



# Extrapolation of the Species Accumulation Curve for Incomplete Species Samplings: A New Nonparametric Approach to Estimate the Degree of Sample Completeness and Decide when to Stop Sampling

Jean Béguinot<sup>1\*</sup>

<sup>1</sup>Biogéosciences, Université de Bourgogne, F 21000 – Dijon, France.

## Author's contribution

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

## Article Information

DOI: 10.9734/ARRB/2015/22351

### Editor(s):

(1) George Perry, Dean and Professor of Biology, University of Texas at San Antonio, USA.

### Reviewers:

(1) Rezzoug Waffa, Ibn Khaldoun University Tiaret, Algeria.

(2) Sholeh Ghollsimod, Birjand University, Iran.

Complete Peer review History: <http://sciencedomain.org/review-history/12125>

Method Article

Received 29<sup>th</sup> September 2015

Accepted 17<sup>th</sup> October 2015

Published 6<sup>th</sup> November 2015

## ABSTRACT

Incomplete species samplings are deemed to remain the common practice in those groups of animals, such as small or micro- invertebrates, with numerous species that often are more or less difficult to detect in the field. Thus, extrapolating the Species Accumulation Curve as far as possible beyond the actual sample size may thus serve as a useful (although imperfect) surrogate to the desired, but practically inaccessible, complete samplings. In this context, several kinds of theoretical or empirical models for the Species Accumulation Curve and also a lot of estimators of the asymptotic limit of the Curve (i.e. total species richness) have been proposed. The practical issue is now to select appropriately among these numerous, different propositions. Here, I show that realistic Species Accumulation Curves are constrained to respect a general mathematical relationship, which, in turn, may serve to discriminate and select among the available models of Species Accumulation Curves and, as well, among the different formulations of the estimators of species richness that are commonly referred to.

As a result of the application of this screening approach, it follows that, for the generality of cases

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

(i.e. ratio singletons/doubletons larger than 0.6), a specific formulation of the Species Accumulation Curve (bi-hyperbolic with exponents -1 and -2 for sample size) complies at best. Accordingly, the more appropriate estimator of total species richness is Jackknife-2. Only when the ratio singletons/doubletons happens to fall beneath 0.6, Chao estimator may then be preferred. This is the case when samplings closely approach exhaustivity or when they address assemblages with unusually homogeneous abundances of species.

**Keywords:** *Extrapolation; species accumulation curve; sample completeness; stop sampling; non-parametric; point estimation; species richness; jackknife; Chao estimator.*

## 1. INTRODUCTION

Incomplete species samplings are deemed to become increasingly frequent as biodiversity surveys are progressively dedicated to those animals groups which usually give rise to assemblages of numerous species represented by individuals of small sizes, more or less hard to detect in the field (such as, for example, assemblages of small- or micro-invertebrates). From a practical point of view, incomplete species samplings raise, in particular, two important questions:

- (i) When “reasonably” stop sampling, given the regularly decreasing recording rate of newly detected species with progressively increasing sampling effort and accounting for usually limited available resources for field investigations and the coexistence of other competing priorities ;
- (ii) After having stop sampling, how many species are missed, still having escape the recording; in other words which level of sampling completeness is actually achieved.

Both questions refer ultimately to a same third issue, i.e. the possibility of reliably extrapolate the Species Accumulation Curve (S.A.C.) beyond the actual sample size. This includes: (i) extrapolation considered at its beginning for treating the first question and (ii) extended extrapolation up to asymptotic limit for the second question.

For both cases, several kinds of solutions are suggested in the literature:

- A series of artificially designed models for S.A.C.s (‘Clench’, ‘Negative Exponential’,...) may help predict the rate at which new species would be collected thanks to supplementary sampling efforts: See for example [1] for a review of such models,

- A series of nonparametric estimators of the expected number of missing species (Chao, Jackknife, ... see [2] for a review) may provide estimates of the total species richness and, accordingly, may predict the asymptotic limit of the S.A.C.

