

Targeting Sites for Biological Collections

**Jorge Soberón**

Instituto de Ecología, UNAM, Departamento de Ecología Evolutiva, Circuito Exterior, Ciudad Universitaria, Apdo. Postal 70–275, 04510 México D.F., México, and Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Avenida Liga Periférico - Insurgentes Sur No. 4903, Vol. Parques del Pedregal, Delegación Tlalpan, 14010 México D.F., México

Patricia Dávila

Escuela Nacional de Estudios Profesionales Iztacala, Unidad de biología, tecnología y prototipos (UBIPRO), Av de los Barrios s/s, Los Reyes, Iztacala Tlalnepantla, Apdo. Postal 314, Estado de México 54090, México

Jordan Golubov

Universidad Autónoma Metropolitana-Xochimilco Departamento El Hombre y su Ambiente, Calzada del Hueso 1100, Col. Villa Quietud, 04960, México D.F., México

Summary

Through the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Mexico has quality controlled, computerised and georeferenced 2.5 million of the 10 million specimens within the natural history collections held in herbaria and museums. Whilst analysis shows the data to be biased in many ways, it also reveals valuable insights into many of the problems that involve biodiversity. The benefits for conservation and the management of biological diversity handsomely repay the databasing effort and add significant value to existing natural history collections.

Introduction

“If maintained properly, specimens in every (natural history) collection can provide a permanent record of life on earth” (Mehrhoff, 1997). The advantages of specimens in natural history collections are several, but the main asset of these collections is a record of a particular species located at a particular place and time. These two groups of data link the specimen to a taxonomic reference (and therefore access to taxonomic, ecological and genetically-based information) and to a spatial component (making GIS technology and spatial dynamics available for analysis). The value of natural history collections is no longer restricted to conventional purposes such as identification, systematics research and teaching. They can now also be used for a variety of other purposes such as documenting changes in species distributional ranges (Favret and DeWalt, 2002), assessing the declining and threats to species over time (McCarthy, 1998, Shaffer *et al.*, 1998), comparing biodiversity between regions (Fagan and Kareiva, 1997), determining the completeness of inventories (Medellin and Soberón, 1999; León Cortes *et al.*, 1998), predicting crop damage (Sánchez-Cordero and Martínez-Meyer, 2000), for setting conservation priorities (Soberón *et al.*, 2000) and predicting species distributions under global environmental change (Peterson *et al.*, 2002). The use of specimen-based databases has expanded rapidly because the number of specimens that can be accessed has increased over time (Soberón, 1999). In the past decade, many specimens have been databased and are accessible in electronic format. Equally, the algorithms and computer technology needed to analyse large data sets has improved tremendously. In Mexico, there are approximately 10 million specimens deposited in herbaria and museums and of these, close to 2.5 million have been computerised, quality controlled and georeferenced (Soberón and Koleff, 1997). In addition, an equal number of Mexican specimens are deposited in foreign institutions. There is therefore, a large amount of information contained in databases that can now be analysed to address many of the problems that involve biodiversity.

Bias in Collection Data

However, the data in collections is not without its own problems. Databases are temporally biased because collection efforts are subject to discontinuous budgets creating “stochastic” fluctuation in collections, leading to differences between years and even variation between seasons. Collections are also spatially biased as many collection efforts are localised in close proximity to roads, research stations, institutions and cities (Soberón *et al.*, 2000). Many species in collections are represented by a small number of specimens. Such low numbers can confound the effects of these sampling and ecological/historical factors. Furthermore, collections may also be biased by taxonomic fashion where collection effort is usually centred on well known taxa or taxa that are currently in vogue (“taxonomist comfort”). As a consequence of these biases, samples are difficult to compare across sites, so it is extremely difficult and sometimes unreliable to obtain true estimates and assign priority areas for conservation or management.

Extensive and exhaustive collections are useful because: (1) species lists can then be compared between collecting sites; (2) they provide a tool to plan collection efforts; (3) they support the use of predictive tools for conservation and biodiversity studies; and (4) they help to predict and monitor changes in species richness in a given area and quantitatively determine the effect of modified habitats. However, even if the methodological tools for analysing collection effort have been developed (Colwell and Coddington, 1994), biases in collections may give unreliable estimates of the collection effort. Therefore, an effort to improve the quality control of collections is needed. To illustrate the bias of collections, we will use a butterfly database that was compiled from 20 museums (Mexican and foreign) that consists of close to 50,000 data points in 2,600 localities restricted to the *Pieridae* and *Papilionidae* families, around or about 70 species (Llorente *et al.*, 1997). Using this database, we can illustrate the spatial (Figure 12.1A) and temporal (Figure 12.1B) bias as well as identify two important aspects of collection efforts. The first bias is associated with the number of records in specific locations (Figure 12.2A) that exemplifies an intensive collection effort in a limited number of sites. The second bias aspect, related to the first, is the number of collected specimens per site which is usually strongly left skewed meaning that few species are collected at each site (Figure 12.2B).

