



Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments

T. PATRICIA FERIA A.* and A. TOWNSEND PETERSON *Facultad de Ciencias, Universidad Nacional Autónoma de México, Apartado Postal 70–399, México, DF 04510, México; and Natural History Museum, University of Kansas, Lawrence, Kansas 66045, U.S.A.*

Abstract. Local biological communities are made up of species, each of which has its own particular relationship with the environment. To the extent that these autecological niches limit species' distributions, and by extension community composition, models of species' ecological niches can predict species composition at particular sites, or at least provide a null hypothesis of potential species composition in the absence of effects of species interactions. We developed distributional predictions (ecological niche models) for 89 species occurring in dry tropical forest in

the Balsas Basin of south-western Mexico using an interpolation technique, and predicted the species likely to occur at 8 sites across the region. Onsite field inventory data were then used to test the community predictions, all of which were statistically significant. These results suggest that inventory efforts can be made more efficient by development beforehand of hypotheses that focus onsite collecting and inventory.

Key words. Community, birds, predicted distributions, ecological niche modelling.

INTRODUCTION

A crisis exists with regard to the survey and inventory of life on Earth (Wilson, 1988). Given the unknown nature of much of biological diversity, development of rigorous approaches for evaluation and understanding of patterns of the distribution and diversity of that life for purposes of conservation becomes critical (Davis *et al.*, 1990; Vane-Wright *et al.*, 1991; Lomolino, 1994; Peyton, 1994; Williams *et al.*, 1996; Prendergast *et al.*, 1999; Peters & Myers, 1991–92). Considerable effort has focused recently on predicting species' geographical distributions based on known occurrences (Nix, 1986; Austin *et al.*, 1990; Jones & Gladkov, 1999). Given the poor state of knowledge of the distribution and natural history

of most species, these interpolation techniques must work even in situations of small numbers of known occurrence points (Peterson *et al.*, in press), as is the case for most species (Peterson *et al.*, 1998). One particular such interpolation technique, the Genetic Algorithm for Rule-set Prediction (GARP) (Stockwell & Noble, 1992; Stockwell & Peters, 1999; Stockwell & Peterson, in press) has seen extensive testing with diverse regions and taxa (Peterson & Cohoon, 1999; Peterson *et al.*, 1999; Peterson & Vieglais, 2001; Peterson, 2001; Peterson *et al.*, 2002; Peterson *et al.*, in press, Stockwell & Peterson, in press).

A critical extension of these approaches, however, is prediction of presences and absences across local communities of species. Single-species tests have proven successful in this challenge (Peterson & Cohoon, 1999; Stockwell & Peters, 1999; Godown & Peterson, 2000; Peterson & Vieglais, 2001; Peterson, 2001; Peterson *et al.*, 2002; Peterson *et al.*, in press, Stockwell & Peterson, in

* Present address: Department of Biology, University of Missouri, St. Louis, 8001 Natural Bridge Road, St. Louis, Missouri 63121-4499, U.S.A. E-mail: tpfz89@admiral.umsl.edu

press), although assumptions regarding historical effects on species' distributions may be necessary (Peterson *et al.*, 1999). However, tests of community-level predictions have not, to our knowledge, been attempted. We report the results of such a test here: we predict local communities of birds across the Balsas Basin of south-western México one species at a time, and test predictions based on independent field inventories. If such a methodology is viable, it may prove invaluable in making future surveys and inventories more efficient, providing preliminary hypotheses of key areas for conservation, and prioritizing areas for detailed study.

