

Extrapolation of total species richness from incomplete inventories: application to the Gastropod fauna associated to coral reefs in 'Mannar Gulf Biosphere Reserve', India.

Original Research article

ABSTRACT

Tropical coral reefs are known to harbor considerable biodiversity, especially among invertebrates. Gastropod fauna, as an important component of this biodiversity, yet remains poorly surveyed across most tropical reefs. Moreover, the few published inventories are generally far from being exhaustive, as is almost inevitable in practice with species-rich faunas. Hence the necessity of implementing a numerical extrapolation of species accumulation, providing both (i) estimates of the total species richness of the partially sampled sites and (ii) a way to predict the additional sampling effort needed to achieve a given additional gain in sampling completeness. Such numerical extrapolation was applied here to three partial inventories of Gastropod fauna associated to coral reefs in 'Mannar Gulf Biosphere Reserve'. To ensure the maximal possible accuracy for extrapolations, a newly derived procedure was implemented, designed to select the least-biased among nonparametric estimators of the number of still unrecorded species. The total species richness of Gastropods at each site was estimated between 49 and 53 species according to sites. Accordingly, the completeness of the reported inventories was estimated between 71% and 78%. This, in turn, calls for further effective sampling but immediately raises the question of how far to extend the extra effort with, in return, a reasonable expected benefit, in terms of ratio between the expected number of newly recorded species and the corresponding additional sampling effort required. The least-biased extrapolation of species accumulation curves proves a convenient tool for rationally addressing this important question.

Keywords: marine snail, diversity, accurate estimation, species accumulation, nonparametric estimator, Tuticorin

INTRODUCTION

Marine Gastropods not only constitute a major component of marine biodiversity *per se*; they also have significant economic importance: their flesh is commercially exploited for food, their shell is used for ornamental items or, more pragmatically, for lime extraction. Also some species (in particular among Conidae) feature emerging promises in medical applications. For all these reasons, addressing Gastropods species diversity, especially now, at *local scales*, is considered important [1-4].

A major basic element to be considered, when quantifying biodiversity at a site, is the corresponding level of true species richness, that is, the total number of species that would be recorded by an ideally exhaustive inventory of the species assemblage under consideration [5-8]. In particular, the level of total species richness at a given site usually features as a positive predictor for places of higher conservation value [9]. Yet, practical evaluation of the total species richness at a site is usually problematic. This is because, in most circumstances, local samplings are doomed to remain far from being exhaustive,

48 due to limited available time devoted to each field investigation when concurrent
49 ongoing investigations are becoming increasingly numerous. Hence, the practical
50 necessity to deal with “quick surveys” of biodiversity in the urgent context of ecological
51 monitoring and conservation planning with, consequently, limited levels of sampling
52 completeness [10]. The trend for incomplete samplings is even more pronounced, as
53 expected, when dealing with invertebrates groups comprising very large numbers of
54 species including a significant proportion of more or less rare taxa.

55 In practice, incomplete inventories can, however, benefit from a kind of compensation,
56 in term of a “numerical extrapolation” of the species accumulation curve towards ideal
57 exhaustivity. As a result, estimates of the number of still unrecorded species and, in turn,
58 the evaluation of true (total) species richness are derived. This is usually achieved by
59 implementing one or the other among a series of nonparametric estimators [11-12].

60 Regrettably, however, many published accounts of local biodiversity still provide only
61 the *as-recorded* data issued from incomplete inventories, without further estimations of
62 the number of unrecorded species and, accordingly, no evaluation of the total species
63 richness of the sampled species assemblages is made possible. This restriction, which
64 deprives inventories of an essential piece of information, likely results from the former
65 confusing situation arising from the multiplicity of different kinds of nonparametric
66 estimators, that provide divergent estimations. Yet, this confusing situation is now over,
67 as a new procedure is now available, which makes it possible to select rationally the
68 least-biased type among all these estimators.

69 Hereafter, we have implemented this new procedure to estimate, as accurately as
70 possible, the total species richness of Gastropod fauna associated to coral reefs at three
71 sites in Mannar Gulf Biosphere Reserve (south India), based on recorded data reported
72 by MOHANRAJ *et al.* [1]. These authors carried out samplings along coral reefs around
73 three small islands in Mannar Gulf, opportunely taking care of noting the respective
74 abundance of each sampled species, which is necessary to compute nonparametric
75 estimators. The proportions of singletons (species recorded only once) amount around
76 20% at each investigated site, thus indicating substantially incomplete samplings [12],
77 which justifies the implementation of a numerical extrapolation procedure.

78 79 MATERIAL & METHOD

80 *Field data*

81 The coral reefs at short distance from the shores of three small islands, located in
82 “Mannar Gulf Biosphere Reserve”, namely ‘Hare’, ‘Vaan’ and ‘Koswari’, were sampled for
83 associated gastropods fauna. The corresponding recorded data issued from these partial
84 surveys (on which numerical extrapolations will be based) is reported in details by
85 MOHANRAJ *et al.* [1] and the reader is invited to browse this publication for contextual
86 information regarding the sites and the implemented sampling methods. The number N_0
87 of collected individuals, the number R_0 of recorded species and the number f_1 of
88 singletons (species recorded only once) are given at Table 1 for each investigated site.
89 As already mentioned, the proportions of singletons (f_1/R_0) recorded at each site are
90 closed to 20%, which denotes *substantial incompleteness* of each of the three samplings,
91 a common situation in practice, as already underlined. The species lists reported by the
92 authors include the respective abundances of the recorded species, which makes
93 possible to implement the extrapolation of species accumulation and to derive least-
94 biased estimates of total species richness.

