

Database records as a surrogate for sampling effort provide higher species richness estimations

Jorge M. Lobo

Received: 14 May 2007 / Accepted: 9 January 2008 / Published online: 23 January 2008
© Springer Science+Business Media B.V. 2008

Abstract The compilation of all the available taxonomic and distributional information on the species present in a territory frequently generates a biased picture of the distribution of biodiversity due to the uneven distribution of the sampling effort performed. Thus, quality protocol assessments such as those proposed by Hortal et al. (Conservation Biology 21:853–863, 2007) must be done before using this kind of information for basic and applied purposes. The discrimination of localities that can be considered relatively well-surveyed from those not surveyed enough is a key first step in this protocol and can be attained by the previous definition of a sampling effort surrogate and the calculation of survey completeness using different estimators. Recently it has been suggested that records from exhaustive databases can be used as a sampling-effort surrogate to recognize probable well-surveyed localities. In this paper, we use an Iberian dung beetle database to identify the 50×50 km UTM cells that appear to be reliably inventoried, using both data derived from standardized sampling protocols and database records as a surrogate for sampling effort. Observed and predicted species richness values in the shared cells defined as well-surveyed by both methods suggest that the use of database records provides higher species richness values, which are proportionally greater in the richest localities by the inclusion of rare species.

Keywords Biodiversity databases · Sampling effort assessment · Database records · Survey completeness · Species accumulation curves

Introduction

Reliable site-related taxonomic information is needed for the study of spatial biodiversity variation, as well as for conservation purposes (Neldner et al. 1995; Ferrier 2002; Brooks et al. 2004; Guralnick et al. 2007). Unfortunately, the compilation of the available information

J. M. Lobo (✉)

Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC),
C/ José Gutiérrez Abascal 2, 28006 Madrid, Spain
e-mail: mcnj117@mncn.csic.es

on species distribution can be seen to reveal a pattern of spatial bias and incompleteness, mainly in hyperdiverse groups such as invertebrates and/or in countries without a long naturalist tradition (Dennis and Hardy 1999; Dennis et al. 1999; Soberón et al. 2000; Hortal et al. 2007; Lobo et al. 2007). Site accessibility, distance from the residence of the collecting taxonomist, or degree of interest to the naturalist are all key parameters able to explain the current location of the available species presence data (Nelson et al. 1990; Peterson et al. 1998; Dennis and Hardy 1999; Dennis et al. 1999; Dennis and Thomas 2000; Zaniewski et al. 2002; Anderson 2003; Parnell et al. 2003; Reddy and Dávalos 2003; Reutter et al. 2003; Graham et al. 2004; Soberón and Peterson 2004; Martínez-Meyer 2005; Dennis et al. 2006; Romo et al. 2006). The pervasive occurrence of these biases across different taxonomic groups would deeply affect our knowledge on distribution of biodiversity and the design of useful conservation strategies (Nelson et al. 1990; Faith 2002; Williams et al. 2002).

Both nonparametric estimators and collector's curves have been widely used to estimate the "true" number of species present in a site (Colwell and Coddington 1994; Gotelli and Colwell 2001; Hortal et al. 2006; Soberón et al. 2007). The former require the use of species abundance data, while the latter, based on the increase of the accumulated number of species with the sampling effort, require species presence data along with a measure of sampling effort. Such a sampling effort measure can be expressed in units as the number of collector-days or trap-days (Soberón and Llorente 1993). Species distribution databases that compile all available distributional information from heterogeneous sources (scientific publications with different aims, together with natural history museums and private collections) do not contain a homogeneous and comparable sampling effort unit. Thus, it does not seem possible to carry out the identification of adequately inventoried territories using collector's curves when such sampling effort units are lacking. However, the number of records of exhaustively compiled databases has been used on some occasions as a surrogate for the sampling effort (see Hortal et al. 2001; Lobo and Martín-Piera 2002; Martín-Piera and Lobo 2003; Romo and García-Barros 2005; Romo et al. 2006), and it was recently demonstrated that their use provides results as good as those obtained using other measures more directly related to the sampling effort (Hortal et al. 2006). In this paper, I show that the use of biodiversity databases coming from unstandardized surveys is not only possible, but can also help in providing the most accurate picture of biodiversity useful for conservation and biogeographical purposes. Observed and predicted species richness values coming from standardized surveys (abundance data) and database records (presence data) are compared, showing that such databases are able to provide supplementary biological information that is difficult to obtain using temporally restricted standardized survey data.

