



**HAL**  
open science

**Inter-annual variations of true species richness in a subtropical butterfly assemblage: an estimation based on least-biased extrapolations of species accumulation curves.**

Jean Béguinot

► **To cite this version:**

Jean Béguinot. 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), pp.1-16. 10.9734/AJOB/2017/33876 . hal-01700765

**HAL Id: hal-01700765**

**<https://hal.science/hal-01700765>**

Submitted on 5 Feb 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Inter-annual Variations of True Species Richness in a Subtropical Butterfly Assemblage: An Estimation Based on *Least-biased* Extrapolations of Species Accumulation Curves

Jean Béguinot<sup>1\*</sup>

<sup>1</sup>Department of 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/AJOB/2017/33876

### Editor(s):

(1) Tulay Askin Celik, Department of Biology, University of Adnan Menderes, Turkey.

### Reviewers:

(1) Thadeu Sobral-Souza, Sao Paulo State University, Brazil.

(2) E. G. Oboho, University of Benin, Benin City, Edo State, Nigeria.

(3) Marcelle Barboza Pacheco Cardoso, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

Complete Peer review History: <http://www.sciencedomain.org/review-history/19568>

Original Research Article

Received 1<sup>st</sup> May 2017  
Accepted 12<sup>th</sup> June 2017  
Published 16<sup>th</sup> June 2017

## ABSTRACT

More or less strong inter-annual variations of species richness are well known in insects in general and in butterfly in particular. Yet, such variations generally rely upon more or less *partial samplings* and are rarely reliably documented in relevant terms of "*true*" total species richness (instead of "simply" observed number of species). In addition, equity of sampling completeness between annual inventories proves hard to ensure in practice and had long remained difficult to check properly. Thus, as neither standardized sampling procedures nor rarefaction procedure at a same sampling size may reliably warrant the equity of sampling completeness, asymptotic estimations of total species richness should imperatively be implemented for each annual inventory. This implies that, among the different estimators of the number of missing species available in the literature, the *least-biased* one should be adequately selected at first, for each annual inventory. As such a procedure of selection is now made available, it becomes possible to tackle relevantly the issue presented above. Applying this procedure to the field data recorded (and already published) by Lee et al. for butterfly assemblages at Mount Gariwang-san (south Korea), surveyed during several years, I show that (i) "*true*" inter-annual variations of species richness may indeed vary in a large range - from simple to

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

double - along successive years; (ii) annual total species richness represents only a limited part - comprised between less than  $\approx 40\%$  and at most  $\approx 80\%$  - of the potential species richness of the site (i.e. the true species richness along several years pooled together) .

*Keywords: Diversity; sampling completeness; nonparametric estimator; Gariwang-san; Korea; Lepidoptera.*

## 1. INTRODUCTION

Most insects assemblages undergo both *intra*-annual (seasonal) and *inter*-annual variations of species richness in temperate as well as tropical regions worldwide [1-8]. Due to their attractiveness and relative easiness of determination at the species level, butterfly assemblages have been particularly considered as a convenient model for studying temporal variations of insects' diversity [3-8]. Under tropical climates, the *seasonal* fluctuations of species diversity may be of variable amplitude while *inter*-annual variations are sometimes considered as clearly prominent [6]. Yet, the temporal fluctuations of species richness has been much more often addressed at the short time scale, i.e. seasonal variations [1-4,7] than during *longer periods*, i.e. substantially longer time series involving several successive years (see [8] however). Higher costs investments and shortage of available time to be devoted to such studies are arguably the main reasons for the current scarcity of long-term investigations. For the same reasons, the levels of completeness of the successively scheduled inventories are usually far from being exhaustive, which seriously hamper the significance to be given to as-recorded results, in terms of "true" variations of *total* species richness. Appropriate *extrapolations* of species accumulation beyond actually achieved sampling-sizes are thus needed to derive reliable estimates of the true total species richness that occurs at any stage of the time series. Therefore, the current difficulties to obtain relevant extrapolations, providing reliable estimates, urge to implement appropriate procedures involving *least-biased* extrapolations (thus following the path first initiated by BROSE and coworkers [9] and more recently improved by Béguinot [10-11].

Hereafter, I consider the field data issued from a long-term study (seven years) of butterfly assemblages at Mount Gariwang-san (South Korea) carried on by Lee and et al. [12]. As mentioned by the authors, each successive samplings performed along the seven investigated years, remain more or less *substantially incomplete*. Yet, they offer valuable

crude data from which to extrapolate species accumulation and estimate total species richness each year, in order to be able to relevantly address the three following questions:

(1) how large are the *inter*-annual variations of the "true" (i.e. total) species richness of the butterfly assemblages at Gariwang-san ?

(2) which proportion of the "overall potential butterfly richness" at the studied site (equated to the true species richness cumulated during the seven years, as a first approximation) actually occurs at any given year ?

Now, as the currently achieved inventories remain substantially incomplete, relevantly answering these two questions requires either (i) to *further continue sampling efforts* until closely approaching sampling exhaustively or, (ii) to *extrapolate, with minimized bias*, the species accumulation process, thereby being in capacity to provide reliable estimations to answer properly these two questions above. Option (i) would, of course, be ideal and, thus, to be privileged insofar as it proves compatible with the available resources in terms of time and costs. Yet, in common practice, option (ii) offers an economic, convenient and straightforward solution, to be considered at first. Then, before possibly considering the ideal option (i), a third question is to be addressed:

(3) which additional sampling efforts would be required to closely approach exhaustively (say, for example, reaching 95% sampling completeness), so as to be in capacity to predict, on a rational basis, the practicability (or not) of option (i): further continue sampling efforts.

## 2. MATERIALS AND METHODS

Lee and coworkers [12] conducted a series of samplings of the butterfly fauna at Mount Gariwang-san (South Korea), during 1987 and from years 2010 to 2015. All details relative to the sampling procedure, the environment context, the list of recorded species with their respective abundances are provided in [12] with free access and, accordingly, will not be recalled

here. Accounting for species *abundances* is of prime interest in the perspective of the extrapolation of partial samplings, since abundance data provides estimates of the numbers  $f_1, f_2, f_3, f_4, \dots, f_x, \dots$  of those species recorded respectively 1-, 2-, 3-, ..., x- times in the realised partial sampling. These numbers are required, in turn, to reliably extrapolate the Species Accumulation Curve, as explained below. As substantial numbers of singletons (i.e. species recorded only once) are retained in the inventories performed during each of the seven years (as well as in the seven inventories pooled together), it follows that all these inventories remain *substantially incomplete*. Extrapolating the species accumulation, beyond the actually achieved sampling sizes, is thus necessary to predict at best the true total species richness of butterfly assemblage for each year.