Procedure of selection of the least-biased estimator

A series of nonparametric estimators of the number of still unrecorded species at the end of partial inventories have been proposed, all of them based on the numbers f_x of those species already recorded x -times (namely, the numbers $f_1, f_2, f_3, \dots$ of singletons, doubletons, tripletons, etc...). The more commonly used types of nonparametric estimators are *Chao's* estimator and the series of *Jackknife* estimators at increasing orders (JK-1, JK-2, JK-3, etc...). A serious problem arises however in practice: each of these different types of estimators - being formulated differently - provides a substantially distinct estimate. No consensus had never been obtained as to which of these estimators would be the more accurate [13-15]. Hence, the traditional practice has become to consider together all of them without making any choice [16], an admittedly rather frustrating situation! This unsatisfactory situation has probably largely contributed to still a certain reserve regarding the use of these estimators.

Yet, more recently, BROSE *et al.* [13, 17] rightly suggested that, although none of the available estimators may consistently remain the more accurate, each of them may prove, *in turn*, being the more appropriate. They further argued that the criteria to select among these estimators might be related to the estimated degree of sampling completeness. Yet, as emphasised by these authors themselves, this quantitative relationship holds only under the explicit restriction of a given theoretical type of species abundance distribution or, at least, under the condition of a given degree of unevenness of the species abundance distribution in the sampled assemblage of species under consideration [18]. Due to this explicit restriction, the (partially distinct) keys successively proposed by BROSE *et al.* [13, 17] cannot benefit from general reliability and undifferentiated applicability, which severely limits their range of practical use. This limitations applies, as well, to other subsequently reported procedures still subordinated to particular types of species abundance distributions, for example: [18, 19]. Nevertheless, this explicit restriction of applicability appears to have been often overlooked by end-users of these procedures.

In fact, it can be demonstrated (see Appendix 1) that, without any particular restriction, the least-biased one, within the set of available nonparametric estimators is, simply, this estimator which provides *the highest estimate*, as compared to the others nonparametric estimators. This, indeed, is the straightforward consequence of the fact that all nonparametric estimators admittedly provide *under-estimates* of the number of unrecorded species, [11, 12, 18, 20]. So that it is the estimator which provide *the highest estimate* which is necessarily the *least-biased one*, among them all.

Selecting this way the least-biased type of estimator thereby provides the best available evaluation of the number Δ of still unrecorded species and, in turn, the best evaluation of the total species richness of the partially sampled assemblage, $S_t (= R_0 + \Delta)$. In addition, the least-biased expression for the *extrapolation* of the species accumulation curve is straightforwardly derived (Appendix 2).

RESULTS

Estimation of the total species richness at each site

The least-biased type of estimator, selected according to the key in Appendix 1, was JK-5 for both Hare and Vaan islands and JK-4 for Koswari island (Figure 1). The resulting least-biased estimates of (i) the number Δ of still unrecorded species, (ii) the total species richness S_t and (iii) the sampling completeness R_0/S_t are provided at Table 1 for each of three locations.

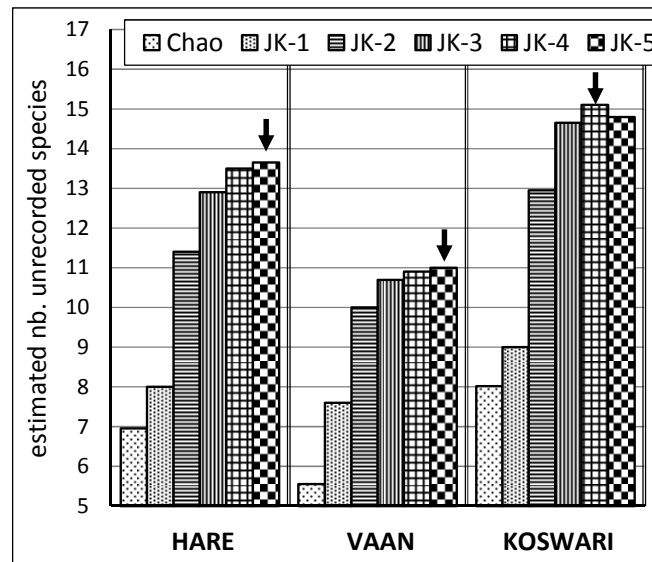


Figure 1 – Estimated numbers Δ of still unrecorded species according to six nonparametric estimators (Chao and the five first Jackknifes at order 1 to 5) for each of the three sites, Hare, Vaan, Koswari islands. The highest estimates are provided by JK-5 for Hare and Vaan and by JK-4 for Koswari, which defines accordingly the least-biased estimators for each site.

Table 1 – The number of collected individuals N_0 , the number of recorded species R_0 , the selected least-biased type of nonparametric estimator, the estimated number Δ of unrecorded species, the resulting estimate of the “true” total species richness S_t , the resulting estimated level of sampling completeness R_0/S_t . Estimations according to the least-biased procedure: selection key in Appendix 1.

	Hare	Vaan	Koswari
nb. collected individuals N_0	310	484	416
nb. recorded species R_0	35	40	38
least-biased estimator	JK-5	JK-5	JK-4
nb. unrecorded species Δ	13.7	11.0	15.1
total species richness S_t	49	51	53
sample completeness R_0/S_t	71%	78%	72%

Table 2 and Figure 2 allow to compare the least-biased estimates of the numbers of unrecorded species implemented here to the corresponding estimations according to the procedure proposed by BROSE *et al.* [13]. The discrepancy between both methods is substantial, with gaps ranging from 45% to 71%.

Table 2 – Comparisons between the estimated number of unrecorded species according to the procedure by BROSE *et al.* [13] and the procedure implemented here.

