

Standardization and optimization of arthropod inventories—the case of Iberian spiders

Pedro Cardoso

Received: 28 February 2008 / Accepted: 23 June 2009 / Published online: 8 July 2009
© Springer Science+Business Media B.V. 2009

Abstract Conservation of species requires accurate knowledge on their distribution. For most groups this can only be achieved through targeted biodiversity assessment programs that must explicitly incorporate comparability and efficiency in their definition. These require the standardization and optimization of sampling protocols, especially for mega-diverse arthropod taxa. This study had two objectives: (1) propose guidelines and statistical methods to improve the standardization and optimization of arthropod inventories, and (2) to propose a standardized and optimized protocol for Iberian spiders based on such guidelines and methods. Definition of the protocol has the following four steps. Firstly, the evaluation of the source data to ensure that the protocol is based on close to complete sampling of a number of sites. Secondly, optimizing the effort per collecting method, using an iterative algorithm that optimizes the combination of methods and samples per method tested in the different sites. Thirdly, defining the overall effort (stop-rules), considering not only desired sampling completeness levels, but also practical strategies during field work. Finally, standardizing the protocol, by finding common results between the optimal options for the different sites. The steps listed were successfully followed in the determination of a sampling protocol for Iberian spiders. A protocol with three sub-protocols of varying degrees of effort (24, 96 and 320 h of sampling) is proposed. I also present recommendations on how to apply the same principles to other regions and taxa.

Keywords Araneae · Mediterranean · Methodology · Optimization algorithms · Portugal · Rapid biodiversity assessment · Sampling · Spain · Species richness

P. Cardoso
Natural History Museum of Denmark, Zoological Museum and Centre for Macroecology,
University of Copenhagen, Copenhagen, Denmark

P. Cardoso
Azorean Biodiversity Group, CITA-A, Universidade dos Açores, Angra do Heroísmo, Portugal

P. Cardoso (✉)
Smithsonian Institution, National Museum of Natural History, P.O. Box 37012, MRC 105,
Room E-509, Washington, DC 20013-7012, USA
e-mail: pcardoso@ennor.org

Introduction

To know how many and which species live in any place on Earth is a daunting task (e.g., Blackmore 1996). However, it is a fundamental one for conservation biology, taxonomy, ecology, biogeography and many other fields. Many projects are currently compiling information on the distribution of species like GBIF (www.gbif.org) and Fauna Europaea (www.faunaeur.org). Other projects are compiling different types of information (taxonomic, ecological) like the Encyclopedia of Life (www.eol.org). However, information on most taxa is still scarce. Even if everything published regarding all the known species would become available in a single database, it would still be poor. Without more and fresh field data, such a database would certainly be incomplete and risk inadequacy (Hortal et al. 2007).

Arthropods, as a megadiverse group, present a formidable challenge for taxonomists, ecologists, biogeographers and conservationists alike. Their sheer numbers make it usually impossible to fully assess their richness, functions in the ecosystem and geographical patterns, even for restricted taxa or confined geographical areas (e.g., Ramos et al. 2001). *Ad hoc*, non-standardized approaches to inventorying an area can be very efficient to compile species lists in many groups (Gordon and Newton 2006), especially easily sampled taxa like birds (Droege et al. 1998; Roberts et al. 2007). Unfortunately, for invertebrates these sampling strategies do not allow reliable comparisons even when made by the same team. Rapid Biodiversity Assessment (RBA) programs were developed as a tool to gather as much information as possible from an area using minimal time and resources (e.g., Oliver and Beattie 1996; Duelli 1997; Jones and Eggleton 2000). Commonly, a combination of methods is required in such protocols (e.g., Delabie et al. 2000; Snyder et al. 2006) and this combination requires standardization for the comparison of sites (Stork et al. 1996; Duelli 1997; Duelli et al. 1999; Jones and Eggleton 2000). The paucity of such standardised protocols is one of the reasons why arthropods are usually relegated to a secondary place or even not considered in conservation programs (Stork et al. 1996; New 1999). However, standard protocols were proposed for large-scale or even global inventories of different taxa like ants (Agosti and Alonso 2000) and butterflies (Pollard and Yates 1993), among others. Also, many protocols are in current use for environmental monitoring purposes using, e.g., carabid beetles (Niemelä et al. 2000) and stream macroinvertebrates (Hering et al. 2004). Such protocols usually address two main issues, (1) the comparability between areas, occasions or even taxa, which can only be guaranteed by the use of standardized inventories and, (2) efficiency, so that scarce resources are not wasted, which can only be guaranteed by the optimization of inventories.

Spiders (order Araneae) are one of the most diverse and abundant arthropod orders. With more than 40,000 known species, it was estimated that one hectare of tropical forest may support between 300 and 800 species of spiders (Coddington et al. 1991, 2009) and that one hectare of Mediterranean forest may support more than 200 species at any given time (Cardoso et al. 2008a, b). They are among the most numerous arthropods of many samples in all kinds of habitats (e.g., Basset 1991; Borges and Brown 2004; Cardoso et al. 2007a). With such richness and abundance it is as challenging to know what lives where as with many other arthropod taxa. However, no standardized and optimized field protocol has been proposed for this taxon in the Mediterranean region, one of the richness hotspots worldwide.

