

Never ending story: a lesson in using sampling efficiency methods with ground beetles

Michał Żmihorski · Paweł Sienkiewicz ·
Piotr Tryjanowski

Received: 6 March 2012 / Accepted: 12 July 2012 / Published online: 27 July 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Biological inventory is a crucial activity in life sciences field research. However, it is sometimes time-consuming and laborious to take representative samplings of communities, especially in the case of invertebrates. In this paper, we address the issue of sampling efficiency and its influence on obtained results. As a study system, we used data on epigeic carabid beetles (Carabidae) collected in 1999–2001 in the Warta River valley of western Poland. We trapped a total of 17,722 individuals belonging to 108 species. However, due to rarefaction methods, the expected number of species was estimated at 134–140, suggesting that from 26 to 32 species are missing from the material, even expressed as a huge number of collected specimens. The estimated probability that another captured individual will represent a new species (i.e. a species that was not already recorded) is 0.0010. In order to record all the species present in the study area, another 193,338 individuals need to be sampled (abundance-based approach) or another 1,871 samples need to be collected (incidence-based approach). This means that the collected material should be 10.9 times greater (or 7.9 times greater for incidence-based data) than what was actually collected in order to record all the species present in the study area. The

results show that, in practice, full inventory is simply nearly impossible to achieve, and this knowledge should be included in inventory planning. Therefore, we argue that species accumulation curves and unseen species estimators need to be carefully examined and threshold probability of detecting a new species should be built into the design of inventory science. The ratio between recorded and estimated species richness and the estimated efficiency of further sampling can be easily computed with available freeware software and should be incorporated when performing biological inventories.

Keywords Biodiversity inventories · Species list · Carabids · Community structure · Faunistic estimators

Introduction

Biological inventory is a crucial point of the life sciences (May 1992). Simply stated, inventories provide the foundation for improving the applied pursuits of sustainable resource management and conservation biology (Magurran 1996). Even from a theoretical point of view, May (2010) interestingly argues that if aliens visited our planet, one of their first questions would be, “How many distinct life forms—species—does your planet have?” He also pointed out that we would be “embarrassed” by the uncertainty of our answer. Our knowledge about the number of species in particular places of the planet is obviously increasing, but on the other hand, as new statistical techniques are developed, we also see how great our ignorance is (Certain et al. 2011). This narrative story by Robert May (2010) well illustrates the fundamental nature of knowing how many species there are on Earth and our limited progress with this research topic thus far (May 1992, 2010, Storks 1993).

M. Żmihorski (✉)
Museum and Institute of Zoology, Polish Academy of Sciences,
Wilcza 64, 00-679 Warsaw, Poland
e-mail: zmihorski@miiz.waw.pl

P. Sienkiewicz
Department of Entomology and Environmental Protection,
Poznań University of Life Sciences, Dąbrowskiego 159, 60-594
Poznań, Poland

P. Tryjanowski
Institute of Zoology, Poznań University of Life Sciences,
Wojska Polskiego 71C, 60-625 Poznań, Poland

However, even among well described taxa, the required size of a sample to both establish local diversity as well as determine the number of species in an entire taxa is still a topic of discussion (Novotny and Basset 2000; Martikainen and Kouki 2003; Chao et al. 2006, 2009). Obviously, indirect estimates remain uncertain due to the use of controversial approaches (just assuming that a number of species is equal to the number of caught species, which is typical in the interpretation of species lists for a particular area). In that way constructed lists, containing major information particularly on species number and species composition with that wrong assuming are afterwards used as the two fundamental characteristics of animal communities. Moreover, they are broadly used as information on the spatio-temporal distribution of natural resources and as an input for biogeographical and macroecological studies (e.g. Brown 1995; Lennon et al. 2004). Species lists are used in selecting areas for conservation (e.g. biodiversity hotspots), as bioindicators and inputs to compare habitats (Myers et al. 2000). For example, in national conservation plans, species pools of different regions must be comparatively assessed and their changes monitored over time. Two specific problems arise: (1) species diversity must be standardized per area, because regions differ in size, and (2) the diversity measure should take into account how common or rare a particular species is at the regional scale (Tista and Fiedler 2011).