## 2.1 Numerical Extrapolation of Species Accumulation beyond the Achieved Sampling Size

As sampling size increases, the number of recorded species is monotonically growing, at first rapidly and then less and less quickly. The so-called 'Species Accumulation Curve'  $R(N)$  accounts for the growth kinetics of the number of recorded species  $R$  with increasing sampling size  $N$  ( $N$ : typically, the number of observed individuals during sampling). The mathematical expression (and thus the details of the shape) of the Species Accumulation Curve are dependent upon both the total species richness of the sampled assemblage of species and the degree of heterogeneity of the species abundance distribution within the sampled assemblage of species [13]. This would apparently make the extrapolation of the Species Accumulation Curve rather difficult to compute, since both preceding factors are unknown *a priori*. Yet, the numbers  $f_1, f_2, f_3, f_4, \dots, f_x, \dots$  of those species recorded respectively 1-, 2-, 3-, 4-, ..., x- times during sampling are directly dependent also upon the total species richness *and* the degree of heterogeneity of the species abundances. This explains why these numbers  $f_1, f_2, f_3, f_4, \dots$ , may serve as an appropriate basis from which to extrapolate the Species Accumulation Curve, beyond the actual size of the sample under consideration. In particular, the most commonly used estimators of the number of unrecorded species (i.e. *non-parametric* estimators such as 'Chao' and the series of 'Jackknife') are all computed from the recorded values of the first numbers  $f_x$  [14]. In practice, a problem remains

however: as already mentioned, each of these different types of estimators provides a *substantially distinct estimate* and none among these estimators remains consistently the more appropriate. Accordingly the traditional practice has become to consider together all of them without making any choice [15], an admittedly frustrating situation!

Yet, it has been shown recently that although none of the available estimators consistently remains the more accurate [9], each of them may prove, in turn, being the less biased, depending on the value taken by  $f_1$  as compared to the other  $f_{x>1}$  [10]. Accordingly, in practice, the most appropriate – i.e. *the least biased* – estimator of the number of unrecorded species may be selected by comparing the value of  $f_1$  to the values of the other  $f_x$  for  $x > 1$  [10-11]. Selecting this way the least-biased type of estimator thereby provides the best possible estimate of the number  $\Delta$  of "missing" species and, in turn, the best estimate of the total species richness  $S_t$  of a partially sampled assemblage. In addition, the least biased expression for the *extrapolation* of the species accumulation curve  $R(N)$  is straightforwardly derived.

In practice, the formulations summarized in Appendix 1 provide: (i) the expressions of  $\Delta$ ,  $S_t$  and  $R(N)$ , according to each of the most commonly used types of nonparametric estimators and (ii) the *key to select*, among these estimators, which one reveals the less biased and, thereby, which expressions for  $\Delta$ ,  $S_t$  and  $R(N)$  are the less-biased. In practice, the selection in favour of the less-biased estimator proceeds among a limited but rather large range of nonparametric estimators, including, not only, three commonly used estimators, Chao and Jackknife at orders 1 and 2, but also the following Jackknife at orders 3, 4, 5 : see reference [10] and also Appendix 1 for more details). Here, for the seven investigated years, the estimators that were selected, for each years, as being the less biased were either Jack-3, Jack-4 or Jack-5. That is, seeking for bias reduction had imposed, here, to rely only on those kind of estimators which remain uncommonly used. In turn, this highlights that conventional practices of estimations, still in current use however, may occasionally lead to substantial bias, as already cautioned, in particular, by Brose et al. [9].

Also, in order to reduce the dispersive influence of drawing stochasticity (which inevitably affects

the as-recorded values of the  $f_x$ ), it is advisable to regress the as-recorded distribution of the numbers  $f_x$  versus  $x$ .

### 3. RESULTS

#### 3.1 Extrapolations and Asymptotic Estimates for Each Sampled Year

The estimated numbers  $\Delta$  of missing (= unrecorded) species, according to each of six types of nonparametric estimators (Jackknife at orders 1 to 5 and Chao), are provided at Table 1 for the seven investigated years. Note that the selected least-biased estimator may differ from year to year: Jackknife type estimators at order 3, 5, 3, 5, 4, 3, 5 are respectively selected for years 1987, 2010, 2011, 2012, 2013, 2014, 2015; while Chao and Jackknife 1 & 2 were never selected. And selecting the least-biased estimator in each case is important since the estimated number of missing species may vary in a wide range for a same sample, typically from simple to double according to estimator type, as shown in Table 1. The least-biased estimations

of the 'true' (total) species richness,  $S_t = R_0 + \Delta$ , and the level of sampling completeness (=  $R_0/S_t$ ) are derived immediately: Table 2. Note, incidentally, that the levels of sampling completeness achieved for each of the seven studied years are very weakly related to either sampling-size or total species richness (Figs. 1 and 2).

In addition to the asymptotic estimates of  $\Delta$ ,  $S_t$  and  $R_0/S_t$ , the full range of extrapolation of the species accumulation curve was computed for each type of estimator (the expression of the least-biased extrapolation of the species accumulation curve being, of course, the one associated to the least-biased estimator). Figs. 3 and 4 provide two examples, for years 2010 and 2011 (the least-biased extrapolation being in bold). Figs. 5 and 6 allow to compare the selected, least-biased extrapolations of the species accumulation curves for six successive years, from 2010 to 2015 inclusively. Note that the six accumulation curves grow at different relative paces and, accordingly, some of them may intersect.

