Biodivers Conserv (2009) 18:3115–3125
DOI 10.1007/s10531-009-9630-6

ORIGINAL PAPER

Use of extensive habitat inventories in biodiversity studies

Audrey Muratet · E. Porcher · J.-C. Abadie · L. Poncet · J. Moret · A. Robert

Received: 25 September 2008 / Accepted: 16 March 2009 / Published online: 28 March 2009

Abstract Large monitoring programs exist in many countries and are necessary to assess present and past biodiversity status and to evaluate the consequences of habitat degradation or destruction. Using such an extensive data set of the floristic richness in the Paris Ile-de-France region (France), we compared different sampling efforts and protocols in different habitat units to highlight the best methods for assessing the actual plant biodiversity. Our results indicate that existing data can be used for a general understanding of site differences, but analysts should be aware of the limitations of the data due to non-random selection of sites, inconsistent observer knowledge, and inconsistent sampling period. The average species diversity recorded in a specific habitat does not necessarily reflect its actual diversity, unless the monitoring effort was very strong. Overall, increasing the sampling effort in a given region allows improvement of the (1) number of habitats visited, (2) the total sampled area for a given habitat type, (3) the number of seasons investigated. Our results indicate that the sampling effort should be planned with respect to these functional, spatial and temporal heterogeneities, and to the question examined. While the effort should be applied to as many habitats as possible for the purpose of capturing a large proportion of regional diversity, or comparing different regions, inventories should be conducted in different seasons for the purpose of comparing species richness in different habitats.

Keywords Data quality · Floristic diversity · Monitoring · Sampling effort · Species richness

A. Muratet (✉) · E. Porcher · J.-C. Abadie · L. Poncet · J. Moret
Conservatoire Botanique National du Bassin Parisien, UMS 2699 CNRS-MNHN Inventaire et Suivi de la Biodiversité, Muséum national d'Histoire naturelle, 61 rue Buffon, 75005 Paris, France
e-mail: muratet@mnhn.fr

E. Porcher · J.-C. Abadie · A. Robert
UMR 7204 CNRS-MNHN-UPMC, Conservation des Espèces, Restauration et Suivi des Populations, Muséum national d'Histoire naturelle, 55 rue Buffon, 75005 Paris, France

Introduction

It is now widely recognized that the current extinction rates of plant and animal species are between a hundred and a thousand times higher than the background rates throughout life's history on Earth (May 2002). However, documenting species extinction only, i.e. the most obvious manifestation of biodiversity loss, is not sufficient to develop effective conservation policies, partly because extinction rates carry no information regarding changes in community composition, which may have dramatic consequences for ecosystem stability (Worm and Duffy 2003). There is an urgent need to quantify the spatiotemporal changes in biodiversity by considering community composition and trends in species abundances (Convention on Biological Diversity in Rio 1992). Such information is necessary to identify the mechanisms (e.g. environmental variables, human-induced disturbances, etc.) controlling the variation in species richness through space and time, as well as to identify sites of conservation concern and appropriate policies to improve the current biodiversity.

Ideally, this quantification would require large scale, long-term surveys based on standardized methodologies to allow comparisons in space and time. Such protocols already exist in a limited number of cases or are just starting to be implemented. The British Countryside Survey (CS) (Firbank et al. 2003; Haines-Young et al. 2003), for example, was established in 1978 in the United Kingdom and focuses on several taxonomic groups, including plants. The Biodiversity Monitoring Program (BDM) in Switzerland (Weber et al. 2004; Plattner et al. 2004) was launched in 1995 and focuses on local plant diversity. Other protocols have been implemented to survey the diversity of particular taxonomic groups, as exemplified by breeding bird surveys in different countries (since 1966 in North America, Sauer et al. 1997; since 1994 in UK, Newson et al. 2005; since 1989 in France, Julliard et al. 2003). Such surveys are based on formatted sampling protocols generally occurring twice a year within different discrete classes of habitat at the national scale. In these examples, the inventory protocol is generally standard and well defined, which allows the sampling effort to be homogeneous among observers, constant in time, or clearly quantified, so that any statistical inference can be made independently of the monitoring effort. Moreover, inventory protocols are designed to ensure that sampling is proportional to the area occupied by each habitat/settlement type in the region of interest.