The difficulty, in both these respects, is *to choose “at best” among the numerous available approaches*, since each of them provides substantially distinct estimates of their own, due to their specific own formulation. A considerable amount of work has been devoted to this question, on an empirical basis, especially regarding nonparametric estimators of species richness [3-9]. As useful as they are, these empirical approaches, however, hardly help to disclose any information of *general* value. In fact, as might be expected, no consensus emerged from these studies. This is because each kind of estimator (Chao, Jackknife order 1 or 2,...) may provide a centered, unbiased prediction *only* when applying to a particular, specific shape of species abundance distribution within the assemblage of species under consideration, that is, in turn, a specific shape of the S.A.C. [1,9,10-12].

Accordingly, it would seem appropriate to address the issue *more generally*, on a *more theoretical* basis. More precisely, it looks desirable to consider a rigorous, purely mathematical approach, involving both probabilistic and algebraic developments appropriate to the subject.

Hereafter, I try to tackle the question this way, making use of a *newly derived relationship* linking (i) the successive derivatives  $\partial^x R_{(N)}/\partial N^x$  of the Species Accumulation Curve  $R_{(N)}$  to (ii) the series of numbers  $f_x$  of those species already recorded  $x$ -times in the currently available sample (number of singletons  $f_1$ , number of doubletons  $f_2$ , etc...). Beyond theoretical aspects, this relationship between the  $\partial^x R_{(N)}/\partial N^x$  and the  $f_x$ , offers the potential advantage of

linking the factors that govern the shape of the Species Accumulation Curve  $R_{(N)}$  (i.e. the series of its successive derivatives  $\partial^x R_{(N)}/\partial N^x$ ) to easily recorded parameters (i.e. the series of numbers  $f_x$  of species respectively recorded  $x$ -times within the ongoing sampling).

In short, this relationship highlights how the data immediately available from the ongoing sampling (the series of recorded  $f_x$ ) allow to extrapolate the Species Accumulation Curve  $R_{(N)}$ , the shape of which is governed by the series of its successive derivatives  $\partial^x R_{(N)}/\partial N^x$ .

## 2. DERIVING THE EQUATIONS OF THE SPECIES ACCUMULATION CURVES RESPECTIVELY ASSOCIATED TO EACH OF THE MORE COMMONLY USED ESTIMATORS OF SPECIES RICHNESS

The method implemented hereafter will consist,

- First, to explicit the general relationship linking the successive derivatives  $\partial^x R_{(N)}/\partial N^x$  of the Species Accumulation Curve  $R_{(N)}$  to the series of numbers  $f_x$  of those species recorded  $x$ -times in the ongoing sample;
- Then to consider the respective expressions of the most commonly used nonparametric estimators of total species richness (Chao, Jackknife-1, Jackknife-2), all these expressions being in terms of the two first terms of the series of  $f_x$  :  $f_1$  &  $f_2$ ;
- Accordingly, to replace, in each of these expressions,  $f_1$  and  $f_2$  by the derivatives,  $\partial R_{(N)}/\partial N$  and  $\partial^2 R_{(N)}/\partial N^2$ , in accordance with the general relationship above. These replacements thus give rise to three specific differential equations, respectively associated to each of the three estimators of species richness considered here;
- In turn, the integration of each of these three differential equations (in terms of  $\partial R_{(N)}/\partial N$  and  $\partial^2 R_{(N)}/\partial N^2$ ) yields the equations respectively governing the Species Accumulation Curves  $R_{(N)}$  associated to each of the three estimators.

Deciphering this way the equation of the Species Accumulation Curve associated to a given estimator provides a mathematically rigorous, nonparametric approach of its extrapolation, that permits:

- (i) To predict at what pace new species are expected to add progressively to the set of already recorded species, would the ongoing sampling be further continued. By the way, comparison may be made between the expected gain of newly recorded species and the required supplementary sampling effort. A decision may thus be made more rationally about the opportunity to continue or to stop an ongoing sampling;
- (ii) Finally, to obtain the less biased estimation of the asymptotic limit of species accumulation, that is, the expected total species richness of the sampled assemblage.