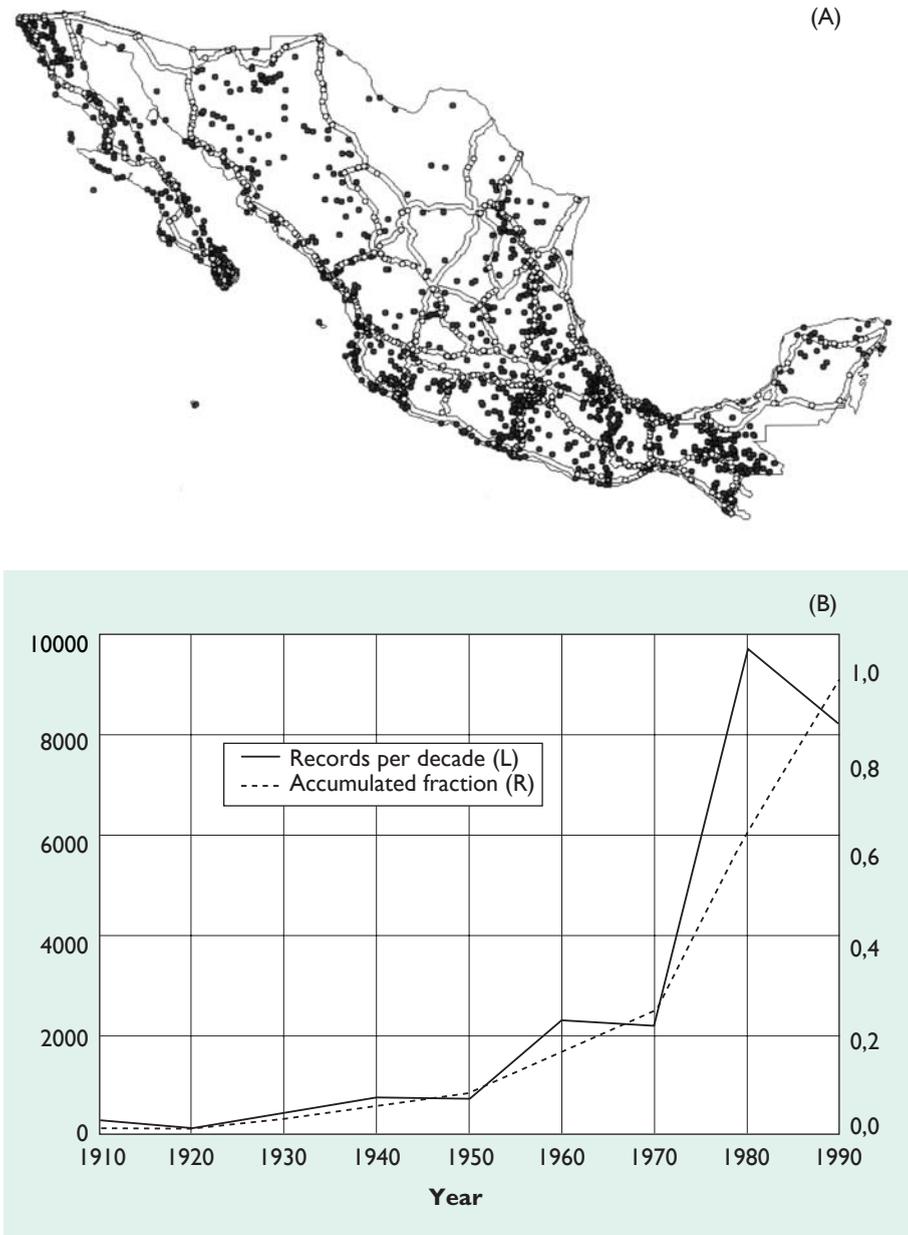


Figure 12.1 (A) Spatial bias of butterfly collections. Points correspond to actual collection points and lines correspond to major highways. (B) Number of collected specimens over time. Lines correspond to records per decade and the dotted lines are the accumulated fraction of specimens collected. See text for more detail.

