

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/329326096>

# Comparing the Complete Hierarchical Structuration of Species Abundances in Reef Fish Communities According to Coral Morphology, Using the Numerical Extrapolation of Only Incomplete...

Article · November 2018

DOI: 10.9734/AJEE/2018/45402

CITATIONS

0

READS

24

1 author:



Jean Béguinot

University of Burgundy

57 PUBLICATIONS 79 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



to popularize knowledge about leaf-mining insects [View project](#)



insects behavior [View project](#)



# Comparing the Complete Hierarchical Structuration of Species Abundances in Reef Fish Communities According to Coral Morphology, Using the Numerical Extrapolation of Only Incomplete Inventories

Jean Béguinot<sup>1\*</sup>

<sup>1</sup>*Société Histoire Naturelle-Bourgogne Nature, 7 bvd H.P. Schneider 71200, Le Creusot, France.*

## *Author's contribution*

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

## *Article Information*

DOI: 10.9734/AJEE/2018/45402

### Editor(s):

(1) Dr. Wen-Cheng Liu, Professor, Department of Civil and Disaster Prevention Engineering, Taiwan Typhoon and Flood Research Institute, National United University, Taiwan.

### Reviewers:

(1) Manoel Fernando Demétrio, Brazil.

(2) M. Thangaraj, Annamalai University, India.

(3) João Pedro Barreiros, Universidade dos Açores, Portugal.

Complete Peer review History: <http://www.sciencedomain.org/review-history/27497>

**Original Research Article**

**Received 16 September 2018**

**Accepted 24 November 2018**

**Published 30 November 2018**

## **ABSTRACT**

Numerous investigations have been conducted trying to unveil some possible correlations between coral-associated fish communities and different gross types of coral morphology that are assumed being differentially attractive to various kinds of fishes. Investigations in this respect are, yet, strewn with pitfalls and exposed to unexpected source of bias. One potential cause of bias comes from having to deal with substantially incomplete samplings, a situation practically unavoidable with species-rich communities having very uneven distribution of abundances. To overcome this difficulty, the *numerical extrapolation* of incompletely sampled communities offers a convenient solution, serving as a surrogate, when having no reasonable possibility to pursue samplings until reaching completion.

As an example, previous work came to the conclusion that relatively faint differences only distinguish the fish communities respectively associated with two distinctly different stony-coral morphologies, namely the regularly convex *Diploria strigosa* and the deeply lobed *Orbicella*

\*Corresponding author: E-mail: [jean-beguिनot@orange.fr](mailto:jean-beguिनot@orange.fr);

*annularis*. This reported conclusion may be considered questionable, however, as it has been drawn from partial inventories only. Accordingly – given the practical impossibility to further improve the levels of completeness of the available samplings – a more soundly established analysis would require the implementation of a procedure of numerical extrapolation of the Species Abundance Distributions of both communities. After completion of the Species Abundance Distributions by numerical extrapolation, it then become clear that, in fact, substantial differences actually distinguish both fish communities, in terms of true (total) species richness and taxonomic distinctiveness. And, rather unexpectedly, it is the coral having lesser gross rugosity (*Diploria strigosa*) that hosts the most species-rich fish community. Also, the degree of unevenness of species abundances substantially differs between both communities while, however, the genuine intensity of the process of hierarchical structuring of species abundances remains practically unchanged. It is argued that this discrepancy between the unevenness pattern and the intensity of the underlying process actually results from the influence of species richness upon the degree of unevenness of species abundances – a purely mathematical influence, regrettably too often ignored and, indeed, devoid of biological significance.

On a more general methodological ground, this case study demonstrates, once more, the usefulness of numerical extrapolation when dealing with incomplete samplings, in order to avoid the potential risk of deriving erroneous conclusions.

**Keywords:** Coral reef; species diversity; species richness; ranked abundance distribution; Caribbean; incomplete inventory; evenness; unevenness.

## 1. INTRODUCTION

Tropical marine ecosystems in shallow waters are of major interest, being considered as embodying remarkably high levels of diversity and biological complexity [1-5]. Reef fish communities are especially important components within these tropical ecosystems, playing significant roles within each feeding guild and having, thus, a decisive influence on the general organisation of these ecosystems [6].

Now, the usually high species diversity of reef fish communities inevitably confronts to a methodological issue: the difficulty and often the virtual impossibility to complete samplings so as to reach (or at least to closely approach) exhaustive species inventories. And this, in turn, can lead to severely unreliable inferences, since sampling incompleteness not only delivers undetermined underestimates of the true species richness but also hampers any comprehensive and unbiased approach to the hierarchical structuration of species abundances within communities [7–10]. Hence, the necessity to implement a reliable procedure of *numerical extrapolation* of partial sampling [11], able to provide estimates with minimised bias regarding: (i) the number of the still unrecorded species and (ii) the distribution of abundance of these unrecorded species. 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 [12–17].

A recently developed procedure of numerical extrapolation takes into account these needs, aiming to provide *relevant* and *comprehensive* inferences regarding both the true (total) species richness and the full range of the species abundance distribution, in spite of having to rely only on incomplete species inventories.

In particular, this new extrapolating tool invites to revisit the already available reported data based upon non-extrapolated partial inventories and to critically reconsider the as-derived interpretations. The purpose being to tentatively establish more relevant interpretations, based on numerically extrapolated – and thus *numerically completed* – samplings. More specifically, once properly numerically completed (and *only* when it is so [9]), the distribution of species abundances can provide synthetic data, in both qualitative and quantitative terms, about the underlying process that drives the hierarchical structuring of species abundances within community [18–22].

Hereafter, I address the true level of species richness and the internal structuration of two reef fish communities respectively associated to distinct dominant coral species exhibiting quite distinct types of morphology: *Diploria strigosa* (Dana 1846) and *Orbicella annularis* Ellis & Solander 1786. Both species are massive stony corals which, yet, strongly differ in their general aspect: *Orbicella annularis* forms large lobes with space in between each lobe that may provide structural refuges for fishes. *Diploria strigosa*, on

the contrary, grows in a very massive, regularly convex bulbous shape, as such less likely to serve as structural refuge for the most shy fish species.

In particular, I examine how this deep difference in physical traits of the surrounding habitat may - or may not - affect the internal structuration of each corresponding fish community; in terms of total species richness and intensity of the hierarchical distribution of species abundances within community. As already emphasised,

relevant comparison requires, however, to consider exhaustive inventories of each fish community.

As the available (reported) data was limited to only *partial* samplings for each of the two compared communities, these partial inventories – and the resulting species abundance distributions – had first to be completed by a procedure of *least-biased numerical extrapolation*.



Illustration of habitat "1" with dominance of *Diploria strigosa* (Dana 1846)  
© Assman & Lenoble



Illustration of habitat "2" with dominance of *Orbicella annularis* Ellis & Solander 1786  
© National Ocean A.A

## 2. MATERIALS AND METHODS

### 2.1 The Reported Field Data

The present study is based on two partial samplings of reef fish communities conducted offshore Bonaire Island in the Dutch Caribbean (12°09'36.3"N - 68°16'55.0"W) and reported by Margaret Meyer [23]. All details regarding the precise locations of compared habitats and the implemented sampling procedure are provided in the open-access reference above and need not being repeated here. The most important point is that the respective absolute abundance (number of individual occurrences) of each species has been recorded, thus making possible to implement numerical extrapolations. These extrapolations were required because the relatively high proportion (around 20%) of species recorded only once ("singletons") suggests that reported samplings remained substantially incomplete, as was confirmed later. The number  $N_0$  of collected individuals and the number  $R_0$  of recorded species in each of the two communities are given in Table 1.