However, quantifying species diversity at a regional scale is quite challenging because of the difficulties in measuring species abundance and distribution. Even in well recognized taxa, it is difficult to take saturated samples. Sampling in the field, in turn, can be characterized by different sampling efforts and the recorded number of species usually does not contain some of the species present in the investigated area (Chao et al. 2006). This is because of imperfect detection and the rarity of some species. Unfortunately, the problem of undetected species seems to be insufficiently addressed in faunistic explorations, both those published in the literature and those collected for biodiversity management purposes. Normally, the main goals of most arthropod inventories commonly fall into one of two categories: strict inventory or community characterization (Longino and Colwell 1997). Strict inventory generates a nearly comprehensive species list for a discrete spatiotemporal unit, which requires species-level identification of samples (Longino and Colwell 1997). On the other hand, the proper construction of some taxa's species list is nearly impossible, especially with invertebrates, because representative sampling of communities is time-consuming and laborious (Longino and Colwell 1997; Tista and Fiedler 2011). Therefore, it is useful to make a trade-off between the time-consuming job of collecting and identification and fully establishing a local species list (Longino and Colwell 1997; Tista and Fiedler 2011).

In the present paper, we attempted to address the efficiency of field sampling of ground beetles in natural habitat. For this purpose, we used large collections of carabids from grasslands in a natural river valley. Ground beetles constitute a species-rich and relatively well known taxonomic group of invertebrates, commonly used for ecological studies and acting as bioindicators (Luff 2007; Pearce and Venier 2006; Rainio and Niemelä 2003). We applied species richness estimations and sampling efficiency assessment methods to show that even using intensive sampling at a small spatial scale, we are still far from able to completely recognize the carabid community. In a more practical context, we aimed to show that methods of endless inventorying need to be revised with the logistical, financial and ecological costs taken into account.

Methods

We used data on epigeic carabid beetles (Carabidae) collected in 1999–2001 in the Warta River valley, western Poland. The sampling was conducted at 4 sites, 300–600 m. apart, and covered by grassland habitat mowed 1–2 times per year. A more detailed description of the habitat characteristics and plant species composition is given in Sienkiewicz (2003) and Sienkiewicz and Konwerski (2004). We used pitfall-traps with a diameter of 18 cm and a height of 14 cm. At each site, 9 such traps were set up and placed in transect every 2 meters. The traps were filled with ethylene-glycol and detergent to reduce surface tension. The glycol was replaced at least every month. The traps were emptied every 7–10 days from the beginning of April to mid-November. As a result, we collected 237 samples, among which 231 contained at least one individual. The samples were used as replications in further statistical analyses.

We conducted an estimation of sample size needed to record all species present in a given area proposed recently by Chao et al. (2009). The proposed estimation method (Chao et al. 2009) is based on estimating undetected species in samples. On the basis of the number of singletons (species with only one individual), doubletons (species with only two individuals) as well as uniques (species that occur in only one sample) and duplicates (species that occur in only two samples), it is possible to assess the number of species still remaining undetected (i.e. are absent from the collected samples). Such species richness estimation is widely used in ecological research (e.g. Chao et al. 2006; Banaszak-Cibicka and Żmihorski 2012). We have plotted the Chao1 and Chao2 estimators against samples size to see whether the estimates were still dependent on sample size or stabilized towards reaching

the full data set. However, recently Chao et al. (2009) provided algorithms that enable the estimation of sample size (expressed by the number of individuals or number of samples) needed to ensure the detection of all unseen species. We used the excel spreadsheet provided with the paper of Chao et al. (2009) for the computations and calculated: (1) the estimated number of undetected species, (2) the probability that another individual will bring a new (formerly unrecorded) species, (3) the estimated sample size (expressed by the number of individuals or number of samples) that need to be collected to record 95 and 100 % of ground beetle species in the study area.

Results

We trapped 17,722 individuals belonging to 108 species. Among the dominants zoophagous were the most abundant [e.g. *Patrobus atrorufus* (Stroem), *Amara lunicollis* Schiödte, *Loricera pilicornis* (F.), *Bembidion biguttatum* (F.), *Chlaenius nigricornis* (F.), *Dyschirius globosus* (Herbst), *Carabus granulatus* L., *Pterostichus melanarius* (Ill.)]. Majority of species were hygrophilous and mesohygrophilous whereas xerophilous and eurytopic were much less common. Among the sampled species several less common or even rare were recorded including *Limodromus longiventris* Mann., *Amara fulvipes* (Aud.-Serv.), *Carabus clatratus* L., *Blethisa multipunctata* (L.), *Pterostichus gracilis* (Dej.) and *Oodes helopioides* (F.).