Materials and methods

An exhaustive Iberian dung beetle (Coleoptera, Scarabaeidae) database (BANDASCA) was used for this study. Its 15,740 database records include all distribution information available on 101,996 individuals of the 53 species present in the Iberian Peninsula (Martín-Piera 2000), mainly from such a variety of sources as museums and private collections, published or unpublished biogeographic or faunistic works, and standardized ecological studies. The compiled information involves more than a century of surveys from 1,872 to 2,004. Each record of this specimen-based database is a pool of specimens from a single species with identical information in the following fields: location, altitude, date of capture, type of habitat, food resource and collector.

This information was used to calculate the number of species accumulated with the increase in the number of database records for each of 255 Iberian 50×50 km Universal Transverse Mercator grid cells (UTM) with more than 15% of land surface. This accumulated species number was estimated after randomizing the order of entrance of the species in the inventory 100 times to smooth the curve of species richness increase with the number of database records (Colwell 2000). This smoothed, randomized accumulation curve was fitted to the recommended Clench function (Colwell and Coddington 1994; Peterson and Slade 1998; Soberon and Llorente 1993) in order to estimate the asymptotic value (i.e. the estimated total richness score for an unlimited number of samples). The estimations derived from the Clench function seem to perform better than other asymptotic estimators (Hortal et al. 2006). The adequately inventoried UTM cells were defined as those with observed species richness scores of 80% or more of the asymptotic predicted scores. The identification of adequately inventoried UTM cells by using database records is called “database-data” in the following.

In the next step, all data from studies carried out with standardized sampling methods, such as baited pitfall traps or dung pats, were taken into account in the calculation of observed species richness, and in the compilation of species abundance data for each UTM cell. Total richness was estimated from the species abundance through the abundance-based coverage estimator (ACE) proposed by Chazdon et al. (1998), a modified version of Chao and Lee's estimator (Colwell and Coddington 1994), designed to compensate for the consistent overestimation of species richness by the latter. The comparative efficiency of ACE estimations has been recently established (Hortal et al. 2006). Adequately inventoried UTM cells were also defined as those with observed species richness scores equal to, or higher than, 80% of the total species richness estimated by ACE, called “standardized-data” in the following. Species richness estimates were calculated using the EstimateS package (Colwell 2000).

Adequately inventoried UTM cells identified by both methods ($n = 18$) were used to exemplify the differences in species richness that can be obtained when different types of sources of biological data are used to derive biodiversity values. Of course, inventory completeness can be measured using different estimators, which can provide different richness estimations that in turn depend on the sampling effort; sampling success estimations vary as sampling effort increases (Chiarucci et al. 2001; Hortal et al. 2006). In this paper, the selection of estimators was carried out only to choose a subgroup of probable well-prospected UTM cells according to the two sampling effort measures (individuals and records). The use of other estimators or other methods to assess the degree of completeness (see Hortal and Lobo 2005) slightly vary the number of probable common well-surveyed cells but not the obvious result: that the use of biological data coming from heterogeneous sources allows us to add supplementary biodiversity information.