Semi-quantitative sampling was proposed as the best option to sample spiders with maximum efficiency in terms of species richness and also to provide results amenable to robust statistical treatment. It has been applied in several parts of the world like Bolivia

(Coddington et al. 1991), Peru (Silva and Coddington 1996), USA (Coddington et al. 1996; Dobyns 1997; Toti et al. 2000), Tanzania (Sørensen et al. 2002), Denmark (Scharff et al. 2003) and Guyana (Coddington et al. 2009). Based on a large body of work using semi-quantitative sampling developed in Portugal, this study has two objectives: (1) to propose guidelines and statistical methods that can be followed to guarantee the standardization and optimization of arthropod inventories, and (2) to propose a standardized and optimized protocol for Iberian spiders based on such guidelines and methods.

Materials and methods

Sampling procedures

From 2004 to 2006, standardised fieldwork was carried out in three protected areas of Portugal explicitly selected to represent the range of Iberian Peninsula climatic regions. These were the National Park of Peneda-Gerês, NP of Arrábida and NP of Vale do Guadiana (Cardoso et al. 2008a, b, 2009). Sampling always occurred during May and June, when spider species and abundance typically are highest (Cardoso et al. 2007b).

At each site, the sampling area was one hectare (100 × 100 m). Five methods with 64 samples each were used: aerial searching, beating trees and branches, ground searching, sweeping and pitfall trapping. Previous results with the same datasets demonstrated that all method/time-of-day (day or night) combinations except pit-fall traps (which run 24/7) were statistically distinct (Cardoso et al. 2008a, b, 2009). Sampling effort was therefore divided equally across all combinations. The sampling design was semi-quantitative, with a sampling unit defined as one person-hour of effective fieldwork. Pitfall traps were left in the field for 2 weeks, after which four adjacent traps were combined to form one “sample”, so that effort in the field per sample was comparable with the other methods. In Peneda-Gerês and Arrábida 320 samples were collected. Guadiana had no arboreal vegetation, making some methods inappropriate: 192 samples were collected. For details of the sampling methods see Cardoso et al. (2008a, b, 2009).

Statistical analysis

To define a standardized and optimized protocol I propose four steps:

Evaluation of the source data

When selecting data used to design a standardized protocol, each site contributing to the design should have been intensively, almost completely, sampled. Completion can be assessed visually in species accumulation curves and by calculating the sampling completeness: the percentage of species estimated to exist in the sampling plots that are actually observed. For the three sites considered here those analyses have been published elsewhere (Cardoso et al. 2008a, b, 2009).

Optimization of the effort per method

Different procedures for combining methods to improve inventory efficiency have been published. One of the most explicit and generally accepted is to compare the slopes of

accumulation curves of different methods alone or in combination (Longino and Colwell 1997; Longino 2000; Ellison et al. 2007; Rohr et al. 2007). I also define inventory efficiency as the slope of the accumulation curve; steepest slopes are optimal. However, the former procedures generally forced any point in the curve to have the same proportion of samples from different methods. Instead I propose an iterative procedure that optimizes the slope of the curve for any overall number or mix of samples. It is applied to each site in turn. The objective is to determine the number of samples per method that maximizes the species richness (slope) as samples are successively added:

1. start with 0 samples (an empty pool);
2. add one randomly chosen sample from any method (without replacement);
3. calculate the species richness;
4. repeat from step 2 for each method in turn;
5. repeat steps 2–4 many times (e.g., 10,000);
6. identify which method, on average, maximizes richness;
7. add 1 sample of that method to the pool;
8. repeat from step 2 until the pool includes all available samples.

As an example, if the optimum combination for a pool of 3 samples is 2 of method A and 1 of method B (2A1B), this procedure calculates the average richness in 3A1B and 2A2B to determine which addition maximizes richness. The best mix will be the new pool for 4 samples, and the procedure is repeated for 5, 6, etc. samples. A small program in Java to perform the procedure is available upon request.

Definition of the overall effort (stop-rules)

The third step calculates how many samples are required to reach a desired level of completion, i. e. when the field workers can stop sampling. Stop-rules depend on the completeness desired, but for an inventory to be considered as “reasonable” I suggest that about 50% of the species ought to be sampled. A “comprehensive” inventory might reach 70–80%, and an “exhaustive” inventory might reach 90%; 100% is impractical because the effort required to reach it increases astronomically, at least for arthropod taxa, even in delimited areas. Such values are of course subject to discussion, but in my view they combine thoroughness with pragmatism. Defining different targets adjusts the protocol to different objectives or available resources. For Iberian spiders I propose 3 nested (sub-) protocols reflecting varying degrees of effort.