METHODS

Distributional data were derived from specimens housed in 32 natural history museums (Peterson *et al.*, 1998) (see Acknowledgements) and summarized in the *Atlas of Mexican Bird Distributions* (Navarro and Peterson, in prep.). All records were georeferenced via direct consultation of 1 : 100 000 and 1 : 50 000 topographic maps (Peterson *et al.*, 1998), and unique locality records accumulated in spreadsheets for each species. As an initial filter, locality records (12 308 records of a total of 354 species in the region 17°–20°N latitude and 97°30'–103°5'W longitude) were separated for analysis. Later, we identified 89 resident species for which ≥ 20 specimens were available (median 61, range 20–399 specimens; median 29, range 8–81 unique latitude-longitude combinations); the range of these species is broadly within the Balsas Basin's dry desert scrub vegetation, based on personal experience and published habitat descriptions (Howell & Webb, 1995). Ecological/environmental base layers (elevation, annual mean precipitation, annual mean temperature and potential vegetation) were obtained from the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (<http://www.conabio.gob.mx>) at a resolution of 0.01° (approximately 1 km) grid cells, and the study region was clipped out for analysis; point-occurrence data were rasterized to this same grid system. Prior studies have shown that all four environmental data sets contribute significantly to the predictions (Peterson & Cohoon, 1999; Peterson, unpubl. data).

Species' ecological niches can be defined as the range of conditions under which they are able

to maintain populations without immigration (Grinnell, 1917), a definition more easily visualized in geographical dimensions than later views (MacArthur, 1972). One approach for modelling these niches is GARP, a machine-learning procedure based on a genetic algorithm (Stockwell & Noble, 1992; Stockwell & Peters, 1999; Stockwell, 1999). Occurrence points are divided evenly into training and test data sets. GARP works in an iterative process of rule selection, evaluation, testing, and incorporation or rejection: a method is chosen from a set of possibilities (e.g. logistic regression and bioclimatic rules), applied to the training data, and a rule is developed or evolved. Predictive accuracy is then evaluated based on 1250 points resampled from the test data and 1250 points sampled randomly from the study region as a whole to provide pseudoabsences (note that these subsamples and tests are internal to GARP, and are not available to the user for model validation, making necessary additional subsampling prior to analysis). Rules (which take the form of IF ... THEN statements) may evolve by a number of means that mimic DNA evolution: point mutations, deletions, crossing over, etc. The change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the model, and the algorithm runs either 2560 iterations or until convergence. When the modelling process ceases, the ecological niche model is projected back onto the landscape to predict the species' geographical distribution in the form of a raster grid, in which each grid square is designated as present or absent. GARP has seen strong statistical support for its predictive abilities under diverse conditions (Peterson & Cohoon, 1999; Peterson & Vieglaiss, 2001; Peterson, 2001; Peterson *et al.*, 2002; Peterson *et al.*, in press; Stockwell & Peterson, in press).

We developed GARP models for all 89 species, exported resulting predictions as ASCII raster grids, and uploaded them as GIS layers into ArcView (version 3.2). In ArcView, the 89 species' predictions were overlaid, permitting queries of expected presence and absence for each species at any point across the study region.

To test these community predictions, we accumulated single-site inventory results from across the Balsas Basin: in all, data from eight localities were available to us. The sample sites were

chosen *a priori* to fulfil three requirements: (a) to represent the variety of environments — and likely biotic communities—across the Balsas Basin, (b) to be accessible to permit efficient surveys, and (c) to be located in areas that were safe for access. Three were based on previous undergraduate thesis studies (Villasenor Gomez, 1985; Feria, 1997; Mendez R., 1997), one was based on an inventory developed as part of a field course (Peterson *et al.*, unpubl. data), and four were surveyed by Feria as part of this study (Feria, 2001). In all cases, inventories consisted of regular surveys by experienced field observers, in which species were recorded by sight and by call; sampling intensity ranged from 4 days (Peterson *et al.*, unpubl. data) to many months (Villasenor Gomez, 1985). For the six inventories in which we were involved, accumulation curves were used to assess inventory completeness (Peterson & Slade, 1998). For model validation, for each site, predicted and observed faunas were organized in 2×2 contingency tables of species predicted present and absent vs. species actually present and absent [= confusion matrices (Fielding & Bell, 1997)], and compared using chi-square tests.

RESULTS AND DISCUSSION

We developed GARP models for all 89 bird species (11 effectively endemic to the region) in the dry desert scrub vegetation of the Balsas Basin (examples in Fig. 1). Although the robustness of these predictions was not tested individually, owing to relatively small sample sizes within the Balsas Basin, prior tests of this type based on well-sampled species across Mexico analysed with the same data sets have indicated excellent predictive ability (Peterson *et al.*, 2002). Here, we analysed all species that met minimal sampling criteria, including many for which sample sizes were quite small. Combining these models across species produced clear concentrations of species richness and endemism (Fig. 2): both were focused in narrow rings around the deepest and driest parts of the basin (see Feria & Peterson, unpubl. data, for a quantitative analysis of conservation priorities).