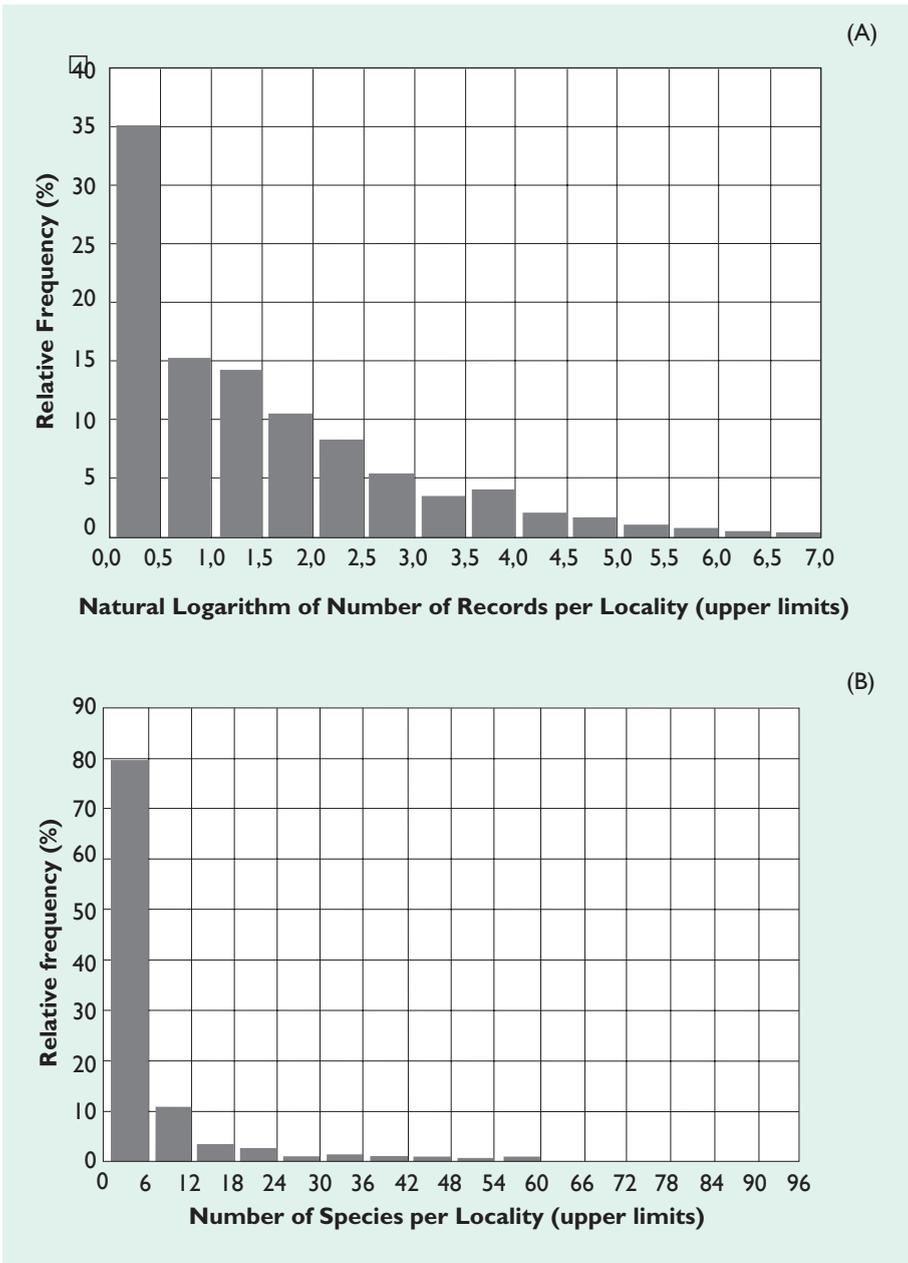


Figure 12.2 (A) Relative frequency of records (% of total records) as a function of the natural logarithm of records per site and (B) Relative frequency (% of total samples) of collected samples as a function of the number of species per site. See text for more detail.