Results

Taking the record data into account, 89 Iberian UTM cells were observed to have species richness scores higher than 80% of the values predicted by the Clench function. Of these, only 18 cells were observed to have species richness scores higher than, or equal to, 80% of the corresponding scores predicted by ACE using standardized-data (Table 1, Fig. 1). In these 18 UTM cells, the predicted mean number of species does not differ significantly from that observed, taken from either the database-data ($F_{(1,34)} = 0.72$; ns) or from the standardized-data ($F_{(1,34)} = 0.18$; ns). However, the observed mean number of species derived from standardized-data differs significantly from that derived from database-data

Table 1 Abundance (N), species richness (S) and number of database records (R) for the 18 Iberian 50×50 km UTM cells with more than 80% species richness as estimated by both methods. S_{ACE} is the Abundance-Based Coverage Estimator species richness score (Chazdon et al. 1998) taking into account the species distribution of the number of individuals collected by standardized methods. S_{CLE} is the asymptotic species richness, estimated by adjusting the Clench function to a smoothed rarefaction curve, a plot of the accumulated number of species as a function of the number of database records for each UTM cell. Non-ecological data are the number of species and records of the database without considering the standardized-data

UTM code	Abundance-based estimation				Database-based estimation				Non-ecological data	
	N	S	S_{ACE}	%	R	S	S_{CLE}	%	R	S
30TUK3	4,010	26	26.3	99	728	33	35.2	94	338	33
29SQA2	3,637	26	27.3	95	400	29	30.4	95	37	19
30TVL2	7,140	24	29.4	82	814	37	42.1	88	489	34
30TUK1	3,701	23	24	96	340	31	32.9	94	152	30
29TQF2	10,182	19	19.5	97	459	24	28.2	85	435	23
30SWG4	1,377	19	20.9	91	161	25	26.7	94	106	21
30SVG4	852	18	18.5	97	222	27	28.5	95	106	24
29SQB2	8,682	17	17.2	99	300	22	23.1	95	73	22
30TUN1	4,719	16	16.4	98	364	19	20.5	93	75	14
30SUG4	2,107	15	15	100	168	18	18.8	96	77	18
30TVL4	735	15	17.4	86	135	29	36.3	80	81	26
30SWG1	2,025	13	13	100	125	26	26.9	97	57	18
30TTM4	1,138	12	12.6	95	209	13	13.6	96	2	1
30SUG3	669	12	13.2	91	154	20	24.5	82	85	20
30SWF3	1,091	10	10	100	86	17	17.4	98	66	14
30SXH3	527	8	9.1	88	39	9	10	90	15	6
30SUG1	1,938	6	6	100	126	20	22.5	89	111	20
30SUH4	885	6	6	100	80	15	16.7	90	53	14

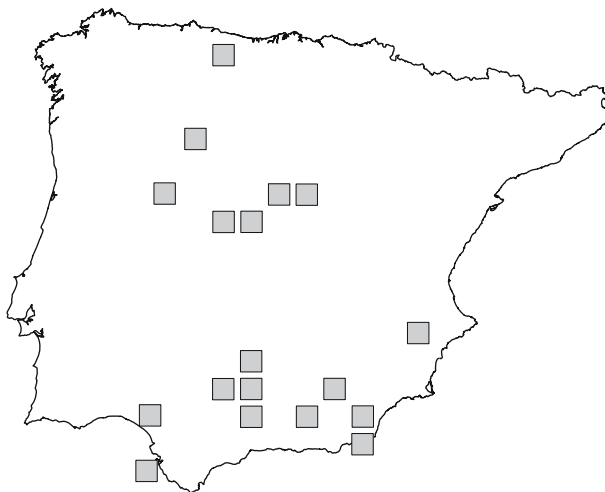


Fig. 1 Iberian location of the eighteen 50×50 UTM km cells with observed species richness scores equal to, or higher than, 80% of both the estimated scores predicted: by fitting the Clench function to the number of species vs. the number of database records (database-data approximation); and by the abundance-based coverage estimator (ACE), considering the abundance of species (standardized-data approximation)