The number of samples also depends on practical issues as well as target completeness values. Previous studies (Coddington et al. 1996) and experience indicate that six samples per collector per day is approximately the maximum sustainable effort possible without excessive fatigue, and without degradation in quality of the later samples. Protocols in multiples of six therefore best accommodate variation in collector number, or days available.

Standardization of the protocol

Given a target completeness value, the individual optimization results of the second step are analyzed to find common outcomes among the results for different sites. All combinations close to optimal per method and per site should be tested, seeking a mix of methods as optimal as possible across all sites. Here again practical issues are important. For

example, sampling only at night may not be feasible even if optimal; the best available mix should be preferred, given practical considerations.

For the Iberian spider dataset, I tested the efficiency of the quasi-optimum (sub-) protocols. For each site and number of samples (24 and 96) I ran 100 sampling simulations comparing a balanced protocol (equal numbers of samples/method) to the optimum protocols for each site individually and to the quasi-optimal protocol for three sites combined. For each simulation I calculated the average and confidence intervals of the species richness observed for each protocol, site, and total sample number.

Results

Evaluation of the source data

The datasets were chosen because of their exhaustiveness (Cardoso et al. 2008a, b, 2009). The three sites presented very high sampling completeness levels and the data was therefore considered as appropriate for the definition of the protocol.

Optimization of the effort per method

The optimization procedure maximizes collecting efficiency and this was reflected in the species accumulation curves (Fig. 1). The optimized curves were more efficient (steeper) than the balanced curves (Fig. 1). Any single method is always inferior to any combination of methods (Fig. 1).

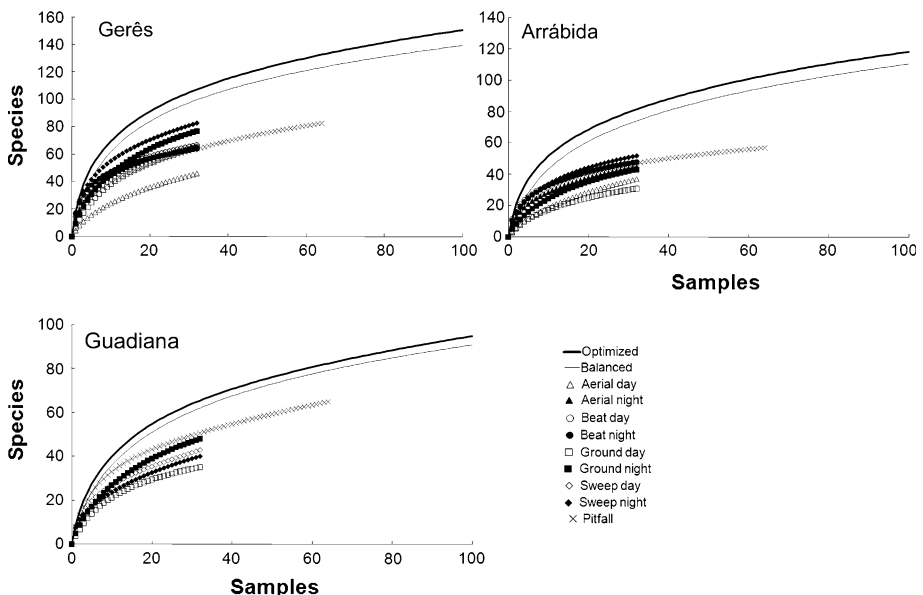


Fig. 1 Randomized species accumulation curves at the three sampled sites. Optimized and balanced curves give average richness for optimal and even mixes of methods, respectively. Only the first 100 samples are presented

Definition of the overall effort (stop-rules)

For these areas 24 samples provided around 50% of the species observed in each area and approximately 50% of estimated species richness, except in Guadiana (Table 1), but Guadiana was also the least well-sampled site (Cardoso et al. 2009). Twenty-four samples is also pragmatic, enabling one person to perform the entire protocol in 4 days, two persons in 2 days or four persons in 1 day.

For these areas 96 samples provided around 80% of the observed species and 70% of the estimated richness (Table 1). As a multiple of 24, 96 samples are also pragmatic. The exhaustive, “high-effort” protocol required an overall effort identical to the one applied in Gerês and Arrábida (320 samples), higher than the one of Guadiana (192 samples). Given the high-quality results obtained for the former two sites, this was the only option that could be recommended with the data available.

Standardization of the protocol

Table 1 reveals the general trend of the distribution of samples by methods. In all cases the optimum combinations required one-third to half the samples being assigned to pitfall traps. Aerial collecting is more productive at night, as well as ground collecting. The latter was chosen by the iterative process only in Guadiana, where aerial sampling was omitted. Beating and sweep methods were ambiguous; either day or night sampling was best depending on the site/habitat (Table 1). These results suggested that the ideal compromise protocol across all sites would include a large proportion of pitfall trapping, night time aerial searching (or ground searching in treeless habitats), and both day and night beating and sweeping.