Application for Targeting Collecting Sites

Accepting that the information found in museum specimens is biased, they are still a very valuable source of information for planning new exploration. Aggregated databases display gaps in collection efforts (Figure 12.3) that can provide a first approximation when targeting collection efforts. In addition to the spatial arrangement of data collection points, using remote sensing technology and GIS tools, the potential of identifying and planning collecting is enhanced significantly. For example, by aggregating data points of plants in a GIS framework (Figure 12.4) it is possible to overlap accessible digitised maps of a variety of different themes, such as vegetation maps that contain habitat type (disturbed and undisturbed vegetation cover), land use (agricultural, e.g., cattle grazing), political divisions, topography, major highways, geology, etc. The type of input data will largely depend on the objectives of the collection effort. If the objective is to identify areas of undisturbed habitat (Figure 12.5A) that have plant species of arid environments (Figure 12.5B), we have a first approximation of where to concentrate future collections. In addition, it is possible to add a further aspect such as political division (municipalities, Figure 12.6A) or compare the collection effort between adjacent areas (Figure 12.6B). These strategies taken together will optimise resources (human and economic), improve our understanding of biodiversity at local, regional and national scales and add even more value to current collections.

After deciding which species have enough occurrence data, it is possible to model the potential area of distribution. Originally, simple bioclimatic models were used to delimit the geographical range of a species (Soto and Gómez-Pompa, 1990; Carpenter *et al.*, 1993). Recently, a suite of algorithms has been developed that can display potential species distributions using a variety of input variables and model comparisons (Stockwell and Peters, 1999; Guisán and Zimmerman, 2000). These spatially explicit models, in essence, attempt to identify regions of high abiotic or biotic (depending on the model) similarity to those where the species has been documented, generating a grid that contains a probability of occurrence. These models have been used for a wide variety of applications: identification of “hot spots”; prioritisation of collection effort; and prediction of the incidence of invasive species (Soberon *et al.*, 2001) and global climate change (Peterson *et al.*, 2002). Specifically, we have used these models jointly with specimen databases to infer the likely distribution of *Opuntia* species in Mexico and the incidence of an invasive species that affects *Opuntia*. *Cactoblastis cactorum* is a South American moth that was successfully used as a biological control agent against several species of *Opuntia* during the 20th Century. After successful introductions into Australia, Hawaii and South Africa, it was introduced to the West Indies and from here has dispersed (either naturally or induced) throughout the