Although such large scale monitoring schemes are crucial to document future changes in biodiversity, they will unfortunately not suffice to quantify the present changes in biodiversity, and specifically to evaluate the 2010 biodiversity target. A complementary approach to quantify changes in biodiversity could be to use the large amounts of existing inventory data collected by various biodiversity stakeholders (some of which are compiled in the Global Biodiversity Information Facility, GBIF 2008). However, because such data come from a very large number of observers and geographic locations, they were generally collected using very different methodologies and are highly heterogeneous in nature. The question that immediately arises is whether such heterogeneous data can be exploited to document reliably the trends in biodiversity.

Here we address this issue using plant inventory data from Paris Basin (France). We analyzed data from thousands of inventories carried out between 2001 and 2005 by botanists who were involve in the same Botanical Conservatory but who were not instructed to follow a given standardized protocol. Focusing on the proportion of total vascular plant species detected as a function of (1) annual number of visits per habitat type and (2) season of data collection, we investigated different options for data analysis and survey protocol, to optimize the use of existing data and improve future monitoring. We specifically addressed the following questions: (1) Are one time surveys of floristic diversity indicative of the total diversity of a region, and do species richness estimated from one time surveys vary across habitats, seasons and years? (2) What is the benefit of increasing survey effort, by increasing either the number of survey habitats or the time span of surveys?

Materials and methods

Study area

The Ile-de-France region, including the city of Paris (48°68′ N; 0°17′ E) and the surrounding area, covers 12,072 km^2 (Fig. 1). The climate is oceanic with continental trends (mean annual temperature 12°C, with a minimum in January and a maximum in July; average monthly rainfall 57 mm) and the relief is relatively flat (elevation between 11 and 217 m a.s.l.). The population density is 952 inhabitants/km^2 (INSEE 2006), which makes Ile-de-France the most densely populated administrative region of France.

A total of 1,225 plant species were encountered in the study area between 2001 and 2005, as calculated from records of the FLORA database [National Botanical Conservatory of the Paris Basin, CBNBP (2008) and see below for a description of the database]. Of these species, 11% were naturalized species, i.e. non-indigenous species that reproduce and sustain populations without direct intervention by humans (Richardson et al. 2000).

Inventory protocol

The data used in this study were collected between 2001 and 2005 by botanists from the National Botanical Conservatory of the Paris Basin (hereafter CBNBP), a French public organization aiming to study and protect the flora of the Paris basin. One central objective of

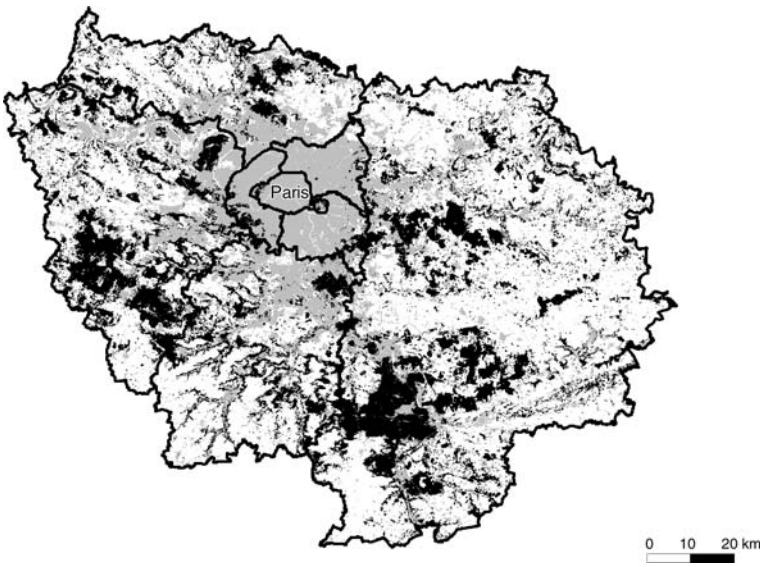

Fig. 1 Map of the study area, the Paris Ile-de-France region. Forests appear in black, cultures and other rural habitats in white and open and built urban area in grey (IAURIF 2003). Dark lines correspond to the district boundaries

CBNBP is to describe the geographical distribution of all species growing in the area, which dictates the methodology used to collect data. Every year, a total of 149 botanists (both professionals and competent amateurs) visited the 'communes' (French administrative municipalities) of the region between March and October and recorded as many plant species as they could observe within a municipality, as well as the spatial locations of each species. There was no standardized protocol: the duration of data collection, sampling locations and total area sampled were left to the appreciation of the observers and varied greatly among individuals. For example, sampling locations within a municipality were not randomly distributed, but were instead usually chosen to maximize the total number of species observed.