### 2.2 The Numerical Extrapolation Procedure and Its Exploitation

#### 2.2.1 Implementation of the procedure of numerical extrapolation

**Total species richness:** The least-biased estimation of the number of still undetected species during partial sampling and the resulting estimation of the total species richness of the partially sampled community are derived according to the procedure defined in Béguinot [24,25] and briefly summarised in Appendix 1, on the basis of the numbers  $f_x$  of species observed  $x$ -times during partial sampling. The same procedure allows to derive the least-biased extrapolation of the "Species Accumulation Curve", which predicts the expected increase in the number of newly recorded species,  $R(N)$ , as a function of the growing sampling size  $N$  ( $N$ : number of currently recorded individuals); see Appendix 1 for computation. In practice, this extrapolation allows to *forecast* the additional sampling efforts that would be required to obtain any desirable increment in sampling completeness.

**Species Abundance Distribution:** As mentioned above, the Species Abundance Distribution ("S.A.D.") is intended to provide the basic data necessary (i) to describe the *pattern*

of structuration of species abundances within community and (ii) to qualify and quantify the underlying *process* that drives this structuration. Yet, to accurately exploit its full potential [26,27], the "S.A.D." requires:

- first, to be *corrected* for the 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 partial "S.A.D." is described in details in reference [27] and briefly summarised in Appendix 2. Also, a concrete example of implementation of the procedure is commented in details in reference [28].

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, ordered by decreasing values of abundance (with, thus,  $a_1$  and  $a_{St}$  respectively standing for the highest and the lowest abundances in a community of  $S_t$  species).

#### 2.2.2 Species abundance structuration: The apparent pattern (abundance unevenness)

The "S.A.D." (either exhaustive or, else, completed by numerical extrapolation) conveys all the relevant quantitative data required to address the internal organisation among member-species within a local community, especially the hierarchical structuration of species abundances. In particular, it is always advisable to use such species-abundance plots to quantify the degree of evenness or, more suggestively, the degree of *unevenness* of species abundances [29]. Indeed, following [30], the degree of unevenness – rather than evenness itself – should be the preferred way to address the hierarchical structuring of species abundances in communities. Optionally, 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 by Grzès [31], that is:

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

### 2.2.3 Species abundance structuration: The underlying process (mechanism and intensity)

Beyond the mere description of the pattern of hierarchical structuration, quantified by the degree of unevenness  $U$ , the complete “S.A.D.” can help addressing several important questions regarding (i) the kind of *mechanism* involved in the process driving the hierarchical structuration of abundances and (ii) the genuine *intensity* of this structuring process.

As regards the kind of mechanism involved, it is appropriate to distinguish between two major alternative hypotheses: schematically, the hierarchical structuration of abundances may result either (i) from the major contribution of *one dominant* factor or (ii) from the combined contributions of *many mutually independent factors* acting together. This 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 [18,32–35].

Now, as regards the genuine *intensity* of the structuring process, it is first necessary to remind that the degree of unevenness  $U$  does not univocally mirror the intensity of the structuring process, since unevenness  $U$  is *also mathematically dependent* negatively upon the species richness  $S_t$  [36,37]; see also Appendix 3. This double dependence of  $U$  means that the unevenness pattern  $U$  cannot be considered as reliably reflecting the genuine – i.e. biologically significant – intensity of the structuring process itself [20,21,38]. To get rid of this mathematical influence of species richness and, thus highlight the genuine intensity of the structuring process, it has been shown appropriate [39] to cancel this mathematical influence by comparing the slope of the actual “S.A.D.” to the slope of the corresponding “broken-stick” distribution [40], computed for the same species richness  $S_t$  [41-43]. Accordingly, the genuine intensity, “ $I_{str}$ ”, of the hierarchical structuring process is relevantly

defined by standardising the degree of unevenness  $U$  of the “S.A.D.” to the degree of unevenness  $U'$  of the corresponding “broken-stick” distribution, computed for the same species richness  $S_t$ , that is:

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

that is:

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

with  $a_1$  and  $a_{St}$  standing for the highest and the lowest abundances in the studied community and  $a'_1$  and  $a'_{St}$  standing for the highest and the lowest abundances in the corresponding “broken-stick” distribution (the abundances being classically log-transformed).

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

## 3. RESULTS

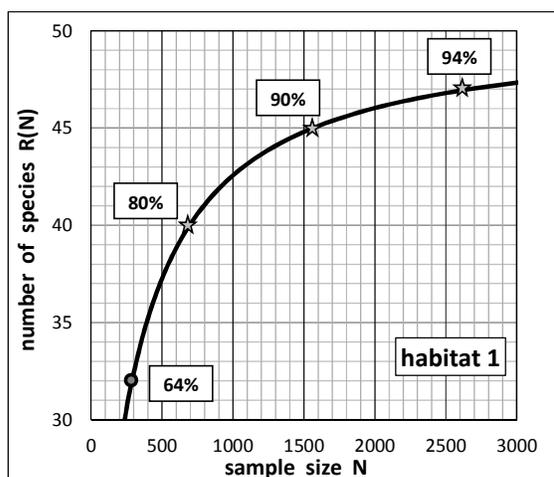
### 3.1 Estimated Total Species Richness of Each Community

The *recorded* species richness of both fish communities, as reported in Meyer [23], are very similar: 32 and 33 species (Table 1). These recorded values do not presume, however, what is the true species richness of each community, since with  $\approx 20\%$  singletons, both inventories are presumably substantially incomplete. In fact, numerical extrapolation predicts appreciably different true (total) species richness between coral habitat “1” (*Diploria*) and coral habitat “2” (*Orbicella*): 50 species and 43.5 species respectively (Table 1).

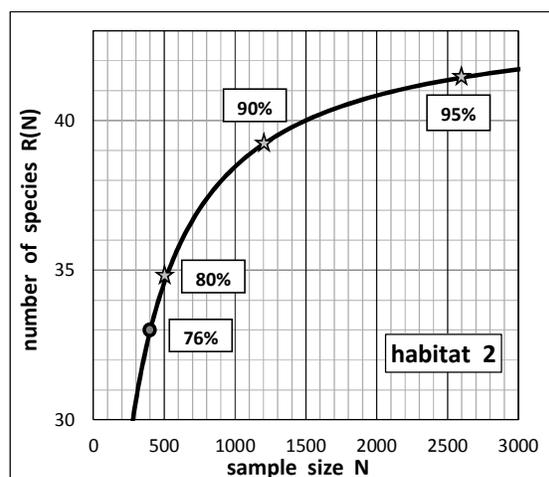
Now, with sampling completeness levels around two third and three quarter respectively (Table 1), one might consider of interest to improve completeness by further pursuing samplings. Yet, to make a rationally based decision as to whether it would be reasonable or not to pursue sampling any further, the extrapolation of the species accumulation curve beyond the actual sampling size ought to be considered. This extrapolation allows to predict the additional sampling effort required to obtain any desired increment in sampling completeness. Figs. 1 and 2 provide the extrapolated part of the species accumulation curve for the inventories of fish communities in habitats “1” and “2”.