We validated these hypotheses of species composition at eight localities across the Basin (Fig. 1). Accumulation curves for the six sites surveyed by us were all asymptotic, or nearly so,

suggesting that the inventories were close to complete (Feria & Peterson, unpubl. data; but see below). At each of the eight sites, presence and absence of species was predicted statistically significantly better than at random ($P < 0.05$), with false-negative error rates ranging 2.3–19.1% and false-positive error rates of 8.5–40.5% (Table 1).

Interestingly, error in predictions was principally in terms of predicting species present that were not detected in the onsite inventories (false-positive error) (Table 1). At five sites, false-positive error exceeded false-negative error, but only at one was the converse the case (two sites were tied between the two error components). This difference suggests non-detections in inventories, rather than model failure, as a principal factor in producing prediction error (Peterson & Slade, 1998).

To understand patterns of model failure in greater depth, we inspected lists of species on which the error statistics were based. The species most commonly predicted present but not detected were low-density (e.g. *Campephilus guatemalensis*), vulnerable to hunting pressure (e.g. *Philortyx fasciatus*), or difficult to detect and identify (e.g. *Cyananthus sordidus*) species. In contrast, species unexpectedly present included species poorly sampled in the specimen data base (e.g. *Cathartes aura*) and species associated with human disturbance (e.g. *Carduelis psaltria*). Hence, although many species were mispredicted at the eight test sites, patterns of error distribution suggest that further refinement of models and inventory data can yield much-improved prediction of species' presences and absences.

The ability to predict many species' distributions simultaneously may seem a minor extension of the ability to predict single species. If one species can be modelled with sufficient accuracy, as GARP has been shown to do efficiently in several studies (Peterson & Cohoon, 1999; Peterson *et al.*, 1999; Peterson *et al.*, 2000; Peterson & Vieglais, 2001; Peterson, 2001; Peterson *et al.*, 2002), at least for birds, then extension to multiple species should be a logical consequence. This methodology may, of course, have different strengths at different spatial scales (Ortega-Huerta & Peterson, unpubl. data). For instance, at regional or continental scales, predictive ability may be high, yet at local scales, where interaction effects may be more common, predictive ability may prove lower.