The expected numbers were 140 (abundance-based approach) and 134 species (incidence-based approach). The Chao1 species richness estimators were still dependent on sample size toward reaching the whole sample size, however the Chao1 estimator reached a plateau at the sample size denoting ca. 170–180 samples (Fig. 1).

According to the two methods, from 26 to 32 species are still missing from the material. The estimated probability

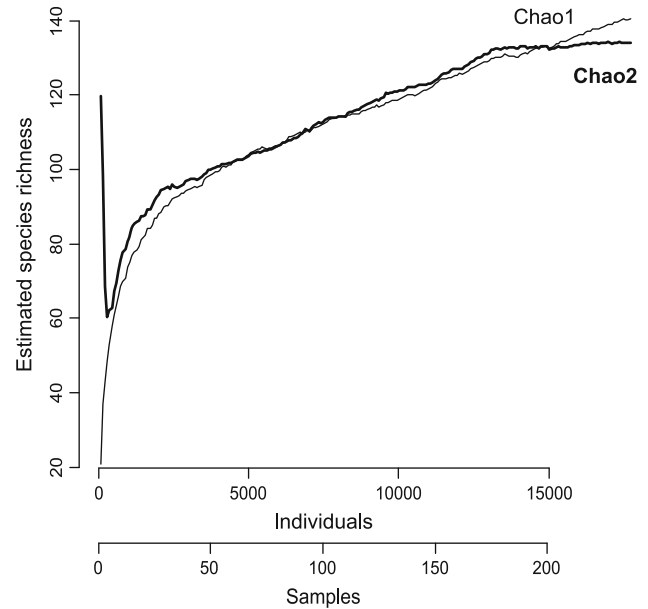


Fig. 1 Species richness estimators (Chao1 and Chao2) as a function of samples size expressed as number of individuals sampled and number of samples

that another captured individual will represent a new species (i.e. a species that was not already recorded) is 0.0010. In order to record all the species present in the study area, another 193,338 individuals need to be sampled (abundance-based approach) or another 1,871 samples need to be collected (incidence-based approach). This means that the collected material should be 10.9 times greater (or 7.9 times greater for incidence-based data) than actually collected in order to record all the species present in the study area. In order to record 95 % of species the sample size needed to collect is substantially smaller. The sampling effort necessary for complete detection is presented in Fig. 2.

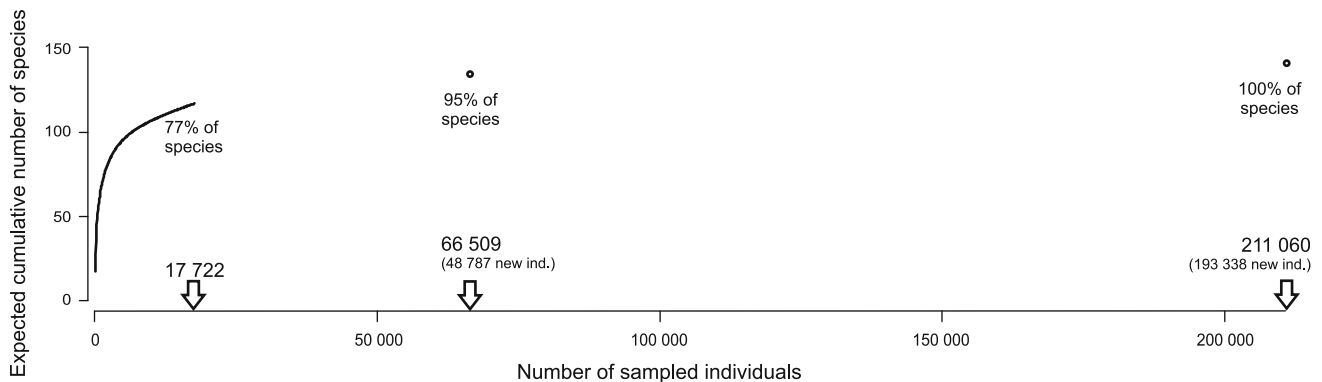


Fig. 2 Expected cumulative number of species as a function of the number of collected individuals (line)—among 17,722 individuals 108 species were observed, which constitutes 76.9 % of all species

estimated for the study area. The estimated sample sizes needed for detecting 95 and 100 % of all species in the study area are marked with arrows

Discussion

Accompanying the rapid loss of biodiversity in many parts of the globe is a crisis in biodiversity knowledge (May 1992, 2010; Mooney and Mace 2009). In many taxonomic groups, the delimitation of species is still unclear, and we understand little of their distributions and potential uses. Making informed management decisions always requires some level of biodiversity data (Mooney and Mace 2009), and many believe that ultimately we have a moral responsibility to know and steward the other taxa with which we share the planet. Yet field inventory proceeds slowly and is uncoordinated. Traditional taxonomic revisionary activity is restricted to a few taxa, and carabids are among this group (Luff 2007; Pearce and Venier 2006; Rainio and Niemelä 2003).