Database contents and study data

Inventory data were pooled in FLORA, a database built by CBNBP. The database includes information on species (scientific and common names), observer, date of observation, location (municipality) and habitat type according to CORINE land cover nomenclature (Bissardon et al. 1997), and contains more than 1 million observations (i.e. one species recorded at a given time and in a given site) for the Ile-de-France region (CBNBP 2008).

We chose to work with data collected between 2001 and 2005, because the quality and quantity of data are much lower before this period. For statistical reasons, we also discarded all observations from rarely sampled habitats, i.e. habitats that were visited less than once a month between 2001 and 2005, so that data from eight habitat types only were retained (see Table 1). For this study, this yielded a total of 237,884 observations corresponding to 7,358 different sites (i.e. the total area covered by a given habitat type in a given place) within the Ile-de-France region.

Data analysis

Because the database contains very little information on species abundance or frequency, and does not allow estimating species detection probabilities, plant communities were characterized by the observed species richness only.

Species richness at the site level

We first analyzed the variation in species richness at the site level by fitting an analysis of variance model using the R software (Core Team 2007), where site richness was a

Table 1 Description of habitats types

Habitat type	Number of visits by surveyors	Proportion of the total study area (%) (IAURIF 2003)
Stagnant fresh water	412	1.2
Circle of water edges	437	
Mesophile meadows	259	Not available
Deciduous forests	2,072	20.5
Cultures (essentially cereals)	257	51.2
Urban parks and gardens	1,012	4
Cities and industrial sites	1,596	15.6
Wastelands	1,313	0.36

The distribution of the number of inventories across habitat types between 2001 and 2005 and the spatial distribution of habitats are given function of (1) habitat type (2) inventory month, (3) inventory year, and (4) all pairwise interactions.

As this analysis showed statistical differences among years on the richness recorded, all years were considered separately in subsequent analysis.

Assessment of optimal monitoring effort

To optimize monitoring programs, monitoring effort should be minimum, but large enough to provide accurate estimates of species richness (and, ideally, other parameters of community composition). To evaluate this, we performed random resampling in the database to simulate various monitoring efforts, by varying the number of sites, habitats, or months sampled.

(a) *Increasing effort within a given habitat*: To estimate the species accumulation curve within each habitat type, we plotted the ratio of observed versus total species richness as a function of the number of inventories, x, as follows. Within a given year, x inventories (= x sites) were sampled at random, each in a different month, and the overall species richness (excluding redundancies) of this sample was computed. This species richness was then divided by the total number of species observed in this habitat type. For each x and each habitat type, the procedure was repeated 50,000 times and the average ratio of observed versus total species richness was plotted.

(b) *Correlation between sampled and total species richness*: To test whether the number of species recorded in x inventories was representative of the "true" floristic richness of the different habitats, we compared the number of species recorded in x inventories within a year in each habitat to the overall number of species in each habitat, using a Spearman rank correlation across habitats. This procedure was performed 50,000 times for each habitat, and the average correlation coefficient, r_s, as well as the proportion of significant correlations at the 5% level were plotted as a function of the number of inventories per habitat, x.

(c) *Optimization of the number of habitats or months sampled*: We compared the benefit of increasing the number of months or the number of habitats sample, given a constant effort. To this end, we plotted the observed species richness as a function of number of habitats (respectively months) visited, with a constant number of inventories. Keeping the number of inventories (8) constant allowed us to test for a habitat or month effect without confounding area effects. Within a given year, eight sites were chosen at random among x habitat types (respectively months) and the overall species richness in these eight inventories (i.e., excluding redundancies) was computed. The procedure was repeated 50,000 times and the average species richness in x habitats (respectively months) was plotted against the number of habitats (months).

Results

Variation in average observed species diversity

Site species richness varied significantly across years, months [maximum species richness in June (36.5), minimum in August (28)], and habitat types [maximum number of species in cities and industrial sites (41), minimum in stagnant freshwater (18), Table 2]. In addition, all interactions were also significant, so that the difference in species richness among habitats were highly variable within and across years (Table 2).