**Table 1.** The number of collected individuals  $N_0$ , the number of recorded species  $R_0$ , the type of nonparametric estimator (Jackknife) selected as being the least-biased one, the estimated number  $\Delta$  of unrecorded species, the resulting estimate of the “true” total species richness  $S_t$  ( $= R_0 + \Delta$ ), the resulting estimated level of sampling completeness  $R_0/S_t$ . Estimations are computed according to the least-biased procedure [24]: selection key provided in Appendix 1

<b>Coral habitat</b>	<b>1 - <i>Diploria</i></b>	<b>2 - <i>Orbicella</i></b>
nb. collected individuals $N_0$	288	402
nb. recorded species $R_0 = R(N_0)$	32	33
selected least-biased estimator	JK-5	JK-5
number unrecorded species $\Delta$	18	10.5
<b>total species richness <math>S_t</math></b>	<b>50</b>	<b>43.5</b>
sample completeness $R_0/S_t$	64%	76%



**Fig. 1.** Extrapolated part of the Species Accumulation Curve for the fish community of coral habitat “1”. This numerical extrapolation highlights the increase in the number of detected species  $R(N)$  as a function of growing sample size  $N$ , beyond the actually achieved sampling ( $N_0 = 288$ ,  $R_0 = R(N_0) = 32$ , sampling completeness 64%). The expected additional sampling effort required to reach higher levels of sampling completeness, say for example 80%, 90% and 94% completeness, are around  $N = 700$ ,  $1550$ ,  $2600$  respectively



**Fig. 2.** Extrapolated part of the Species Accumulation Curve for the fish community of coral habitat “2”. This numerical extrapolation highlights the increase in the number of detected species  $R(N)$  as a function of growing sample size  $N$ , beyond the actually achieved sampling ( $N_0 = 402$ ,  $R_0 = R(N_0) = 33$ , sampling completeness 76%). The expected additional sampling effort required to reach higher levels of sampling completeness, say for example 80%, 90% and 95% completeness, are around  $N = 500$ ,  $1200$ ,  $2600$  respectively

### 3.2 Degree of Taxonomic Dissimilarity between the Two Fish Communities

According to the recorded data reported by Meyer [23], the fish communities in coral habitats “1” and “2” have respectively 32 and 33 detected species, among which 25 are shared in common. Thus, the recorded value of the Jaccard similarity index is  $J = 0.63$  (i.e.  $25/(32 + 33 - 25)$ ). Yet, due to the levels of sampling incompleteness highlighted above, this inference remains quite hypothetical, since either all, part or none of the undetected species in each community may be shared by the other community.

Fortunately, although the taxonomic identities of the undetected species remain unknown, the numerical extrapolation can partially clear up the issue. Among the 18 undetected species in community “1”, from none of them to all of them may possibly be shared with community “2”. Thus, the total number of shared species between both communities – including the 25 shared species that were already detected – is comprised between 25 and 43 (= 25+18) species. The extrapolated value of the Jaccard similarity index is therefore comprised between  $J = 0.36$  (that is  $25/(50+43.5-25)$ ) and  $J = 0.85$

(that is  $43/(50+43.5-43)$ ). Accordingly, in place of the hypothetical value derived above from incomplete inventories, it is now possible to securely infer that the Jaccard similarity index is in any case less than 0.85 and possibly as low as 0.36. Thus, taxonomic differences actually exist.

In another respect, and based on recorded species at least, it is to be noticed that in term of feeding guilds, both habitats similarly show a weak contribution of herbivores as compared to the contribution of omnivores and carnivores:  $\approx$  20% only in species richness and no more than 2% to 4% only in number of individuals.

### 3.3 Three Significantly Discriminant Species

In addition to taxonomic distinctiveness in term of the presence/absence of species according to

the habitat type, three species – although co-occurring in both habitats – yet show highly significant differences in terms of preference for one or the other type of coral habitat.

Bicolor Damsel, *Stegastes partitus* (Poey 1868), shows a very strong preference for habitat “2” (coral habitat *Orbicella annularis*):  $\chi^2$  with Yates correction = 8.5,  $p < 0.005$ .

Conversely, two other species, Bluehead Wrasse, *Thalassoma bifasciatum* (Bloch 1791), and Secretary Blenny, *Acanthemblemaria maria* J.E. Böhlke 1961, show a marked preference for habitat “1” (coral habitat *Diploria strigosa*):  $\chi^2$  with Yates correction = 7.3,  $p < 0.01$  and  $\chi^2$  with Yates correction = 5.4,  $p < 0.02$ , respectively.



**Bicolor Damsel in front of the coral *Orbicella annularis***  
*Stegastes partitus* (Poey 1868)  
© Dino Kanlic



**Bicolor Damsel *Stegastes partitus* (Poey 1868)**  
© F. Herranz Martin



**Bluehead Wrasse *Thalassoma bifasciatum***  
(Bloch 1791) © James St John



**Secretary Blenny *Acanthemblemaria maria***  
J.E. Böhlke 1961 © D. Ross Robertson

### 3.4 Completed Species Abundance Distribution for Each Community

Figs. 3 and 4, provide the graphical representations of the Species Abundance Distribution (“S.A.D.”) for each of the two communities under study. These “S.A.D.s” have been (i) bias-corrected and (ii) numerically extrapolated according to the procedure reported in Béguinot [27]: see Appendix 2, equations (A2.1) and (A2.2) respectively.

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

- along the abscissa, the member-species are ranked by decreasing values of their respective abundances (relative abundance ‘ $a_i$ ’ for the species of rank ‘ $i$ ’);
- along the ordinate, the relative abundances are plotted according to a logarithmic scale ( $\log_{10}$ ) (although, for specific purpose, an ordinary untransformed scale may be adopted, as suggested by MacArthur [40]).

The abundances of the *recorded* species are plotted as grey circles, while the *extrapolated* part of the abundance distribution is plotted as a thick double line.

### 3.5 Testing for the Type of Process Involved in the Structuring of Species Abundances

In Figs. 5 and 6, the “S.A.D.” of each community is compared to the corresponding “*log-series*” and “*log-normal*” models, computed for the same species richness. As mentioned in Methods section, these comparisons allow to infer which kind of process is more likely involved in the hierarchical structuration of species abundances in each community. Here, for both communities, the species abundance distributions clearly fit the “*log-normal*” model best than the “*log-series*” model.

### 3.6 Beyond the Apparent Unevenness of Species Abundances: The Genuine Intensity of the Hierarchical Structuring Process

According to the definitions provided in Methods section:

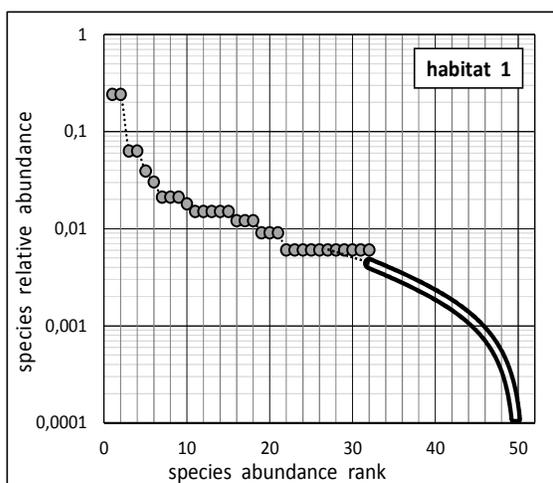
- the degree,  $U$ , of unevenness of species abundance distribution (resulting from the

hierarchical structuring process – but also influenced mathematically by the level of species richness) is computed according to equation (1);

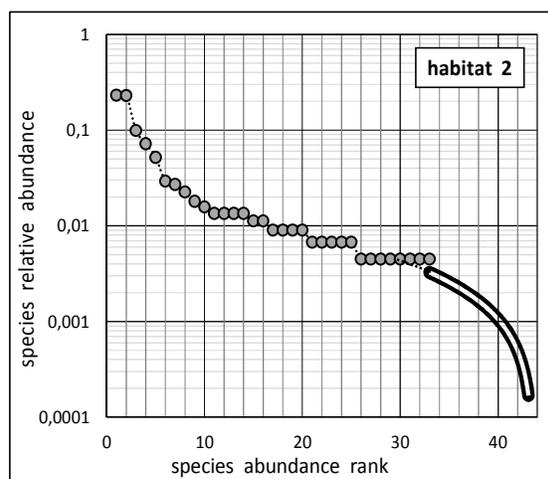
- the genuine intensity,  $I_{str}$ , of the process driving the hierarchical structuration of abundances is relevantly appreciated by comparing the “S.A.D.” of the studied community to the corresponding “*broken-stick*” model, computed for the same species richness. Figs. 7 to 10 provide these comparisons, from which the genuine intensity  $I_{str}$  of the structuring process is derived according to equation (2).