Now, consider the progressive sampling of an assemblage of  $S$  species, providing a steadily increasing number  $R_{(N)}$  of collected species with growing sample size  $N$  and, thus, a steadily decreasing number  $\Delta = S - R$  of missing (unrecorded) species. Let  $f_{1(N)}, f_{2(N)}, f_{3(N)}, \dots, f_{x(N)}$ , be the numbers of species recorded respectively one, two, three, ...,  $x$ -times, within a sample of given size  $N$ . A bi-univocal relationship may be derived algebraically between each of the successive derivatives  $\partial^x R_{(N)}/\partial N^x$  of the number of recorded species  $R_{(N)}$  on the one hand and the value of each  $f_{x(N)}$  on the other hand:

$$f_{x(N)} = (-1)^{x-1} C_{N,x} [\partial^x R_{(N)}/\partial N^x] \quad (1)$$

with  $C_{N,x} = N!/(N-x)!/x!$  and  $[\partial^x R_{(N)}/\partial N^x]$  as the  $x^{\text{th}}$  derivative of  $R_{(N)}$  with respect to  $N$ , at point  $N$ . The detailed derivation of this general relationship has been described and published previously [13,14].

As the first terms only of the series of  $f_x$  will be considered in practice,  $x$  will remain quite negligible as compared to  $N$ , so that, with good approximation:

$$f_{x(N)} = (-1)^{x-1} (N^x/x!)[\partial^x R_{(N)}/\partial N^x] \quad (2)$$

In particular:

$$f_{1(N)} = N \cdot \partial R_{(N)}/\partial N \quad (3)$$

and

$$f_{2(N)} = -\frac{1}{2} N^2 \cdot \partial^2 R_{(N)}/\partial N^2 \quad (4)$$

Let  $\Delta_{(N)}$  be the number of missing (still unrecorded) species in a sample of size  $N$ , that is  $\Delta_{(N)} = S - R_{(N)}$ . The estimates of  $\Delta_{(N)}$ , respectively

associated to each of the three most popular estimators of species richness, Chao, Jackknife-1 and Jackknife-2 are:

$$\begin{aligned} \Delta_{(N)} &= f_{1(N)}^2 / (2 f_{2(N)}) \text{ for Chao;} \\ \Delta_{(N)} &= f_{1(N)} \text{ for Jack-1;} \\ \Delta_{(N)} &= 2 f_{1(N)} - f_{2(N)} \text{ for Jack-2.} \end{aligned}$$

Here,  $f_{1(N)}$  and  $f_{2(N)}$  are the numbers of singletons and doubletons among the  $R_{(N)}$  recorded species within a sample of size  $N$ .

According to equations (3) and (4) and the definition of  $\Delta_{(N)} = S - R_{(N)}$ , the three arithmetic relations defining  $\Delta_{(N)}$  above directly convert to the following three differential equations:

$$\Delta_{(N)} = (\partial \Delta_{(N)} / \partial N)^2 / (\partial^2 \Delta_{(N)} / \partial N^2) \quad (5)$$

$$\Delta_{(N)} = -N \cdot (\partial \Delta_{(N)} / \partial N) \quad (6)$$

$$\Delta_{(N)} = -2 \cdot N \cdot (\partial \Delta_{(N)} / \partial N) - \frac{1}{2} \cdot N^2 \cdot (\partial^2 \Delta_{(N)} / \partial N^2) \quad (7)$$

Accounting for  $R_{(N)} = S - \Delta_{(N)}$  and  $\partial^x R_{(N)} / \partial N^x = -\partial^x \Delta_{(N)} / \partial N^x$ , the successive integrations of each of these three differential equations yield the equations of the corresponding theoretical Species Accumulation Curves  $R_{(N)}$  respectively associated to each of the three estimators:

$$R_{(N)} = S \cdot (1 - \exp(-k \cdot N)) \text{ for Chao} \quad (8)$$

$$R_{(N)} = S \cdot (1 - k' / N) \text{ for Jack-1} \quad (9)$$

$$R_{(N)} = S \cdot (1 - k'' / N - k''' / N^2) \text{ for Jack-2} \quad (10)$$

For a given sample of size  $N_0$  with  $R_0 (= R(N_0))$  recorded species, the estimated total species richness  $S$  of the sampled assemblage is equal to  $R_0 + \Delta_{(N_0)}$  with  $\Delta_{(N_0)}$  respectively equal to  $f_1^2 / (2f_2)$ ,  $f_1$  and  $2f_1 - f_2$  in equations (8-10) with  $f_1 = f_{1(N_0)}$  and  $f_2 = f_{2(N_0)}$ .