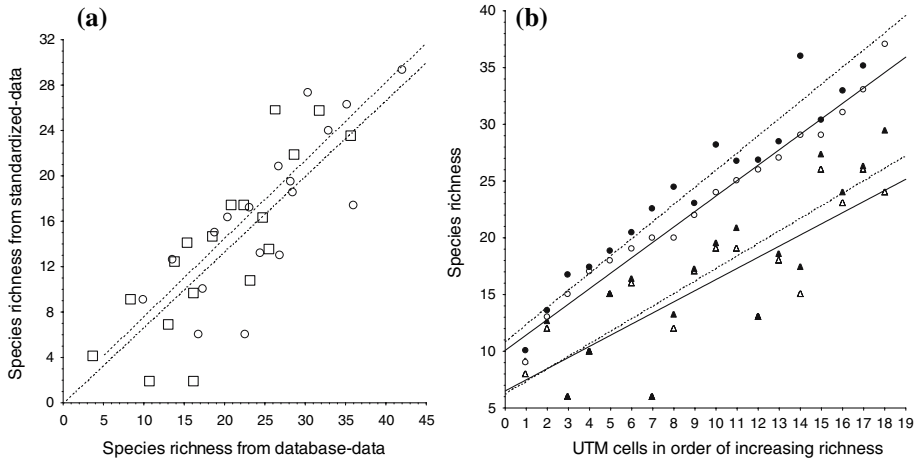


Fig. 2 (a) Relationship between standardized-data and database-data species richness values for the well-surveyed Iberian UTM cells; observed (squares) and predicted (circles) scores. Species richness predicted using abundance values and the Abundance-Based Coverage Estimator (ACE), and database values fitted to a Clench function after smoothing the curve in which the accumulated number of species is plotted against the number of database records. Broken lines are the fitted straight lines with slopes significantly different from 0 both for observed ($t_{(16)} = 5.65, P < 0.001$) and predicted data ($t_{(16)} = 5.33, P < 0.001$), but not significantly different between them (0.701 and 0.667 respectively; $F_{(1,33)} = 0.19, ns$). (b) Variation in observed (open circles) and predicted (black circles) number of species from database records; and in observed (open triangles) and predicted (black triangles) number of species from standardized surveys. Fitted straight lines for observed (broken) and predicted scores (solid) have significantly different slopes ($P < 0.001$) between both methods ($F_{(1,33)} = 36.7$ and 38.0 , respectively). Predicted scores calculated using the Abundance-Based Coverage Estimator from standardized ecological studies data, and using the asymptotic species richness obtained by fitting the Clench function to a smoothed accumulation curve in which the increase in number of species is plotted against the number of database records. The 18 UTM cells are plotted in order of increasing species richness, according to the observed database records

($F_{(1,34)} = 9.84; P = 0.003$). This number is larger when derived from all the information available in the database (23.0 ± 1.7 ; mean \pm SE), and smaller when derived from standardized-data (15.8 ± 1.5). Clench-fitted species richness scores (25.2 ± 2.0) also differ from ACE (16.8 ± 1.6) ($F_{(1,34)} = 10.89; P = 0.002$). Thus, although both estimations are positively and significantly correlated, the species richness scores derived from the database records are clearly higher than those from standardized-data (Fig. 2a). This difference is greater in those cells with greater species richness values (Fig. 2b); database data allows inclusion of around four or five more species in the poorer grid cells, and between 10 and 12 more species in the richer grid cells.

A comparison of the survey data species richness values against those coming from the remaining information shows that the median percentage of species collected by the standardized-data for total species (75%) also significantly differs from the median percentage of species coming from non-standardized-data (91%; Mann-Whitney U test: $U = 73.5, N1 = N2 = 18, P = 0.005$).

Discussion

Both observed and predicted species richness values are higher when all the available information is considered. Why this should be so is intimately linked with the question of