		Hare	Vaan	Koswari
selection according to BROSE <i>et al.</i> [13]	selected estimator	JK-1	JK-1	JK-1
	nb. unrecorded species Δ_{Brose}	8.0	7.6	9.0
selection key implemented here	selected estimator	JK-5	JK-5	JK-4
	nb. unrecorded species Δ	13.7	11.0	15.1
relative gap (%)	$(\Delta - \Delta_{\text{Brose}})/\Delta_{\text{Brose}}$	71%	45%	68%

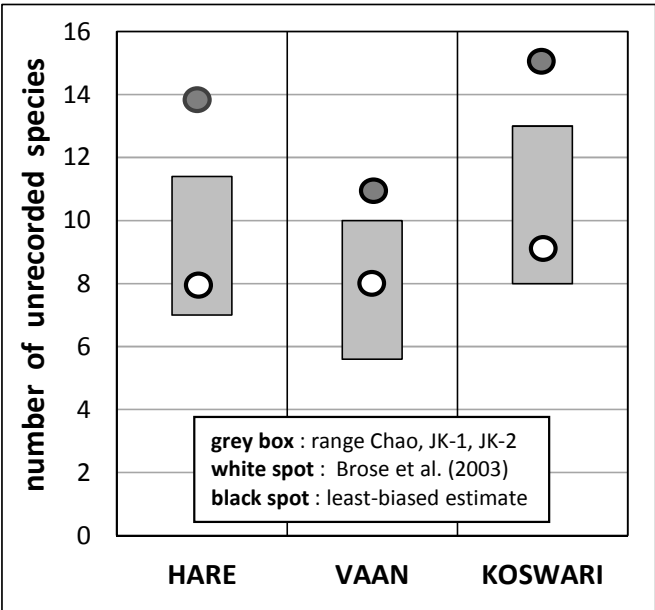


Figure 2 – Estimations of the number Δ of still unrecorded species according to (i) the still common practice consisting in considering indistinctly the full range of values taken by Chao, Jack-1 and Jack-2 [16]: *light grey boxes*; (ii) the selection keys proposed by BROSE *et al.* [13]: *white points* ; (iii) the selection key in favour of the least-biased type of estimator, among Chao and the five first Jackknifes, as implemented here: *grey points*.

Evaluations of the additional sampling effort that would be required to improve sampling completeness at each site

For each sampled site, Figures 3 and 4 provide the least-biased extrapolations of the Species Accumulation Curves, in order to show how the number $R(N)$ of recorded species increases with growing sampling-size N , beyond the actually achieved sampling-size. Computation conducted according to Appendix 2. For example, Figure 3 predicts the sampling efforts which would be necessary to reach, say, 90% or 95% sampling completeness for Hare, Vaan and Koswari islands (instead of the actually achieved levels 71%, 78%, 72%, respectively, Table 1). Clearly the additional sampling effort required to detect one new species is accelerating very quickly as sampling is going on further, which is logically expected since (i) less and less species remain still unrecorded as sampling is progressing and (ii) those species still remaining unrecorded are – statistically – among the less abundant. Besides, Figure 4 – a zoom of Figure 3 focused on the beginning of extrapolations – shows that the slopes of the species accumulation curves may notably differ according to

sites, so that the curves may *intersect*. This is the case, here, for the accumulation curve at Koswari reef, which increases more rapidly than the two others and thus came to intersect the accumulation curve of Vaan at a sample-size around $N = 800$. Accordingly, the recorded richness of Koswari, which was less than that of Vaan as long as sampling-size N remains less than 800, then exceeds that of Vaan beyond $N = 800$.

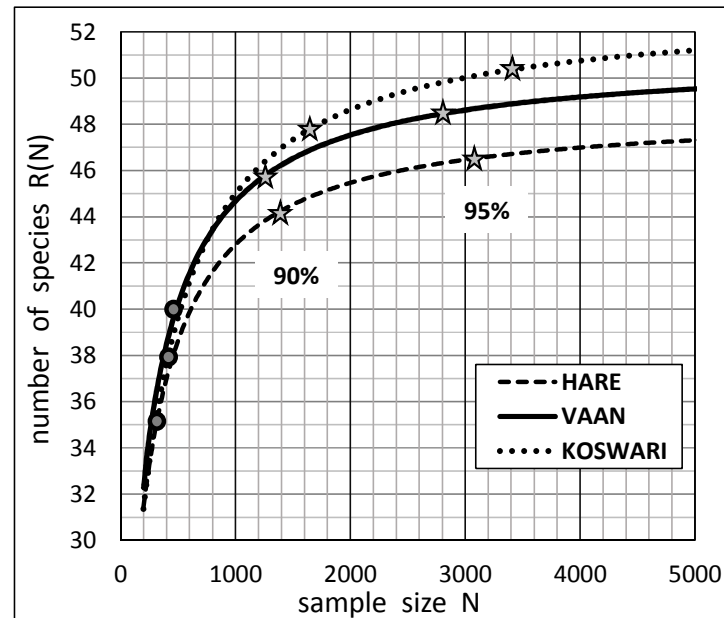


Figure 3 – Least-biased extrapolations of the Species Accumulation Curves beyond the actually reached numbers of recorded species (the latter marked by grey points). The sampling-efforts (sampling-size N) predicted to reach 90% or 95% sampling completeness are represented by grey stars for each of the three sites.

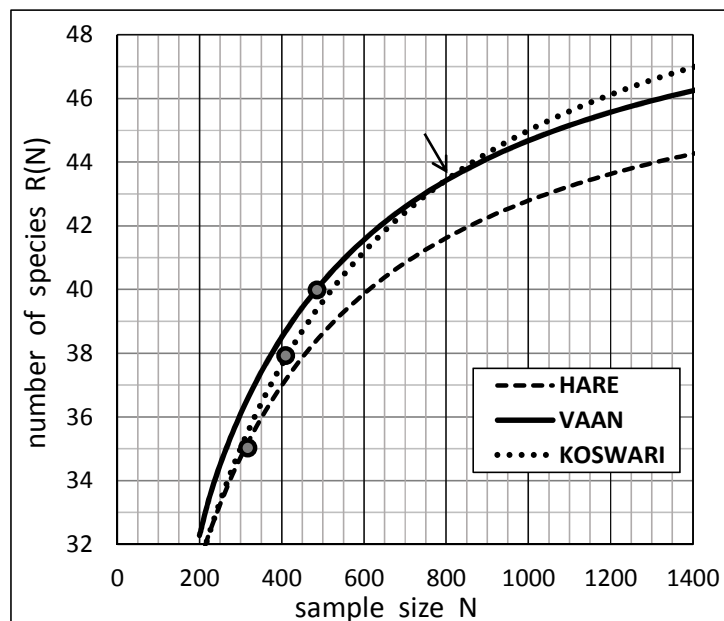


Figure 4 – A zoom of Figure 3, focused on the beginning of extrapolations. Note the steeper slope of the species accumulation curve of Koswari as compared to the curves of Vaan and Hare, resulting in the intersection of the curves of Koswari and Vaan around sampling-size $N = 800$.