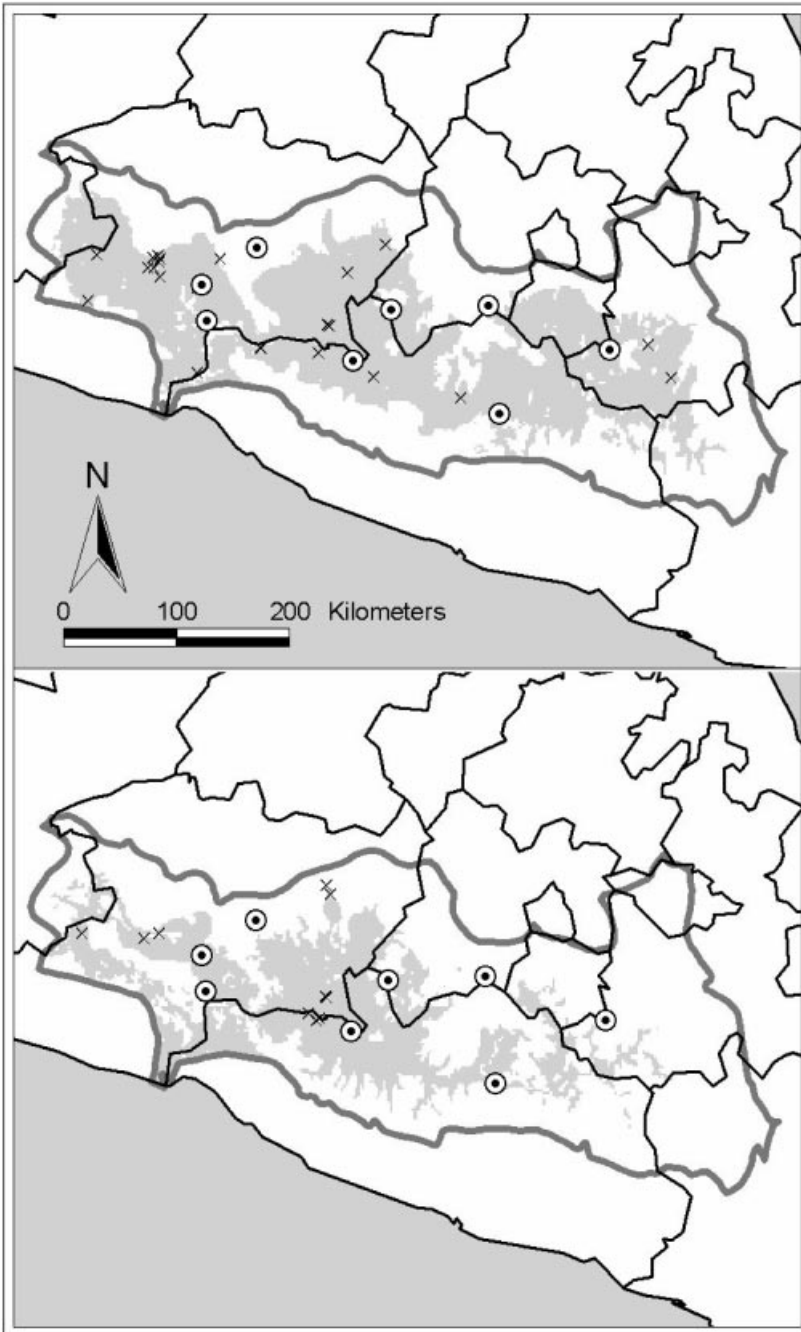


Fig. 1 Example maps showing modelled geographical distributions of *Passerina leclancherii* (top) and *Cacicus melanicterus* (bottom) in the portion of their distributions that falls in the Balsas Basin of south-western Mexico (thick, grey line shows limits of study region). Gray shading indicates predicted presence, and crosses represent known occurrences (note that some are very close to others and not all are visible); circles indicate the eight test inventory sites.