whether a species list in finished form can actually be associated with a particular site. Standardized surveys are an important source of biodiversity information. However, they are frequently temporally limited, being generally designed to obtain standardized inventories able to generate between site comparisons. Thus, as most of these studies do not aim to inventory the complete species list of a territory, biological data from non-standardized studies are necessary to include some supplementary species. Such an increase in the number of species when non-standardized studies are used may be due to the incorporation of species with a low detectability, due to the inclusion of vagrant species incapable of producing viable descendants in the site (Pulliam 1988; Gaston 1994), or even due to the meta-population character of many species and the relevance of dispersal limitation processes that cause a species to possibly be absent from suitable sites (Pulliam 2000). In our study, the number of species from standardized survey data is even lower than the number of species in the remaining non-standardized data, and the estimated number of species from the complete database is higher even though the method used to derive species richness estimations from records (Clench asymptotic functions) frequently underestimate the “true” species richness (Hortal et al. 2006). Hence, the incorporation of faunistic data compiled along a wide temporal period allows an increase in the number of species of a locality, and this increase is even higher in cells with the greatest species richness probably because the environment, history and geography of these sites have allowed the survival of a greater number of geographically restricted species (Lamoreux et al. 2005). The species inventory of a site varies with time, as does the species-area relationship (Adler and Laurenroth 2003), depending substantially on the regional species pool because local and regional diversity are intimately connected (Ricklefs and Schluter 1993). Thus, at a given time, one can never be sure if the species inventory of a site is complete. Collector’s curves are not always asymptotic, and estimations depend on the survey effort (Chiarucci et al. 2001). Repeated sampling of some very mobile groups can produce the addition of new species seemingly indefinitely (Dennis 2001; Gotelli and Colwell 2001). Some authors hold that these rare species added to inventories by increased sampling, not belonging to a stable population, should not be added to species richness estimates (Dennis and Hardy 1999; Dennis 2001). However, it seems reasonable that these species should be considered members of the potential group that may inhabit a site and that they may help to guarantee the stability of ecosystem functions (Adler and Lauenroth 2003; Lyons et al. 2005). Hence, the probable contributions of these species are better reflected when exhaustive databases are used that offer higher richness scores more closely related to the group of species that, over a period of time, are able to inhabit a site and, therefore, form part of the biodiversity contributing to the functioning of the ecosystem.

Previous studies suggest that exhaustive faunistic information compiled from all available sources can be used to identify well-surveyed inventories if database records are used as a sampling effort surrogate (see also Hortal et al. 2001; Lobo and Martín-Piera 2002; Martín-Piera and Lobo 2003; Romo and García-Barros 2005; Hortal et al. 2006; Romo et al. 2006). In this paper, we also use the database records as a sampling effort surrogate, showing that the obtained estimations are correlated with those provided by the use of measures more directly related to the sampling effort. Furthermore and importantly, the use of all the biodiversity information available can achieve a better determination of the sites with the greatest species richness by the inclusion of some rare species that are difficult to detect. The use of database records as a sampling effort surrogate can be an opportunity to use the taxonomic and biogeographic information stored in multi-sourced databases (Soberón et al. 1996; GBIF 2003; Graham et al. 2004; Guralnick et al. 2007) for basic and applied purposes. As Hortal et al. (2007) propose, such databases should be built in an

exhaustive way and be used to: (i) discriminate probable well-surveyed localities; (ii) examine if these localities represent the environmental and spatial heterogeneity of the analyzed territory; and (iii) identify the location of future sampling sites able to optimize the coverage of data on biodiversity patterns. Only the previous accomplishment of a quality data assessment protocol like this can allow us to guarantee the success of alternative strategies such as those suggested by the use of predictive modelling techniques (Guisan and Zimmermann 2000) because the inaccuracy of absences, a consequence of spatially biased sampling, leads one to suppose that the resulting distributions obtained by such methods suffer from the same or greater biases than those in the information from which they are derived (Ferrier 2002; Hortal and Lobo 2005; Gu and Swihart 2004; Kadmon et al. 2004; Funk et al. 2005; Lobo 2008).

Acknowledgements This paper was supported by the Spanish MEC project CGL2004-0439/BOS, a Fundación BBVA Project, and The European Distributed Institute of Taxonomy (EDIT) project.