Accordingly, equations (8-10) define those particular shapes of Species Accumulation Curves  $R_{(N)}$  for which (and only for which) Chao, Jack-1, Jack-2 formulations, respectively, deliver unbiased (centered) estimates of the number  $\Delta_{(N_0)}$  of missing (unrecorded) species.

The coefficients  $k$ ,  $k'$ ,  $k''$ ,  $k'''$  in equations (8-10) are defined by applying equations (3) and (4) at  $N = N_0$ . The Appendix provides the details of derivation of these coefficients and, accordingly, specifies to what extent each of the three corresponding Species Accumulation Curves actually satisfies the requirements of equations (3) and (4):

- For Chao,  
 $k = - [\text{Ln}((f_1^2/2f_2)/(R_0+f_1^2/2f_2))]/N_0$  and, accordingly, the associated Species Accumulation Curve does not satisfy in general the requirements (3) and (4) and should not be selected, except for the specific case when species abundances are approximately homogeneously distributed, as already argued previously [13,15];
- For Jack-1, the coefficient  $k'$  should simultaneously equal  $(f_1 \cdot N_0)/S$  and  $(f_2 \cdot N_0)/S$ ; accordingly, the Species Accumulation Curve associated to Jack-1 cannot satisfy both equations (3) and (4), except when it happens that  $f_1 = f_2$ .
- For Jack-2, it comes:  
 $k'' = (3f_1 - 2f_2) N_0 / (R_0 + 2f_1 - f_2)$  and  
 $k''' = (f_2 - f_1) N_0^2 / (R_0 + 2f_1 - f_2)$

and thus, for Jack-2, there is no restriction to the satisfaction of both equations (3) and (4).

### 3. SELECTING THE MORE APPROPRIATE (LESS BIASED) ESTIMATOR OF SPECIES RICHNESS

In summary, the Species Accumulation Curve (S.A.C.) associated to Jackknife-2 always complies with equations (3) and (4) (as required for any proper formulation of S.A.C.), while the S.A.C.s respectively associated to Jackknife-1 and to Chao estimators do not in general. The S.A.C. associated to Jack-1 is only satisfying when  $f_1 = f_2$ , that is in the very particular case when Jack-2 confounds with Jack-1. The S.A.C. associated to Chao estimator should be considered only for approximately homogeneous distributions of species abundances. The latter condition is quite uncommon in species assemblages in nature (only when the ratio of maximum to minimum values of species abundances in the sample is weak, less than  $\approx 5$ ) and, otherwise, is met only when the sampling closely approaches completeness (since then, the still unrecorded species all share relatively similar (low) abundance levels).

The more appropriate choice of estimator, among the three formulations considered here, thus depends upon the estimated degree of completeness of sampling, which, in turn, is related to the decrease of the ratio  $f_1/f_2$ .

Jack-2 is to be selected, at first, when  $f_1 > f_2$ . Then, Jack-1 may also be considered when  $f_1 \approx f_2$ . For samplings approaching completeness, i.e.

when  $f_1 < f_2$ , both Jack-2 and Chao may be considered. More specifically, when  $f_1 < \frac{1}{2} f_2$ , Jack-2 becomes negative and then nonsense (which incidentally shows that satisfying equations (3) and (4), although a necessary condition, is yet not always sufficient). Thus, for such sub-complete samplings, Chao estimator should be selected instead. In fact, for the sake of continuity, it is advisable to select Chao (instead of Jack-2) a little earlier, as soon as Jack-2 estimates falls beneath Chao's, that is when  $f_1 < 0.6 f_2$  (exactly when  $f_1 < (2-\sqrt{2})f_2$ ).