**Table 1. The estimated number  $\Delta$  of missing (= unrecorded) species, according to six types of nonparametric estimators, Jackknife at orders 1 to 5 and Chao. Estimates are provided for each of the seven annual samplings and, at last, for the seven samples pooled together. The least-biased estimator, selected according to the procedure described, is in bold for each sample**

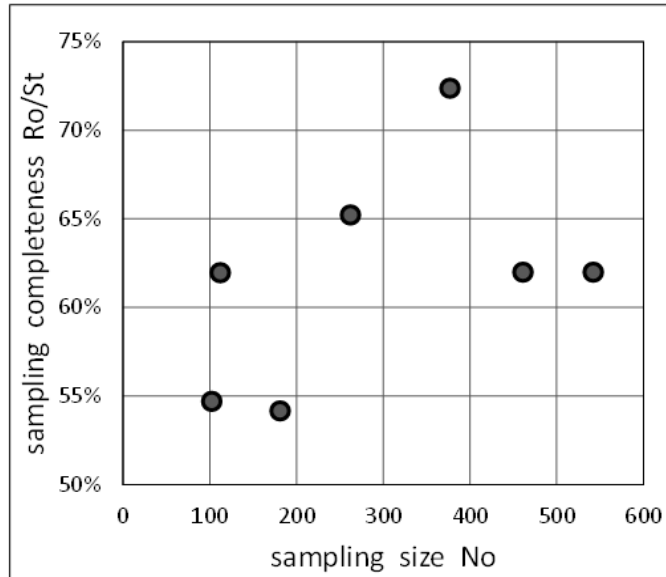
	1987	2010	2011	2012	2013	2014	2015	pooled
Jack-1	19	12	23	14	16	11	17	14
Jack-2	26	18	34	18	25	15	26	<b>15</b>
Jack-3	<b>27</b>	21	<b>38</b>	19	29	<b>16</b>	31	14
Jack-4	26	23	37	20	<b>31</b>	16	33	13
Jack-5	24	<b>24</b>	33	<b>21</b>	30	15	<b>33</b>	14
Chao	15	12	22	10	17	9	18	7

**Table 2. The sample size (number of sampled individuals)  $N_0$ , the number of recorded species  $R_0$  [=R( $N_0$ )], the selected, *least-biased* type of estimator, the corresponding estimated number  $\Delta$  of missing (unrecorded) species, the resulting estimated total species richness  $S_t$  and the resulting sampling completeness  $R_0/S_t$  (%), as computed for each of the seven annual inventories and the pooled inventory. Butterfly fauna at Mount Gariwang-san, field records according to Lee et al. [12] for years 1987, 2010, 2011, 2012, 2013, 2014, 2015**

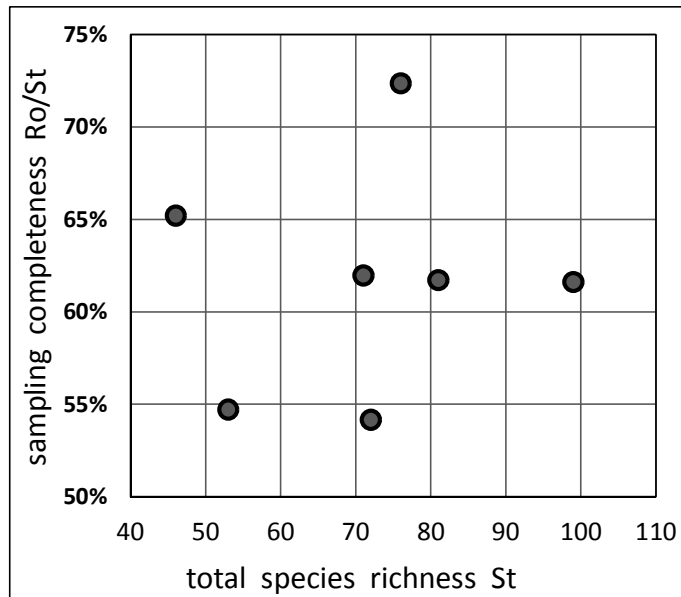
	1987	2010	2011	2012	2013	2014	2015	pooled
*Sample size $N_0$	112	102	542	377	461	262	181	2037
*Nb. recorded sp. $R_0$	44	29	61	55	50	30	39	105
*Selected estimator	<i>JK-3</i>	<i>JK-5</i>	<i>JK-3</i>	<i>JK-5</i>	<i>JK-4</i>	<i>JK-3</i>	<i>JK-5</i>	<i>JK-2</i>
*Nb. missing sp. $\Delta$	27	24	38	21	31	16	33	15
*Total sp. richness $S_t$	<b>71</b>	<b>53</b>	<b>99</b>	<b>76</b>	<b>81</b>	<b>46</b>	<b>72</b>	<b>120</b>
*Completeness $R_0/S_t$	<b>62%</b>	<b>55%</b>	<b>62%</b>	<b>72%</b>	<b>62%</b>	<b>65%</b>	<b>54%</b>	<b>88%</b>

Incidentally, the substantially distinct rates of growth of the numbers of recorded species (Figs. 5 and 6) have important practical consequence. *Contrary to what is still too frequently advocated,* comparing the numbers of recorded species

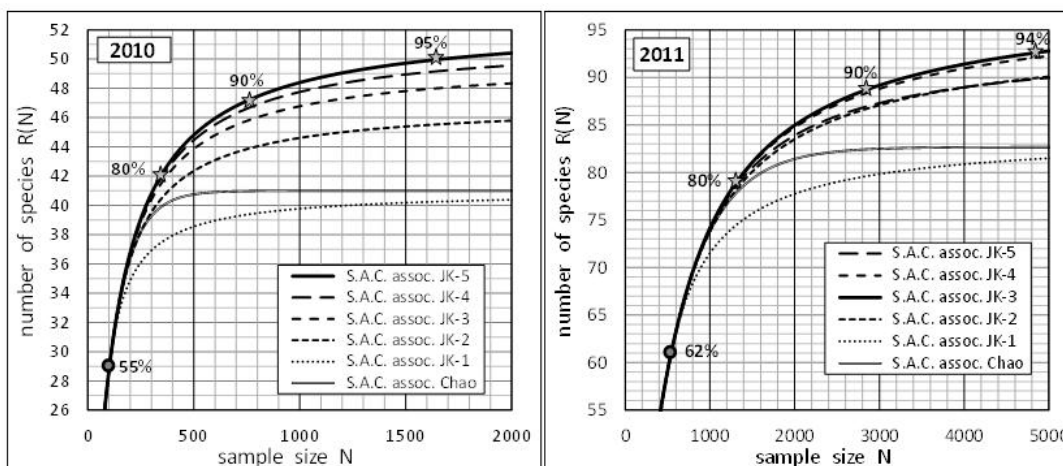
between several samples usually provides no relevant information in terms of comparison of true species richness, *even when samples have the same size,* (as already underlined [16] and more extensively argued by Béguinot [17].