References

- Adler PB, Lauenroth WK (2003) The power of time: spatiotemporal scaling of species diversity. *Ecol Lett* 6:749–756
- Anderson RP (2003) Real vs artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *J Biogeogr* 30:591–605
- Brooks T, da Fonseca GAB, Rodrigues ASL (2004) Species, data, and conservation planning. *Conserv Biol* 18:1682–1688
- Chiarucci A, Maccherini S, De Dominicis V (2001) Evaluation and monitoring of the flora in a nature reserve by estimation methods. *Biol Conserv* 101:305–314
- Colwell RK (2000) EstimateS: statistical estimation of species richness and shared species from samples (Software and User's Guide). Version 6.0b1. Available at <http://viceroy.eeb.uconn.edu/Estimates6/>
- Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philos Trans R Soc Lond B Biol Sci* 345:101–118
- Chazdon RL, Colwell RK, Denslow JS, Guariguata MR (1998) Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of north-eastern Costa Rica. In: Dallmeir F, Comiskey JA (eds) *Forest biodiversity research, monitoring and modelling. Conceptual background and Old World case studies*. Parthenon Publishing, Paris, France, pp 285–309
- Dennis RLH (2001) Progressive bias in species status is symptomatic of fine-grained mapping units subject to repeated sampling. *Biodivers Conserv* 10:483–494
- Dennis RLH, Hardy PB (1999) Targeting squares for survey: predicting species richness and incidence for a butterfly atlas. *Global Ecol Biogeogr Lett* 8:443–454
- Dennis RLH, Thomas CD (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *J Insect Conserv* 4:73–77
- Dennis RLH, Sparks TH, Hardy PB (1999) Bias in butterfly distribution maps: the effects of sampling effort. *J Insect Conserv* 3:33–42
- Dennis RLH, Shreeve TG, Isaac NJB, Roy DB, Hardy PB, Fox R, Asher J (2006) The effects of visual apparancy on bias in butterfly recording and monitoring. *Biol Conserv* 128:486–492
- Faith D (2002) Those complementarity analysis do not reveal extent of conservation conflict in Africa. *Science debate*. <http://www.sciencemag.org/cgi/eletters/293/5535/1591/#381>
- Ferrier S (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst Biol* 51:331–363
- Funk VA, Richardson KS, Ferrier S (2005) Survey-gap analysis in expeditionary research: where do we go from here? *Biol J Linn Soc Lond* 85:549–567
- Gaston KJ (1994) *Rarity*. Chapman & Hall, London
- GBIF (2003) Global biodiversity information facility strategic plan. Available at www.gbif.org/GBIF_org/documents
- Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett* 4:379–391
- Graham CH, Ferrier S, Huettner F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol Evol* 19:497–503