Table 2 Result of the analysis of variance, where site richness was a function of (1) habitat type (2) inventory month, (3) inventory year, and (4) all pairwise interactions

Parameters	Degree of freedom	F value	Pr(>F)
Habitat	7	135.82	$<10^{-4}$
Month	7	22.94	$<10^{-4}$
Year	4	31.20	$<10^{-4}$
Habitat × month	49	4.20	$<10^{-4}$
Habitat × year	27	6.14	$<10^{-4}$
Month × year	27	5.26	$<10^{-4}$
Habitat × month × year	165	1.78	$<10^{-4}$

Species accumulation curves within habitats

The shape of the species accumulation curves varied greatly across habitats (Fig. 2). The proportion of total species recorded appeared to reach a plateau at five inventories per habitat in mesophile meadows, cultures, cities and industrial sites or wastelands. Note however that the fraction of total species observed remained low (between 15% and 25%). In contrast, the species accumulation curves did not appear to saturate in stagnant fresh water, circle of water edges, deciduous forest or urban parks and gardens.

Correlation between observed and total species richness across habitats

As expected, the correlation between observed and total species richness across habitats was close to zero and non-significant when the number of inventories per habitat was small ($x < 6$, Fig. 3). However, seven or eight inventories per habitat provided a better picture of the total species richness (Spearman correlation coefficient significantly different from 0,

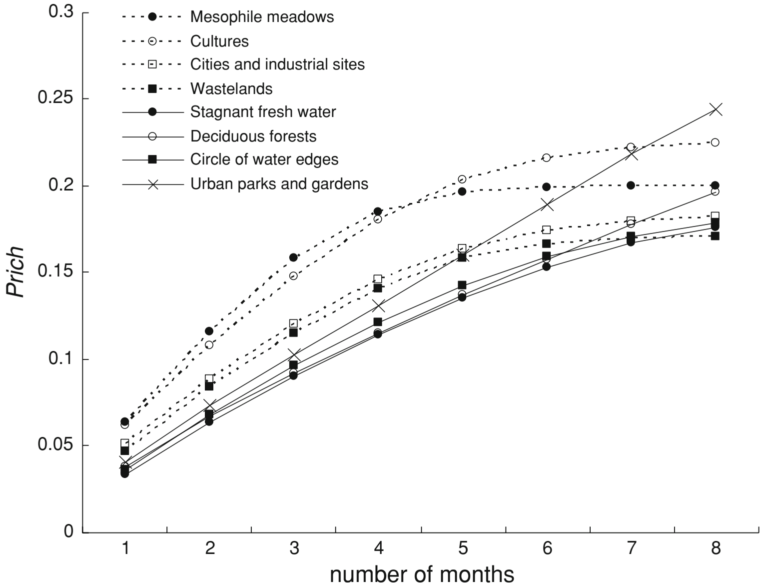

Fig. 2 Proportion of total species richness (Prich) as a function of the number of seasons sampled (number of months)