In the case of invertebrates, sampling is most commonly related to killing animals. This is because sampling methods commonly use killing traps—pitfall-traps as in our case, window traps and many others, where animals are placed in containers with killing and fixing substances (alcohol, glycol and others). This also holds true in the case of our study. Second, the determination of individual to species level usually is based on detailed features of its morphology, and in many cases, preparation of e.g. copulatory organs is necessary. It is impossible to determine the species of several ground beetles without a precise inspection of morphological traits (Müller-Motzfeld 2004 and see also: Pearce and Venier 2006), which in turn means that individuals have to be killed before identification. The exact determination to species level of the majority of invertebrates from various groups in the field is difficult, despite the fact that in some groups, e.g. butterflies, this is possible to some extent (e.g. butterfly monitoring in UK—Asher et al. 2001). As a consequence, simple faunal explorations and taxonomic studies may have a relatively high ecological footprint (Rodríguez-Estrella and Blázquez-Moreno 2006; Tista and Fiedler 2011). Therefore, knowledge on the sampling efficiency and expected rate of gain of the number of species with increasing sampling effort is crucial in order to optimize sampling.

Our study shows that the actual probability of another collected individual belonging to a new species is extremely low and denotes 0.001. Moreover, this value will decrease with increasing sample size. We detected 108 species among 17,722 individuals and computed that adding an additional 17,722 individuals (i.e. twofold increase in sample size) will bring us just 14 new species. A further 17,722 individuals (i.e. 35,444 additional individuals in total) will let us detect just another 8 species whereas another 17,722 (53,166 additional individuals in total) just 4 more new species. The important question is whether such a low probability of detecting new species is

high enough to continue the sampling. Of course, the aim of the investigation, availability of financial and time resources, and importance of a given study need to be taken into account to address this question. However, estimating the probability is invaluable when one needs to decide whether further field work is still profitable.

Another interesting issue is the problem of reaching the complete species lists. However, one may ask: is it really possible? On one hand, the number of species in a given time and place is constant and, theoretically, total species number can be indeed surveyed, which in turn gives us highly valuable information on species number and composition. On the other hand, number of species changes over time, and it is very likely that new species immigrated and old species disappeared from their sites in the course of the survey. In this regards, it may make more practical sense to set e.g. 95 % of species as desired target for surveys, rather than 100 % of species which, arguably, will be also inaccurate—once achieved, it may already include some species no longer present.

The collection and species level determination of another nearly 200,000 carabids is logistically and financially difficult, or even simply impossible in our case. Therefore, our study shows that even at a restricted spatial scale, the complete investigation of carabidofauna is unrealistic in practice. As this seems to be the general pattern in entomological studies (Novotny and Basset 2000; Chao et al. 2009), one can conclude that we should correct all further analysis (e.g. ecological, biogeographical, macroecological, etc.) and interpretation of results (e.g. faunal similarity) for the undetected species. What is important, it seems that sampling designs and plans of zoological explorations should take into account current knowledge on sampling efficiency. The presented example of carabids in the Warta valley leads to a conclusion that the advanced statistical tools available for planning zoological sampling should be used, if possible, as they can be helpful in designing zoological explorations. Unfortunately, despite the high practical meaning of the studies of Chao et al. (2009) that appeared in April 2009 in the ISI Web of Science database, it has been cited <10 times in strictly entomological studies (checked in January 2012).

As endless zoological sampling may be destructive to local fauna (Rodríguez-Estrella and Blázquez-Moreno 2006; Tista and Fiedler 2011) and requires huge financial and time resources, we recommend that more attention be paid to sampling efficiency and to control the sampling efficiency during the field work. More specifically, we recommend computing available estimators for part of the material that is planned to be collected and deciding (with the help of the procedures described in Chao et al. 2009) whether the expected probability that another captured individual will represent a new species is high enough to

continue the sampling. We are sure that knowledge about the predicted efficiency (expressed as the probability of another individual belonging to a new species) of further sampling is much lower than is acceptable from an economic and ecological point of view in many of the biodiversity surveys actually conducted. In such cases, fieldwork should be stopped. As a consequence, currently ongoing explorations and monitoring programmes need to be checked for expected sampling efficiency in the context of reducing their ecological footprint (see also: Longino and Colwell 1997; Rodríguez-Estrella and Blázquez-Moreno 2006).