At last, Figure 5 provides quantitative prediction regarding the *marginal* additional sampling effort needed to increase by one the number of recorded species.

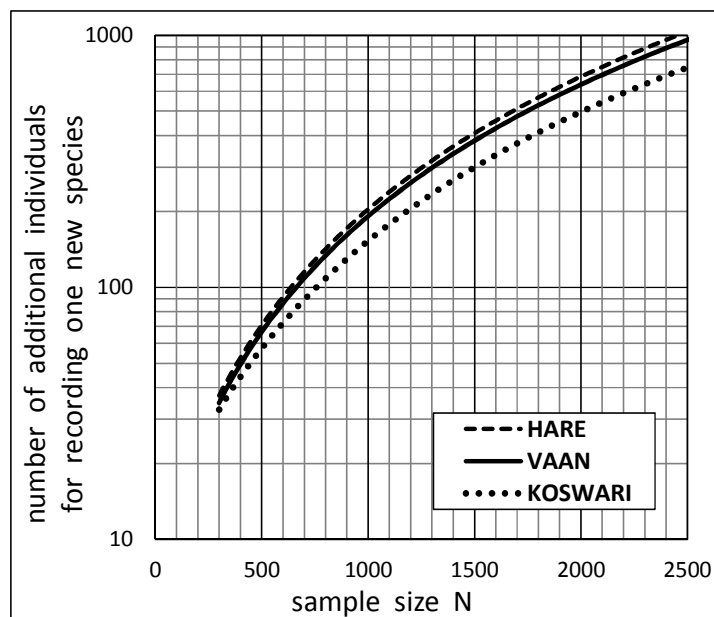


Figure 5 – Additional (average) number of individuals necessary to detect one new species. This highlights how the additional sampling effort needed to increase by one the recorded species richness (i.e. the “marginal effort”) is dramatically increasing with growing sampling size.

DISCUSSION

Pristine tropical coral reefs are well known for hosting an exceptionally high level of associated biodiversity. Reefs provide unequalled variety of resources in terms of food and shelter for an incredibly large number of animals, all across size and taxonomical ranges. Shelled macro-gastropods just make a subset of this wide variety of animals’ forms, with, yet, large biological and commercial importance. This justifies special extra efforts in investigating the gastropods fauna of tropical coral reefs along Indian coast, as elsewhere, because still too few reports are made available on the subject.

Underwater samplings are often less easy to achieve than usual, which, coupled with limited available time for field investigation, almost inevitably results in more or less incomplete underwater inventories. With around 20% of singletons (species only recorded once), each of the three surveys conducted by MOHANRAJ and coworkers [1] on coral reefs in the Gulf of Mannar Biosphere Reserve (India) are expected to remain incomplete and this to some unknown extent. This arises two complementary questions: (i) **what** might be the true, total species richness of shelled macro-gastropods in each of the three investigated sites and (ii) **what** additional sampling efforts would be required **to increase the level of completeness of these inventories to some given extent.**

Both questions call for implementing an appropriate procedure of numerical extrapolation of the species accumulation curves, beyond the already achieved sampling-size, leading to minimal bias estimate of true species richness. Such procedure is necessary since neither the still traditional procedure suggested by COLWELL & CODDINGTON [16] nor even the more promising procedure proposed by BROSE *et al.* [13, 17] reveal finally satisfactory (Table 2, Figure 2).

251 ***The estimated total Gastropod species richness of each of the three investigated***
252 ***reefs***

253 The selected, least-biased type of estimator of the number of unrecorded species were
254 Jackknife-4 (for Koswari) and Jackknife-5 (for Vaan and Hare). Accordingly, the *total*
255 *species richness* amounts to 53, 51, 49 species for the shelled macro-gastropods
256 respectively associated to the coral reefs along Koswari, Vaan and Hare islands (instead
257 of 38, 40, 35 recorded species respectively). Thus, the three investigated reefs have fairly
258 similar total species richness, around fifty. In turn, this confirms that the three
259 inventories were actually incomplete, with 72%, 78%, 71% completeness levels
260 respectively (Table 1). Now, this local account from coral reefs fringing three small
261 islands should be considered by comparison to the more extensive and ecologically
262 diverse coral reefs of Mannar Gulf as a whole, which encompasses still other types of
263 marine habitats and ecosystems, in particular seagrass habitats and mangrove habitats
264 [2]. According to the general survey by MELKANI *et al.* [21], the entire Gulf of Mannar is
265 host to some 260 species of Gastropods, a figure, yet, which moreover includes *non-*
266 *shelled* species. On the other hand, a survey of coral reefs fringing two other
267 neighbouring islands, in addition to Koswari and Vaan islands [2], numbers a total of 34
268 species of shelled macro-gastropods out of which no less than 23 were not listed among
269 the 40 species recorded in the inventories by MOHANRAJ *et al.* [1]. Thus, adjoining the two
270 surveys by MOHANRAJ *et al.* [1 & 2] leads to a total recorded species richness of 63. This
271 figure substantially exceeds the average 38 recorded species per individual reef (Table
272 1), thus suggesting a rather patchy distribution of Gastropod fauna, differing appreciably
273 from one coral reef to another one. Accordingly, with (i) a figure of 38 recorded species
274 per individual reef on average and (ii) a figure of 63 species for 4 reefs pooled together,
275 it follows that a potential of 70 to 80 recorded species seems likely for Mannar Gulf as a
276 whole. At last, accounting for the level of completeness of inventories (around 75%,
277 Table 1), this finally suggests a true total richness around 100 species for the Gastropod
278 fauna associated to coral reefs all across the Gulf of Mannar. This, indeed, highlights the
279 biodiversity conservation value of the “Gulf of Mannar Biosphere Reserve” for this
280 specific fauna as well.

