

Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela

Robert P. Anderson* *Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West at 79th Street, New York, NY, USA*

Abstract

Aim Numerous geographical information system (GIS)-based techniques for estimating a species' potential geographical distribution now exist. While a species' potential distribution is more extensive than its documented range, the lack of records from some suitable regions may simply derive from inadequate sampling there. Using occurrence records of both the study species and the more inclusive overall target group, I propose a progression of statistical models to evaluate apparent absences in species distributions.

Location Northern Venezuela.

Methods Employing data from the Smithsonian Venezuelan Project (a large set of standardized mammalian inventories undertaken across Venezuela), I tested distributional hypotheses for the sigmodontine rodent *Oryzomys albigularis* (Tomes, 1860). Those inventories collected *O. albigularis* in two of the five major montane regions of northern Venezuela (the Cordillera de Mérida/Macizo de El Tamá and Cordillera de la Costa Central). I used the Genetic Algorithm for Rule-Set Prediction (GARP) to estimate the species' potential distribution in northern Venezuela. Then, based on all collection localities from the Smithsonian Venezuelan Project, I determined the probability that the absence of *O. albigularis* from the three regions of potential presence where it was not documented (the Serranía de Perijá, Lara–Falcón highlands, and Cordillera de la Costa Oriental) could be the result of inadequate sampling.

Results and main conclusions All statistical models indicated that the sampling efforts of the Smithsonian Venezuelan Project were insufficient to demonstrate conclusively the absence of *O. albigularis* from any of the three regions lacking records. Indeed, a subsequent compilation of specimens from ten natural history museums confirmed its presence in the Serranía de Perijá and the Lara–Falcón highlands. Tests using empirical sampling effort and taking human modification of the landscape into account most closely fulfilled the assumptions required for the tests. By providing a framework for bringing additional quantitative rigour to studies of species distributions, these methods will probably prove of wide applicability to other systems.

Keywords

Absence, distribution, genetic algorithm, geographical information system, locality, modelling, museum specimens, presence, range, sampling.

Resumen

Hoy en día existen varias técnicas asociadas con sistemas de información geográfica (SIG) para estimar la distribución geográfica potencial de una especie. Aunque la distribución potencial de una especie es más extensa que su distribución documentada, la

*Correspondence: Robert P. Anderson, Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA. E-mail: rpa@amnh.org

falta de registros de la especie en algunas regiones de presencia potencial puede simplemente derivar de un inadecuado esfuerzo de muestreo. Con base en localidades tanto de la especie de estudio como del grupo funcional (taxonómico-ecológico) más inclusivo, se propone una progresión de modelos estadísticos para evaluar las ausencias aparentes en la distribución de una especie. Usando datos del Smithsonian Venezuelan Project (una serie grande de inventarios uniformizados para mamíferos, llevados a cabo en todas las regiones de Venezuela), se probaron varias hipótesis sobre la distribución del roedor sigmodontino *Oryzomys albigularis* (Tomes, 1860). Tales inventarios recolectaron a *O. albigularis* en dos de las cinco principales regiones montañas del norte de Venezuela (la Cordillera de Mérida/Macizo de El Tamá y la Cordillera de la Costa Central). Se utilizó el Genetic Algorithm for Rule-Set Prediction (GARP) para estimar la distribución potencial de la especie en el norte de Venezuela. Después, con base en todas las localidades de colección del Smithsonian Venezuelan Project, se determinó la probabilidad de que la ausencia de *O. albigularis* en cada una de las tres regiones de presencia potencial en las cuales no fue documentada (la Serranía de Perijá, las tierras altas de Lara–Falcón y la Cordillera de la Costa Oriental) podría haberse derivado de un esfuerzo de muestreo inadecuado. Los resultados de todos los modelos estadísticos indicaron que los esfuerzos de muestreo del Smithsonian Venezuelan Project fueron insuficientes para demostrar de manera concluyente la ausencia de *O. albigularis* en cualquier de las regiones que carecen de registros de la especie. Efectivamente, una reciente compilación de ejemplares de referencia pertenecientes a diez museos de historia natural confirmó la presencia de la especie en la Serranía de Perijá y en las tierras altas de Lara–Falcón. Las pruebas en las cuales se integraron medidas empíricas de esfuerzo de muestreo y se tomaron en cuenta modificaciones del paisaje por el hombre fueron las que más adecuadamente cumplieron los supuestos requeridos. Los métodos acá propuestos probablemente encontrarán aplicación amplia en otros sistemas, ya que presentan una estructura lógica para facilitar mayor rigor cuantitativa en estudios de las distribuciones de las especies.