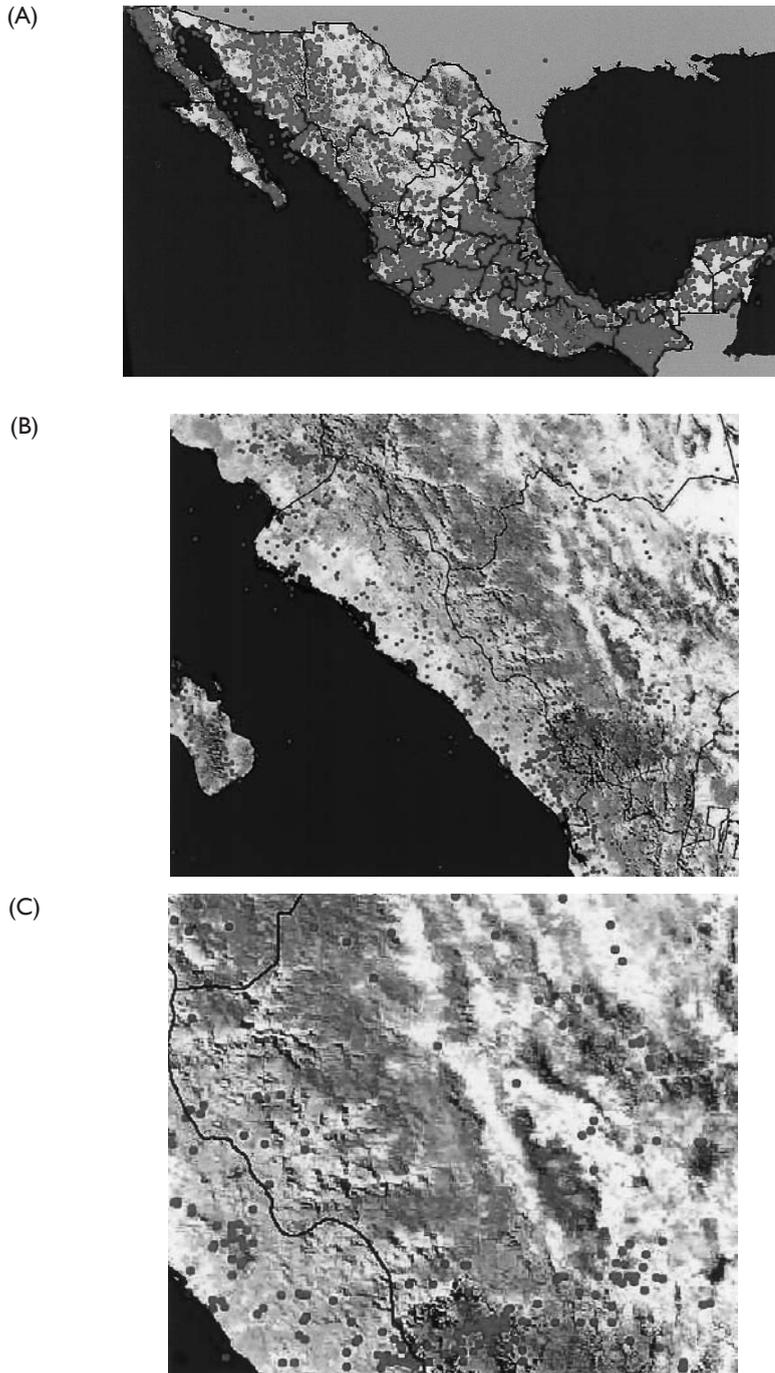


Figure 12.3 Data points of butterflies at national (A), regional (B) and local (C) scales illustrating the gaps in the collection effort at small spatial scales.

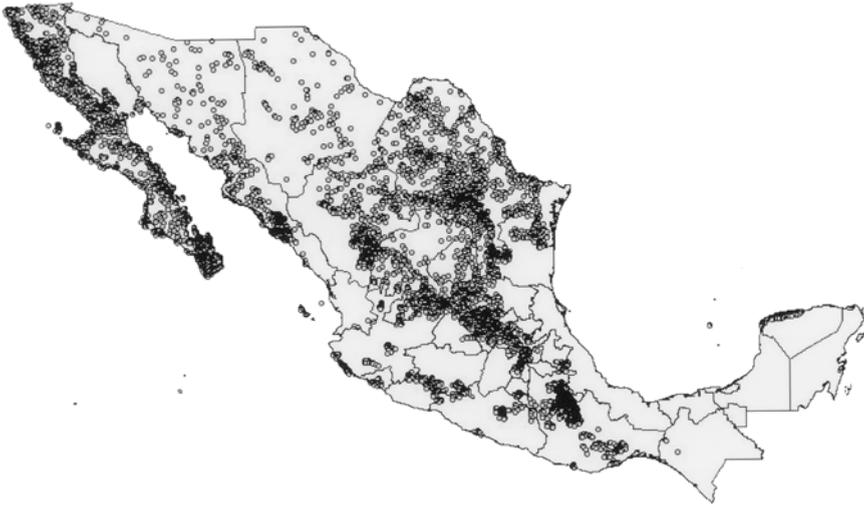


Figure 12.4 Data points of plants found in arid environments. Points represent a subsample from plant data base that contains close to 1.5 million specimens found in 320,000 localities, compiled from 100 different herbaria. The subsample was selected using a GIS and a classification of vegetation types of Mexico, by J. Rzedowsky and available in electronic format in www.conabio.gob.mx.

Caribbean islands. By 1989, it was documented in the Florida Keys. Since then, *C. cactorum* has spread in Florida attacking 6 species of *Opuntia* (Johnson and Stiling 1996). The impact that *C. cactorum* can have on native *Opuntia* populations in Mexico and the US is unprecedented. In Mexico alone, there are close to 3 million hectares of wild *Opuntia* populations that could be affected. We have used these algorithms to help plan the possible routes of invasion and most heavily affected regions in Mexico (Soberón *et al.*, 2001). The results of these studies could form the basis for targeting collections or other conservation actions.