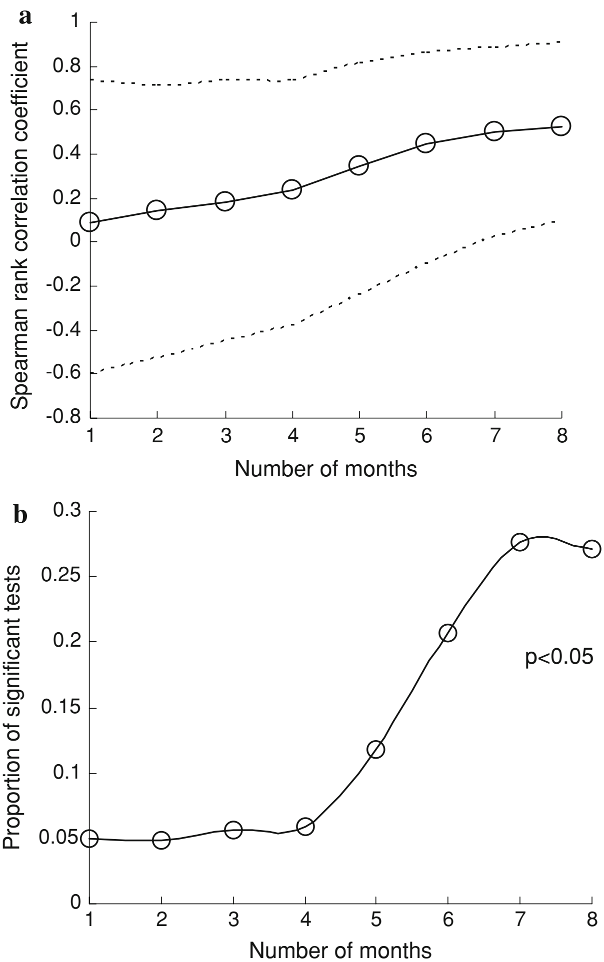

Fig. 3 Correlation between overall and recorded species richness in the different habitats, as a function of the monitoring effort (increase of the number of inventory months x). Protocol presented in method section. **a** Average (*open circles*) and 95% confidence intervals (*dashed lines*) Spearman coefficients of rank correlation r_s. **b** Proportion of significant one-tailed correlations between overall and recorded species richness among 50,000 independent computations of recorded species richness

Fig. 3a). Note however that mean correlation coefficients remained relatively low (Fig. 3a), suggesting that yearly monitoring protocols with few inventories in each habitat do not allow to compare species richness in the different habitats.

Optimization of monitoring effort by increasing the number of habitats or months

As expected, observed species richness increased (+12%) when the number of habitats increased for a constant monitoring effort. Similarly, there was a lower but non negligible benefit (+7%) of increasing the number of inventory months.

Discussion

Biodiversity inventories are costly in time and money, and maximizing the number of species observed during a given monitoring effort is therefore an important task. Our study focuses on the use of existing, non standardized inventory data to address the optimization of monitoring effort.

Non-standardized data and minimal monitoring effort

Our results reflect the well-known heterogeneity of plant communities in time (year) and among habitat types: the observed species richness depends on the habitat, season, year and their interactions. When dealing with non-standardized data, this raises the issue of how to disentangle actual ecological sources of heterogeneity (e.g. true differences among habitat, seasons, years…) from sampling or methodological sources of variation. In particular, owing to the lack of randomization and to observer variability, among-inventory differences in species richness were not only due to differences in the period of sampling (month and year), but also to differences in sites themselves (inventories performed in different months were not necessarily conducted on the same sites). This for example implies that classical methods to estimate species richness (e.g. those derived from the CAPTURE program; Rexstad and Burnham 1991) cannot be used with such non-standardized inventory data. Hence, total species richness in a given habitat was estimated as the total number of species observed over a large number of inventories. Although this probably results in an underestimation of species richness, we nonetheless believe that it provided a reasonably good picture of community composition.

General guidelines about minimal monitoring efforts can be inferred from the results above. We showed that one to five yearly inventories per habitat do not provide an accurate picture of habitat richness (Fig. 4), at least in the semi-natural habitats commonly encountered in Île-de-France. Sampling effort is clearly an important issue regardless of the survey method used (Metcalfe-Smith et al. 2000; Walther and Martin 2001), and other studies have reached similar conclusions regarding minimal sampling efforts. For example, De Solla et al. (2005) showed that, in anuran monitoring programs, the average observed species richness was only 25.1% of the total richness with a single sampling night, but reached an average of 80% of the total species richness with 12 sampling nights. Archaux et al. (2006) showed that on 400 m^2 forest quadrats, the level of exhaustiveness of plant censuses increased in a semi-logarithmic way with sampling time. The study of Estevez

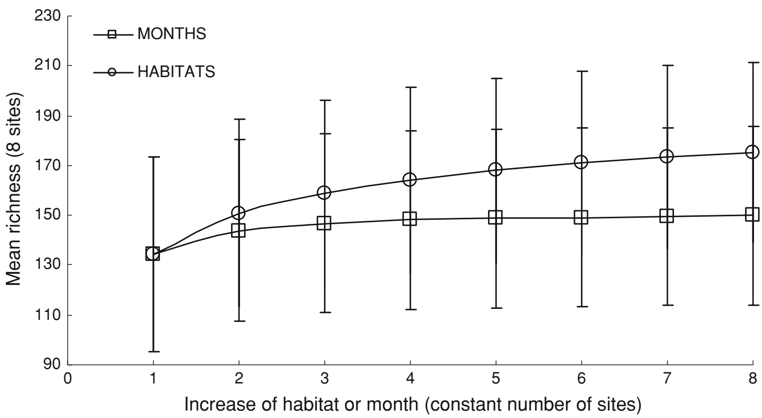

Fig. 4 Observed species richness as a function of the number of months or habitats visited for a constant effort (eight sites sampled). Error bars represent standard errors