These results are summarised synthetically in Table 2 which highlights the variations of (i) the true total species richness  $S_t$ , (ii) the ratio  $a_1/a_{S_t}$  between the abundances of the commonest and the rarest species, (iii) the degrees of unevenness  $U$  and  $U'$  of species abundances for the actual SAD and the corresponding “*broken-stick*” model and, finally, (iv) the genuine intensity  $I_{str}$  of the process driving the hierarchical structuration of species abundances. The parameters  $U$  and  $I_{str}$  respectively quantify (i) the *apparent pattern* of species abundance structuration and (ii) the *true intensity* of the underlying *process* driving this structuration.

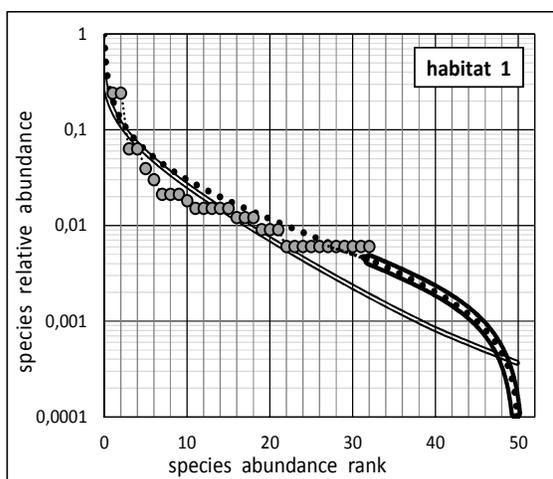
The complete species abundance distributions of both fish communities (including the distribution of abundance of the undetected species derived by numerical extrapolation) are plotted together in Fig. 11 to allow an easy direct comparison of the respective influences of the two different coral habitats on the associated fish communities. The main trends derived from this comparison are graphically highlighted in Fig. 12, where both the apparent unevenness  $U$  and the genuine intensity of the structuring process  $I_{str}$  are plotted together against the species richness  $S_t$ . While the intensity  $I_{str}$  of the structuring process slightly increases along the range of variation of species richness  $S_t$ , the degree of unevenness  $U$ , on the contrary, strongly decreases with  $S_t$ . This discrepancy between the *process* (quantified by its intensity  $I_{str}$ ) and the corresponding *pattern* (quantified by the degree of unevenness  $U$ ) might seem paradoxical at first sight. In fact, this discrepancy is entirely due to the already underlined negative mathematical dependence of  $U$  upon  $S_t$ .



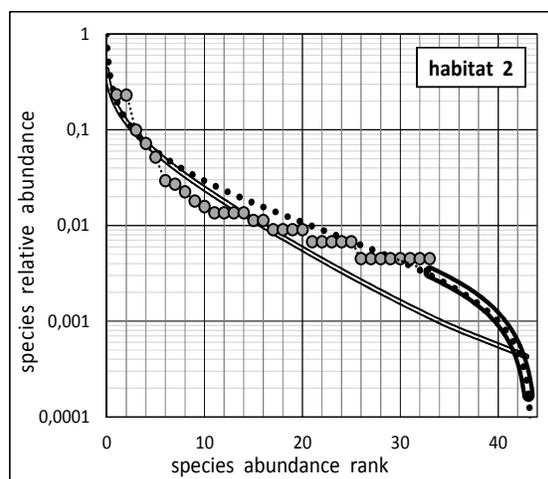
**Fig. 3.** The bias-corrected and numerically extrapolated Species Abundance Distribution (“S.A.D.”) for the fish community in coral habitat “1” (*Diploria*). Grey circles: recorded part of the “S.A.D.” after correction for bias. Coarse double line: least-biased extrapolation of the abundance distribution for the set of species remaining unrecorded (sampling completeness: 64%)



**Fig. 4.** The bias-corrected and numerically extrapolated Species Abundance Distribution (“S.A.D.”) for the fish community in coral habitat “2” (*Orbicella*). Grey circles: recorded part of the “S.A.D.” after correction for bias. Coarse double line: least-biased extrapolation of the abundance distribution for the set of species remaining unrecorded (sampling completeness: 76%)



**Fig. 5.** Two classical models: “log-normal” (dotted line) and “log-series” (double line) fitted to the Species Abundance Distribution of community in coral habitat “1” (*Diploria*), duly corrected and extrapolated. Best fit is with the “log-normal” distribution

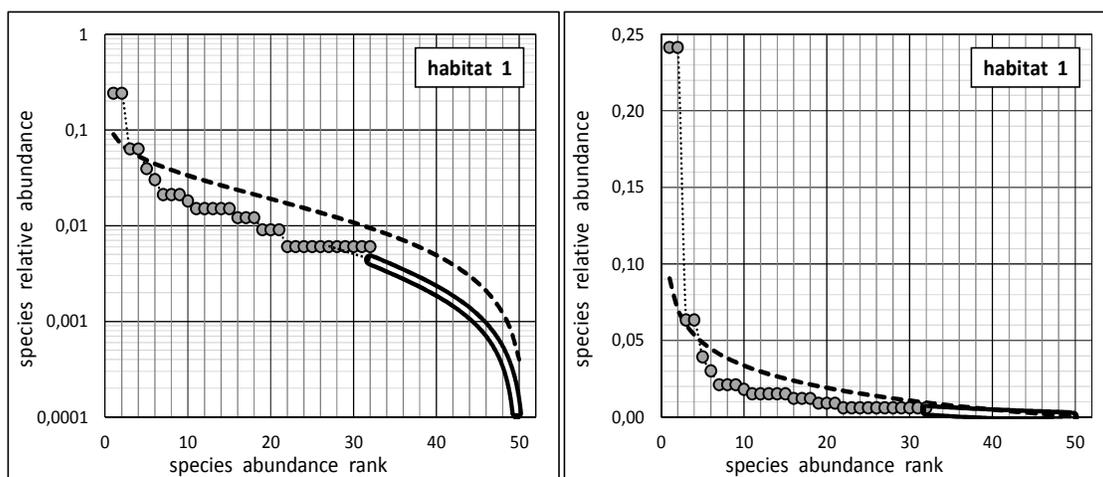


**Fig. 6.** Two classical models: “log-normal” (dotted line) and “log-series” (double line) fitted to the Species Abundance Distribution of community in coral habitat “2” (*Orbicella*), duly corrected and extrapolated. Best fit is with the “log-normal” distribution

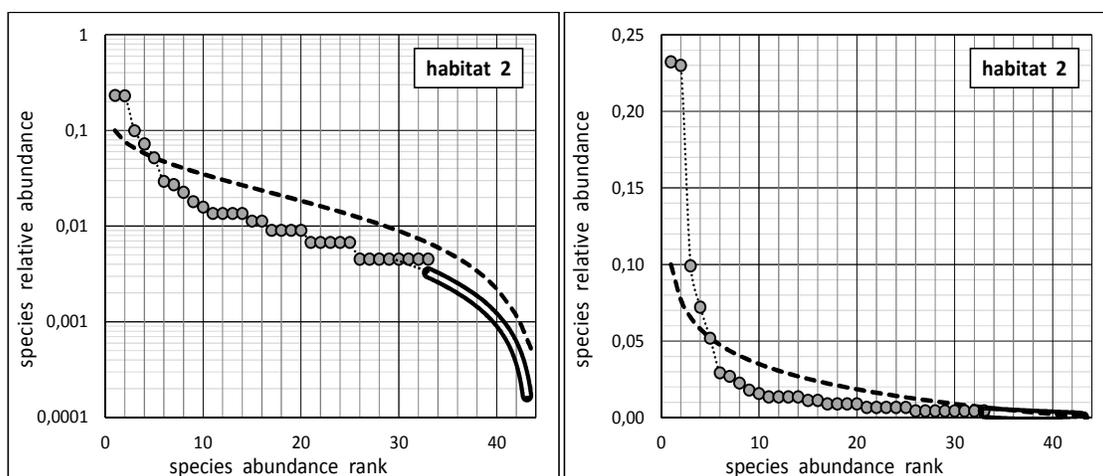
#### 4. DISCUSSION

It has been argued that fish assemblages associated to tropical coral reefs may exhibit close relationships with the particular type of

coral species which structure topographically their nearby environment. In particular, the overall “rugosity” of the coral display has been advocated as being a major determinant of these relationships [44,45].