In summary, Mexico possesses between 10 and 20% of all living organisms and is therefore considered to occupy fourth place in world biodiversity (Mittermeier and Goettsch, 1992). The complex topography, climate variability and biological/geological history have contributed to this level of diversity (Sarukhán *et al.*, 1996). Until a decade ago, there was no easy means of gathering information in Mexico. At the present, Mexico, through CONABIO, has ready access to over 2 million specimens in electronic format that have gone through exhaustive quality controls to ensure the reliability of the data. This extensive data set, coupled with the use of remote sensing technology and GIS tools will help us plan future collection efforts and to have a more comprehensive idea of what to collect and where to emphasise future

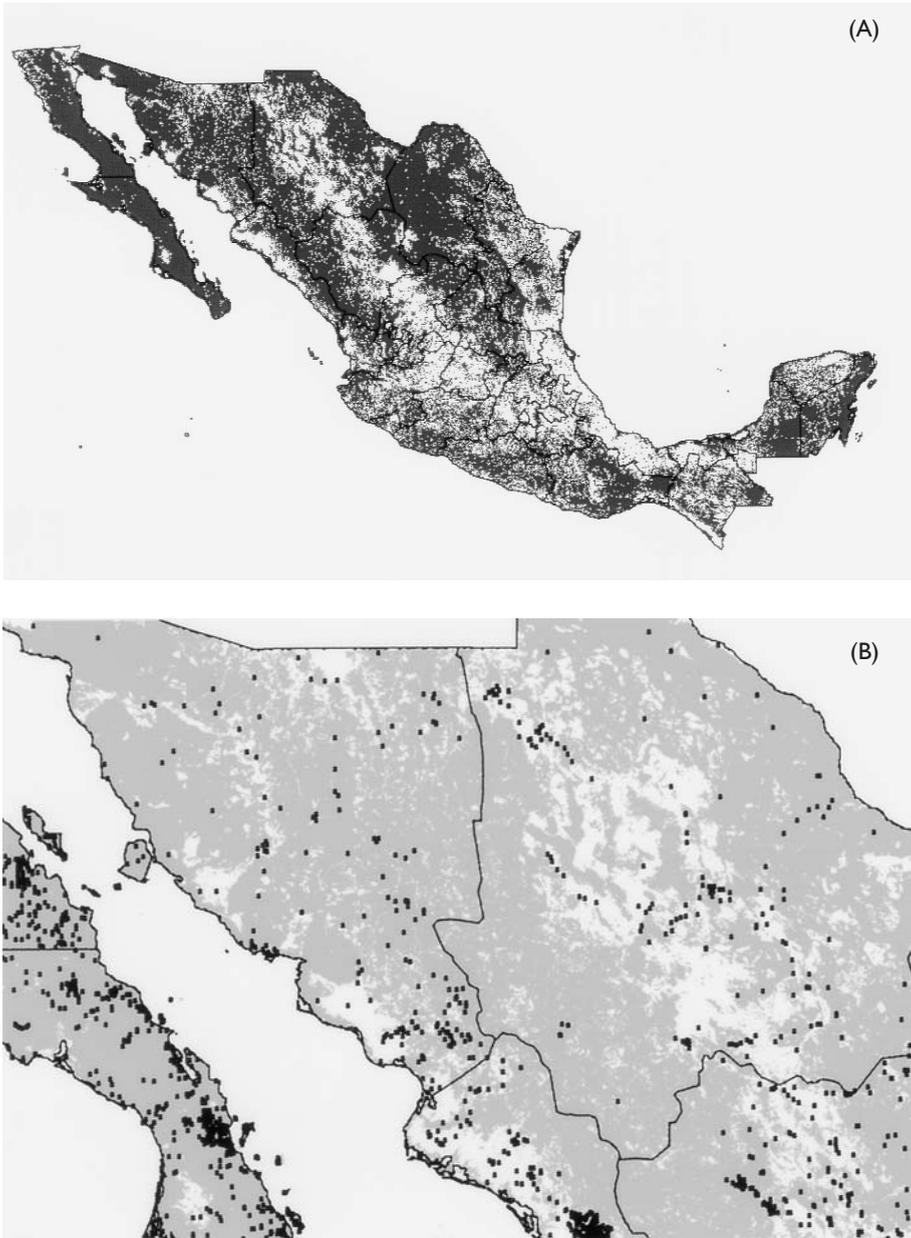


Figure 12.5 (A) Digitization of Landsat ETM map (National Forestry Inventory of Mexico 1:250,000) highlighting, by dark shading the undisturbed vegetation cover. (B) Data points of plant collections restricted to arid environments (dots) overlapped with undisturbed vegetation cover (grey shading).