Finally, as Jack-2 ( $= 2f_1 - f_2$ ) identifies to Jack-1 ( $= f_1$ ) when  $f_1 \approx f_2$ , the *rule of selection* of the more appropriate estimator of the number of missing species turns to be quite simple:

- \* As  $f_1$  remains  $> 0.6 f_2$ , *Jackknife-2* estimator and its associated Species Accumulation Curve should be preferred as providing the less biased estimates and extrapolation;
- \* When  $f_1 < 0.6 f_2$ , *Chao* formulation and its associated Species Accumulation Curve should be preferred as providing the less biased estimates and extrapolation.

Thus, in short, Jackknife-2 estimator features *more appropriate in general*, while Chao is to be substituted in more *specific cases*, when  $f_1 < 0.6 f_2$ , that is mainly when samplings closely approach completeness or for those unusual cases when species abundances are sub-evenly distributed in the studied assemblage (typically, when the ratio of maximum to minimum values of species abundances in the ongoing sampling is less than  $\approx 5$ ).

Accordingly, to *minimise the bias of extrapolation of the Species Accumulation Curve*, it is suggested to use the following nonparametric expressions, according to the value of the recorded ratio  $f_1/f_2$ :

- \* For general case, i.e. when  $f_1 > 0.6 f_2$ :
 
$$R(N) = (R(N_0) + 2f_1 - f_2) - \frac{(3f_1 - 2f_2) N_0/N - (f_2 - f_1) N_0^2/N^2}{(f_2 - f_1) N_0^2/N^2} \quad (11)$$

- \* For the particular case when  $f_1 \leq 0.6 f_2$ :
 
$$R(N) = (R(N_0) + f_1^2/2f_2)(1 - \exp(-\frac{[\ln((f_1^2/2f_2)/(R_0+f_1^2/2f_2))][N/N_0]}{N/N_0})) \quad (12)$$

with  $N_0$  as the size of the ongoing sampling and  $R(N_0)$  as the number of recorded species.

It is important to notice that the bias that affects nonparametric estimators such as Jackknife tends to increase significantly when the ratio  $R/S$  decreases towards low values, less than  $\approx 0.5$ . Thus, extrapolations involving nonparametric estimators of species richness should preferably be considered when the ongoing sampling has reach about half of the estimated total species richness (in accordance with [16]).

#### 4. DISCUSSING THE SELECTION OF APPROPRIATE ESTIMATORS OF SPECIES RICHNESS

Apart from the puzzling diverging results provided by the series of empirical studies cited in Introduction, a semi-empiric approach [16] provided more firmly developed results that deserve consideration. On the basis of a large series of simulated samplings with various degrees of completeness, the authors proposed to select Jackknife-2 for estimated sampling completeness from 50% up to 74%, then to retain Jackknife-1 for completeness between 75% and 96% and, finally, to simply rely on the number of recorded species for completeness above 96%. This, in fact is in general agreement with the purely theoretical approach developed here. I also propose to select Jackknife-2 for low to medium levels of completeness and also consider Jackknife-1 when completeness goes higher, i.e. when the ratio  $f_1/f_2$  decreases toward 1 (i.e.  $f_1 \approx f_2$ ). But, as Jackknife-2 gradually rejoins Jackknife-1 when  $f_1$  approaches  $f_2$ , my proposal of consistently keeping with Jackknife-2 (including when  $f_1 \approx f_2$ ) prevents the somewhat unrealistic brutal transition from Jack-2 to Jack-1 at 74% completeness threshold, as proposed in [16]. Similarly, considering Chao estimator for very high levels of completeness (that is, when  $f_1 < 0.6 f_2$ ) avoids the sharp transition between Jack-1 estimation and zero (no missing species) at 96% completeness, as in [16].

Besides these considerations in favor of continuity, the complementary, *purely theoretical* basis of the present approach provides supplementary soundness to the rule of selection in favour of the most appropriate estimator of species richness proposed here.