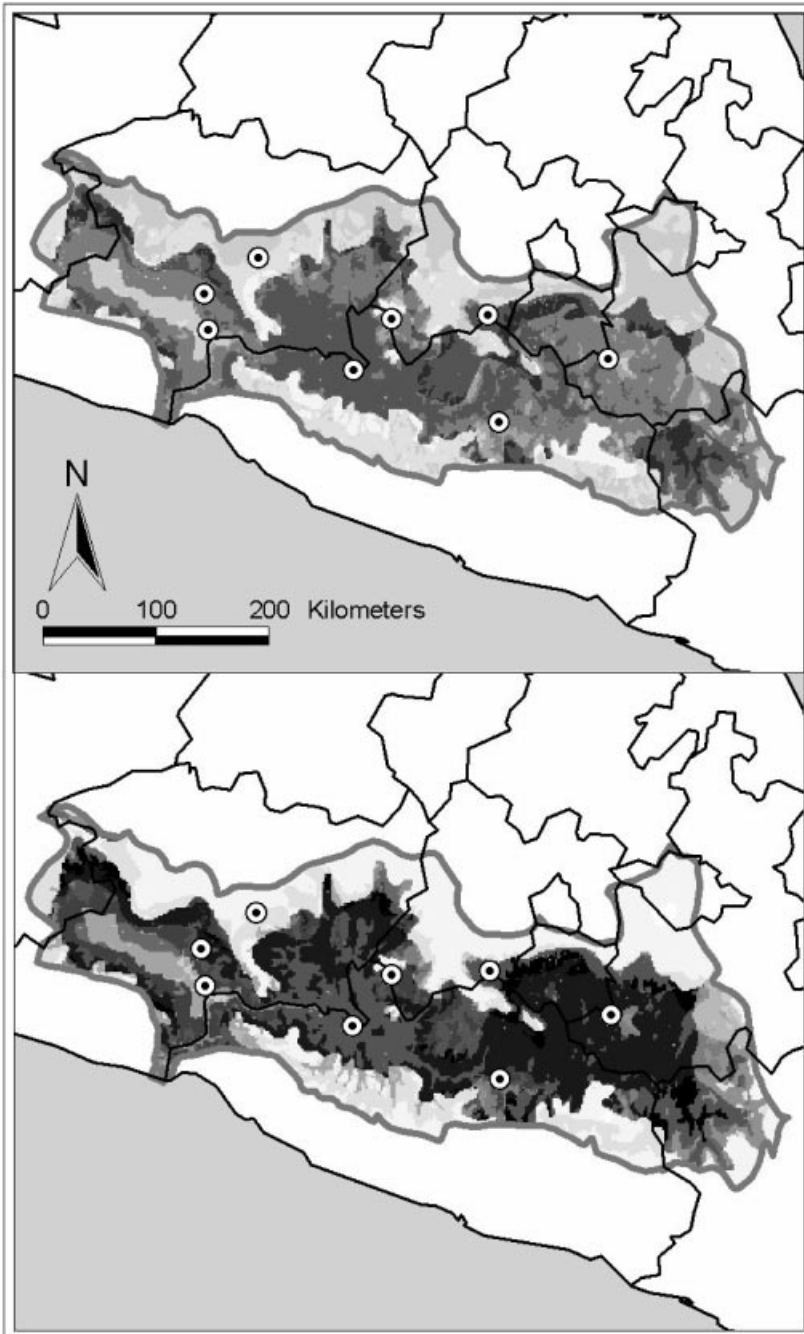


Fig. 2 Geographic patterns summed across 89 bird species in dry tropical forests in the Balsas Basin of southwestern Mexico: top, species richness (grey scale: 0 species = white, 1–9 species = lightest grey, 10–19 species = light grey, ... 60–66 species = black); bottom, richness of 16 species endemic to Mexico (Escalante-Pliego *et al.*, 1993; Peterson & Navarro-Siguenza, 1999; Navarro-Siguenza & Peterson, 2000) (grey scale: 0 species = white, ... 16 species = black).

Table 1 Summary of test sites and predictive accuracy of community predictions at the eight test sites across the Balsas Basin, south-western Mexico

No.	Locality and source	Data source	Latitude	Longitude	Species predicted	Species observed
1	Zicuiran, Michoacán	Villaseñor Gomez (1985)	18°57'	101°54'	74	50
2	Tacámbaro, Michoacán	Méndez (1997)	19°14'	101°28'	71	70
3	Infiernillo, Michoacán	Feria (2001)	18°40'	101°52'	43	36
4	Las Tinajas, Guerrero	Feria (2001)	18°21'	100°43'	40	37
5	Bejucos, Mexico	Feria (2001)	18°45'	100°25'	38	38
6	Valente Trujillo, Guerrero	Feria (2001)	17°56'	99°34'	32	37
7	Huautla, Morelos	Peterson & Feria (unpubl. data)	18°47'	98°99'	44	44
8	Tlancualpicán, Puebla	Feria (1997)	18°26'	98°42'	77	43

Nevertheless, this demonstration — GARP community predictions corroborated with actual field inventory data — opens the way for several new capabilities for biodiversity surveys, inventories and analyses. Community composition can be assessed based on existing information across large regions before initiation of fieldwork, and key sites (hotspots of diversity and endemism, areas with unexpected species combinations, etc.) can be chosen for field study. These key sites can then be surveyed to produce more reliable onsite field data, but less critical or less interesting sites can be accorded lower priority. In this way, the always-limited resources available for biodiversity survey and inventory (Peterson *et al.*, 1998) can be put to maximum use, and the information obtained from results can be maximized. Although error in our predictions was non-trivial, additional testing under diverse conditions combined with more robust inventory data may produce a methodology that has the necessary predictive ability.

An additional theoretical implication also arises. To the extent that species composition can be predicted from individual, niche-based models for each species, then autecology would emerge as the force that dominates the shaping of biological communities, as opposed to widespread interactions among species. Of course, this conclusion will depend largely on the spatial scale at which the question is asked, so this first result must await detailed analysis and testing. Nevertheless, our approach presents a new opportunity to attack this fundamental, but very difficult, ecological question. Obviously, the evidence remains equivocal, but parallel tests of community composition vs. species' autecology in diverse

regions for varied taxa would present a very useful insight.