I tried all possible combinations of 24 and 96 samples that followed these trends to identify the highest efficiency options. From these I chose those that best balanced the effort per method (Table 2). These final combinations of methods, the quasi-optimal solutions, are nicknamed the Cobra protocol (for Conservation Oriented Biodiversity Rapid Assessment). The simulation results (Fig. 2) suggested that Cobra was always significantly or close to significantly more efficient than a protocol that equalized the number of samples per method. The Cobra protocol also yielded close to the maximum theoretical richness for each site, given a predetermined number of samples (Fig. 2).

Table 1 Optimal combinations of samples per method for each site given 24 or 96 samples per site

Samples	Site	Ad	An	Bd	Bn	Gd	Gn	Sd	Sn	Pf	%S	%S*
24	Gerês	0	3	2	2	0	0	0	7	10	53	46
	Arrábida	0	3	5	0	0	0	0	7	9	49	45
	Guadiana	–	–	–	–	0	2	9	0	13	53	38
96	Gerês	0	10	3	2	0	0	0	32	49	81	70
	Arrábida	0	14	23	0	0	0	9	19	31	78	72
	Guadiana	–	–	–	–	0	9	32	17	38	85	60

Ad/An Aerial searching day/night, *Bd/Bn* beating day/night, *Gd/Gn* ground searching day/night, *Sd/Sn* sweeping day/night, *Pf* pitfall trapping (4 traps/sample), *S* observed species richness, *S** estimated species richness

Table 2 Proposed sub-protocols that simultaneously optimize sampled richness and obey to practical reasoning, with their application in the three tested sites

Samples (%)	Ad	An	Bd	Bn	Gd	Gn	Sd	Sn	Pf	Gerês		Arrábida		Guadiana	
										%S	%S*	%S	%S*	%S	%S*
24 (7.5)	0	4 (0)	2 (0)	2 (0)	0	0 (4)	2 (4)	2 (4)	12	52	45	48	44	52	37
96 (30)	0	16 (0)	8 (0)	8 (0)	0	0 (16)	8 (16)	8 (16)	48	79	69	76	70	85	60
320 (100)	32 (0)	32 (0)	32 (0)	32 (0)	32 (64)	32 (64)	32 (64)	32 (64)	64	100	87	100	92	–	–

Number of samples in parenthesis refers to alternative protocols when sites do not present arboreal cover, e.g., Guadiana

Ad/An Aerial searching day/night, *Bd/Bn* beating day/night, *Gd/Gn* ground searching day/night, *Sd/Sn* sweeping day/night, *Pf* pitfall trapping (4 traps/sample), *S* observed species richness, *S** estimated species richness

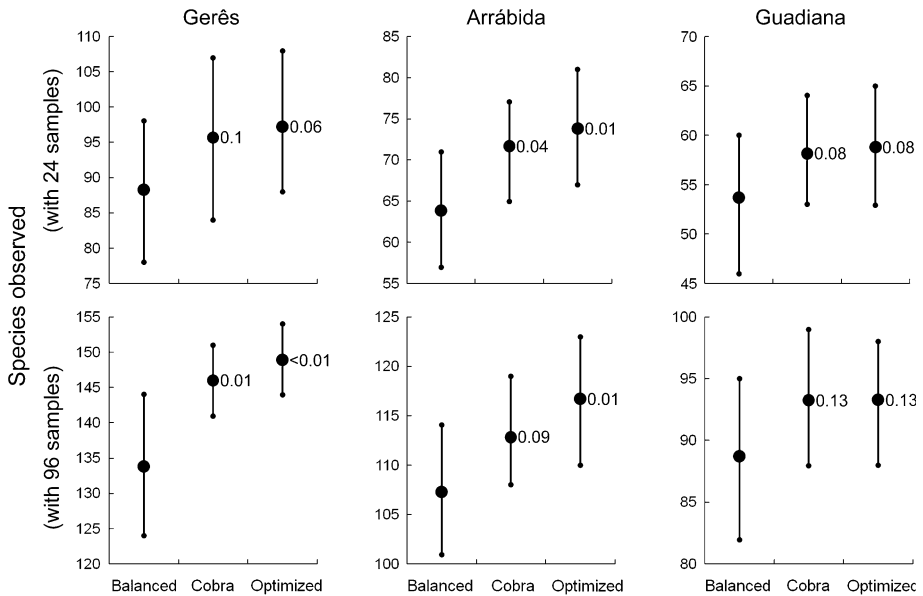


Fig. 2 Number of species observed in each case of 100 simulated sampling protocols. The balanced protocol evenly distributes samples by the different methods. The quasi-optimum Cobra protocol is the final combination of samples per method here proposed to follow in the future in areas other than the three sampled for this work. The optimized protocol optimally distributes samples between methods, reflecting the theoretical maximum number of species possible to obtain at each site. Represented are the average, 95% confidence intervals and the *P* value between the balanced protocol and the other two, as determined by the simulations