281 Now, at a methodological view point, it has been noticed that the kinetics of species
282 accumulation notably differs according to sites, so that the species accumulation curves
283 of Koswari and Vaan come to intersect at a given a sample-size (around N = 800). The
284 occurrence of such situations – which are far from being uncommon – highlights the fact
285 that comparisons between non-exhaustive inventories cannot be reliably extrapolated
286 towards the total species richness, even when their sampling-sizes are equal [22]. In
287 particular, this invalidates the yet still currently admitted opinion that rarefaction
288 procedure would guarantee for reliable predictions in terms of total species richness of
289 the compared species assemblages: see also other similar examples [23-25].

291 ***Predicting the additional sampling-effort required for a given improvement of***
292 ***sampling completeness***

293 Both (i) the *total* sampling effort necessary to reach a given improvement of
294 completeness and (ii) the *marginal* sampling effort for increasing by one the number of
295 recorded species (Figures 3 and 5 respectively) show that both appreciations of
296 required sampling effort dramatically increase when still higher levels of completeness
297 are targeted. The practical interest of numerical extrapolations is, precisely, to be able to

accurately quantify this dramatic increase and, thus, be able to address rationally the inevitable question: when to “reasonably” stop additional sampling effort?

When to reasonably stop sampling effort?

Clearly, the answer to this question is a matter of compromise and, as such, a problem debatable, especially when the determinants involved in the question can only be appreciated qualitatively. Hence, the valuable contribution **to be** expected from *quantifying* the terms of the balance between the gain (i.e. the expected increase of the number of recorded species) and the cost (i.e. the corresponding additional sampling effort that is required). As just mentioned, the question can be handled two ways, according to either the *total* or the *marginal* costs (i.e. the ‘integral’ or the ‘derivative’).

In the first case, the criterion to be considered is the ratio, $N/R(N)$, of sampling-size N to the corresponding number $R(N)$ of recorded species, which should not exceed an “acceptable” maximum threshold value, above which the “sampling yield” would be judged insufficient.

In the second option, the criterion to be considered is the derivative, $\partial N/\partial R$ ($= 1/(\partial R(N)/\partial N)$). This derivative is easily valued in practice, since, according to equation [A.1] (see Appendix 1), it is equal to the ratio, N/f_1 , of the sample size N to the corresponding number f_1 of recorded singletons. Here again, the criterion $\partial N/\partial R$ ($= N/f_1$) should not exceed an “acceptable” maximum threshold value above which the “sampling yield” would be judged insufficient.

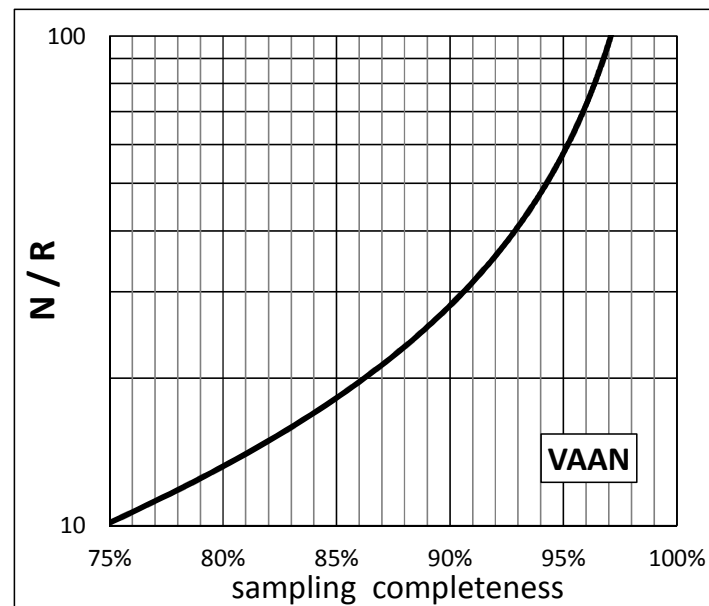


Figure 6 – The *total* sampling effort N related to the corresponding number of recorded species R (ratio N/R), plotted against the level of sampling completeness (here for the site of Vaan). In practice, the graph provides the predicted maximum sampling completeness that may be reached for a pre-determined maximum acceptable value of the ratio N/R . For example, a maximum accepted ratio N/R of 20 will allow to reach $\approx 86\%$ level of sampling completeness. Conversely, reaching 95 % sampling completeness would require to accept at least a ratio N/R of ≈ 60 .

Let consider, for example, the case of Vaan Island. Figures 6 and 7 (both derived from Figure 3) show the variations of the criteria N/R and $\partial N/\partial R$ with sampling completeness. In practice, these figures allow to predict the maximum level of sampling completeness that may be reached as a function of the maximum “reasonably

acceptable” value for one or the other of these two criteria. That is the level of completeness obtained when sampling has to be reasonably stopped on the rational basis of the minimum “sampling yield” considered acceptable in practice, in the context of study.