ACKNOWLEDGMENTS

We thank Dr A.G. Navarro S. for his kind assistance in many phases of this study. Museums kindly providing data include: American Museum of Natural History, New York; Academy of Natural Sciences, Philadelphia; Bell Museum of Natural History, Minneapolis; British Museum (Natural History), Tring; Carnegie Museum of Natural History, Pittsburg; Canadian Museums of Nature, Ottawa; Denver Museum of Natural History; Delaware Museum of Natural History; Fort Hays State University; Field Museum of Natural History, Chicago; Iowa State University, Ames; University of Kansas Natural History Museum, Lawrence; Los Angeles County Museum of Natural History; Natuurhistorische Museum, Amsterdam; Louisiana State University Museum of Zoology, Baton Rouge; Museum of Comparative Zoology, Harvard University, Cambridge; Moore Laboratory of Zoology, Occidental College, Los Angeles; Museum Nationale d'Histoire Naturelle, Paris; Museum of Vertebrate Zoology, Berkeley; Museo de Zoología, Facultad de Ciencias, Universidad Nacional Autónoma de México; University of Nebraska, Lincoln; Royal Ontario Museum, Toronto; San Diego Natural History Museum; Southwestern College, Kansas; Texas Cooperative Wildlife Collections; University of Arizona; University of British Columbia Museum of Zoology; University of California at Los Angeles; Universidad Michoacana de San Nicolás de Hidalgo; US National

Museum of Natural History; Western Foundation of Vertebrate Zoology; and Peabody Museum, Yale University. This work was supported by the US National Science Foundation, and TPFA was supported by a graduate fellowship from the Consejo Nacional de Ciencia y Tecnología.

