. Wang X, Ellwood F, Ai D, Zhang R & Wang G Species abundance distributions as a
- 21 proxy for the niche-neutrality continuum. *Journal of Plant Ecology*. 2017; rtx 013.
- 22
- 23 10. Wells JW Coral reefs: 609-632; in J.W. Hedspeth editor, *Treatise on marine ecology*
- 24 *and paleoecology*. Geological Society of America. 1957; Mem. 67.
- 25
- 26 11. Glynn PW High complexity food webs in low-diversity eastern Pacific reef-coral
- 27 communities. *Ecosystems*. 2004; 7: 358-367.
- 28
- 29 12. Jeeva C, Mohan PM, Sabith DB, Ubare VV, Muruganantham M & Kumari RK
- 30 Distribution of Gastropods in the intertidal environment of south, middle and north
- 31 Andaman Islands, India. *Open Journal of Marine Science*. 2018; 8: 173-195.
- 32
- 33 13. Rao SN & Dey A Catalogue of marine Molluscs of Andaman and Nicobar Islands.
- 34 Records of Zoological survey of India, Kolkatta, India. Occ. 2000; Paper n° 187: 1-323.
- 35
- 36 14. Coddington JA, Agnarsson I, Miller JA, Kuntner M & Hormiga G Undersampling
- 37 bias : the null hypothesis for singleton species in tropical arthropod surveys. *Journal of*
- 38 *Animal Ecology*. 2009; 78: 573-584.
- 39
- 40 15. Gotelli NJ & Colwell RK Estimating species richness. pp. 39-54 in: *Biological*
- 41 *Diversity: Frontiers In Measurement And Assessment*. A.E. Magurran and B.J. McGill
- 42 (eds.). 2010; Oxford University Press, Oxford. 345 pp.
- 43
- 44 16. Gotelli NJ & Chao A Measuring and Estimating Species Richness, Species Diversity,
- 45 and Biotic Similarity from Sampling Data. In: Levin S.A. (ed.) *Encyclopedia of*

- 1 Biodiversity. 2013; second edition, volume 5, pp. 195-211. Waltham, MA: Academic
2 Press.
3
- 4 17. Béguinot J Theoretical derivation of a bias-reduced expression for the
5 extrapolation of the Species Accumulation Curve and the associated estimation of total
6 species richness. *Advances in Research*. 2016; 7(3): 1-16. doi:
7 10.9734/AIR/2016/26387 ; <hal-01367803>
8
- 9 18. Béguinot J Extrapolation of the Species Accumulation Curve associated to “Chao”
10 estimator of the number of unrecorded species: a mathematically consistent derivation.
11 *Annual Research & Review in Biology*. 2016; 11 (4): 1-19 doi:
12 10.9734/ARRB/2016/30522; <hal 01477263 >
13
- 14 19. Chao A, Hsieh T, Chazdon RL, Colwell RK & Gotelli NJ Unveiling the species-rank
15 abundance distribution by generalizing the Good-Turing sample coverage theory.
16 *Ecology*. 2015; 96(5): 1189-1201.
17
- 18 20. Béguinot J How to extrapolate species abundance distributions with minimum bias
19 when dealing with incomplete species inventories. *Advances in Research*. 2018; 13(4):
20 1-24. doi: 10.9734/AIR/2018/39002.
21
- 22 21. Loreau M Species abundance patterns and the structure of ground-beetle
23 communities. *Ann. Zool. Fennici*. 1992; 28: 49-56.
24
- 25 22. Magurran AE & Henderson PA Explaining the excess of rare species in natural
26 species abundance distributions. *Nature*. 2003; 422: 714-716.
27
- 28 23. Connolly SR, Hughes TP, Bellwood DR & Karlson RH Community structure of corals
29 and reef fishes at multiple scales. *Science*. 2005; 309: 1363-1365.
30
- 31 24. Ulrich W, Soliveres S, Thomas AD, Dougill AJ & Maestre FT Environmental
32 correlates of species rank-abundance distributions in global drylands. *Europe PMC*
33 *Funders Group*. 2016; 20: 56-64.
34
- 35 25. MacDonald ZG, Nielsen SE & Acorn JH Negative relationships between species
36 richness and evenness render common diversity indices inadequate for assessing long-
37 term trends in butterfly diversity. *Biodiversity Conservation*. 2017; 26: 617-629.
38
- 39 26. Novotny V & Basset Y Rare species in communities of tropical insect herbivores:
40 pondering the mystery of singletons. *Oikos*. 2000; 89: 564-572.
41
- 42 27. Harte J Tail of death and resurrection. *Nature*. 2003; 424: 1006-1007.
43
- 44 28. Henderson PA & Magurran AE Direct evidence that density-dependent regulation
45 underpins the temporal stability of abundant species in a diverse animal community.
46 *Proceedings of The Royal Society B*. 2014; 281 doi: 10.1098/rspb.2014.1336.

- 1
2 29. MacArthur RH On the relative abundance of bird species. Proceedings of the
3 National Academy of Sciences U.S.A. 1957; 43: 293-295.
4
- 5 30. Béguinot J Extrapolation of total species richness from incomplete inventories:
6 application to the Gastropod fauna associated to coral reefs in 'Mannar Gulf Biosphere
7 Reserve', India. Asian Journal of Environment and Ecology. 2017; 4(3): 1-14. doi:
8 109734/AJEE/2017/36831.
9
- 10 31. Magurran AE Species abundance distributions: pattern or process? Functional
11 Ecology. 2005; 19: 177-181.
12
- 13 32. Matthews TJ & Whittaker RJ On the species abundance distribution in applied
14 ecology and biodiversity management. Journal of Applied Ecology. 2015; 52: 443-454.
15
- 16 33. Henderson PA & Magurran AE Linking species abundance distributions in
17 numerical abundance and biomass through simple assumptions about community
18 structure. Proceedings of The Royal Society B. 2009; doi: 10.1098/rspb.2009.2189.
19
- 20 34. Béguinot J An algebraic derivation of Chao's estimator of the number of species in a
21 community highlights the condition allowing Chao to deliver centered estimates. ISRN
22 Ecology. 2014; vol. 2014 : article ID 847328, doi:10.1155/2014/847328; <hal-
23 01101415>
24
- 25 35. Béguinot J When reasonably stop sampling? How to estimate the gain in newly
26 recorded species according to the degree of supplementary sampling effort. Annual
27 Research & Review in Biology. 2015; 7(5): 300-308. doi : 10.9734/ARRB/2015/18809 ;
28 <hal-01228695>
29
- 30 36. O'Hara RB Species richness estimators: how many species can dance on the head
31 of a pin? Journal of Animal Ecology. 2005; 74: 375-386.
32
- 33 37. Brose U, Martinez ND & Williams RJ Estimating species richness: sensitivity to
34 sample coverage and insensitivity to spatial patterns. Ecology. 2003; 84(9) : 2364-2377.