**Figs. 7 and 8.** Comparison between the bias-corrected and numerically extrapolated “S.A.D.” for the community at habitat “1” (grey circles and coarse double line) and the corresponding “broken-stick” model (dashed line). Fig. 7: ordinate with *logarithmic* scale; Fig. 8: ordinate with *arithmetic* scale



**Figs. 9 and 10.** Comparison between the bias-corrected and numerically extrapolated “S.A.D.” for the community at habitat “2” (grey circles and coarse double line) and the corresponding “broken-stick” model (dashed line). Fig. 9. ordinate with *logarithmic* scale; Fig. 10. ordinate with *arithmetic* scale

**Table 2.** A synthetic summary of the main quantitative features of the hierarchical organisation of species abundances within community, as derived from each numerically completed “S.A.D.” : (i) the total species richness  $S_t$  of the community ; (ii) the relative abundances  $a_1$  and  $a_{S_t}$  of the most and least abundant species (species rank 1 and  $S_t$ ) ; (iii) the unevenness of abundances in the community:  $U = \log(a_1/a_{S_t})/(S_t-1)$ ; (iv) the unevenness of abundances in the corresponding “broken-stick” distribution:  $U' = \log(a'_1/a'_{S_t})/(S_t-1)$  and, at last, (v) the genuine intensity of the structuring process  $I_{str} = U/U'$

Habitat types	$S_t$	$a_1$	$a_{S_t}$	$a_1/a_{S_t}$	$U$	$U'$	$I_{str}$
habitat “1” : <i>Diploria strigosa</i>	50	0.241	0.00011	2190	0.0686	0.0482	1.42
habitat “2” : <i>Orbicella annularis</i>	43.5	0.232	0.00017	1373	0.0738	0.0536	1.38
habitat “1” / habitat “2”	1.16	1.04	0.65	1.60	0.92	0.90	1.03

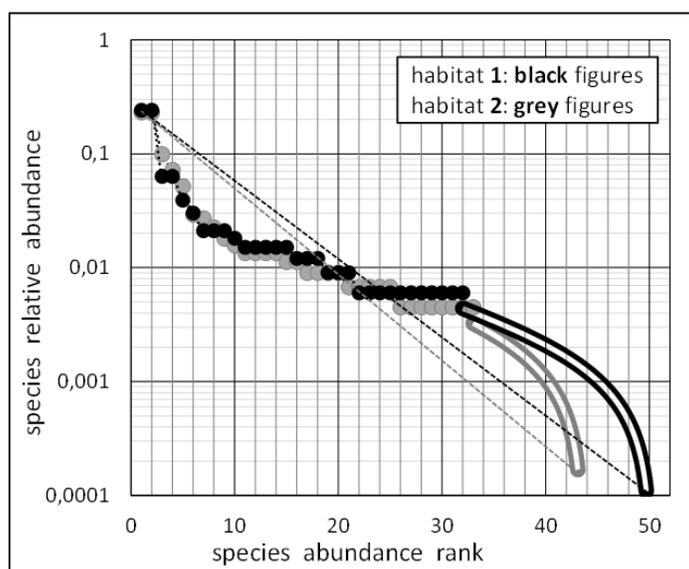


Fig. 11. The Species Abundance Distributions of both fish communities plotted together to allow direct comparison of the respective influences of the two different coral habitats: coral habitat “1” structured by *Diploria strigosa* and coral habitat “2” structured by *Orbicella annularis*. The abundance distributions have been numerically completed, thus including the undetected species as well

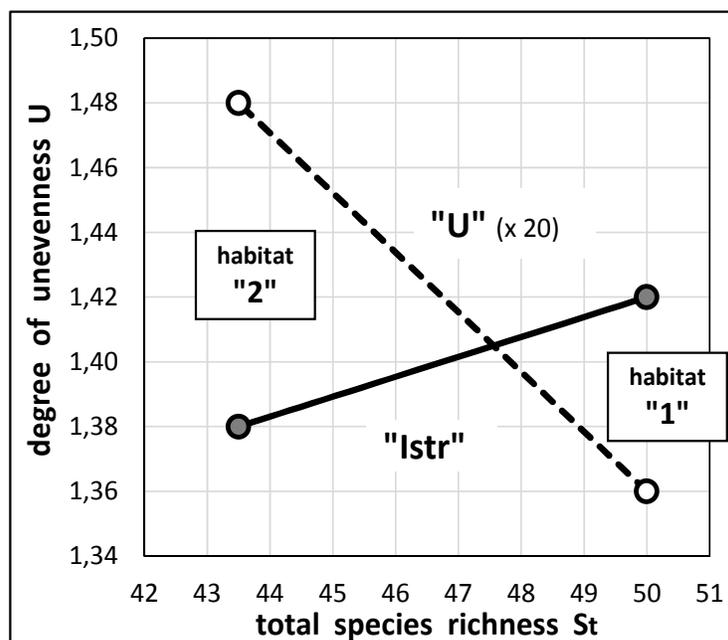


Fig. 12. The degree  $U$  of abundance unevenness (*dashed line*) and the intensity  $I_{str}$  of the underlying structuring process (*solid line*) plotted against the total species richness  $St$ , for the two studied communities. While unevenness  $U$  clearly decreases with increasing species richness, the intensity of the underlying structuring process  $I_{str}$  does not (even slightly increases): comments in text as concerns this apparent paradox. *Note that for commodity of graphical comparison between  $U$  and  $I_{str}$ , the values of  $U$  are uniformly multiplied by a same factor 20*

Thus, while *Diploria strigosa* corals have a smooth convex shape, *Orbicella annularis* corals, on the opposite, display a very irregular and tormented general shape, with many concavities of various sizes that can provide as many refuges for fishes. Accordingly, by comparing the fish communities respectively associated to two morphologically very distinct coral species – *Diploria strigosa* and *Orbicella annularis* – Margaret Meyer [23] rightly expected substantial differences between both fish communities.

Yet, as acknowledged by Margaret Meyer, her reported observations do not comply with this expectation [23]; in particular the recorded species richness of both fish communities reveal being virtually the same. This accordingly, calls for more thorough analysis.

Indeed, one possible pitfall encountered in the interpretation of the reported investigations might be the substantial levels of incompleteness of the samplings of both fish communities. As already mentioned, such incompleteness is almost inevitable because fish communities associated to tropical coral reefs usually include very many species, moreover having strongly uneven abundance distributions, with many rare species, difficult to be detected after reasonably limited sampling efforts. To make this practical difficulty more concrete, the numerical extrapolation of the species accumulation curve crudely highlights the “exponentially” growing efforts that would be required to further improve sampling completeness (Figs. 1 and 2), beyond what has already been carried out by Margaret Meyer.