**Fig. 1.** The level of sampling completeness,  $R_0/S_t$ , plotted against the sampling size, along seven investigated years. While sampling completeness increases of course with sampling size  $N_0$  *within a same given assemblage*, this trend becomes fairly inconsistent when several different assemblages are compared ( $r = 0.47$ ,  $n = 7$ ,  $p = 0.29$ )



**Fig. 2.** The level of sampling completeness,  $R_0/S_t$ , plotted against the estimated total species richness  $S_t$ , for the seven investigated years. The levels of sampling completeness achieved at each year are substantially independent of the corresponding estimated total species richness ( $r = 0.12$ ,  $n = 7$ ,  $p = 0.8$ )



**Figs. 3 and 4. Extrapolations of the species accumulation curve associated to six types of estimators (Jackknife orders 1 to 5 and Chao; the least-biased extrapolation being in bold) for years 2010 and 2011. The actually achieved sampling is indicated by the grey point. The least-biased estimator (and the associated extrapolation) is Jackknife-5 for year 2010 and Jackknife-3 for year 2011. One particular interest of considering the (least-biased) extrapolation is the possibility to reliably predict the additional sampling effort that would be required to reach any given level of sampling completeness beyond the already reached level (here are indicated the 80%, 90%, 95% completeness levels)**

**Table 3. The additional sampling effort N required to reach 95% level of sampling completeness. Predictions differ very strongly according to the type of estimator associated to the extrapolation of the species accumulation curve. The least-biased expectation for the sampling effort required to reach 95% completeness is in bold. Once more, selecting the least-biased estimator features decisive to obtain reliable predictions**

	1987	2010	2011	2012	2013	2014	2015	Pooled
Jack-1	680	598	2970	1530	2235	1405	1100	4800
Jack-2	1030	1020	4970	2190	3970	2190	1900	<b>5300</b>
Jack-3	<b>1100</b>	1300	<b>5940</b>	2470	5205	<b>2435</b>	2465	4600
Jack-4	960	1480	5660	2720	<b>5820</b>	2335	2785	4200
Jack-5	770	<b>1645</b>	4030	<b>3200</b>	5640	2255	<b>2900</b>	4700
Chao	254	285	1390	664	1265	570	540	2350

### 3.2 Estimation of the Additional Sampling Efforts Required to Improve Completeness

One major interest of extrapolating species accumulation curves is the possibility to predict the additional sampling effort that would be required to reach any given level of sampling completeness beyond the already achieved completeness. This prediction provides, in turn, a rational basis to decide whether or not it seems worth to continue any further the sampling operations, putting in balance the additional effort required and the expected benefit in terms of newly recorded species. And, in this respect, as for the asymptotic estimates above, selecting the least-biased extrapolation is very important to

derive reliable predictions, as shown in Table 3 which exemplifies the large scatter of predictions according to the type of estimator involved. Accordingly, Fig. 7 and Table 4 highlight the least-biased estimates of the additional sampling efforts that would have been required to reach different higher levels of sampling completeness, for each of the seven investigated years.

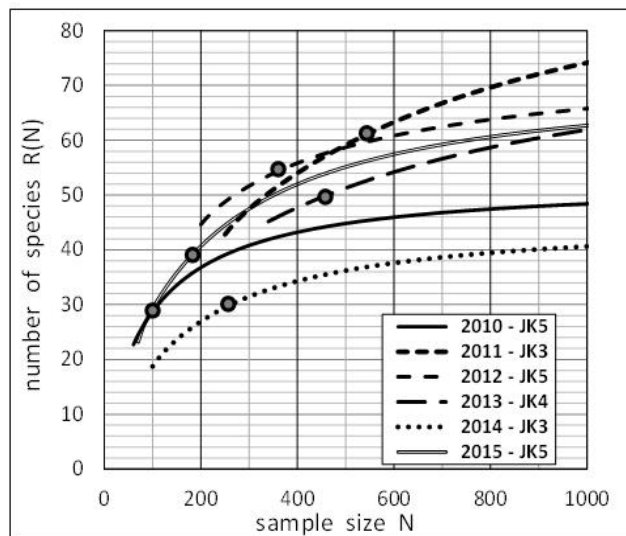
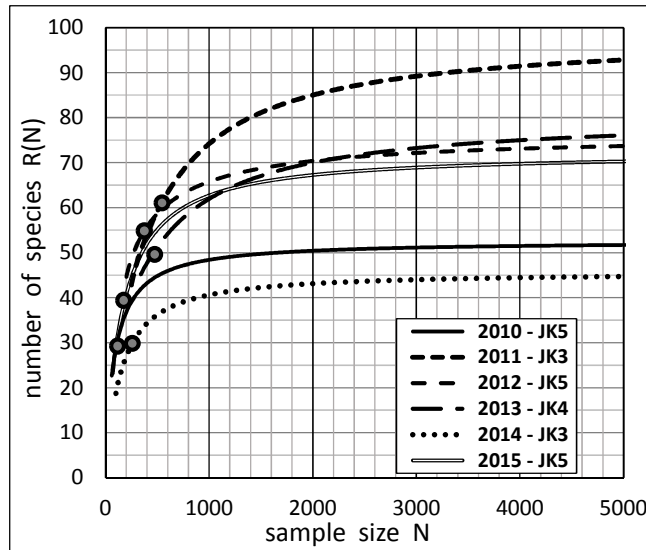
### 3.3 Inter-annual Variations of the Total Species Richness of the Butterfly Assemblage at Mount Gariwang-san

Figs. 8, 9, 10, 11 provide graphical representations of the data given in Table 2, which help visualizing the *range of inter-annual*

variations of the *true* species richness of butterfly fauna at Mount Gariwang-san. The estimated total species richness per year may vary substantially (up to a factor 2) and apparently erratically, according to years: Fig. 10.