Discussion

This proposed protocol for Iberian spiders, with its four explicit steps, meets a series of tests that, in my view, any standardized and optimized protocol should pass:

1. **Suitability.** The methods and effort should be designed for the target organisms and the questions to be answered, and should sample a substantial and known part of the community. This requisite, even if met, is usually not tested. In most cases, sampling is done without knowing beforehand the approximate percentage of species that are captured and how much is missed.
2. **Efficiency.** The chosen methods and their proportions should gather the best information with the least effort. This requisite is often ignored if researchers use only familiar methods, rather than the most efficient.
3. **Feasibility.** The methods and effort employed must be adjusted to available resources (human, financial, time). Although resource limitations strongly affect inventory quality, inventory goals are rarely adjusted to this reality (e.g., suitability, above), which therefore can result in relatively poor quality inventories.
4. **Flexibility.** Different sampling occasions or projects will have different resources available. Most protocols recommend one mix and quantity of samples and therefore do not consider this requisite.

5. Transparency. If the logic behind the inventory design is clearly explained, different teams will be better able to replicate it. Too often the reasoning behind a protocol is opaque so that it is less easily adapted to different situations.
6. Accountability. The results from a chosen protocol must be adequate to support *a posteriori* evaluation and review. Frequently, the density and quantity of data obtained are too few, so that the reliability of the results and methods cannot be adequately critiqued.

How to sample?

Previous analyses with these same datasets found that for spiders:

1. method is the single most important factor in inventory design;
2. each combination of method and time of day is statistically distinct and must be treated separately;
3. collector experience often has no statistical effect or is, at most, less important, although I recommend that at least one collector should be experienced in semi-quantitative sampling and proficient with the methods to maximize consistency between different teams;
4. effort normally invested in methods that are inappropriate for some habitats should be redirected towards other methods. Overall effort therefore remains constant.

In this case the proportion of samples per method did not change from 24 to 96 samples, because the accumulation curves for the different methods were similar and did not cross (but see the steeper curves of ground sampling at night in Gerês and Guadiana). If some curves asymptote much faster than others, the ideal proportions could change with increasing effort.

The unbalanced design of the protocol (i.e. unequal numbers of samples/method) may raise concerns that an unbalanced statistical design implies a biased sample of the target community. On the contrary, this iterative process shows that balanced designs in general do not sample communities in proportion to their abundances, because methods differ in efficiency. Carefully chosen unbalanced designs can in fact balance what misleadingly looked like a fully balanced design, equalizing the fractions of the community that are sampled. For example, ground searching in the proposed low and medium effort protocols is omitted because of its low productivity per sample and because the sub-communities it samples largely overlap with the sub-communities sampled by more efficient methods like pitfall trapping and sweeping. Likewise, aerial searching, beating and sweeping also strongly overlap, but pitfall trapping accesses a rather different sub-community (Cardoso et al. 2008a, b, 2009). Method overlap wastes resources better devoted to different methods that capture different species, hence the large proportion of pitfall traps in the optimal solutions. In sum, carefully chosen unbalanced designs can result in an overall more accurate sample of the target community.

When to sample?

The collecting season is as important as method or time of day (Churchill and Arthur 1999; Jiménez-Valverde and Lobo 2006). Cardoso et al. (2007b) already identified the single optimum season for sampling spiders, given geographical areas and habitat types in Portugal, which should extrapolate to the Iberian Peninsula.

Short term sampling during the richness peak, if done in a thorough way, is likely to capture around 50% of the annual spider diversity of a site in the Iberian Peninsula (Jiménez-Valverde and Lobo 2006; Cardoso et al. 2007b) and probably the rest of Europe (e.g., Scharff et al. 2003). I therefore recommend that this protocol always occur during the phenological richness peak (May–June in the Iberian Peninsula). Although short-term surveys certainly underestimate the total diversity of a site (e.g., Riecken 1999), if all sites are sampled during the same season they will be comparable.

If the objectives require a complete annual picture of the community I recommend to allocate most of the effort (medium to high-effort sub-protocols) during the peak season and to allocate lower effort sub-protocols (low to medium-effort) during the September/October and January/February seasons. I am hence suggesting a “96 + 24 + 24” or a “320 + 96 + 96” strategy. This will capture species whose adults are active mostly during autumn (e.g., most mygalomorphs, many large lycosids) or winter (e.g., linyphiids). This approach, with a series of independent collecting events at each site, is comparable to the one of Duelli (1997); Duelli et al. (1999) for agricultural areas of Switzerland.

Where to sample?

The question of “where to sample” is of critical importance. Most studies will have predefined areas that are to be sampled. But some intend to study patterns on a large-scale for biogeography or conservation purposes and have no defined areas. In such case, a good option is to use an approach that maximizes the environmental diversity encompassed by a predetermined number of areas (e.g., Hortal and Lobo 2005). Because environmental diversity is usually positively correlated with biodiversity (Faith and Walker 1996), this allows maximizing the overall diversity captured and, most importantly, the range of variation in biodiversity of a region.

