Estimating biodiversity through extrapolation: a better function

Franco Strumia

Physics Department, Pisa University, Largo Pontecorvo 3, 56127 Pisa, Italy; e-mail: franco.strumia@unipi.it

ABSTRACT A more general hyperbolic function is proposed for interpolating the number of accumulated species when observations are extended to long time intervals. The suggested function take also in account the insects phenology and the rate in the species accumulation number.
KEY WORDS Extrapolation, Biodiversity; Hymenoptera; Mutillidae; Chrysididae; Gasteruptiidae; Apoidea.

Received 05.03.2020; accepted 26.05.2020; published online 24.07.2020

INTRODUCTION

When the species richness is sampled and investigated by means of any kind of traps, visual collection or literature study, the accumulated number of observed species increase monotonically as a function of investigation efforts. The number of accumulated species, (Number Accumulated Species NAS: S(x)) is expected to approach an asymptotic value corresponding to the extrapolated number (Sm) of species present in the studied biotope during the investigation time interval.

To estimate Sm, we need to interpolate the S(x) data as a function of sampling numbers: from the initial value of 0 for x=0 (no sampling) to the asymptotic value Sm. To perform interpolation we need a function with a convenient shape to represent the time evolution of biodiversity. The function is adjusted to experimental data by the a set of parameters: S(x) = f(x, A, B, C, ...). The parameters are real numbers to be calculated from regression.

Unfortunately, the number of available sample data S(x) is usually limited by constrains on available time for fieldwork: therefore, the function S(x)

must be carefully selected in order to have the right properties, including a fast asymptotic convergence and a minimum number of parameters.

The choice of the function may follows from the solution of a differential equation describing the population dynamic. In any case, the function S(x) must be monotonic, increasing from zero (no sampling) to the asymptotic value Sm. It is worth to note that NAS is not symmetrical in time. There is a difference between past and future. Considering long time intervals the extinction of one species do not decrease the NAS, while the observation of a new species, truly new or occasional visitor, increases the NAS and the asymptotic value. This asymmetry is not present when the richness is estimated from the Lognormal distribution (Magurran 1995). In fact, the Lognormal distribution estimates only the number of species present during the time interval used to perform samplings.

In seminal papers Colwell & Coddington (1994) and Colwell et al. (2004) discussed the biodiversity evaluation through extrapolation. Colwell & Coddington suggested to interpolate the S(x) data to an Hyperbolic function (the well known Michaelis-Menten function):

1) S(x) = x/(K+x) Sm

where Sm is the estimated species number and K a constant (the number of samplings needed to observe Sm /2).

Function 1) increases monotonically from zero to Sm. On the contrary, the derivate of function 1) decreases monotonically from a maximum (for x=0) to zero for large x values (Fig. 1).

Thus, the above eq.1) cannot take in account possible changes in the rate of new records accumulation following from fluctuations in phenology. Of course this is not adequate when the sampling is performed all year around. In fact, a convenient and practical protocol to perform samplings is to capture specimens continuously for at least (possibly) one full year, especially if traps are positioned on sites not easy to be visited, such as isolate islands or dangerous countries, or when special permits are required.

It is well known that the individuals emergence rate is variable and reach its maximum in spring (at least in Palaearctic climate), depending on the species-group and influenced by seasonal fluctuations in temperature, humidity, etc. For practical reasons, at least in Palaearctic region, is convenient to continuously operate traps from early spring to late fall. In such a case, it is well known that the species emergence begins slowly and reaches its maximum in June-July, then significantly decreases from July-August, especially in the case of hot and dry summers as in mediterranean sites. This heuristically suggests the need of a simple hyperbolic function, with a derivate not monotonic. We suggest a generalized hyperbolic function as:

2) $S(x) = x^{n}/(K+x^{n}) Sm$

where "n" is an additional constant and a real positive number. When n>1, eq. 2) has the needed inflexion point (a derivative maximum) and, in addition, approaches faster the asymptotic value Sm (Fig. 2). This is also an additional favorable property because in a year interval, the samplings number may be reduced due to constrains in operating the trap, possibly in remote locations or not easy to be visited (Fig. 3).



Figure 1. Plot of the Michaelis-Menten function for Sm=20 (red line) or of its derivative (blue line). Note the slow approach to the asymptotic value of 20.



Figure 2. Plot of the generalized hyperbolic function (eq.2) (red line) and of its derivative (blue line) for Sm=20 and n= 2. Note the faster approach to the asymptotic value. The derivative is 0 for x=0, and it reaches a maximum and then decreases monotonically to zero.