Fortunately, a procedure has been recently derived, which allow to extrapolate numerically the number of unrecorded species [24,25] and still further, their respective abundances, so that the full-range of the Species Abundance Distribution can be estimated, including the whole set of undetected species as well [27,41]. This new procedure can therefore efficiently help to deal with the preceding issue and, consequently, and, consequently, has been implemented to go further in the preceding analysis (Figs. 3 and 4). In particular, three major traits of the internal organisation within community are highlighted this way: the true species richness  $S_t$ , the unevenness pattern  $U$  and the genuine intensity  $I_{str}$  of the hierarchical structuration of species abundances. This is of major importance, as these traits are likely to be *strongly biased* when inferred from incomplete sampling only, as indeed is clearly confirmed by

the present work and other recent studies as well [28,42,43].

#### **4.1 The Effect of Habitat Type on the True Species Richness of Fish Communities**

Due to the more complex topography of coral structure in habitat “2” (dominated by the coral *Orbicella annularis*), a higher species richness of the associated fish community was anticipated, since the tormented topography of this coral species is expected to offer far more refuges for fishes. Yet, the *recorded* data does not comply with this expectation, since species richness revealed similar for both habitats (33 and 32 species). However, completing the abundance distributions by numerical extrapolation does correct this unreliable interpretation based on incomplete data and demonstrates clearly different levels of species richness between both fish communities. But, rather unexpectedly, it is the community associated with habitat “1” (*Diploria strigosa* dominant coral), characterised by smoother and regular overall topography, which is favored, with an estimated total species richness of 50 species, as compared to the community associated with habitat “2” (*Orbicella annularis*), including only 43.5 species (Table 1).

The unexpected higher attractiveness of habitat “1”, while confirming the discriminating role of coral environment on the species richness of associated fish communities, therefore no longer supports the hypothesised cause initially put forward to explain such differential attractiveness of various coral environments. Here, the relevant cause thus remained to be discovered.

#### **4.2 The Effect of Habitat type on the Degree of Unevenness of Species Abundances and the Intensity of the Underlying Structuring Process**

The *origin* and the *intensity* of the *process* of hierarchical structuration of species abundances, as well as the resulting *unevenness pattern* of abundance distribution, constitute other major factors highlighting the internal organisation of species communities.

##### **4.2.1 The origin of the hierarchical structuration of species abundances**

The shape of the Species Abundance Distribution informs on whether the process of

hierarchical structuration of species abundances is driven by one major dominant factor or, on the contrary, is ruled by the multiplicative contributions of many mutually independent factors. Here, the sigmoidal shape of the abundance distributions in both studied communities conforms fairly well to the *log-normal* model and quite poorly fits the *log-series* model (Figs. 5 and 6). This clearly points in favor of the second mechanism i.e. the multifactorial driving of species abundance structuration. This, indeed, was expected since a fairly good fit with the *log-normal* model seems to be common in most species-rich communities, at least when they are not subjected to harsh environmental stresses (pollutions, etc ...) [32,34,35].

#### 4.2.2 The intensity of the process of hierarchical structuring of species abundances and the apparent degree of unevenness

The genuine intensity  $I_{str}$  of the structuring process appears quasi insensitive to the difference in coral environment between habitats "1" and "2" (Figs. 7 to 10; Table 2). Accordingly, the stronger unevenness of species abundances in the community associated to habitat "2", as compared to habitat "1" ( $U = 0.074$  and  $0.069$  respectively, Table 2; Figs. 11 and 12) does not result from a stronger intensity of the structuring process itself. In fact, the difference of unevenness of abundances between both communities is simply the banal consequence of the negative mathematical influence of species richness upon abundance unevenness [28, 41, 42, 43, see also Appendix 3] – once more highlighted here. Thus, this difference of unevenness  $U$  has no biological origin.

Fig. 11 summarises the main quantitative differences between the internal structuration of both communities in relation with habitat type. Highlighted is the fact that:

- differences are slight for the subset of *most* abundant species, in particular those that were actually recorded: this explains, in turn, the erroneous conclusions reported in [23] due to their being based on incomplete samplings;
- differences are more conspicuous within the subset of *less* abundant species. And as most of the latter were remaining undetected by incomplete samplings, only numerical extrapolation are actually able to unveil these differences.

### 4.3 The Effect of Habitat Type on the Dissimilarities in Taxonomic Composition of Associated Fish Communities

Apart from the marked difference in true (total) species richness, the fish communities in habitat "1" and "2" also differ from each-other as regard their taxonomic compositions (sections 3.2 & 3.3). This difference, already suggested by recorded data, yet remained hypothetical until species abundance distributions were numerically completed. Numerical extrapolations now allow to conclude that taxonomic difference does exist, with a Jaccard similarity index between both fish communities being less than 0.85 and probably much lower.

In addition to taxonomic distinctiveness, three species – although co-occurring in both habitats – yet show highly significant differences in terms of preference for one or the other habitat (section 3.3). Namely, Bicolor Damsel (*Stegastes partitus*) shows a strong preference towards habitat "2" (coral habitat *Orbicella annularis*) while, on the opposite, Secretary Blenny (*Acanthemblemaria maria*) and Bluehead Wrasse (*Thalassoma bifasciatum*) show a marked preference towards habitat "1" (coral habitat *Diploria strigosa*).

At last, a common feature regarding feeding guilds is shared by both habitats: the associated fish communities *similarly* include a very weak contribution of herbivores, as compared to omnivores and carnivores:  $\approx 20\%$  only in species richness and still far less in number of individuals:  $\approx 2\%$  to  $4\%$  only. This, indeed, is a rather common trait to healthy coral habitats in general, where algal cover is systematically very low [6,28,42,43,46].

As a whole, the results derived above after numerical completion of the Species Abundance Distributions, tend to support the important hypothesis already put forward in reference [45], according to which most reef fish species are able to distinguish their surrounding habitat at the level of coral species, when the latter have contrasted morphologies.

## 5. CONCLUSIONS

At first, on a methodological point of view, this case study demonstrates, once again, the risk of deriving erroneous conclusions from incomplete

species inventories and, accordingly, the usefulness of numerical extrapolations in order to escape such pitfalls. This deserves all the more being emphasised that partial inventories often become quasi unavoidable when having to deal with species-rich communities, which moreover show strongly uneven distributions of species abundances, a very common situation indeed in invertebrates and even in some vertebrates groups, in particular under tropical climate.

Thus, here, the numerical extrapolation of the Species Abundance Distributions of the two fish communities reveals quite a different panorama from what was suggested by the partial data issued from the reported incomplete inventories. Contrary to the misleading inferences that could be derived from these incomplete inventories, substantial differences actually distinguish the fish communities associated with two distinct coral habitats, dominated respectively by *Diploria strigosa* and *Orbicella annularis*. Highlighted differences involve both the true (total) species richness, the taxonomic composition of these fish communities and the degree of abundance unevenness. On the other hand, the genuine intensity of the process of hierarchical structuring of species abundances proves remaining substantially unchanged. This, in turn, demonstrates that the observed difference in abundance unevenness between both communities has, in fact, no biological significance, being only the trivial mathematical consequence of the difference in the levels of true species richness between the two compared fish communities.

## ACKNOWLEDGEMENTS

The author appreciates the comments of three anonymous Reviewers and addresses special thanks to Margaret Meyer for her reported field data.

## COMPETING INTERESTS

Author has declared that no competing interests exist.