The 24-sample Cobra protocol, optimized for capturing 50% of the spider species present at one hectare areas, has already been applied in a number of sites in mainland Portugal (Litoral Norte Nature Park, Paúl de Arzila Nature Reserve) and the Azores archipelago (Biscoito da Ferraria Forest Reserve, Terceira). Its application allowed the identification of many new species for science and the country, including some with apparently restricted and declining distributions due to land use change and consequent habitat destruction (Cardoso et al. in prep., Carvalho et al. in prep., Crespo et al. *subm.*). New collecting events are already planned directed to other protected areas in the country (Lagoas de Santo André e da Sancha Nature Reserve). All these collections were made by different teams with diverse objectives (conservation, inventory, taxonomy, and land use effects on diversity) and yet the continued application of the protocol promises to reveal new, fully comparable data at a large-scale in the near future.

What to sample?

Results from the sites now studied suggest that plotless sampling in a uniform habitat allows capturing higher species richness with similar effort than plot-based sampling, although species composition does not differ (Cardoso, unpublished results). However, this is not always the case. Sørensen et al. (2002) did not find any differences between plot and plotless sampling. The opposite conclusions were also verified for ants (James 2004) and partially for spiders (Dobyns 1997). In both cases, this was due to missing rare or cryptic species when effort is low by unit area, even if sampling a single site implies that many low-range species with a scattered distribution in space are not sampled (Gaston 1999). In view of

contradictory results, for comparative purposes between habitats it is probably better to compare similar areas. In this case, I therefore recommend the standardization of one hectare as an adequate area for plot sampling.

Adapting to different regions

The protocol now presented is optimized for spiders in the Iberian Peninsula, although it should be possible to use it in other locations of the Mediterranean. Other regions will certainly need a different approach as the richness numbers and the composition of habitats and communities varies dramatically. Sampling spiders in tropical (e.g., Coddington et al. 1991; Sørensen et al. 2002) or temperate forests (e.g., Coddington et al. 1996; Scharff et al. 2003) probably requires different methods and certainly different sample proportions. Adding a method for cryptic species, hiding under leaf litter, in tree trunks or other specific microhabitats, may be useful in many regions other than the Mediterranean, where such hiding places are often scarce. Canopy fogging is also important to obtain a complete representation of spider species in tropical forests (Sørensen 2004).

Adapting to different taxa

Different taxa require different methods. For capturing Hymenoptera, Malaise traps proved their efficiency (e.g., Fraser et al. 2008). For ants specifically, litter sifting, Winkler traps and pitfall traps are advised (Agosti and Alonso 2000). Flight-interception and yellow pan traps are also suitable for a large range of arthropods (e.g., Missa et al. 2009). Aquatic arthropods obviously require a set of completely different methods (e.g., Gillies et al. 2009). Many examples could be given for different arthropod taxa, but even vertebrates could potentially benefit from the process now presented (e.g., Carassou et al. 2009). Despite all the options, for most taxa a combination of methods is usually better than a single method to capture maximum diversity with minimum effort. The process here applied for spiders is adaptable for any combination of methods as long as effort can be reasonably compared between them, for example, by standardizing effort by person-hour of fieldwork (Coddington et al. 1991). All methods should be amenable to such treatment and it should be possible to calculate the best combination of methods and samples per method to obtain predetermined percentages of species richness per site. Additionally season and sampled area should be standardized. This way all the six previously mentioned requisites for a standardized and optimized protocol can be met.

Concluding remarks

One of the main points of this work is to demonstrate one way to standardize and optimize biodiversity assessment protocols. Many diverse taxa are as difficult to sample as spiders and pose the same challenges. This strategy, however, is general: it ought to work for any diverse arthropod group, or any biome, although, of course, such a prediction must be thoroughly evaluated and tested.

Protocols that follow these suggestions present several advantages to their users. Firstly, it guarantees maximum diversity with minimum effort, if trends detected from preliminary data in fact extrapolate. Given scarce resources, this is essential. Secondly, it offers an alternative to rather weak species richness estimators in situations where empirical parameters are relatively well understood. If we know the approximate percentage of

species captured in each area, estimators become less important, as they can easily be more biased than multiplying the observed richness by an empirically measured factor. Thirdly, studies made with the protocol, independently of the particular objectives or teams, will be easily comparable so that the same data can be re-used a number of times with different purposes. Lastly, it allows future studies in cooperation between different research teams. Large-scale studies are facilitated by comparable previously or newly collected data.