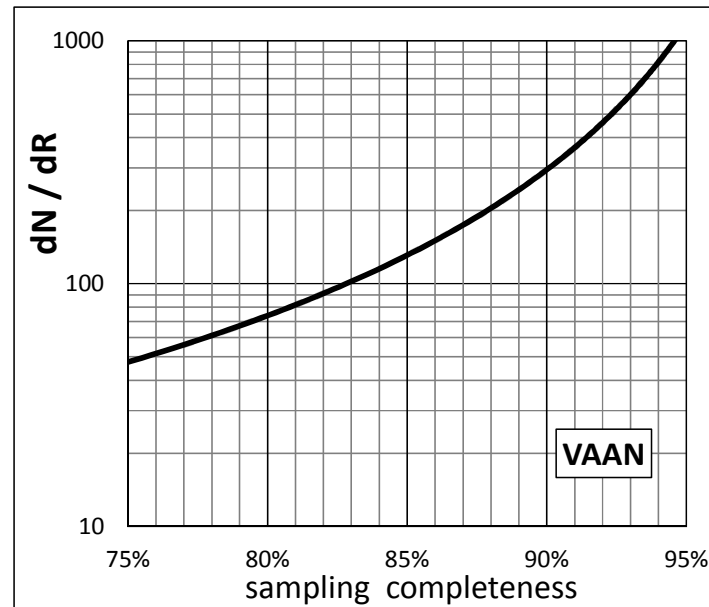


Figure 7 – The marginal cost $\partial N/\partial R$ (i.e. the number of additionally observed individuals required to detect one new species) plotted against the level of sampling completeness (here for the site of Vaan). In practice, the graph provides the predicted maximum sampling completeness that may be reached for a pre-determined maximum acceptable value of the marginal cost $\partial N/\partial R$. For example, a maximum marginal cost of 100 (new individuals needed for increasing by one the number of recorded species) will allow to reach a $\approx 83\%$ level of sampling completeness. Conversely, reaching 90 % sampling completeness would require to accept at least a marginal cost as high as ≈ 300 .

CONCLUSION

An accurate procedure of numerical extrapolation of the partial inventories carried out on the Gastropods fauna associated to the three investigated coral reefs in “Mannar Gulf Biosphere Reserve” [1] provided similar least-biased estimates of the true (total) species richness for each site, that is around *fifty* species at each site, an appreciably higher figure than suggested by the numbers of recorded species only. The 11 to 15 still unrecorded species (according to sites) call for further sampling in each site, to get access to the identities of these still unrecorded species. Yet, since additional sampling efforts required to detect new species unfortunately increases exponentially with further sampling, it is essential to be able to predictively quantify the additional sampling efforts/costs that would be needed as a function of the targeted increase in sampling completeness. Or, conversely, to predict the maximum completeness expected from a given granted additional sampling effort. Here also, the implementation of an accurate procedure of numerical extrapolation of species accumulation has proved being an essential tool.

More generally, the accurate estimation of true (total) species richness – using least-biased extrapolation of the species accumulation during progressive sampling [26, 27] – features all the more necessary that species richness tends to be recognized – now again – as the best numerical parameter to qualify local biodiversity [30].

ACKNOWLEDGEMENTS

The author thanks two anonymous reviewers for their appreciation and comments on the manuscript.

REFERENCES

1. Mohanraj J, Johnson JA, Ranjan R, Johnson L, Pandi U & Shunmugaraj T Coral reef associated gastropods in Tuticorin coast of Gulf of Mannar Biosphere Reserve, India. Indian Journal of Science and Technology. 2010; 3(2): 204-206.
2. Mohanraj T, Jagadis I & Sivanesh H Association of molluscan fauna with the coral reefs of Gulf of Mannar. European Journal of Experimental Biology. 2011; 1(3): 44-48.
3. Mohanraj J, Chelladurai G, Balakrishnan S & Kuamr IV Gulf of Mannar island coral reef associated gastropods assemblage: distribution and diversity pattern. Journal of Coastal life Medicine. 2015; 3(9): 691-695.
4. Pawar PR & Al-Tawaha AR Biodiversity of marine gastropods along the Uran coast, Navi Mumbai, west coast of India. American-Eurasian Journal of Sustainable Agriculture. 2017; 11(2):19-30.
5. Williams PH & Gaston KJ Measuring more of biodiversity: Can higher-taxon richness predict wholesale species richness? Biological conservation. 1994; 67(3):211-217.
6. Pogue MG Preliminary estimates of Lepidoptera diversity from specific sites in the Neotropics using complementarity and species richness estimators. Journal of the Lepidopterists Society. 1999; 53(2): 55-71.
7. Stirling G & Wilsey B Empirical Relationships between Species Richness, Evenness and Proportional Diversity. The American Naturalist. 2001; 158(3): 286-299.
8. Garcia D & Martinez D Species richness matters for the quality of ecosystem services: a test using seed dispersal by frugivorous birds. Proceedings of the Royal Society B. 2012 ; 279: 3106–3113.
9. Bock CE, Jones ZF & Bock JH Relationships between species richness, evenness, and abundance in a southwestern savanna. Ecology. 2007; 88(5): 1322-1327.
10. DeVries PJ, Walla TR & Greeney HF Species diversity in spatial and temporal dimensions of fruit-feeding butterflies from two Ecuadorian rainforests. Biological Journal of the Linnean Society. 1999; 68: 333-353.
11. Gotelli NJ & Colwell RK Estimating species richness. pp. 39-54 in: Biological Diversity: Frontiers In Measurement And Assessment. A.E. Magurran and B.J. McGill (eds.). Oxford University Press, Oxford. 345 pp.; 2010.