## REFERENCES

1. Wells JW. Coral reefs: 609-632; in J.W. Hedspeth editor, Treatise on marine ecology and paleoecology. Geological Society of America. 1957; Mem. 67.
2. Connell J. Diversity in tropical rain forests and coral reefs. Science. 1978;199:1302-1310.
3. Glynn PW. High complexity food webs in low-diversity eastern Pacific reef-coral communities. Ecosystems. 2004;7:358-367.
4. Graham NAJ, Wilson SK, Jennings S, Polunin NVC, Bijoux JP, Robinson J. Dynamic fragility of oceanic coral reef ecosystems. Proceedings of the National Academy of Sciences USA. 2006;103(22): 8425-8429.
5. Jeeva C, Mohan PM, Sabith DB, Ubare VV, Muruganantham M, Kumari RK. Distribution of Gastropods in the intertidal environment of south, middle and north Andaman Islands, India. Open Journal of Marine Science. 2018;8:173-195.
6. Ceccarelli DM, Jones GP, McCook LJ. Territorial damselfishes as determinant of the structure of benthic communities on coral reefs. Oceanography and Marine Biology: An Annual Review. 2001;39:355-389.
7. 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.
8. 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.
9. Connolly SR, Hughes TP, Bellwood DR. A unified model explains commonness and rarity on coral reefs. Ecology Letters. 2017;20:477-486.
10. Chen Y, Shen TJ. Rarefaction and extrapolation of species richness using an area-based Fisher's logseries. Ecology and Evolution. 2017;7:10066-10078.
11. Kery M, Royle JA. Inference about species richness and community structure using species-specific occupancy models in the National Swiss Breeding Bird survey MUB. Proceedings of the 2007 EURING Technical Meeting and Workshop, Dunedin, New Zealand; 2007.
12. Bracken M, Low N. Realistic losses of rare species disproportionately impact higher trophic levels. Ecology Letters. 2012;15: 461-467.
13. Mouillot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelin-Vivien M, Kulbicki M, Lavergne S, Lavorel S,

- Mouquet N, Paine CET, Renaud J, Thuiller W. Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol.* 2013;11(5):e1001569.
14. Jain M, Flynn DFB, Prager CM, Hart GM, DeVan CM, Ahrestani FS, Palmer MI, Bunker DE, Knops JHM, Jouseau CF, Naeem S. The importance of rare species: A trait-based assessment of rare species contribution to functional diversity and possible ecosystem function in tall-grass prairies. *Ecology and Evolution.* 2014;4(1): 104-112.
  15. Low-Decarie E, Kolber M, Homme P, Lofano A, Dumbrell A, Gonzalez A, Bell G. Community rescue in experimental communities. *Proceedings of the National Academy of Sciences USA.* 2015;112(46): 14307-14312.
  16. Leitao RP, Zuanon J, Villéger S, Williams SE, Baraloto C, Fortunel C, Mendonça FP, Mouillot D. Rare species contribute disproportionately to the functional structure of species assemblages. *Proceedings of The Royal Society B.* 2016;283:2016-0084; DOI:10.1098/rspb.2016.0084
  17. Violle C, Thuillier W, Mouquet N, Munoz F, Kraft NJB, Cadotte MW, Livingstone SW, Mouillot D. Functional rarity: The ecology of outliers. *Trends in Ecology;* 2017. DOI:10.1016/j.tree.2017.02.002
  18. May RM. Patterns of species abundance and diversity. In Cody M.L. & Diamond JM. *Ecology and Evolution of Communities.* The Belknap Press of Harvard University. 1975;81-120.
  19. McGill BJ, Etienne RS, Gray JS, et al. Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters.* 2007;10:995-1015.
  20. Ulrich W, Ollik M, Ugland KI. A meta-analysis of species-abundance distributions. *Oikos.* 2010;119:1149-1155.
  21. Komonen A, Elo M. Ecological response hides behind the species abundance distribution: community response to low-intensity disturbance in managed grasslands. *Ecology and Evolution.* 2017;7:8558-8566.
  22. Wang X, Ellwood F, Ai D, Zhang R, Wang G. Species abundance distributions as a proxy for the niche-neutrality continuum. *Journal of Plant Ecology.* 2017;rtx 013.
  23. Meyer M. Comparing the diversity, total abundance and richness of fish species associated with two stony corals: *Diploria strigosa* and *Orbicella annularis*. *Physis.* 2015;18:61-69.
  24. Béguinot J. Theoretical derivation of a bias-reduced expression for the extrapolation of the Species Accumulation Curve and the associated estimation of total species richness. *Advances in Research.* 2016;7(3):1-16. DOI:10.9734/AIR/2016/26387; <hal-01367803>
  25. Béguinot J. Extrapolation of the Species Accumulation Curve associated to “Chao” estimator of the number of unrecorded species: A mathematically consistent derivation. *Annual Research & Review in Biology.* 2016;11(4):1-19 DOI:10.9734/ARRB/2016/30522; <hal 01477263 >
  26. Chao A, Hsieh T, Chazdon RL, Colwell RK, Gotelli NJ. Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology.* 2015;96(5):1189-1201.
  27. Béguinot J. How to extrapolate species abundance distributions with minimum bias when dealing with incomplete species inventories. *Advances in Research.* 2018; 13(4):1-24. DOI:10.9734/AIR/2018/39002.
  28. Béguinot J. Numerical extrapolation of the species abundance distribution unveils the true species richness and the hierarchical structuring of a partially sampled marine gastropod community in the Andaman Islands (India). *Asian Journal of Environment and Ecology.* 2018;6(4):1–23. DOI:10.9734/AJEE/2018/41293 <hal-01807454>
  29. Heip CHR, Herman PMJ, Soetaert K. Indices of diversity and evenness. *Océanis.* 1998;24(4):61-87.
  30. Strong WL. Assessing species abundance unevenness within and between plant communities. *Community Ecology.* 2002;3(2):237-246. DOI:10.1556/ComEc.3.2002.2.9
  31. Grzès IM. Ant species richness and evenness increase along a metal pollution gradient in the Boleslaw zinc smelter area. *Pedobiologia.* 2009;53:65-73.
  32. Loreau M. Species abundance patterns and the structure of ground-beetle communities. *Ann. Zool. Fennici.* 1992;28: 49-56.

33. Magurran AE, Henderson PA. Explaining the excess of rare species in natural species abundance distributions. *Nature*. 2003;422:714-716.
34. Connolly SR, Hughes TP, Bellwood DR & Karlson RH. Community structure of corals and reef fishes at multiple scales. *Science*. 2005;309:1363-1365.
35. Ulrich W, Soliveres S, Thomas AD, Dougill AJ, Maestre FT. Environmental correlates of species rank-abundance distributions in global drylands. *Europe PMC Funders Group*. 2016;20:56-64.
36. Smith B, Wilson JB. A consumer's guide to evenness indices. *Oikos*. 1996;76:70-82.
37. Loiseau N, Gaertner JC. Indices for assessing coral reef fish biodiversity: The need for a change in habits. *Ecology and Evolution*. 2015;5(18):4018-4027.
38. MacDonald ZG, Nielsen SE, Acorn JH. Negative relationships between species richness and evenness render common diversity indices inadequate for assessing long-term trends in butterfly diversity. *Biodiversity Conservation*. 2017;26:617-629.
39. Wilson JB. Would we recognise a Broken-Stick community if we found one? *Oikos*. 1993;67(1):181-183.
40. MacArthur RH. On the relative abundance of bird species. *Proceedings of the National Academy of Sciences U.S.A.* 1957;43:293-295.
41. Béguinot J. The hierarchical structuring of species abundances within communities: disentangling the intensity of the underlying structuring process behind the apparent unevenness pattern. *Advances in Research*. 2018;16(1):1-12. DOI:10.9734/AIR/2018/43918.
42. Béguinot J. The full hierarchical structuration of species abundances reliably inferred from the numerical extrapolation of still partial samplings: A case study with marine snail communities in Mannar Gulf (India). *Asian Journal of Environment and Ecology*. 2018;7(3):1-27. DOI:109734/AJEE/2018/36831.
43. Béguinot J. Analyzing the role of environmental stresses on species richness and the process of hierarchical structuring of species abundances in marine Gastropods communities at Suva (Fiji Islands). *International Journal of Environment and Climate Change*. 2018;8(3):200-233. DOI:109734/IJECC/2018/44913.
44. Danilowicz B. Choice of coral species by naïve and field-caught Damselfish. *Copeia*. 1996;735-739.
45. Messmer V, Jones GP, Munday PL, Holbrook SJ, Schmitt RJ, Brooks AJ. Habitat biodiversity as a determinant of fish community structure on coral reefs. *Ecology*. 2011;92:2285-2298.
46. Meekan MG, Choat JH. Latitudinal variation in abundance of herbivorous fishes: A comparison of temperate and tropical reefs. *Marine Biology*. 1997;128:373-383.
47. Béguinot J. An algebraic derivation of Chao's estimator of the number of species in a community highlights the condition allowing Chao to deliver centered estimates. *ISRN Ecology*. 2014; vol. 2014 : article ID 847328. DOI:10.1155/2014/847328; <hal-01101415>
48. Béguinot J. When reasonably stop sampling? How to estimate the gain in newly recorded species according to the degree of supplementary sampling effort. *Annual Research & Review in Biology*. 2015;7(5):300-308. DOI:10.9734/ARRB/2015/18809; <hal-01228695>
49. O'Hara RB. Species richness estimators: How many species can dance on the head of a pin? *Journal of Animal Ecology*. 2005;74:375-386.
50. Gotelli NJ, Colwell RK. Estimating species richness. pp. 39-54 in: *Biological Diversity: Frontiers In Measurement And Assessment*. A.E. Magurran and B.J. McGill (eds.). 2010; Oxford University Press, Oxford. 345 pp.
51. Gotelli NJ, Chao A. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In: Levin S.A. (ed.) *Encyclopedia of Biodiversity*. 2013; second edition, volume 5, pp. 195-211. Waltham, MA: Academic Press.
52. Brose U, Martinez ND, Williams RJ. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*. 2003;84(9):2364-2377.