Acknowledgments Many thanks to all the people that helped in fieldwork, logistics and advice: Clara Gaspar, Sérgio Henriques, Nikolaj Scharff, Jesper Schmidt, Tamás Szűts, Israel Silva, Rui Carvalho, Pedro Sousa, Luis Crespo, Luis Carlos Pereira, Pedro Humberto Castro, Ricardo Silva, Ana Filipa Gouveia and Alberto de Castro. Thanks to Nikolaj Scharff, Jonathan Coddington, Clara Gaspar and Paulo A.V. Borges for fruitful discussion and comments on previous versions of the manuscript. Jonathan Coddington has extensively edited English usage. Support was given by the Portuguese Foundation for Science and Technology (grants no. SFRH/BD/1195/2000 and SFRH/BPD/17351/2004), the Danish Bøje Benzons Foundation, the Danish Natural Science Research Council (grant no. 272-05-0431), the Portuguese Institute for Nature Conservation, Peneda-Gerês National Park and Vale do Guadiana Nature Park.

References

- Agosti D, Alonso LE (2000) The ALL protocol—a standard protocol for the collection of ground-dwelling ants. In: Agosti D, Majer JD, Alonso LE, Schultz TR (eds) *Ants—standard methods for measuring and monitoring biodiversity*. Biological Diversity Handbook Series. Smithsonian Institution Press, Washington
- Basset Y (1991) The taxonomic composition of the arthropod fauna associated with an Australian rain-forest tree. *Aust J Zool* 39:171–190
- Blackmore S (1996) Knowing the Earth's biodiversity: challenges for the infrastructure of systematic biology. *Science* 274:63–64
- Borges PAV, Brown VK (2004) Arthropod community structure in pastures of an island archipelago (Azores): looking for local-regional species richness patterns at fine-scales. *Bull Entomol Res* 94:111–121
- Carassou L, Mellin C, Ponton D (2009) Assessing the diversity and abundances of larvae and juveniles of coral reef fish: a synthesis of six sampling techniques. *Biodivers Conserv* 18:355–371
- Cardoso P, Borges PAV, Gaspar C (2007a) Biotic integrity of the arthropod communities in the natural forests of Azores. *Biodivers Conserv* 16:2883–2901
- Cardoso P, Silva I, Oliveira NG, Serrano ARM (2007b) Seasonality of spiders (Araneae) in Mediterranean ecosystems and its implications in the optimum sampling period. *Ecol Entomol* 32:516–526
- Cardoso P, Gaspar C, Pereira LC et al (2008a) Assessing spider species richness and composition in Mediterranean cork oak forests. *Acta Oecologica* 33:114–127
- Cardoso P, Scharff N, Gaspar C et al (2008b) Rapid biodiversity assessment of spiders (Araneae) using semi-quantitative sampling: a case study in a Mediterranean forest. *Insect Conserv Divers* 1:71–84
- Cardoso P, Henriques SS, Gaspar C et al (2009) Species richness and composition assessment of spiders in a Mediterranean scrubland. *J Insect Conserv* 13:45–55
- Churchill TB, Arthur JM (1999) Measuring spider richness: effects of different sampling methods and spatial and temporal scales. *J Insect Conserv* 3:287–295
- Coddington JA, Griswold CE, Davila DS (1991) Designing and testing sampling protocols to estimate biodiversity in tropical ecosystems. In: Dudley EC et al (eds) *The unity of evolutionary biology: proceedings of the fourth International Congress of Systematic and Evolutionary Biology*. Dioscorides Press, Portland
- Coddington JA, Young LH, Coyle FA (1996) Estimating spider species richness in a southern Appalachian cove hardwood forest. *J Arachnol* 24:111–128
- Coddington JA, Agnarsson I, Miller JA et al (2009) Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *J Anim Ecol* 78:573–584
- Delabie JHC, Fisher BL, Majer JD, Wright IW (2000) Sampling effort and choice of methods. In: Agosti D, Majer JD, Alonso LE, Schultz TR (eds) *Ants—standard methods for measuring and monitoring biodiversity*. Biological Diversity Handbook Series. Smithsonian Institution Press, Washington
- Dobyns JR (1997) Effects of sampling intensity on the collection of spider (Araneae) species and the estimation of species richness. *Environ Entomol* 26:150–162