Acknowledgments We are grateful to P. Chylarecki and to two anonymous referees for helpful comments and valuable discussion. B. Przybylska kindly improved the English.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Asher J, Warren MS, Fox R, Harding PT, Jeffcoate G, Jeffcoate S (2001) The Millennium atlas of butterflies in Great Britain and Ireland. Oxford University Press, Oxford
- Banaszak-Cibicka W, Żmihorski M (2012) Wild bees along an urban gradient: winners and losers. *J Insect Conserv* 16:331–343
- Brown JH (1995) Macroecology. University of Chicago Press, Chicago
- Certain G, Skarpaas O, Bjerke JW, Framstad E, Lindholm M, Nielsen JE, Norderhaug A, Oug E, Perderson HC, Schartau AK, Storaunet KO, Van der Meeren G, Aslaksen I, Engen S, Garnåsjordet PA, Kvaløy P, Lillegård M, Yoccoz G, Nybø S (2011) The nature index: a general framework to synthesize knowledge on the state of biodiversity. *PLoS One* 6(4):e18930. doi:10.1371/journal.pone.0018930
- Chao A, Li PC, Agatha S, Foissner W (2006) A statistical approach to estimate soil ciliate diversity and distribution based on data from five continents. *Oikos* 114:479–493
- Chao A, Colwell RK, Lin CW, Gotelli NJ (2009) Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90:1125–1133
- Lennon JJ, Koleff P, Greenwood JJD, Gaston KJ (2004) Contribution of rarity and commonness to patterns of species richness. *Ecol Lett* 7:81–87
- Longino JT, Colwell RK (1997) Biodiversity assessment using structured inventory: capturing the ant fauna of a tropical rain forest. *Ecol Appl* 7:1263–1277
- Luff ML (2007) The Carabidae (ground beetles) of Britain and Ireland. Handbooks for the identification of British insects, volume 4, part 2, 2nd edn. Royal Entomological Society, St Albans
- Magurran AE (1996) Ecological diversity and its measurement. Princeton University, Princeton
- Martikainen P, Kouki J (2003) Sampling the rarest: threatened beetles in boreal forest biodiversity inventories. *Biodiv Conserv* 12:1815–1831
- May RM (1992) How many species inhabit the earth? *Sci Amer* 10:18–24
- May R (2010) Tropical arthropod species, more or less? *Science* 329:41–42
- Mooney H, Mace G (2009) Biodiversity policy challenges. *Science* 325:1474
- Müller-Motzfeld G (eds) (2004) Bd. 2 Adepfaga 1: Carabidae (Laugkäfer). In: Freude H, Harde KW, Lohse GA, Klausnitzer B. Die Käfer Mitteleuropas. Spectrum-Verlag, Heidelberg, 1. Auflage
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858
- Novotny V, Basset Y (2000) Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. *Oikos* 89:564–572
- Pearce JL, Venier LA (2006) The use of ground beetles (Coleoptera: Carabidae) and spiders (Araneae) as bioindicators of sustainable forest management: a review. *Ecol Indicators* 6:780–793
- Rainio J, Niemelä J (2003) Ground beetles (Coleoptera: Carabidae) as bioindicators. *Biodiv Conserv* 12:487–506
- Rodríguez-Estrella R, Blázquez-Moreno MC (2006) Rare, fragile species, small populations, and the dilemma of collections. *Biodiv Conserv* 15:1621–1625
- Sienkiewicz P (2003) Ground beetles (Coleoptera, Carabidae) of the seasonally flooded meadows in the valley of the middle course of the Warta—qualitative analysis. *Baltic J Coleopterol* 3:129–136
- Sienkiewicz P, Konwerski S (2004) Znaczenie rezerwatu „Krajkowo” koło Poznania dla ochrony chrząszczy (Coleoptera) środowisk dolin rzecznych. *Wiad entomol* 23(Supl.):189–191
- Storks N (1993) How many species are there? *Biodiv Conserv* 2:215–232
- Tista M, Fiedler K (2011) How to evaluate and reduce sampling effort for ants. *J Insect Conserv* 15:547–559