Palabras claves

algoritmo genético, ausencia, distribución, ejemplares de referencia, localidad, modelaje, muestreo, presencia, SIG.

INTRODUCTION

The geographical distribution and artefactual absences

Species geographical ranges constitute the fundamental basis for subsequent biogeographical research, and for many evolutionary studies and important current applications to conservation biology, invasive-species management and public health (e.g. Udvardy, 1969; MacArthur, 1972; Nelson & Platnick, 1981; Brooks & McLennan, 1991; Ricklefs & Schluter, 1993; Mace & Stuart, 1994; Funk *et al.*, 1999; Ron, 2000; Zalba *et al.*, 2000; Peterson & Vieglais, 2001; Lizcano *et al.*, 2002; Rodrigues-Coura *et al.*, 2002). However, the distributions of many taxa are extremely poorly known, particularly in the tropics (Raven & Wilson, 1992; Voss & Emmons, 1996; Peterson *et al.*, 1998). Observation or collection of a species at particular geographical localities documents its geographical distribution. In contrast, it is more difficult to substantiate a species' absence from sites where it has not been documented (MacKenzie *et al.*, 2002). This asymmetry derives primarily from inadequate sampling effort, which leads to artefactual absences in species

distributions (Voss & Emmons, 1996, pp. 42–43; Ponder *et al.*, 2001). Most regions have been only minimally and unevenly surveyed biologically, even for relatively well-known groups (Prendergast *et al.*, 1993; Fagan & Kareiva, 1997; Kress *et al.*, 1998). Furthermore, sampling remains incomplete at the vast majority of sites that have been sampled at all (Soberón-M. & Llorente-B., 1993; Colwell & Coddington, 1994; Nichols *et al.*, 1998; Peterson & Slade, 1998; MacKenzie *et al.*, 2002). Two lines of current research (single-species distributional modelling and multi-species geographical analyses) have converged on the problem of artefactual absences in species distributions, and each offers a key element lacked by the other.

A single-species modelling approach provides a map showing the species' potential distribution but does not incorporate data regarding collection effort for other taxonomically related and ecologically similar species. Recent geographical information system (GIS)-based techniques that model a species' requirements using environmental characteristics of localities of known occurrence represent a great advance in the production of range maps (e.g. Busby, 1986; Walker & Cocks, 1991; Carpenter *et al.*, 1993; Skov, 2000;

Peterson, 2001; Anderson *et al.*, 2002a; for a broader overview of geographical modelling tools with more diverse objectives, see James & McCulloch, 2002). These applications operationally divide the study region into habitable vs. uninhabitable areas for the species. However, the species' predicted potential distribution is consistently more extensive than its documented range (Skov & Borchsenius, 1997). Some geographical regions of potential distribution from which it is not known may be inhabited by closely related species (which have similar requirements; Peterson *et al.*, 1999; Anderson *et al.*, 2002b) or represent suitable areas from which the species has been excluded historically by barriers or in which it has gone locally extinct (Anderson *et al.*, 2002a). Alternatively, they may simply reflect regions where sufficient sampling efforts have not yet been made. Rejecting the latter possibility is necessary to make strong inferences regarding the species' true distribution and allow consideration of biotic and historical causal factors.

Conversely, multi-species geographical analyses have been undertaken to estimate the probability of a particular species' absence from a given geographical region – but without including information regarding the suitability of environmental conditions there. Biological surveys and inventories typically employ similar suites of collection techniques for a given target taxonomic or ecological group (such as various kinds of traps for small non-volant mammals; Voss & Emmons, 1996; Wilson *et al.*, 1996). Thus, collection records of the more inclusive target group as a whole (e.g. non-volant mammals; see 'background group' of Ponder *et al.*, 2001) can serve as a measure of sampling effort. Data regarding known localities of the study species (e.g. a particular rodent) can then be combined with those of the target group to determine whether the study species' absence in a given area is probably real or, rather, may well be artefactual (because of inadequate sampling; Heyer *et al.*, 1999; see also spatial methodologies of Ponder *et al.*, 2001). However, this approach depends on the presence of equally suitable environmental conditions for the species in both areas considered, an unstated and untested assumption to date. Predictive modelling of a species' potential distribution provides precisely the geographically explicit regions of suitable environmental conditions where a multi-species sampling-effort approach holds relevancy.