## APPENDIX 1

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

Consider the survey of an assemblage of species of size  $N_0$  (with sampling effort  $N_0$  typically identified either to the number of recorded individuals or to the number of sampled sites, according to the inventory being in terms of either species abundances or species incidences), including  $R(N_0)$  species among which  $f_1, f_2, f_3, f_4, f_5$ , of them are recorded 1, 2, 3, 4, 5 times respectively. The following procedure, designed to select the less-biased solution, results from a general mathematical relationship that constrains the theoretical expression of *any* theoretical Species Accumulation Curves  $R(N)$  [see [24, 47, 48]:

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

Compliance with the mathematical constraint (equation (A.1)) warrants *reduced-bias* expression for the extrapolation of the Species Accumulation Curves  $R(N)$  (i.e. for  $N > N_0$ ). Below are provided, accordingly, the polynomial solutions  $R_x(N)$  that respectively satisfy the mathematical constraint [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 [24]:

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

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

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

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

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

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

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

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

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

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

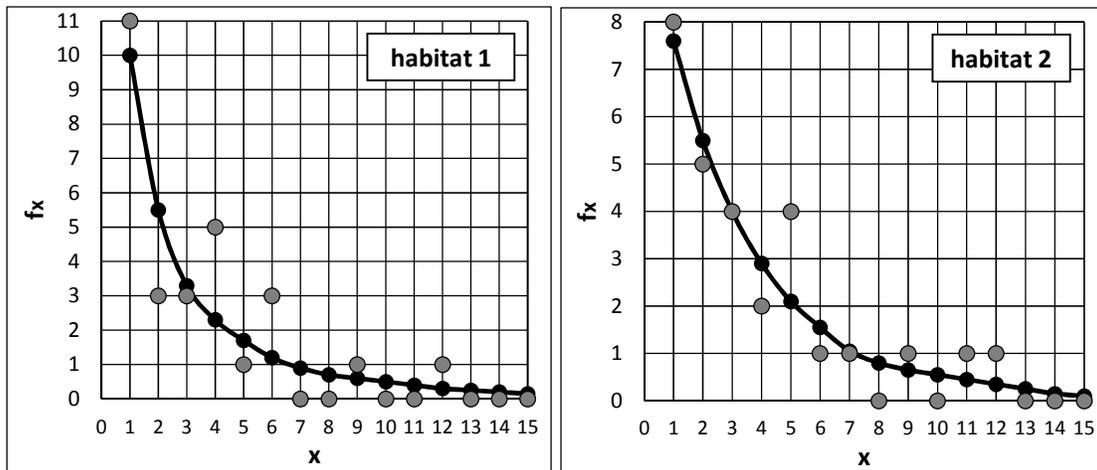
$$* f_1 > 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{j5} = 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5 ; R_5(N)$$

**N.B. 1:** As indicated above (and demonstrated in details in [24]), this series of inequalities define the ranges that are best appropriate, respectively, to the use of each of the five estimators, JK-1 to JK-5. That is the respective ranges within which each estimator will benefit of minimal bias for the predicted number of missing species.

Besides, it is easy to verify that another consequence of these preferred ranges is that the selected estimator will *always* provide the highest estimate, as compared to the other estimators. Interestingly, this mathematical consequence, of general relevance, is in line with the already admitted opinion that all non-parametric estimators provide *under*-estimates of the true number of missing species [8, 10, 49 – 51]. Also, this shows that the approach initially proposed by [52] – 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  $x$ : see Figs. A1 and A2.

**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 [25].



**Figs. A1 and A2.** 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 (when required) of the as-recorded S.A.D.**

N.B.: details regarding the derivation of the following expressions are provided in [27].

#### *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:

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

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

#### *2) Extrapolation of the recorded part of the S.A.D. accounting for the complementary abundance distribution of the set of unrecorded species*

The following expression stands for the estimated abundance,  $a_i$ , of the unrecorded species of rank i (thus for  $i > R_0$ ):

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

which, in practice, comes down to:  $a_i \approx (2/N_i) / (1 + R(N_i)/N_i)$ , as  $f_1(N)$  already becomes quite negligible as compared to  $N$  for the extrapolated part.

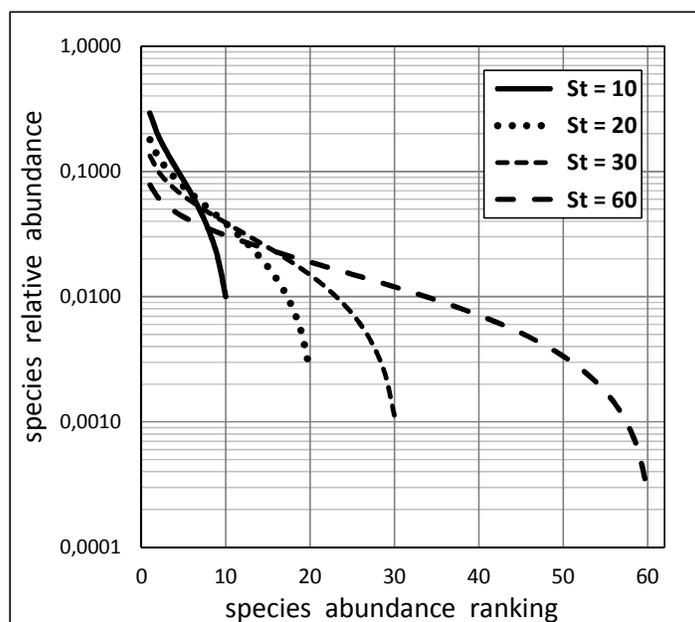
This equation provides the extrapolated distribution of the species abundances  $a_i$  (for  $i > R(N_0)$ ) as a function of the least-biased expression for the extrapolation of the species accumulation curve  $R(N)$  (for  $N > N_0$ ), 'i' being equal to  $R(N_i)$ . The key to select the least-biased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

## 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: Fig. A3.



**Fig. 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  $S_t$

© 2018 Béguinot; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Peer-review history:*

*The peer review history for this paper can be accessed here:*  
<http://www.sciencedomain.org/review-history/27497>