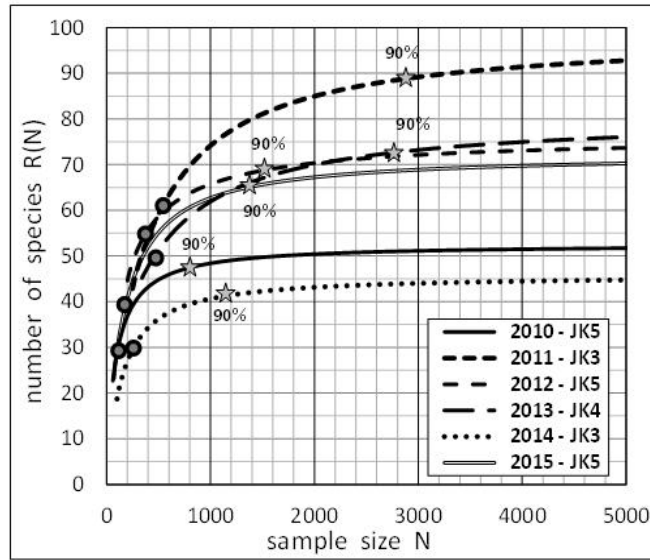
Still more interesting is the estimated *proportion* of the “potential species richness” of the site which actually occurs each year: Fig. 11. Here, the “potential species richness” is approximately equated, as a surrogate, to the total species richness estimated for the seven years (1987 &

2010 to 2015) taken *together*, i.e. 120 species. This proportion of the “potential species richness” occurring each year is comprised between a little less than 40% and a little more than 80% of the “potential species richness” of the site. In fact, this figure might be slightly overestimated since the cumulated number of species along seven years might well remain a slight underestimation of the real “potential species richness” of the site. Therefore, a figure comprised between one third and two third of the “potential species richness” might be perhaps more realistic.



**Figs. 5 and 6. The selected (least-biased) extrapolations of the species accumulation curves for six successive years, from 2010 to 2015 inclusively. For each year, the actually achieved sampling is indicated by the grey point; the selected estimator and the associated extrapolation are Jackknife at order 5, 3, 5, 4, 3, 5 for years 2010, 2011, 2012, 2013, 2014, 2015 respectively. Fig. 6 is a zoom of Fig. 5, focusing on the beginnings of the extrapolations**





**Fig. 7.** The selected, least-biased extrapolations of the species accumulation curves for six successive years, from 2010 to 2015 inclusively. The actually achieved sampling is indicated by the grey point. The estimated additional sampling efforts that would have been required to reach a same 90% level of sampling completeness are indicated for each year. These required additional sampling efforts differ very substantially from one year to the other

**Table 4.** Estimated sampling sizes that would have been required to reach increasing levels of sampling completeness (80%, 85%, 90%, 95%) for the inventories conducted during successive years (1987, 2010, 2011, 2012, 2013, 2014, 2015) and for the pooled inventories. Accordingly, approaching sampling exhaustively would require to consent considerably higher sampling efforts than actually performed. For example, aiming at reaching 95% completeness would have required to multiply the actually achieved sampling sizes by a factor ranging from 8.5 (year 2013) to 16.2 (year 2010). This would have probably exceed by far the resources in time and credit made available to Lee et al. [12]

	1987	2010	2011	2012	2013	2014	2015
Achieved sampling-size No	112	102	542	377	461	262	181
Selected estimator type	JK-3	JK-5	JK-3	JK-5	JK-4	JK-3	JK-5
Achieved completeness Ro/St	62%	55%	62%	72%	62%	65%	54%
Size for 80% completeness	250	340	1300	610	1220	530	625
Size for 85% completeness	340	490	1820	890	1730	740	870
Size for 90% completeness	540	775	2850	1470	2750	1160	1385
Size for 95% completeness	1100	1645	5940	3200	5820	2435	2900

Note also that, according to Fig. 10, the estimated total species richness does not show any clear temporal trend: the very slight decrease that appears between 1987 and 2015 contributes for less than 2% to the inter-annual variance of species richness ( $r^2 = 0.016$ ). Yet, the large gap between 1987 and the years 2010 to 2015 calls for some precaution and interpretation should limit more carefully to the time series 2010 to 2015.

A final remark regarding the results above: no estimation of error and confidence interval are provided for the estimated total species richness at each years. This is admittedly regrettable but results from the current lack of formulation of standard deviation for Jackknife's estimators at higher orders ( $\geq 3$ ), to my knowledge. This being said, I emphasise that *reducing bias* (as was the main object of the procedure implemented here) is a *priority* over trying to estimate the confidence interval.

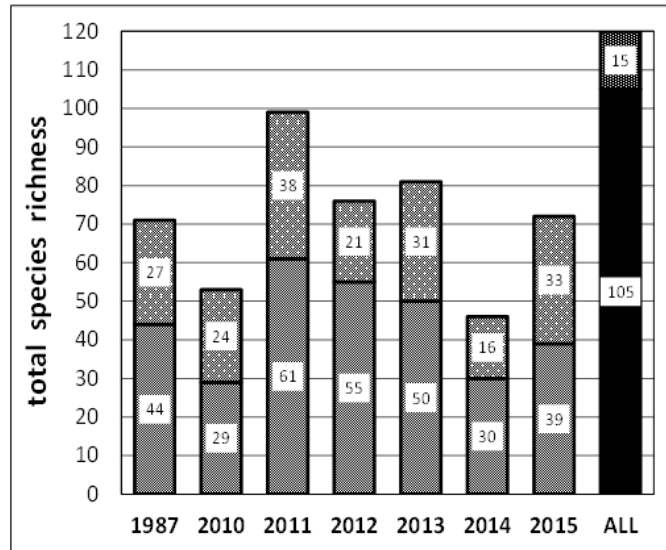


Fig. 8. Histogram providing: (i) *below*, the number  $R_0$  of recorded species, (ii) *above*, the estimated number  $\Delta$  of still unrecorded species, (iii) the sum  $R_0 + \Delta$ , that is the estimated total species richness  $S_t$ , for the seven investigated years and for all seven years pooled together. Butterfly fauna at Mount Gariwang-san, field records according to LEE *et al.* [12] for years 1987, 2010, 2011, 2012, 2013, 2014, 2015. Data directly derived from Table 2

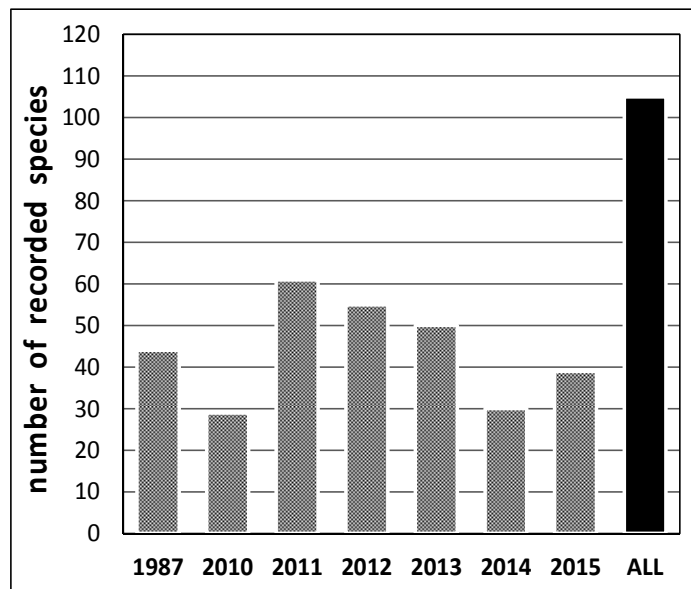


Fig. 9. Histogram of the number  $R_0$  of recorded species for the seven investigated years and for all seven years pooled together

#### 4. DISCUSSION

Lee and coworkers [12] have conducted a series of samplings of the butterfly fauna at Mount Gariwang-san (South Korea), during 1987 and from 2010 to 2015. As part of their report, the

authors provide, in particular, a detailed account of the inter-annual variations of the *recorded* species richness. Yet, as mentioned by the authors, these annual samplings are more or less incomplete, as confirmed by the significant proportion of 'singletons' in the species lists.