Tests of absence and adequate sampling

Here, I integrate these perspectives, formalizing and implementing a progression of statistical models for determining whether a species' absence from one of the two regions of potential distribution could likely be artefactual (because of inadequate sampling). The models differ in the data requirements and assumptions involved, but all employ simple binomial calculations (Sokal & Rohlf, 1995, pp. 71–74) and depend in part on the occurrence records for the study species. The second and third models also require data for the overall target taxonomic or ecological group. The models represent a progression from simple to more complex; subsequent models integrate additional empirical data and rely on fewer assumptions. Heyer *et al.* (1999) provide

distinct logic and calculations with a similar intent, and Anderson & Jarrín-V. (2002, p. 15) present qualitative antecedents of these particular approaches.

Model 1: area approach

Null hypothesis: The study species occurs in both regions of potential distribution.

Data requirements: Potential distribution of the study species; occurrence records of the study species.

Assumptions: (1) Uniform average environmental suitability for the study species in both regions of potential distribution, (2) uniform number of localities of the target group per unit area in both regions of potential distribution, (3) uniform average sampling intensity at localities of the target group in both regions of potential distribution and (4) similar sampling techniques employed at localities of the target group in both regions of potential distribution.

Model: Under these conditions, occurrence records of the study species should exist in the two regions in proportion to their areas. Then, if A_1 is the area of region 1 (where the species has been documented) and A_2 is the area of region 2 (from which no records exist), the probability P of all k occurrence records of the study species falling in region 1 is simply, $P = [A_1/(A_1 + A_2)]^k$. A small binomial probability ($P \leq 0.05$) indicates a highly non-random pattern and leads to rejection of the null hypothesis. In such a case, we infer that the lack of records of the study species in region 2 is not likely artefactual. When the test yields a larger binomial probability, the null hypothesis cannot be rejected, and we infer that the absence of records of the species from region 2 could be artefactual.

Model 2: locality approach

Null hypothesis: As in model 1.

Data requirements: As for model 1, and occurrence records of the target group.

Assumptions: (1), (3) and (4) above.

Model: Here, the assumption of a uniform number of localities of the target group per unit area in both regions of potential distribution (assumption 2 of model 1) is replaced by empirical data (localities of the target group). Under these conditions, occurrence records of the study species should exist in the two regions in proportion to the number of localities of the target group present in each. Then, if L_1 is the number of occurrence localities of the target group in region 1 and L_2 is the number of occurrence localities of the target group in region 2, the probability P of all k occurrence records of the study species falling in region 1 is, $P = [L_1/(L_1 + L_2)]^k$.

Model 3: weighted locality approach

Null hypothesis: As in models 1 and 2.

Data requirements: As for model 2, and information for calculating a weighting index (e.g. number of specimens of the target group known from each locality).

Assumptions: (1) and (4) above, and the assumption that the weighting index adequately reflects sampling intensity.

Model: Here, the assumption of uniform average sampling intensity at localities of the target group in both regions of potential distribution (assumption 3 of models 1 and 2) is replaced by an empirical index of sampling intensity. Under these conditions, occurrence records of the study species should exist in the two regions in proportion to the number of localities of the target group present in each, with localities weighted by the index of sampling intensity. Then, if WL_1 is the number of occurrence localities of the target group in region 1, each weighted by the index of sampling intensity, and WL_2 is the number of occurrence localities of the target group in region 2, each weighted by the index of sampling intensity, the probability P of all k occurrence records of the study species falling into region 1 is, $P = [WL_1 / (WL_1 + WL_2)]^k$. When the number of specimens collected at each locality (a simple measure suggested by Heyer *et al.*, 1999; see also Voss & Emmons, 1996, pp. 62–63) is used as the index of sampling intensity (as here), let S_1 be the number of specimens of the target group collected in region 1 and S_2 , the number of specimens of the target group collected in region 2. Calculation of P then mathematically reduces to, $P = [S_1 / (S_1 + S_2)]^k$. Many alternative indices of sampling intensity (or inventory completeness) are possible as well (e.g. see Soberón-M. & Llorente-B., 1993; Colwell & Coddington, 1994; Peterson & Slade, 1998).