REFERENCES

- Austin, M.P., Nicholls, A.O. & Margules, C.R. (1990) Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* **60**, 161–177.
- Davis, F.W., Stoms, D.M., Estes, J.E., Szepean, J. & Scott, J.M. (1990) An information systems approach to the preservation of biological diversity. *International Journal of Geographic Information Systems* **4**, 55–78.
- Escalante-Pliego, P., Navarro-Siguenza, A.G. & Peterson, A.T. (1993) A geographic, historical, and ecological analysis of avian diversity in Mexico. *Biological diversity of Mexico: origins and distribution* (ed. by T.P. Ramamoorthy, R. Bye, A. Lot and J. Fa), pp. 281–307. Oxford University Press, New York.
- Feria, A.T.P. (1997) Diversidad y distribución avifaunística en una localidad del Municipio de Chiautla de Tapia, Puebla. Tesis Profesional. FES Zaragoza, UNAM. Mexico, D.F.
- Feria, A.T.P. (2001) Patrones de distribución de las aves residentes de la Cuenca del Balsas. Tesis de maestría. Facultad de Ciencias, UNAM. Mexico, D.F.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**, 38–49.
- Godown, M.E. & Peterson, A.T. (2000) Preliminary distributional analysis of U.S. endangered bird species. *Biodiversity and Conservation* **9**, 1313–1322.
- Grinnell, J. (1917) Field tests of theories concerning distributional control. *American Naturalist* **51**, 115–128.
- Howell, S.N.G. & Webb, S. (1995) *A guide to the birds of Mexico and northern Central America*. Oxford University Press, Oxford.
- Jones, P.G. & Gladkov, A. (1999) *Floramap: a computer tool for predicting the distribution of plants and other organisms in the wild*. Centro Internacional de Agricultura Tropical, Cali, Colombia.
- Lomolino, M.V. (1994) An evaluation of alternative strategies for building networks of nature reserves. *Biological Conservation* **69**, 243–249.
- MacArthur, R. (1972) *Geographical ecology*. Princeton University Press, Princeton, NJ.
- Mendez, R.M.E. (1997) Contribución al conocimiento de las aves de la región de Tacámbaro, Centro del Estado de Michoacán, México. Tesis Profesional. Escuela de Biología. UMSNH. Morelia, Michoacán, México.
- Navarro-Siguenza, A.G. & Peterson, A.T. (2000) Western Mexico: a significant center of avian endemism and challenge for conservation action. *Cotinga* **14**, 42–46.
- Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. *Atlas of elapid snakes of Australia* (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.
- Peters, R.L. & Myers, J.P. (1991–92) Preserving biodiversity in a changing climate. *Issues in Science and Technology* **1991–92**, 66–72.
- Peterson, A.T. (2001) Predicting species' geographic distributions based on ecological niche modeling. *Condor* **103**, 599–605.
- Peterson, A.T., Ball, L.G. & Cohoon, K.C. (2002) Predicting distributions of tropical birds. *Ibis* (in press).
- Peterson, A.T. & Cohoon, K.C. (1999) Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling* **117**, 159–164.
- Peterson, A.T., Egbert, S.L., Sanchez-Cordero, V. & Price, K.P. (2000) Geographic analysis of conservation priorities using distributional modelling and complementarity: endemic birds and mammals in Veracruz, Mexico. *Biological Conservation* **93**, 85–94.
- Peterson, A.T. & Navarro-Siguenza, A.G. (1999) Alternate species concepts as bases for determining priority conservation areas. *Conservation Biology* **13**, 427–431.
- Peterson, A.T., Navarro-Siguenza, A.G. & Benitez-Diaz, H. (1998) The need for continued scientific collecting: a geographic analysis of Mexican bird specimens. *Ibis* **140**, 288–294.
- Peterson, A.T. & Slade, N.A. (1998) Extrapolating inventory results into biodiversity estimates and the importance of stopping rules. *Diversity and Distributions* **4**, 95–105.
- Peterson, A.T., Soberon, J. & Sanchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science* **285**, 1265–1267.
- Peterson, A.T., Stockwell, D.R.B. & Kluza, D.A. (in press) Distributional prediction based on ecological niche modeling of primary occurrence data. *Predicting species occurrences: issues of scale and accuracy* (ed. J.M. Scott). Island Press, Washington, DC.
- Peterson, A.T. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modeling. *BioScience* **51**, 363–371.
- Peyton, B. (1994) Conservation in the developing world: ideas on how to proceed. *International Conference on Bear Reserve Management* **9**, 115–127.

- Prendergast, J.R., Quinn, R.M. & Lawton, J.H. (1999) The gaps between theory and practice in selecting nature reserves. *Conservation Biology* **13**, 484–492.
- Stockwell, D.R.B. (1999) Genetic algorithms II. *Machine learning methods for ecological applications* (ed. A.H. Fielding), pp. 123–144. Kluwer Academic Publishers, Boston.
- Stockwell, D.R.B. & Noble, I.R. (1992) Induction of sets of rules from animal distribution data: a robust and informative method of analysis. *Mathematics and Computers in Simulation* **33**, 385–390.
- Stockwell, D.R.B. & Peters, D.P. (1999) The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographic Information Systems* **13**, 143–158.
- Stockwell, D.R.B. & Peterson, A.T. (in press) Controlling bias in biodiversity data. *Predicting species occurrences: issues of scale and accuracy* (ed. J.M. Scott). Island Press, Washington, DC.
- Vane-Wright, R.I., Humphries, C.J. & Williams, P.H. (1991) What to protect? Systematics and the agony of choice. *Biological Conservation* **55**, 235–254.
- Villasenor, L.E. (1985) Avifauna de la Presa Zicuiran, Depresion del Balsas inferior, Michoacan, Mexico. Tesis Profesional, Escuela de Biologia. U.M.S.N.H. Morelia, Michoacan.
- Williams, P.H., Prance, G.T., Humphries, C.J. & Edwards, K.S. (1996) Promise and problems in applying quantitative complementary areas for representing the diversity of some Neotropical plants (families Dichapetalaceae, Lecythidaceae, Caryocaraceae, Chrysobalanaceae and Proteaceae). *Biological Journal of the Linnean Society* **58**, 125–157.
- Wilson, E.O. (1988) The current state of biological diversity. *Biodiversity* (ed. E.O. Wilson). National Academy Press, Washington.