Accordingly, any attempt to directly extrapolate the inter-annual variations of *as-recorded* species richness in terms of “*true*” inter-annual variations of total species richness would be at least questionable. Specifically, the involved issue would be: which part of the recorded inter-

annual variations is really attributable to *true* inter-annual variations of total species richness and which part may (artificially!) results from inter-annual inequity of sampling completeness between successive annual inventories?

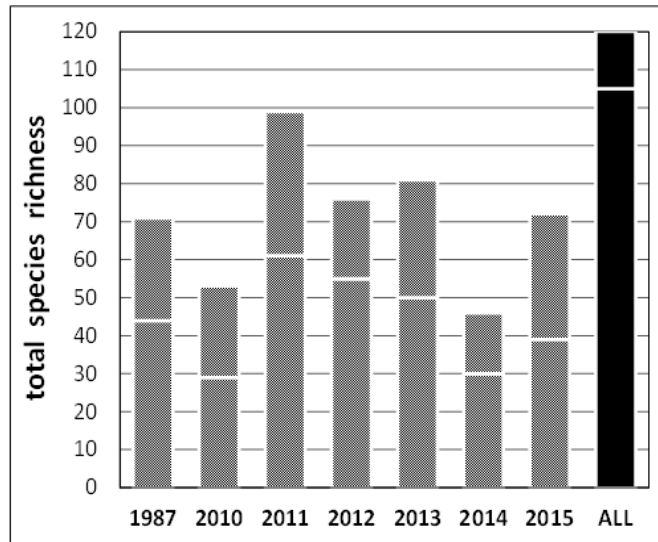


Fig. 10. Histogram of the estimated total species richness  $S_t$  for the seven investigated years and for all seven years pooled together

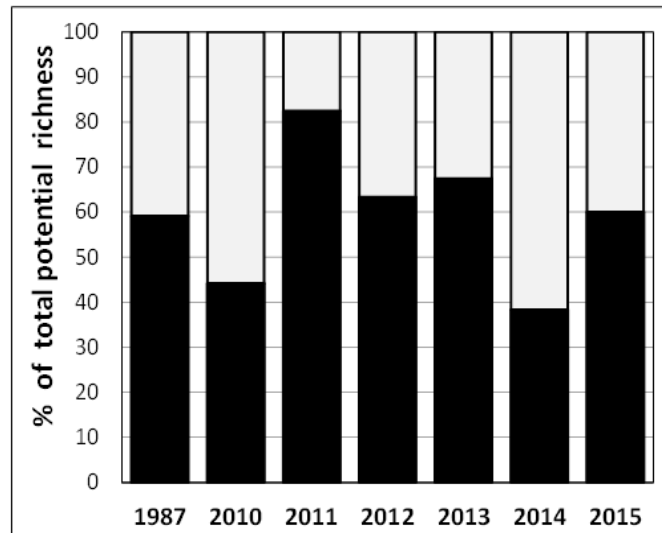


Fig. 11. The estimated proportion (%) of the “potential species richness” of the site (estimated to 120 species [at least]) which actually occurs each year. Data directly derived from Fig. 8, i.e. on the basis of least-biased extrapolations of total species richness for each year. Along these seven years the proportion of the “potential species richness” of the site which *actually occurs annually* varies between 38% and 82% (average: 59%)

One alternative solution that may be considered to answer this issue would be, of course, to further continue sampling operations, each year, so as to closely approach sampling exhaustively. While not strictly impossible this ideal procedure would have involved huge sampling efforts, year after year (Figs. 3, 4, 7 and Table 4), which, in practice, would have probably exceeded available resources in terms of both time and costs expenditure.

It is therefore necessary to consider the less satisfying, but practically unavoidable alternative solution: to extrapolate numerically the species accumulation curves up to their asymptotic levels. Nonparametric estimators may help in this respect, provided the *least-biased* type of estimator is selected, separately, for each annual survey. This precaution proves being indispensable, as highlighted by the large scatter between the estimates issued from the different types of available nonparametric estimators: Figs. 5 and 6 and Tables 1 and 3.

Thus, implementing the selection procedure in favour of the least-biased extrapolation ([10-11], see also Appendix 1) provides more reliable expectations of the “true” (total) species richness, year by year, for the surveyed butterfly assemblages of Mont Gariwang-san.

As expected, the estimated sampling completeness actually achieved vary according to years: from 54% (in 2015) to 72% (in 2012): Table 4. Incidentally, the level of sampling completeness proves weakly (and non-significantly) related to the sampling size (Fig. 1) and independent of the estimated total species richness (Fig. 2). The selected least-biased estimator also differs from year to year: Jackknife estimators at order 3, 4, 5 were selected according to studied years and Jackknife order 2 was selected for the pooled inventories along the seven studied years, while Jackknife 1 and Chao were never retained (Table 2). And the appropriate selection of the type of Jackknife estimator proves being important since the estimation of the number of missed species by inventories most often vary from simple to double (Table 1, Fig. 8) according to estimator type. In particular, this seriously question the traditional approaches that consist in either choosing *a priori* one given type of estimator, on

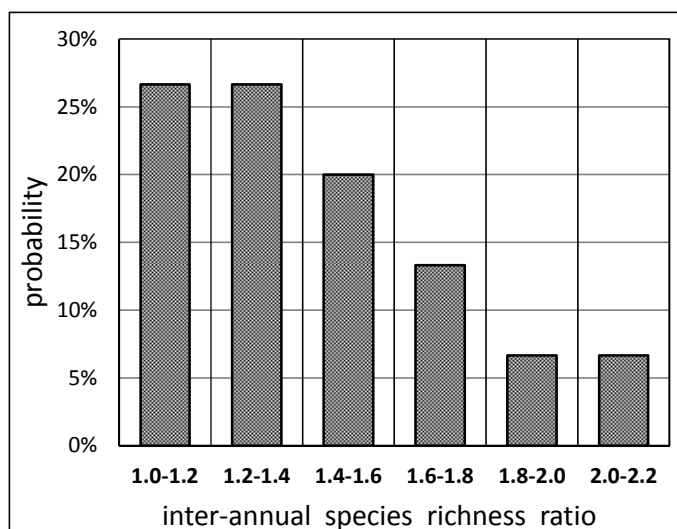
the basis of its alleged particular appropriateness, or considering all types of estimators together without choosing among them [15].