#### 5. EXTRAPOLATION OF THE SPECIES ACCUMULATION CURVE (S.A.C.)

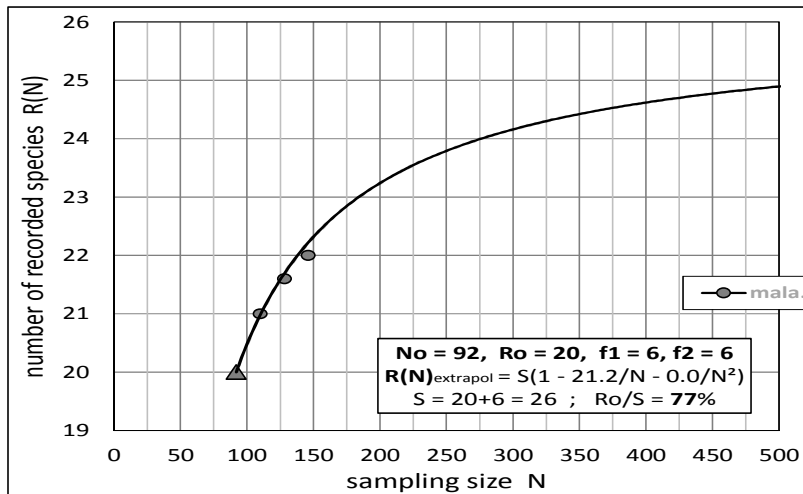
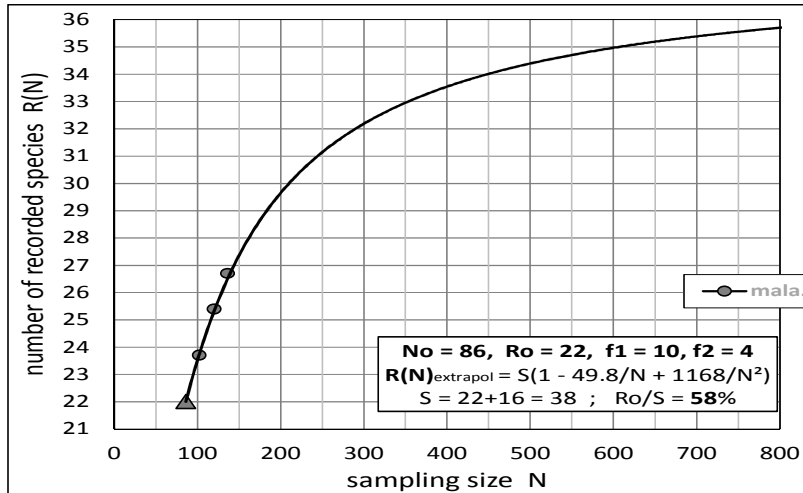
In a previous work, I derived a strictly nonparametric and unbiased procedure for the extrapolation of the Species Accumulation Curve

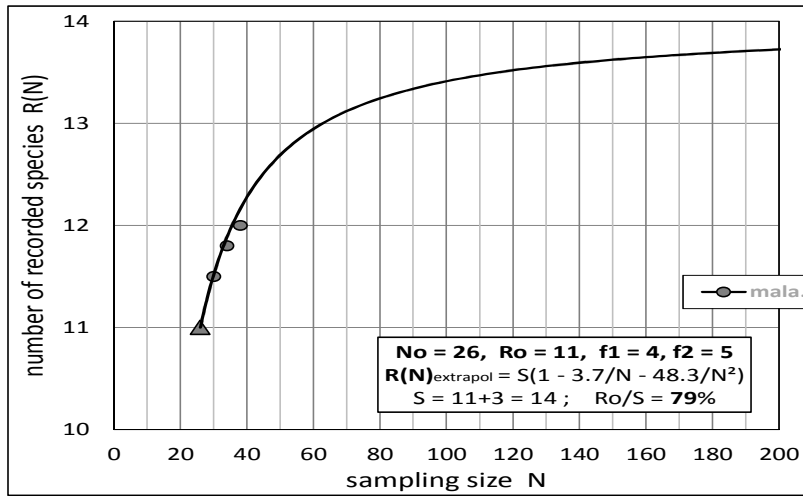
[14]. This extrapolation, however, encompasses a limited domain of additional sampling sizes, typically less than twice the actual sample size  $N_0$ . Further addressing substantially larger extrapolations of S.A.C.s requires in addition:

- 1) The estimation of the asymptotic limit of the S.A.C.; this is one of the main object of the preceding section, i.e. to provide a key to select, among classical formulations, the more appropriate (less biased) nonparametric estimator of species richness, providing accordingly an estimation of the asymptotic limit of the extrapolated S.A.C.;
- 2) The selection of the more appropriate mathematical form for the extrapolated part of the S.A.C., i.e. the particular mathematical form specifically associated

to the selected estimator of species richness; this being also one of the main point of the preceding sections.