12. Gotelli NJ & Chao A Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In: Levin S.A. (ed.) Encyclopedia of Biodiversity, second edition, volume 5, pp. 195-211. Waltham, MA: Academic Press; 2013.
13. Brose U, Martinez ND & Williams RJ Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*. 2003; 84(9): 2364-2377.
14. Basualdo CV Choosing the best non-parametric richness estimator for benthic macroinvertebrates databases. *Revista Sociedad Entomologica Argentina*. 2011;70(1-2): 27-38.
15. 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.
16. Colwell RK & Coddington JA Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society London B*. 1994; 345 : 101-118.
17. Brose U, Martinez ND Estimating the richness of species with variable mobility. *Oikos*. 2004; 105: 292-300.
18. Ulrich W & Ollick M Limits to the estimation of species richness: The use of relative abundance distributions. *Diversity and Distributions*. 2005; 11: 265-273.
19. Reese GC, Wilson KR & Flather CH Performance of species richness estimators across assemblage types and survey parameters. *Global Ecology and Biogeography*. 2014; 23: 585-594.
20. 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.
21. Melkani VK, Edward JKP, Murugan A & Naganathan A Capacity building in identification of scheduled animals: training cum information manual. Gulf of Mannar Biosphere Reserve Trust. Publication n° 8; 2009.
22. Lande R, DeVries PJ & Walla TR When species accumulation curves intersect: implications for ranking diversity using small samples. *Oikos*. 2000; 89(3): 601-605.
23. Béguinot J Least-biased estimations of true species richness of butterfly fauna in sub-urban sites around Jhansi (India) and the range of inter-annual variation of species richness. *Asian Journal of Environment & Ecology*. 2017;2(1): 1-12. doi: 10.9734/AJEE/2017/32040
24. Béguinot J Inter-annual variations of true species richness in a subtropical butterfly assemblage: an estimation based on least-biased extrapolations of species accumulation curves. *Asian Journal of Biology*. 2017; 2(4): 1-16. doi: 10.9734/AJOB/2017/33876

25. Béguinot J & Nidup T Least-biased extrapolation of a partial inventory of butterfly fauna in Manas Range (Royal Manas National Park, Bhutan). *Asian Journal of Environment & Ecology*. 2017; 2(2): 1-14. doi: 10.9734/AJEE/2017/32701
26. 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>
27. 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 >
28. 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, 6 pages, doi:10.1155/2014/847328 ; <hal-01101415>
29. 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>
30. 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.

Appendix 1

Demonstration that among the nonparametric estimators of the number of unrecorded species (Chao, the Jackknife series), the estimator providing the highest estimate is also the least-biased one

This statement is demonstrated, below, according to three different ways.

At first, as a reminder, the expressions of the Chao and Jackknife estimators are as follows:

$$\text{Chao: } \Delta = f_1^2 / (2 f_2) ;$$

JK-1: $\Delta = f_1$; JK-2: $\Delta = 2f_1 - f_2$; JK-3: $\Delta = 3f_1 - 3f_2 + f_3$; JK-4: $\Delta = 4f_1 - 6f_2 + 4f_3 - f_4$; ... ; and, more generally, for the Jackknife estimator at order ‘i’:

$$\text{JK-[i]: } \Delta = \sum_{x=1 \text{ to } i} (-1)^{(x-1)} \cdot C_{(i, x)} \cdot f_x$$

with $C_{(i, x)}$ as the number of combinations of x items among i (see BÉGUINOT 2016a for a demonstration).

*** first demonstration**

Let first consider *as axiomatic* the generally recognized fact that nonparametric estimators all provide under-estimates of the number of unrecorded species, whatever the type of estimator being considered [11, 12, 20]. It immediately follows that the *least-biased estimator* is also the one which *provides the highest estimate among them all*. In turn, this allows to define explicitly, and in all generality, the respective domains within

which each estimator performs best. For example, the domain associated to Jackknife JK-3, when the latter actually provides the highest estimate, is such that, within this domain JK-3 > JK-2 and JK-3 > JK-4. Accounting for the respective expressions of JK-2, JK-3, JK-4, it immediately follows that this domain associated to JK-3 is defined as follows, in terms of values of the f_x : $3f_1 - 3f_2 + f_3 > 2f_1 - f_2$ and $3f_1 - 3f_2 + f_3 > 4f_1 - 6f_2 + 4f_3 - f_4$; that is:

$$2f_2 - f_3 < f_1 < 3f_2 - 3f_3 + f_4$$

which defines the boundaries of the domain where JK-3 provides the least-biased estimate, since the latter exceeds the estimate of the neighbouring estimators.

Extending the same reasoning to other types of nonparametric estimators leads to the following general key of selection of the *least-biased estimator*, based on the value of the recorded number f_1 of singletons as compared to the recorded numbers $f_2, f_3, f_4, \dots$ of doubletons, tripletons, quadrupletons, etc ... :

* if $f_1 < 0.5 f_2 \rightarrow$ select **Chao**
 * if $0.5 f_2 < f_1 \leq f_2 \rightarrow$ select **JK-1**
 * if $f_2 < f_1 \leq 2f_2 - f_3 \rightarrow$ select **JK-2**
 * if $2f_2 - f_3 < f_1 \leq 3f_2 - 3f_3 + f_4 \rightarrow$ select **JK-3**
 * if $3f_2 - 3f_3 + f_4 < f_1 \leq 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow$ select **JK-4**
 * if $f_1 > 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow$ select **JK-5**

This selection key – in terms of values of f_1 compared to f_2, f_3, f_4, f_5 – is, thus, strictly equivalent to the selecting procedure according to which the estimator that provides the highest estimate is the least-biased one.