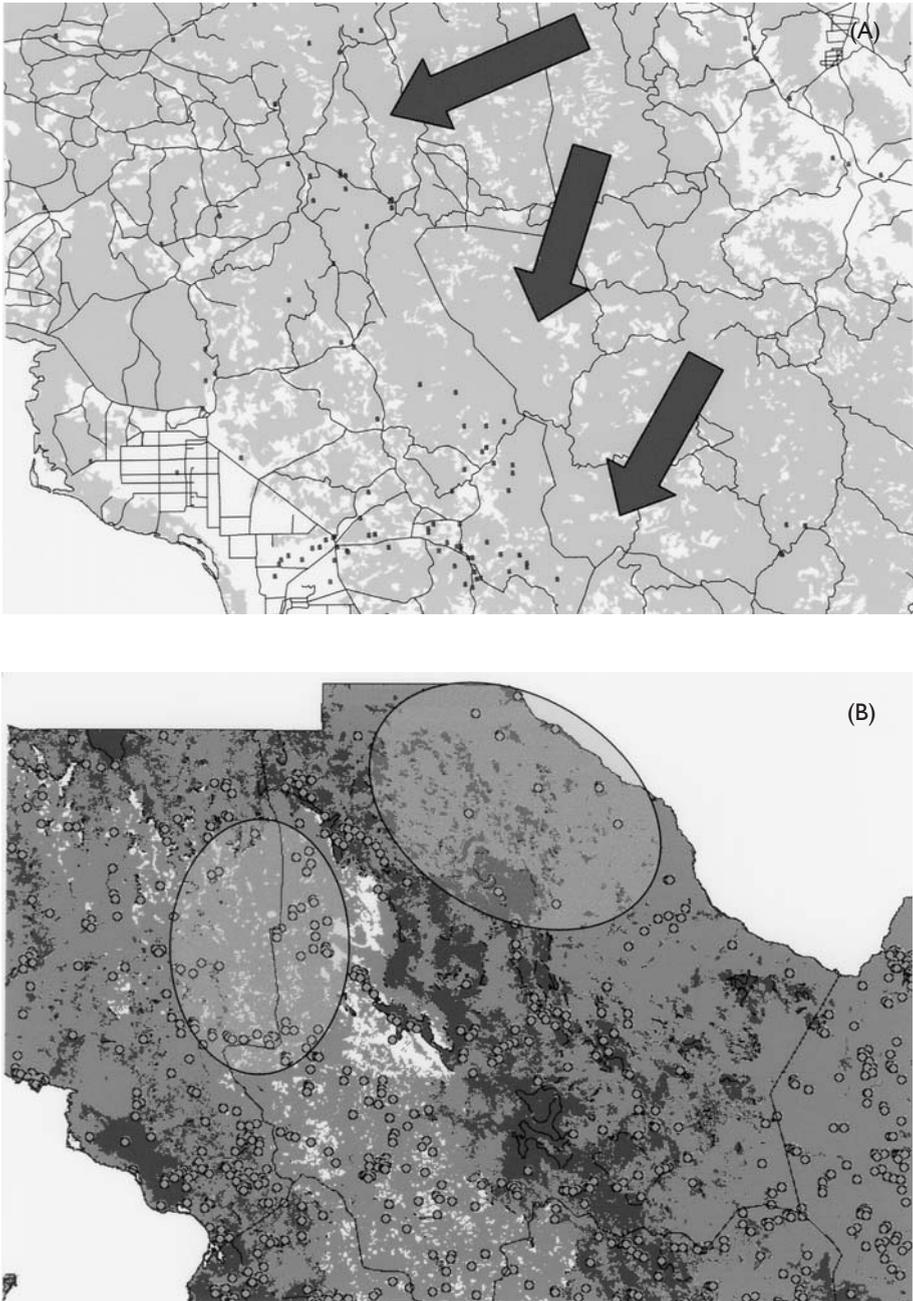


Figure 12.6 (A) Overlap of political boundaries, natural vegetation cover and data points of species restricted to arid environments. Arrows highlight areas of natural cover with very low or non-existent collecting efforts, and (B) differences in collection effort in adjacent areas of the Mexican northwest.

collection efforts. This task will undoubtedly yield new species and change our current knowledge of species distributions and how these can change in time and space. Access to information and analytical tools will eventually aid in the conservation and management of biological diversity and will add a rational view to collection efforts worldwide.

Acknowledgements

We would like to express our gratitude to the countless scientists and institutions around the world that believe that specimen data should be available in electronic format for everyone to use. Without their generosity our work could not be accomplished.