Three illustrative examples of extrapolations of S.A.C.s beyond the actual, ongoing sample size are provided at Figs 1 to 3, involving varying degrees of sampling completeness  $R_0/S$ , within the range 55% to 80%. Both (i) limited extrapolations (based on reference [14] and thus profiting by being both nonparametric & unbiased) and (ii) unlimited extrapolations (based on the present work and thus nonparametric and profiting by minimised bias) are considered. The fair agreement between both types of extrapolations suggests that the unlimited extrapolations, based on Jack-2 estimations of the asymptotic levels, are not only nonparametric but also seems not significantly biased.





**Figs. 1, 2 and 3. Extrapolations of the species accumulation curves for three assemblages of land snails with varying degrees of sample completeness  $R_0/S$**

Triangles: Denote actual samples characteristics ( $N_0$  and  $R_0 = R(N_0)$ ); Dots: limited extrapolation according to reference [14]. Curve: Unlimited extrapolation based on estimation of the asymptotic limit. Completeness:  $R_0/S = 58\%$ ,  $77\%$ ,  $79\%$  ( $f_1/f_2 = 2.5, 1.0, 0.8$ ) for Figs 1, 2, 3 respectively

## 6. CONCLUSION

Extrapolating the Species Accumulation Curve (S.A.C.) beyond the actual size of an ongoing sampling not only provides a “talking” graphical representation of the estimated degree of completeness of the actual sample. Reliable extrapolation of the S.A.C. also offers the possibility of weighing the expected gain in number of newly recorded species, in term of the required additional sampling effort. Thus, a *more informed decision* may be made as to whether it seems opportune either to continue or to stop an ongoing species sampling. A more extensive appreciation in this respect is offered by the unlimited character of the extrapolating procedure described here.

In this perspective and to achieve this goal, I have derived above a new appropriate tool designed to deliver conveniently:

- \* A first step of extrapolation of the Species Accumulation Curve, still limited in its range of validity to less than twice the size of the already performed sampling, but profiting, in compensation, of the advantage of being not only a nonparametric but also an *unbiased* extrapolation.
- \* An extended, nonparametric extrapolation of the Species Accumulation Curve, this time unlimited in additional sampling size,

with *minimum bias*, obtained by selecting the more appropriate – less biased – estimator of total species richness of the sampled assemblage. Most importantly, this selection for the less biased estimator is not based on any subjective, empirical criteria but on the basis of a *rigorous, mathematical requirement*. That the associated Species Accumulation Curve satisfies the general relationship [equations (1 or 2)] linking the derivatives of the S.A.C.  $\partial^x R_{(N)}/\partial N^x$  to the recorded numbers  $f_x$  of species already recorded  $x$  times.

It turns out that, among three most commonly used nonparametric estimators of total species richness (Chao, Jackknife-1, Jackknife-2), the latter, Jackknife-2, features as the only one for which the associated Species Accumulation Curve actually satisfies the required mathematical relationship in all generality. Yet, when  $f_1 \leq 0.6 f_2$  (i.e. when samplings closely approach completeness or when species abundances are unusually homogeneously distributed), the Species Accumulation Curve associated to Chao estimator is specifically appropriate and should be retained.

Accordingly, in conclusion, the Species Accumulation Curve associated to Jackknife 2 (equation (11)) is to be selected when  $f_1 > 0.6 f_2$  (which is the far more common case for

incomplete samplings), while the Species Accumulation Curve associated to Chao (equation (12)) is to be selected in the particular case when  $f_1 < 0.6 f_2$ . These extrapolations of the Species Accumulation Curve should preferably be considered when the ongoing sampling has already reach about half the estimated total species richness of the sampled assemblage.