The levels of total species richness, derived from estimates of the number of missed species, substantially vary from year to year: from 46 (in 2014) to 99 (in 2011). That is more than simple to double (Table 2 and Figs. 8, 10).

The degree of inter-annual variability of true species richness may be quantified by the “inter-annual species richness ratio”, defined as the ratio ( $>1$ ) of the estimated total species richness between two years, successive or not. The histogram of values taken by this inter-annual species richness ratio at Mount Gariwang-san along the years 2010 to 2015 (Fig. 12) provides an estimate of the probabilities for the inter-annual variability of total species richness being more or less large. Thus, the estimated probabilities that the ratio exceeds 1.2, 1.4, 1.6, 1.8, 2.0, are 73%, 46%, 26%, 13%, 7%, respectively (Fig. 12).

At last, one additional interesting question is which proportion of the “overall potential butterfly richness” of the site (equated, as a first approximation, to the total species richness considering the seven years pooled together) actually occurs at any given year? From the estimations above, the proportion of the “potential species richness” at the sampled site which *actually occurs annually* varies between 38% and 82% (average: 59%), along these seven years of field survey (Fig. 11). This appreciably differs from what would be (inappropriately!) deduced from the crude (non-extrapolated) inventories: based on *non-extrapolated* data, the proportion of the “potential species richness” would vary from 28% to 58% (average: 42%): Table 2 and Figs. 8, 9.

Now, from a strictly local point of view, the effective representation of only a limited part (38% to 82%) of the “potential species richness” during each year might raise the paradox of the local perpetuation of those species that are not locally represented during one or several successive years. The likely answer lies, of course, in the concept of meta-population, according to which more or less regular flows of exchanges actually take place between adjacent localities.



**Fig. 12. Histogram of the values taken by the “inter-annual species richness ratio” for the time interval 2010 to 2015, at Mount Gariwang-san. The inter-annual species richness ratio is defined as the ratio (>1) of the estimated total species richness within any couple of years (successive or not)**

## 5. CONCLUSION

While the as-recorded annual species richness of the butterfly fauna at the sampled locality of Mount Gariwang-san was comprised between 29 and 61 species, the actual annual species richness was estimated ranging from 46 to 99 species, using the “least-biased” procedure of extrapolation of species accumulation curves. The total species richness for the seven years pooled together is estimated around 120 species. This figure may arguably be considered as likely approaching the “potential species richness” of the studied locality.

Accordingly, on the basis of these bias-reduced estimates, it becomes possible to address the difficult question of reliably assessing the proportion of the potential species richness of a site which is actually occurring each year on the site. It turns out that, here, 38% to 82% (average: 59%) of the potential species richness actually occur in any given year.

## COMPETING INTERESTS

Author has declared that no competing interests exist.

## REFERENCES

- Ahrens D, Gc YD, Lago PK, Nagel P. Seasonal fluctuation, phenology and turnover of chafer assemblages: Insight to the structural plasticity of insect communities in tropical farmlands. *Agricultural and Forest Entomology*. 2009; 11:265–274.
- Grimbacher PS, Stork NE. Seasonality of a diverse beetle assemblage inhabiting lowland tropical rain forest in Australia. *Biotropica*. 2009;41:328–337.
- Grøtan V, Lande R, Engen S, Sæther BE, DeVries PJ. Seasonal cycles of species diversity and similarity in a tropical butterfly community. *Journal of Animal Ecology*. 2012;81:714–723.
- Hamer KC, Hill JK, Mustaffa N, Benedick S, Sherratt TN, Chey VK, Maryati M. Temporal variation in abundance and diversity of butterflies in Bornean rain forests: Opposite impacts of logging recorded in different seasons. *Journal of Tropical Ecology*. 2005;21:417–425.
- Libert M. Evolution temporelle des populations de rhopalocères de deux collines boisées des environs de Yaoundé (Cameroun). *Revue d'Ecologie la Terre et la Vie*. 1994;49:151–175.
- Kishimoto-Yamada K, et al. Population fluctuations of light-attracted chrysomelid beetles in relation to supra-annual environmental changes in a Bornean rainforest. *Bulletin of Entomological Research*. 2009;99:217–227.
- Owen DF, Chanter DO. Species diversity and seasonal abundance in *Charaxes*

- butterflies (Nymphalidae). *Journal of Entomology*. 1972;SeriesA46:135–143.
8. Valtonen A, Molleman F, Chapman CA, Carey JR, Ayres MP, Roininen H. Tropical phenology: Bi-annual rhythms and inter annual variation in an Afrotropical butterfly assemblage. *Ecosphere*. 2013;4(3):36. DOI: 10.1890/ES12-00338.1
  9. Brose U, Martinez ND, Williams RJ. Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*. 2003;84(9): 2364-2377.
  10. 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>
  11. 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;1(4):1-19. DOI: 10.9734/ARRB/2016/30522
  12. Lee CM, Kim SS, Kwon TS. Butterfly fauna in Mount Gariwang-san, Korea. *Journal of Asia-Pacific Biodiversity*. 2016;9:198-204. DOI: 10.1016/j.japb.2016.02.005
  13. 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.
  14. Gotelli NJ, Chao A. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In: Levin SA (ed.) *Encyclopedia of Biodiversity*. Academic Press, Second Edition. 2013;5:195-211.
  15. Colwell RK, Coddington JA. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society London B*. 1994;345:101-118.
  16. Lande R, DeVries PJ, Walla TR. When species accumulation curves intersect: Implications for ranking diversity using small samples. *Oikos*. 2000;89(3): 601-605.
  17. 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
  18. 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: Article ID 847328, DOI:10.1155/2014/847328 <hal-01101415>
  19. 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>
  20. Béguinot J. On general mathematical constraints applying to the kinetics of species discovery during progressive sampling: Consequences on the theoretical expression of the species accumulation curve. *Advances in Research*. 2016;8(5):1-17. DOI:10.9734/AIR/2016/31791 <hal-01516141>