and Christman (2006) on the movement of animals in confinement clearly indicated that sampling effort had a tremendous impact on the study outcome. Nonetheless, several European countries have started to implement floristic monitoring programs, generally based on one or two inventories per year. For instance, in the United Kingdom, the British Countryside survey (Haines-Young et al. 2000) is based on annual inventories of several hundred of randomly sampled fix plots classified into 32 land use classes. In Switzerland, The Biodiversity Monitoring Program (Hintermann et al. 2002) consists in a grid-sampling program based on five settlement types within which plots are randomly drawn. The local plant diversity is inventoried in these plots every 5 years.

Although the information collected in the aforementioned monitoring programs is useful to document long-term trends, or to compare trends among habitat types (especially for the most frequent species, and when directional variations in species abundances are high), our results suggest that it will not be sufficient to compare the absolute species numbers present in the different habitat types. In the present data set, the variability across observers and sites tended to overwhelm the differences among habitat types when there were fewer than six inventories per year (Fig. 4), which represents a large monitoring effort in comparison with most survey programs.

Optimization of sampling effort

The outcome of a given protocol depends, among others, on the area sampled as well as on seasonal and habitat effects, so that the sampling effort should be judiciously planned and implemented to optimize the number of species recorded. In general, the financial and time costs of a field inventory do not vary across seasons or habitat types and protocols can be optimized via a selection of seasons and habitats visited. For example, with a constant effort, the observed species richness was increased by 6.5% if inventories were conducted in two different seasons versus a single season, and by 11% if they were conducted in two versus one habitats. This is consistent with the generally accepted idea that plant functional beta diversity is larger than seasonal beta diversity. However, the choice of maximizing either the number of seasons or habitats sampled should depend on the question investigated.

If a monitoring program aims to maximize recorded species richness in the study region (e.g. for the purpose of comparing biodiversity across regions or examining annual trends), maximizing habitat types would be the most efficient strategy. In fact, our results indicate that (1) increasing the number of habitats is always more efficient than increasing the number of months; (2) beyond 3 months, any further increase in the number of months sampled has no notable effect on the observed species richness for a constant number of sites visited (Fig. 4). In contrast, to compare species richness across habitats, inventories should be conducted throughout as many sites as possible to ensure that actual differences among habitats can be detected. Assuming that the total species richness was a proxy for true total species richness, we showed that the average species richness observed during a single inventory per habitat was not representative of total richness. First, the average species richness observed in a single inventory was only $4.24 \pm 2.84\%$ of total richness on average. Second, the observed richness is not representative of the total richness of the habitat unless the sampling effort is extremely strong (>5 inventories a year, Fig. 3). It follows that for a constant sampling effort, among habitat comparisons require to use few habitats with many inventories per habitat.

Conclusion

There is general agreement that biodiversity conservation should be guided by biodiversity assessment. As an important part of this assessment, inventory protocols should be designed with care, to identify the specific conservation target that a project ultimately would like to influence (Salafsky et al. 2002). Ideally inventories should include (1) sites randomly sampled according to a standard protocol (for example, using a sampling effort stratified by habitat types), (2) observers with a knowledge level as uniform as possible (3) identical observation periods. As we promote these goals we will promote high quality data for monitoring and other purposes. Existing large data sets collected by various biodiversity stakeholders do not generally meet these criteria, and they should be used with caution to infer biodiversity trends, e.g. in combination with resampling methods to correct for their heterogeneity. The large number of existing inventory data can however be exploited to address other conservation issues, e.g. to quantify floristic index over a homogeneous region (Muratet et al. 2008).

Acknowledgments We thank Gérard Arnal and Sébastien Filoche of the CBNBP, coordinators of the floristic inventories in the Paris Ile-de-France region. Emmanuelle Porcher was partly funded by Agence Nationale de la Recherche (ANR Grant #2006-JCJC-0032).