Figure 3. Plot of eq. 2). for four n values, from bottom to top: n=1, 1.5, 2 and 3 respectively. The asymptotic value is fixed at Sm= 20. For n=3 the asymptote is reasonably approached after only 4 to 6 samples.

EXPERIMENTAL RESULTS

Some examples are discussed to illustrate the better interpolations available from eq.2), when the constant n is let free to better adjust the function to the experimental data. When the experimental points are too few the interpolation with eq.2) (three free parameters instead of two) may become inadequate and the calculated Sm value becomes unstable. In such a case, a better and easer interpolation can be obtained by fixing n to have any value in the range 1.5 < n < 2.5, thus reducing the free parameters to two. In such a case, a good extrapolation is always available, as it will be shown. In the following examples, the non-linear interpolation is calculated with the fast converging Marquardt method and commercial software.

In recent years, a number of Malaise traps operated in Tuscany, Sardinia, Sicily, Piedmont and small Tyrrhenian islands in order to investigate Hymenoptera biodiversity. Major evidence was given to Aculeate Hymenoptera, mainly for the relevance as impollinators (Abrol, 2012).

In 2016, two Malaise traps were positioned in the Monterufoli-Caselli Nature Reserve in southern Pisa Province ("Colline metallifere" Tuscany), in order to investigate the influence of a humid zone on entomocoenosis (Strumia et al., 2017). The trap captured continuously from April 28th to September 8th and the material was sorted and identified in nine samples. The observed number of the accumulated species of Hymenoptera Mutillidae is shown in figure 4, together with the best interpolations available from eqs.2 and 1, respectively. The predicted asymptotic values are respectively 16< Sm=18 <19 (95%) confidence interval, n= 2.83) and 12<S=33<54 with the Michaelis-Menten function (n=1, 95% confidence interval, n=1). Not only the interpolation with the Michaelis-Menten function is clearly unable to properly interpolate the experimental data, but the predicted richness is too large and in disagreement with the results already observed in Tuscany (Strumia, 2007; Strumia & Pagliano, 2007).

In the same site, the species of Hymenoptera Gasteruptiidae were also counted. The interpolated data gives a similar result as shown in figure 5. The calculated Sm values are respectively Sm =19 +/- 6 (n=1) and Sm =14 +/-1 (n=4). In Italy, only 20 species of Gasteruptiidae are known (Pagliano & Scaramozzino, 2000), thus the prediction with

the Michaelis-Menten function is unreasonable. On the contrary, the interpolation with eq.2 is inside the acceptable range and shows that the Caselli reserve hosts a rich Gasteruptiidae biodiversity, as already observed for others insects



Figure 4 . Number of Mutillidae species captured in the "Caselli Nature Reserve" in 2016 by Malaise traps; the blue line shows the best interpolation from eq.2 and the red line the best interpolation from eq.1; the broken lines show the 90% confidence range of eq.2; eq. 1 is clearly unable to properly interpolate the experimental data.



Figure 5. Number of the accumulated Gateruptiidae species observed in the "Caselli Nature. Reserve" in 2016 and captured by a Malaise trap (blue dots); blue line shows the best interpolation with eq.2; red line those with eq.1.

groups: Apoidea, Chrysididae, Pompilidae (Filippi & Strumia, 2018).

Malaise traps operated in the "San Rossore Nature Reserve" near Pisa town. San Rossore extends along the Mediterranean coast and has several humid and wooded zones on the back of coastal sand dunes. These sites are favorable for sand nesting insects. From the captures of 2015 we studied the Mutillidae wasps obtaining a similar result as Gateruptiidae (Fig. 6). In this case, a larger number of sampling data is available and the interpolations is done whit n=1 and n=3 respectively. For n=3 a good interpolation is obtained (R2 = 0.98). The richness prediction is in agreement with the more recent and repeated observations of Mutillidae around Pisa (Strumia, 2007). On the contrary, the result from the Michaelis-Menten function (n=1) badly fits data and give an unreasonable large value for the local richness: 14<Sm= 27< 42 (95% confidence intervals).

In 2002, a Malaise trap operated into "Nebrodi Nature Reserve", Sicily (Galati Mamertino, 38.01'.63"N, 14.47'.30"E, 950 m. a.s.l.). Hymenoptera Crysididae were captured in 8 times intervals and the results from the interpolation with eqs. 1 and 2 are shown in figure 7. The result is similar to that of Mutillidae and Gateruptiidae. The predicted Crysididae richness is Sm = 17 (+/- 2 species, 95% confidence interval). Again, the interpolation with eq. 1 is inadequate.