References

- Carpenter, G., Gilliason, A.N. and Winter, J. (1993). Domain: a flexible modeling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation* **2**: 667–680.
- Colwell, R.K. and Coddington, J.A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London, Series B-Biological Sciences*. **345**: 101–118.
- Fagan, W.F. and Kareiva, P.M. (1997). Using species lists to make biodiversity comparisons among regions: a test using Oregon butterflies. *Biological Conservation* **80**: 249–259.
- Favret, C. and Dewalt, R.E. (2002). Comparing the *Ephemeroptera* and *Plecoptera* specimen databases at the Illinois Natural History Survey and using them to document changes in the Illinois fauna. *Annals of the Entomological Society of America* **95**: 35–40.
- Guisian, A. and Zimmerman, N.E. (2000). Predictive habitat destruction models in ecology. *Ecological Modelling* **135**: 147–186.
- Johnson, D.M. and Stiling, P.D. (1996). Host specificity of *Cactoblastis cactorum* (*Lepidoptera: Pyralidae*), an exotic *Opuntia*-feeding moth, in Florida. *Environmental Entomology* **25**: 743–748.
- León-Cortés, L.L., Soberón-Mainero, J. and Llorente-Bousquets, J. (1998). Assessing completeness of Mexican sphinx moth inventories through species accumulation functions. *Diversity and Distributions* **4**: 37–44.
- Llorente, J., Oñate-Ocaña, L., Luis-Martínez, A. and Vargas-Fernández, I. (1997). *Papilionidae y Pieridae de México: Distribución geográfica e ilustración*. CONABIO-UNAM, México, D.F., México

- McCarthy, M.A. (1998). Identifying declining and threatened species with museum data. *Biological Conservation* **83**: 9–17.
- Medellín, R.A. and Soberón, J. (1999). Predictions of mammal diversity on four land masses. *Conservation Biology* **13**: 143–149.
- Mehrhoff, L.J. (1997). Museum, research collections, and the biodiversity challenge, pp. 447–465. In: M.L. Reaka-Kudla, D.E. Wilson and E.O. Wilson (eds). *Biodiversity II*. Joseph Henry Press, Washington DC., USA.
- Mittermeier, R. and Goetsch, C. (1992). La importancia de la diversidad biológica de México, pp. 62–73. In: J. Sarukhán and R. Dirzo (comps). *México ante los retos de la Biodiversidad*. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. México D.F., México.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sanchez-Cordero, V., Soberón, J., Buddemeier, R.H. and Stockwell, D.R.B. (2002). Future projections for Mexican faunas under global change scenarios. *Nature* **416**: 626–629.
- Sánchez-Cordero V. and Martínez-Meyer, E. (2000). Museum specimen data predict crop damage by tropical rodents. *Proceedings of the National Academy of Sciences* **97**: 7074–7077.
- Sarukhán, J., Soberón, J. and Larson-Guerra, J. (1996). Biological conservation in a high beta-diversity country, pp. 246–281. In: F. Di Castri and T. Younès (eds). *Biodiversity science and development: towards a new partnership*. CAB International, Wallingford, Oxon, UK.
- Shaffer, H.B., Fisher, R. and Davidson, C. (1998). The role of natural history collections in documenting species declines. *Trends in Ecology and Evolution* **13**: 27–30.
- Soberón, J. (1999). Linking biodiversity information sources. *Trends in Ecology and Evolution* **14**: 291.
- Soberón, J. and Koleff, P. (1997) The National Biodiversity Information System of Mexico, pp. 586–595. In: P.H. Raven (ed). *Nature and human society*. National Research Council, Washington, DC., USA.
- Soberón, J., Llorente J. and Oñate, L. (2000). The use of specimen label databases for conservation purposes: an example using mexican Papilionid and Pierid butterflies. *Biodiversity and Conservation* **9**: 1441–1466.
- Soberón, J., Golubov J. and Sarukhán, J. (2001). The importance of *Opuntia* in Mexico and the routes of invasion and impact of *Cactoblastis cactorum* (Lepidoptera: Pyralidae). *Florida Entomologist* **84**: 486–498.
- Soto, E.M. and Gómez-Pompa, A. (1990). *Climatología de la flora de Veracruz. Vol. I*. Instituto de Ecología A. C., Xalapa, México.
- Stockwell, D. and Peters, D. (1999). The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*. **13**: 143–158.