## APPENDICES

### Appendix 1- Bias-reduced extrapolation of the Species Accumulation Curve and the associated bias-reduced estimation of the number of missing species, based on the numbers of species recorded 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 references [10,11,18,19,20]):

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

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 [10]):

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

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

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

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

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

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:

$$* 0.6 f_2 < f_1 \leq f_2 \rightarrow \Delta_{J1} = f_1; \quad R_1(N)$$

$$* f_2 < f_1 \leq 2f_2 - f_3 \rightarrow \Delta_{J2} = 2f_1 - f_2; \quad R_2(N)$$

$$* 2f_2 - f_3 < f_1 \leq 3f_2 - 3f_3 + f_4 \rightarrow \Delta_{J3} = 3f_1 - 3f_2 + f_3; \quad R_3(N)$$

$$* 3f_2 - 3f_3 + f_4 < f_1 \leq 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{J4} = 4f_1 - 6f_2 + 4f_3 - f_4; \quad R_4(N)$$

$$* f_1 > 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{J5} = 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5; \quad R_5(N)$$

**N.B.1:** As indicated above (and demonstrated in details in [10]), 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.

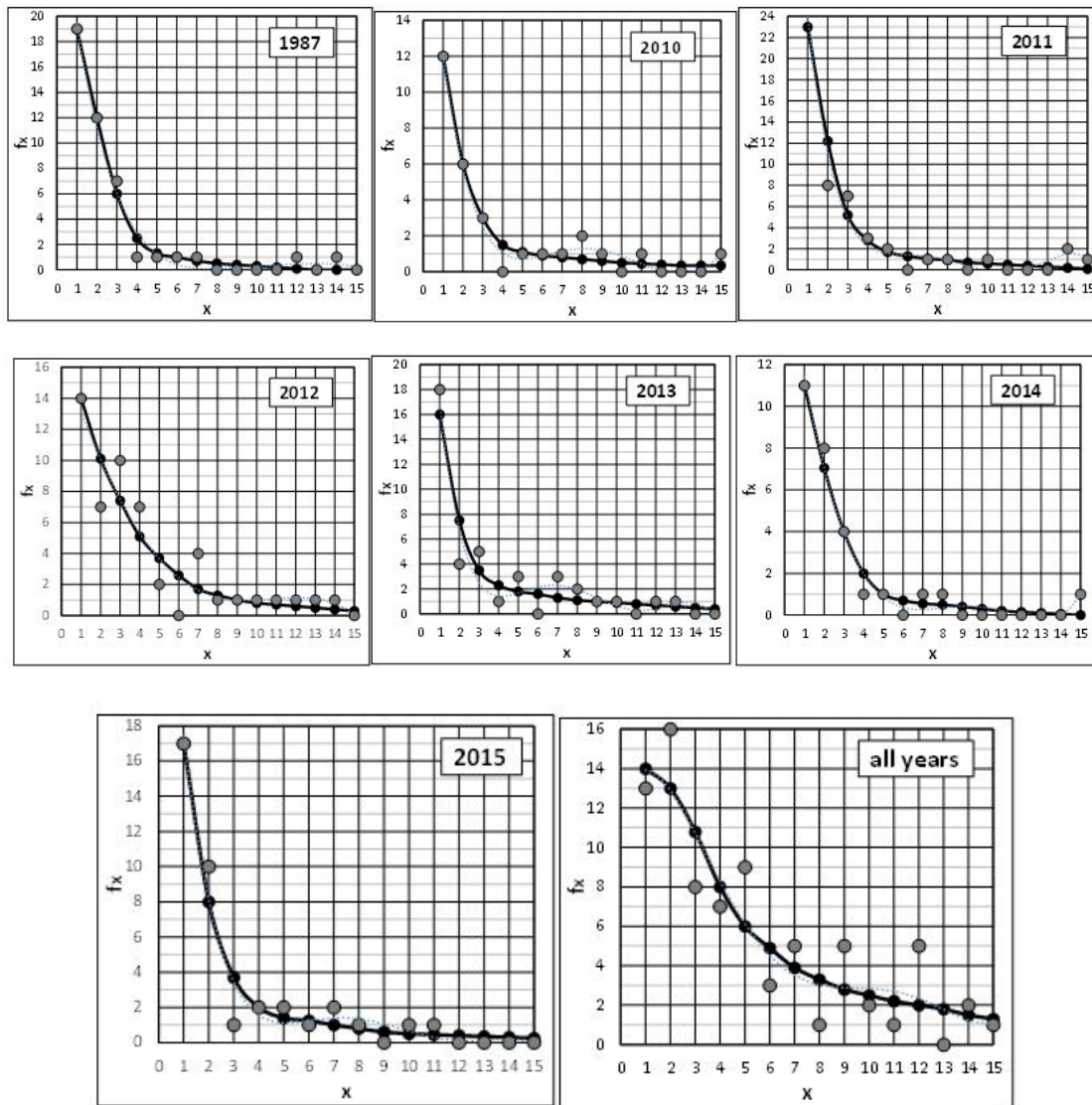
Besides, it is easy to verify that another consequence of these preferred ranges is that the selected estimator will *always* provide the highest estimate, as compared to the other estimators. Interestingly, this mathematical consequence, of general relevance, is in line with the already admitted opinion that all non-parametric estimators provide *under*-estimates of the true number of missing species [13,14]. Also, this shows that the approach initially proposed by BROSE et al. [9] – which has regrettably suffered from its somewhat difficult implementation in practice – might be advantageously reconsidered, now, in light of the very simple selection key above, of *far much easier practical use*.

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

**N.B.3:** For  $f_1$  falling beneath  $0.6 \times f_2$  (that is when sampling completeness closely approaches exhaustivity), then Chao estimator may be selected: see reference [11].



**Appendix 2- Regressions on the distributions of recorded values of  $f_x$  so as to reduce the consequences of drawing stochasticity**



**Figs. A2.1 to A2.8—The recorded values of the numbers  $f_x$  of species recorded  $x$ -times (grey discs) and the regressed values of  $f_x$  (black discs) so as to reduce the consequence of stochastic dispersion, for the seven studied years and for all years pooled together**

© 2017 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/19568>