In previous examples, Hymenoptera families with a moderate number of species were considered. Now, a more species-rich group is considered: the superfamily Apoidea of the Hymenoptera (about 1800 species in Italy: Pagliano 1994). More then 140 species of Apoidea were captured in 2017 in "Caselli Nature Reserve" by a Malaise trap. The Apoidea NAS interpolation is shown by using the functions 1 and 2 respectively (Fig. 8). Apparently, the differences are small, but the numerical results shown in Table 1 demonstrates the better accuracy available from function two (for R² index and the "Absolute sum of squares").

MULTIPLE YEARS DATA COMBINATION

In favorable conditions is possible to repeat the observations for two or more full years. In such a case experimental data can be ordered and combined to obtain a better interpolation from a larger data set. If all data are combined and ordered following the beginning date of each sampling interval, a larger number of S(x) data points becomes available, and a more accurate interpolation is possible.

This is illustrated in the following study on the Hymenoptera Chrysididae biodiversity in a vineyard of southern Piedmont (north-western Italy).



Figure 6. Interpolations of Mutillidae NAS observed in San Rossore Nature Reserve in 2015 (black dots); black line with n=3; red line with n=1.



Figure 7. Number of Hymenoptera Crysididae species captured in Nebrodi Nature Reserve in 2002; blue line interpolation from eq.2; red line interpolation from eq.1; broken lines show the 90% confidence range from eq.2 (Sm =17 + 2 species).

In 2016 and 2017, the Pisa University operated a Malaise trap in a vineyard near Barbaresco (Cuneo, Piedmont, 170 meters a.s.l., 44°4'26"N, 08°04'26"E). The trap was positioned near the Ecotone between the vineyard and a wood along Tanaro river (Strumia et al., 2018). In 2016 the Malaise trap captured 24 species (132 individuals) of Hymenoptera Chrysididae, operating from May 18th up to October 17th. In 2017, a notable larger number of Chrysididae species was captured (39 species and 216 individuals) between May 13th -October 10th.

The corresponding S(x) data were combined, as proposed, giving a larger number (16). The interpolation of all data is shown in figure 9, both separately for each year and for the combination of the two years. The numerical results, obtained with and without combination of data, are also shown in Table 2.

In the combined NAS, the general shape, with an inflexion point, is retained, as predicted by eq.2). The statistical confidence and also the fit goodness is also improved.

The remarkable difference in the number of captured species (about plus 162% in 2017) by a trap operating in the same position and orientation is unexplained and suggests the need of further observations. Possibly meteorological fluctuations in temperature, humidity, etc. can largely influence the local biodiversity. Only data recording over a longer time span would provide some statistical correlation and inference.

In any case, this result highlights that the biodiversity measure, based on observations on one single year, remains uncertain, even if statistically sound. In fact, the result estimates the local richness for the studied time interval. Only repeated observations and data accumulation would allow to observe long term trends in populations richness.

CONCLUSIONS

From the discussed experimental results, it follows that the simple hyperbolic Michaelis-Menten function does not describe properly the time evolution of observed number of accumulated species (NAS), especially when the measurements are extended to a full year, from spring to autumn. The use of more or less automated traps, like Malaise traps, is particularly convenient if traps are positioned on sites not easy to be visited, such as isolate islands or dangerous countries, or when special permits are required.

In such cases, the number of available data is likely to be small and an efficient function to per-



Figure 8. NAS Interpolation of Apoidea speciose group captured in the "Caselli Nature. Reserve" in 2017; red line with eq.1, blue line with eq.2.



Figure 9. Interpolation of NAS of Hymenoptera Chrysididae captured at Barbaresco (CN) in 2016 (round black dots and red broken line) and in 2017 (pink square dots and pink broken line). The combined S(x) data of 2016 and 2017 are shown in blue (triangular dots and blue line).

Franco Strumia

Funct. 1 (Michaelis- Menten)		Function 2	
Best-fit values		Best-fit values	
Vm	169.7	Vm	153.8
n	1	n	1.348
Km	1.465	Km	1.394
95% Confidence		95% Confidence	
Intervals		Intervals	
Vmax	158.1 to 181.2	Vmax	137.4 to 170.1
		n	0.8516 to 1.845
Km	1.070 to 1.860	Km	1.064 to 1.723
Goodness of Fit		Goodness of Fit	
R ²	0.9751	R ²	0.9833
Absolute Sum of Squares	137.1	Absolute Sum of Squares	92.21

Table 1. Comparison of interpolation results ontained from functions 1 and 2 for theHymenoptera Apoidea captured in the "Caselli Nat. Res." in 2017.