## ACKNOWLEDGEMENTS

Advises from two anonymous reviewers are greatly acknowledged.

## COMPETING INTERESTS

Author has declared that no competing interests exist.

## REFERENCES

1. Thompson GG, Withers PC, Pianka ER, Thompson SA. Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. *Austral Ecology*. 2003;28: 361–383.
2. Colwell RK, Coddington JA. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society London B*. 1994;345: 101-118.
3. Poulin R. Comparison of three estimators of species richness in parasite component communities. *Journal of Parasitology*. 1998;84(3):485-490.
4. Herzog SK, Kessler M, Cahill TM. Estimating species richness of tropical bird communities from rapid assessment data. *The Auk*. 2002;119(3):749-769.
5. Chiarucci A, Enright NJ, Perry GLW, Miller BP, Lamont BB. Performance of nonparametric species richness estimators in a high diversity plant community. *Diversity and Distributions* 2003;9:283-295.
6. Fogo A, Attrill MJ, Frost MT, Rowden AA. Estimating marine species richness: An evaluation of six extrapolative techniques. *Marine Ecology Progress Series*. 2003; 248:15-26.
7. Hortal J, Borges PAV, Gaspar C. Evaluating the performance of species richness estimators: Sensitivity to sample grain size. *Journal of Animal Ecology*. 2006;75:274-287.
8. Unterseher M, Schnittler M, Dormann C, Sickert A. Application of species richness estimators for the assessment of fungal diversity. *FEMS Microbiology Letters*. 2008;282:205-213.
9. Basualdo CV. Choosing the best non-parametric richness estimator for benthic macroinvertebrates databases. *Revista Sociedad Entomologica Argentina*. 2011; 70(1-2):27-38.
10. Soberon MJ, Llorente BJ. The use of species accumulation functions for the prediction of species richness. *Conservation Biology*. 1993;7:480–488.
11. Walter BA, Morand S. Comparative performance of species richness estimation methods. *Parasitology*. 1998; 116:395-405.
12. Hellmann JJ, Fowler GW. Bias, precision and accuracy of four measures of species richness. *Ecological Applications*. 1999; 9(3):824-834.
13. 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. *International Scholarly Research Notices – Ecology*. 2014;article ID 847328. DOI: 10.1155/2014/847328
14. 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
15. Chiu CH, Wang YT, Walther BA, Chao A. An improved nonparametric lower bound of species richness via a modified Good-Turing frequency formula. *Biometrics*. 2014;70(3). DOI: 10.1111/biom.12200
16. 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

### Determination of **coefficients $k$ , $k'$ , $k''$ , $k'''$** in equations (8), (9), (10)

- \* **Coefficient  $k$** : According to equation (8) for  $N = N_0$ , it comes:

$k = - [\text{Ln}((f_1^2/2f_2)/(R_0+f_1^2/2f_2))]/N_0$ . Accordingly, it is easily verified that the corresponding S.A.C. cannot satisfy the requirements of equations (3) and (4) in full generality.

- \* **Coefficient  $k'$** : According to equation (9) on the one hand and equations (3) and (4) on the other hand, it comes  $k' = (f_1 \cdot N_0)/S$  and  $k' = (f_2 \cdot N_0)/S$ , respectively. Thus equations (3) and (4) may be both satisfied only when it happens that  $f_1 = f_2$ .

- \* **Coefficients  $k''$  and  $k'''$** : According to equation (10),

Satisfying equation (3) yields  $k'' = (2S - 2R_0 - f_1) N_0/S = (3f_1 - 2f_2) N_0/(R_0 + 2f_1 - f_2)$

Satisfying equation (4) yields  $k''' = (R_0 - S + f_1) N_0^2/S = (f_2 - f_1) N_0^2/(R_0 + 2f_1 - f_2)$

Thus, without any restriction to the satisfaction of requirements of both equations (3) and (4).

---

© 2015 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://sciencedomain.org/review-history/12125>*