References

Archaux F, Gosselin F, Bergès L (2006) Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. J Veg Sci 17:299–306

Bissardon M, Guibal L, Rameau J-C (1997) CORINE biotopes. Ecole Nationale du Génie Rural, des Eaux et des Forêts, Nancy

Conservatoire botanique national du Bassin parisien (CBNBP) Muséum national d'Histoire naturelle 2008. URL: http://cbnbp.mnhn.fr/cbnbp/

Convention on Biological Diversity in Rio (1992) URL: http://www.cbd.int/convention/convention.shtml

De Solla SR, Shirose LJ, Fernie KJ, Barrett GC, Brousseau CS, Bishop CA (2005) Effect of sampling effort and species detectability on volunteer based anuran monitoring programs. Biol Conserv 121:585–594. doi:10.1016/j.biocon.2004.06.018

Development R Core (2007) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Estevez I, Christman MC (2006) Analysis of the movement and use of space of animals in confinement: the effect of sampling effort. Appl Anim Behav Sci 97:221–240. doi:10.1016/j.applanim.2005.01.013

Firbank LG, Barr CJ, Bunce RGH et al (2003) Assessing stock and change in land cover and biodiversity in GB: an introduction to Countryside Survey 2000. J Environ Manage 67:207–218. doi:10.1016/S0301-4797(02)00174-3

Global Biodiversity Information Facility (2008) URL: http://www.gbif.org

Haines-Young RH, Barr CJ, Black HIJ et al (2000) Accounting for nature: assessing habitats in the UK countryside. DETR, London

Haines-Young R, Barr CJ, Firbank LG et al (2003) Changing landscapes, habitats and vegetation diversity across Great Britain. J Environ Manage 67:267–281. doi:10.1016/S0301-4797(02)00179-2

Hintermann U, Weber D, Zangger A, Schmill J (2002) Biodiversity monitoring in Switzerland, BDM—interim report. Swiss Agency for the Environment, Forests and Landscape SAEFL

IAURIF (2003) Institute for Planning and Development of the Paris Ile-de-France Region. URL: http://www.iaurif.org

INSEE (2006) National Institute for Statistics and Economic Studies. http://www.insee.fr

Julliard R, Jiguet F, Couvet D (2003) Common birds facing global changes: what makes a species at risk? Glob Change Biol 10:148–154. doi:10.1111/j.1365-2486.2003.00723.x

May RM (2002) The future of biological diversity in a crowded world. Curr Sci 82:1325–1331

Metcalfe-Smith J, Di Maio J, Staton S, Mackie G (2000) Effect of sampling effort on the efficiency of the timed search method for sampling freshwater mussel communities. J North Am Benthol Soc 19:725–732. doi:10.2307/1468129
Muratet A, Porcher E, Devictor V et al (2008) Evaluation of floristic diversity in urban areas as a basis for habitat management. Appl Veg Sci 11:451–460
Newson SE, Woodburn R, Noble DG, Baillie SR (2005) Evaluating the Breeding Bird Survey for producing national population size and density estimates. Bird Study 52(1):42–54
Plattner M, Birrer S, Weber D (2004) Data quality in monitoring plant species richness in Switzerland. Community Ecol 5:135–143. doi:10.1556/ComEc.5.2004.1.13
Rexstad E, Burnham K (1991) User's guide for interactive program CAPTURE. Colorado State University, Fort Collins
Richardson DM, Pyšek P, Rejmánek M, Barbour MG, Panetta D, West CJ (2000) Naturalization and invasion of alien plants: concepts and definitions. Divers Distrib 6:93–107. doi:10.1046/j.1472-4642.2000.00083.x
Salafsky N, Margoluis R, Redford KH, Robinson JG (2002) Improving the practice of conservation: a conceptual framework and research agenda for conservation science. Conserv Biol 16:1469–1479. doi:10.1046/j.1523-1739.2002.01232.x
Sauer JR, Hines JE, Gough G, Thomas I, Peterjohn BG (1997) The North American breeding bird survey results and analysis. Version 96.4. Patuxent Wildlife Research Center, Laurel
Walther B, Martin J-L (2001) Species richness estimation of bird communities: how to control for sampling effort? Ibis 143:413–419. doi:10.1111/j.1474-919X.2001.tb04942.x
Weber D, Hintermann U, Zangger A (2004) Scale and trends in species richness: considerations for monitoring biological diversity for political purposes. Glob Ecol Biogeogr 13:97–104. doi:10.1111/j.1466-882X.2004.00078.x
Worm B, Duffy JE (2003) Biodiversity, productivity and stability in real food webs. Trends Ecol Evol 18(12):628–632. doi:10.1016/j.tree.2003.09.003