			0 1 1
Best-fit values	2016	2017	Combined
Dest-int values			2016 and 2017
N° of species predicted	24.72	47.53	52.16
n	2.319	2.282	1.966
K	1.769	29.03	30.11
Std. Error:			
Vm	0.2425	2.767	1.705
n	0.1106	0.2031	0.1255
K	0.08142	5.619	4.841
95% Confidence Intervals			
Vm	24.05 to 25.39	40.76 to 54.30	48.47 to 55.84
n	2.012 to 2.626	1.785 to 2.779	1.695 to 2.237
Km	1.543 to 1.996	15.28 to 42.78	19.65 to 40.57
Goodness of Fit			
Degrees of Freedom	4	6	13
R ²	0.9984	0.9966	0.9957
Absolute Sum of Squares	0.3019	5.176	13.67
N° of analyzed points	7	9	16

Table 2. Best values from the interpolation, with eq.2, of Chrysididae NAS species number captured at Barbaresco (CN) in 2016 and 2017.

form the statistical interpolation is mandatory. The generalized hyperbolic function as proposed by equation 2 is a more convenient solution when samples are collected during a full year. This is illustrated for locations with Palaearctic climate but it will work properly in any climate with large seasonal fluctuations (monsoon). The suggested function is reliable even if the number of samplings is minimal (five or six).

The number of accumulated species describes an open system where occasional visiting species increase the real richness of the studied site. Large yearly fluctuations have been recently reported for the local estimated richness (as large as 100%). In alternative to the number of accumulated species, we can use two additional and mathematically independent algorithms, namely: the "Lognormal Distribution" and the "Species Area Relationships" (SARS) (Magurran, 1995). These mathematically independent algorithms may provide a crosscheck for results correctness.

All this statistical inference methods give local prediction strongly correlated to the time intervals used to perform the experimental observations. In any case, such count of the species number is essential to observe and study temporal evolution and fluctuations in populations richness.

In some cases, the proposed function, having three free parameters, do not produce a good fit when the available data are too few. In this case, we can obtain an improvement by fixing n to be a number between 2 and 3, thus reducing the free parameters to two.

ACKNOWLEDGEMENTS

The comments of my friends and colleagues Pier Luigi Scaramozzino (Pisa, Italy), Luciano Filippi (Cecina, Livorno, Italy), and Guido Pagliano (Torino, Italy) stimulated and produced the ideas presented and discussed in this contribution. Pier Luigi Scaramozzino (Pisa University, Italy), provided also a text review.

REFERENCES

- Abrol D.P., 2012. Pollination Biology. Springer Heidelberg, 792 pp.
- Colwell R.K. & Coddington J.A., 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society, Series B 345: 101–118. http://dx.doi.org/10.1098/rstb.1994. 0091
- Colwell R.K., Chang Xuan Mao & Jing Chang, 2004. Interpolating, Extrapolating, and Comparing Incidence-Based Species Accumulation Curves. Ecology, 85: 2717–2727. http://dx.doi.org/10.1890/03-0557
- Filippi L. & Strumia F., 2018. Imenotteri Pompilidi della riserva naturale di Monterufoli-Caselli (Pisa). Proceedings of "Codice Armonico, 2018", vol. VII: 34– 39.
- Magurran A., 1995. Measuring Biological Diversity. Blackwell Publishing, 236 pp.
- Pagliano G. & Scaramozzino P., 2000. Gasteruptiidae italiani (Hymenoptera: Evanioidea). Bollettino del Museo Regionale di Scienze Naturali di Torino, 17: 5–38.
- Pagliano G., 1994. Hymenoptera Apoidea. In: Minelli A., Ruffo S. & La Posta S. (Eds.), Checklist delle specie della fauna italiana, 106. Calderini, Bologna
- Strumia F., 2007. Gli Imenotteri Mutillidi della Riserva Naturale di San Rossore (Pisa) e dei dintorni di Pisa. Frustula entomologica, 30: 93–102.
- Strumia F. & Pagliano G., 2007. Contributo alla conocienza dei Mutillidi italiani (Hymenoptera Scolioidea). Bollettino del Museo Regionale di Scienze Naturali di Torino, 24: 25–110.
- Strumia F., Gianasso D., Filippi L., Massa B. & Pagliano G., 2017. Influence of a humid zone on entomocoenosis diversity. Il Naturalista siciliano, 41: 123– 138
- Strumia F., Loni A., Lucchi A., Renato Ricciardi R. & Scaramozzino P., 2018. La biodiversità degli Imenotteri Crisididi dei vigneti del Piemonte meridionale. Proceedings of "Codice Armonico 2018", vol. VII: 119–123.

.