- Gu W, Swihart RK (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biol Conserv* 116:195–203
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Modell* 135:147–186
- Guralnick RP, Hill AW, Lane M (2007) Towards a collaborative, global infrastructure for biodiversity assessment. *Ecol Lett* 10:663–672
- Hortal J, Lobo JM (2005) An ED-based protocol for optimal sampling of biodiversity. *Biodivers Conserv* 14:2913–2947
- Hortal J, Lobo JM, Martín-Piera F (2001) Forecasting insect species richness scores in poorly surveyed territories: the case of the Portuguese dung beetles (Col. Scarabaeinae). *Biodivers Conserv* 10:1343–1367
- Hortal J, Borges PA, Gaspar C (2006) Evaluating the performance of species richness estimators: sensitivity to sample grain size. *J Anim Ecol* 75:274–287
- 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
- Kadmon R, Oren F, Avinoam D (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol Appl* 14:401–413
- Lamoreux JF, Morrison JC, Ricketts TH, Olson DM, Dinerstein E, McKnight MW, Shugart HH (2005) Global tests of biodiversity concordance and the importance of endemism. *Nature* 440:212
- Lobo JM (2008) More complex distribution models or more representative data? *Biodiv Inform* (in press)
- Lobo JM, Martín-Piera F (2002) Searching for a predictive model for Iberian dung beetle species richness based on spatial and environmental variables. *Conserv Biol* 16:158–173
- Lobo JM, Baselga A, Hortal J, Jiménez-Valverde A, Gómez JF (2007) How does the knowledge about the spatial distribution of Iberian dung beetle species accumulates over time? *Divers Distrib* 13:772–780
- Lyons KG, Brigham CA, Traut BH, Schwartz MW (2005) Rare species and ecosystem functioning. *Conserv Biol* 19:1019–1024
- Martín-Piera F, Lobo JM (2003) Database records as a sampling effort surrogate to predict spatial distribution of insects in either poorly or unevenly surveyed areas. *Acta Entomol Ibérica Macaronésica* 1:23–35
- Martín-Piera F, López-Colón JI (2000) Fauna Ibérica, vol 14. In: Ramos MA et al (eds) *Coleoptera, Scarabaeoidea* 1. Museo Nacional de Ciencias Naturales, CSIC, Madrid, 526 pp
- Martínez-Meyer E (2005) Climate change and biodiversity: some considerations in forecasting shifts in species potential distributions. *Biodivers Inform* 2:42–55
- Neldner VJ, Crossley DC, Cofinas M (1995) Using Geographic Information Systems (GIS) to determine the adequacy of sampling in vegetation surveys. *Biol Conserv* 73:1–17
- Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML (1990) Endemism centers, refugia and botanical collection density in Brazilian Amazonia. *Nature* 345:714–716
- Parnell JAN, Simpson DA, Moat J, Kirkup DW, Chantaranonthai P, Boyce PC, Bygrave P, Dransfield S, Jebb MHP, Macklin J, Meade C, Middleton DJ, Muasya AM, Prajaksood A, Pendry CA, Poomar R, Suddee S, Wilkin P (2003) Plant collecting spread and densities: their potential impact on biogeographical studies in Thailand. *J Biogeogr* 30:193–209
- Peterson AT, Slade NA (1998) Extrapolating inventory results into biodiversity estimates and the importance of stopping rules. *Divers Distrib* 4:95–105
- Peterson AT, Navarro-Sigüenza AG, Benítez-Díaz H (1998) The need for continued scientific collecting: a geographic analysis of Mexican bird specimens. *Ibis* 140:288–294
- Pulliam HR (1988) Sources, sinks and population regulation. *Am Nat* 132:652–661
- Pulliam HR (2000) On the relationship between niche and distribution. *Ecol Lett* 3:349–361
- Reddy S, Dávalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *J Biogeogr* 30:1719–1727
- Reutter BA, Helfer V, Hirzel AH, Vogel P (2003) Modelling habitat-suitability using museum collections: an example with three sympatric *Apodemus* species from the Alps. *J Biogeogr* 30:581–590
- Ricklefs RE, Schluter D (1993) Species diversity: regional and historical influences. In: Ricklefs RE, Schluter D (eds) *Species diversity in ecological communities*. University of Chicago Press, Chicago, pp 350–363
- Romo H, García-Barros E (2005) Distribución e intensidad de los estudios faunísticos sobre mariposas diurnas en la Península Ibérica e Islas Baleares (Lepidoptera, Papilionoidea y Hesperoidea). *Graellsia* 61:37–50
- Romo H, García-Barros E, Lobo JM (2006) Identifying recorder-induced geographic bias in an Iberian butterfly database. *Ecography* 29:873–885
- Soberón J, Llorente J (1993) The use of species accumulation functions for the prediction of species richness. *Conserv Biol* 7:480–488
- Soberón J, Peterson AT (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philos Trans R Soc Lond B* 359:689–698

- Soberón J, Llorente J, Benítez H (1996) An international view of national biological surveys. *Ann Miss Bot Garden* 83:562–573
- Soberón JM, Llorente J, Oñate L (2000) The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. *Biodivers Conserv* 9:1441–1466
- Soberón J, Jiménez R, Golubov J, Koleff P (2007) Assessing completeness of biodiversity databases at different spatial scales. *Ecography* 30:152–160
- Williams PH, Margules CR, Hilbert DW (2002) Data requirements and data sources for biodiversity priority area selection. *J Bioscience* 27:327–338
- Zaniewski AE, Lehmann A, Overton JM (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecol Modell* 157:261–280