- Droege S, Cyr A, Larivée J (1998) Checklists: an under-used tool for the inventory and monitoring of plants and animals. *Conserv Biol* 12:1134–1138
- Duelli P (1997) Biodiversity evaluation in agricultural landscapes: an approach at two different scales. *Agricult Ecosys Environ* 62:81–91
- Duelli P, Obrist MK, Schmatz DR (1999) Biodiversity evaluation in agricultural landscapes: above-ground insects. *Agricult Ecosys Environ* 74:33–64
- Ellison AM, Record S, Arguello A, Gotelli NJ (2007) Rapid inventory of the ant assemblage in a temperate hardwood forest: species composition and assessment of sampling methods. *Environ Entomol* 36:766–775
- Faith DP, Walker PA (1996) Environmental diversity: on the best-possible use of surrogate data for assessing the relative biodiversity set of areas. *Biodivers Conserv* 5:399–415
- Fraser SEM, Dytham C, Mayhew PJ (2008) The effectiveness and optimal use of Malaise traps for monitoring parasitoid wasps. *Insect Conserv Divers* 1:22–31
- Gaston KJ (1999) Implications of interspecific and intraspecific abundance-occupancy relationships. *Oikos* 86:195–207
- Gillies CL, Hose GC, Turak E (2009) What do qualitative rapid assessment collections of macroinvertebrates represent? A comparison with extensive quantitative sampling. *Environ Monit Assess* 149:99–112
- Gordon JE, Newton AC (2006) Efficient floristic inventory for the assessment of tropical tree diversity: a comparative test of four alternative approaches. *Forest Ecol Manag* 237:564–573
- Hering D, Moog O, Sandin L, Verdonschot P (2004) Overview and application of the AQEM assessment system. *Hydrobiologia* 516:1–20
- Hortal J, Lobo JM (2005) An ED-based protocol for optimal sampling of biodiversity. *Biodivers Conserv* 14:2913–2947
- Hortal J, Lobo JM, Jiménez-Valverde A (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conserv Biol* 21:853–863
- James CD (2004) Trapping intensities for sampling ants in Australian rangelands. *Austral Ecol* 29:78–86
- Jiménez-Valverde A, Lobo JM (2006) Establishing reliable spider (Araneae, Araneidae and Thomisidae) assemblage sampling protocols: estimation of species richness, seasonal coverage and contribution of juvenile data to species richness and composition. *Acta Oecologica* 30:21–32
- Jones DT, Eggleton P (2000) Sampling termite assemblages in tropical forests: testing a rapid biodiversity assessment protocol. *J Appl Ecol* 37:191–203
- Longino JT (2000) What to do with the data. In: Agosti D, Majer JD, Alonso LE, Schultz TR (eds) *Ants—standard methods for measuring and monitoring biodiversity*. Biological Diversity Handbook Series. Smithsonian Institution Press, Washington
- Longino JT, Colwell RK (1997) Biodiversity assessment using structured inventory: capturing the ant fauna of a tropical rain forest. *Ecol Appl* 7:1263–1277
- Missa O, Basset Y, Alonso A et al (2009) Monitoring arthropods in a tropical landscape: relative effects of sampling methods and habitat types on trap catches. *J Insect Conserv* 13:103–118
- New TR (1999) Untangling the web: spiders and the challenges of invertebrate conservation. *J Insect Conserv* 3:251–256
- Niemelä J, Kotze J, Ashworth A et al (2000) The search for common anthropogenic impacts on biodiversity: a global network. *J Insect Conserv* 4:3–9
- Oliver I, Beattie AJ (1996) Designing a cost-effective invertebrate survey: a test of methods for rapid assessment of biodiversity. *Ecol Appl* 6:594–607
- Pollard E, Yates TJ (1993) *Monitoring butterflies for ecology and conservation*. Chapman and Hall, London
- Ramos MA, Lobo JM, Esteban M (2001) Ten years inventorying the Iberian fauna: results and perspectives. *Biodivers Conserv* 10:19–28
- Riecken U (1999) Effects of short-term sampling on ecological characterization and evaluation of epigeic spider communities and their habitats for site assessment studies. *J Arachnol* 27:189–195
- Roberts RL, Donald PF, Green RE (2007) Using simple species lists to monitor trends in animal populations: new methods and a comparison with independent data. *Anim Conserv* 10:332–339
- Rohr JR, Mahan CG, Kim KC (2007) Developing a monitoring program for invertebrates: guidelines and a case study. *Conserv Biol* 21:422–433
- Scharff N, Coddington JA, Griswold CE et al (2003) When to quit? Estimating spider species richness in a northern European deciduous forest. *J Arachnol* 31:246–273
- Silva D, Coddington JA (1996) Spiders of Pakitza (Madre de Dios, Perú): species richness and notes on community structure. In: Wilson DE, Sandoval A (eds) *The Biodiversity of Southeastern Perú*. Smithsonian Institution, Washington
- Snyder BA, Draney ML, Sierwald P (2006) Development of an optimal sampling protocol for millipedes (Diplopoda). *J Insect Conserv* 10:277–288

- Sørensen LL (2004) Composition and diversity of the spider fauna in the canopy of a montane forest in Tanzania. *Biodivers Conserv* 13:437–452
- Sørensen LL, Coddington JA, Scharff N (2002) Inventorying and estimating subcanopy spider diversity using semiquantitative sampling methods in an afro-montane forest. *Environ Entomol* 31:319–330
- Stork NE, Samways MJ, Eeley HAC (1996) Inventorying and monitoring biodiversity. *Trends Ecol Evolut* 11:39–40
- Toti DS, Coyle FA, Miller JA (2000) A structured inventory of Appalachian grass bald and heath bald spider assemblages and a test of species richness estimator performance. *J Arachnol* 28:329–345