* second, alternative demonstration

It is not even necessary to call upon the axiom above (according to which nonparametric estimators all provide under-estimates). The same domains of optimality as those just defined above may be derived, *independently of this axiom*, by simply accounting for the constraint of *continuity* at the borders of the domains of optimality respectively associated to each type of estimator. This rule of continuity implies that the values taken by two successive Jackknives (say JK-[i] and JK-[i+1]) should be equal at the boundary between their respective domains. For example, at the boundary between JK-2 and JK-3, both Jackknives should be equal: $2f_1 - f_2 = 3f_1 - 3f_2 + f_3$, that is, finally, $f_1 = 2f_2 - f_3$, at the boundary between JK-2 and JK-3. Similarly, at the boundary between Chao and JK-1, both estimators should provide the same estimate: $f_1^2/(2f_2) = f_1$, that is, finally, $f_1 = 0.5 f_2$ at the boundary between Chao and JK-1. Extending the same reasoning to the other types of estimators leads to the same domains, as defined above, for the values of f_1 compared to f_2, f_3, f_4, f_5 . Thus, according to this second demonstration, the same key of optimality (in terms of values of f_1 compared to f_2, f_3, f_4, f_5) is derived as above, without resorting to the axiomatic preliminary. In fact, this second alternative approach, based on continuity, actually provides a demonstration of the “axiom”, considered as such, in the first approach. And this second approach provides, thus, an alternative independent demonstration that the least-biased nonparametric estimator is the *one which provides the higher estimate*.

* third demonstration

A third, yet less straightforward, demonstration was derived previously [26], leading to the same key of selection of the least-biased type of nonparametric estimator. As a

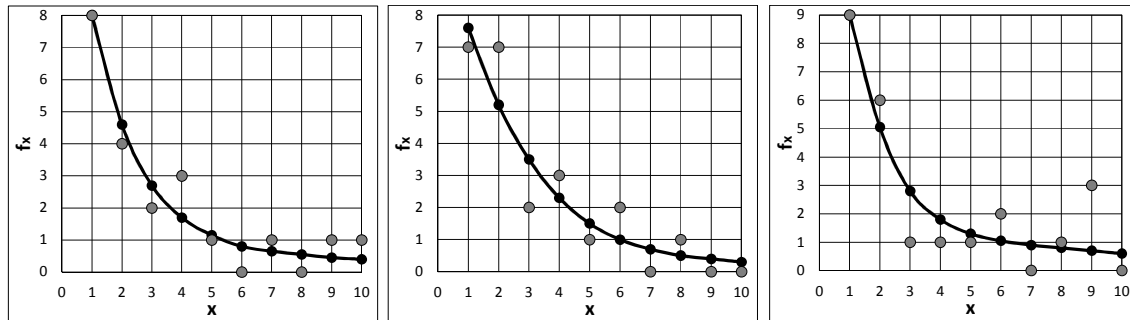
compensation of its substantially longer development, this third demonstration has the advantage of addressing also two additional points of interest:

(i) it proves that Chao and the Jackknife series are the only nonparametric estimators, expressed in terms of the f_x , which complies with the compulsory rule of *additivity*, according to which, in an assemblage of species that encompasses several mutually exclusive categories (that is categories that share *no species in common*; for example taxonomic subsets such as, genus, families, orders, etc...), the estimated number of unrecorded species for the whole assemblage should be equal to the sum of the estimated numbers of unrecorded species in each of the member categories;

(ii) it provides the expressions for the *extrapolation* of the Species Accumulation Curves respectively associated to each type of nonparametric estimator [26-27], based on the general mathematical relationship that constrains the theoretical expression of any theoretical Species Accumulation Curves $R(N)$ [26, 28-29]:

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

N.B. – At a practical point of view, 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 or by regression of the as-recorded distribution of the f_x versus x . Accordingly, for the present study, the regressions carried out to reduce the consequences of drawing stochasticity on the recorded distributions of values of the f_x are given at Figures A.1, A.2, A.3. It is these corrected values of the f_x which are implemented in the computations of the nonparametric estimators and in the numerical extrapolations of the associated species accumulation curves.



Figures A.1, A.2, A.3 – 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, for the sites of Hare, Vaan and Koswari (from left to right).

Appendix 2

Least-biased extrapolation of the Species Accumulation Curve

Compliance with equation [A.1] provides the following least-biased expressions for the extrapolations of the Species Accumulation Curves $R(N)$ (i.e. for $N > N_0$), respectively associated to each nonparametric estimator in its domain of optimality (detailed demonstration in [26]:

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

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

607 $(f_2 - f_1).N_0^2/N^2$
 608
 609 * *for f_1 up to $3f_2 - 3f_3 + f_4 \rightarrow R_3(N) = (R(N_0) + 3f_1 - 3f_2 + f_3) - (6f_1 - 8f_2 + 3f_3).N_0/N$*
 610 *$- (-4f_1 + 7f_2 - 3f_3).N_0^2/N^2 - (f_1 - 2f_2 + f_3).N_0^3/N^3$*
 611
 612 * *for f_1 up to $4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow R_4(N) = (R(N_0) + 4f_1 - 6f_2 + 4f_3 - f_4) - (10f_1 -$*
 613 *$20f_2 + 15f_3 - 4f_4).N_0/N - (-10f_1 + 25f_2 - 21f_3 + 6f_4).N_0^2/N^2 - (5f_1 - 14f_2 + 13f_3$*
 614 *$- 4f_4).N_0^3/N^3 - (-f_1 + 3f_2 - 3f_3 + f_4).N_0^4/N^4$*
 615
 616 * *for f_1 larger than $4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow R_5(N) = (R(N_0) + 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5)$*
 617 *$- (15f_1 - 40f_2 + 45f_3 - 24f_4 + 5f_5).N_0/N - (-20f_1 + 65f_2 - 81f_3 + 46f_4$*
 618 *$- 10f_5).N_0^2/N^2 - (15f_1 - 54f_2 + 73f_3 - 44f_4 + 10f_5).N_0^3/N^3 - (-6f_1 + 23f_2 - 33f_3$*
 619 *$+ 21f_4 - 5f_5).N_0^4/N^4 - (f_1 - 4f_2 + 6f_3 - 4f_4 + f_5).N_0^5/N^5$*
 620
 621 N.B.: for f_1 falling beneath $0.5 \times f_2$, the extrapolation of the Species Accumulation Curve is
 622 given in [27].
 623