METHODS

As a pilot implementation of these methods for assessing the strength of evidence for a species' absence from a biogeographical region, I here use a large published data set representing the results of the Smithsonian Venezuelan Project, a comprehensive set of mammalian inventories in Venezuela conducted from 1965 to 1968 (Handley, 1976; *c.* 38,000 total capture records). These seventy-nine unique primary collection localities are spread across the country and were sampled under similar protocols (Handley, 1976; see also Voss & Emmons, 1996, p. 63 for additional personal communication from Handley). This data set provides the information necessary for all three proposed models and is likely to fulfil many of the required assumptions. In addition to the benefits of such an unusually intensive and uniform collection regime, this also represents an excellent study system because of the presence of several disjunct mountain chains in the northern part of the country, each separated by lowland depressions (Huber, 1997; Fig. 1A). Many species show documented distributions restricted to one or a few of the ranges (Handley, 1976; Linares, 1998). Here, I model the potential distribution of one montane species, *Oryzomys albigularis* (Tomes, 1860), and use information from all collection localities to determine the

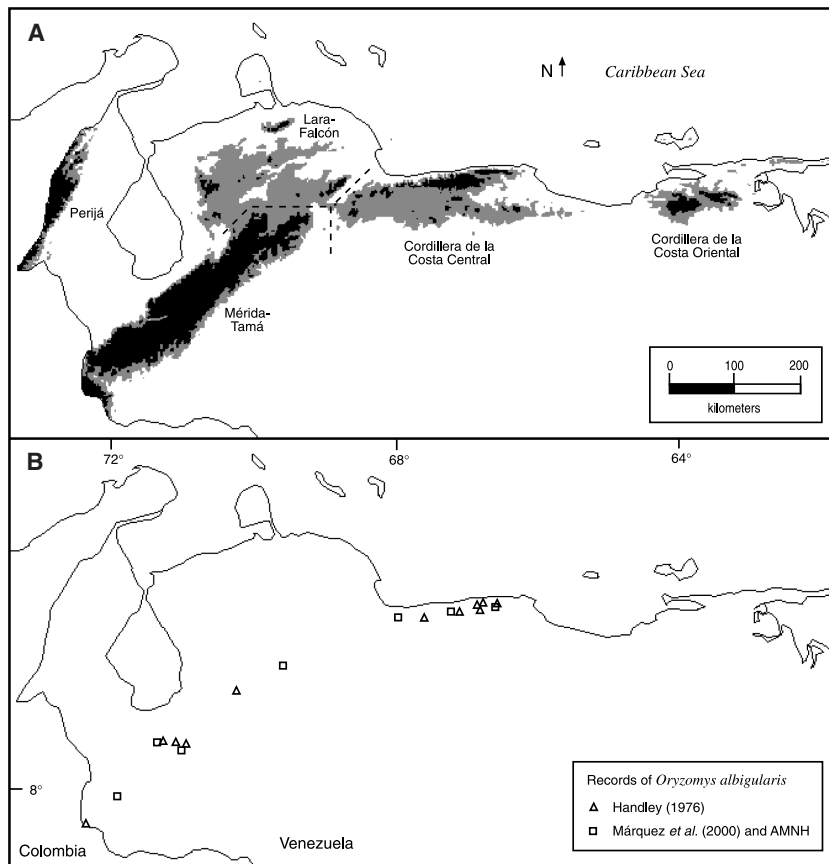


Figure 1 Map showing (A) major highland regions in northern Venezuela, as defined in the text, and (B) collection localities for *Oryzomys albigularis* (Tomes, 1860). In (A), areas above 500 m are shown in grey shading, and regions above 1000 m appear in black. In (B), triangles denote collection localities of *O. albigularis* from the Smithsonian Venezuelan Project (Handley, 1976), and squares mark other records of the species from Márquez *et al.* (2000) and the American Museum of Natural History (see Appendix 1). A highland connection between the Perijá and Mérida–Tamá montane formations exists in Colombia. Highland regions in Venezuela south of the Río Orinoco are not shown.

probability of its absence from areas of potential distribution in which it was not encountered. I then compare these distributional conclusions reached based on data from Handley (1976) with occurrence records from ten museums subsequently compiled by Linares (1998).

Genetic Algorithm for Rule-Set Prediction and distributional modelling

Several approaches to predictive modelling of species geographical distributions in a GIS environment now exist. Such tools have been applied to problems in biogeography, conservation, evolutionary ecology, invasive-species management and public health (e.g. Nicholls, 1989; Walker & Cocks, 1991; Sindel & Michael, 1992; Wilson *et al.*, 1992; Box *et al.*, 1993; Carpenter *et al.*, 1993; Huntley *et al.*, 1995; Austin & Meyers, 1996; Kadmon & Heller, 1998; Yom-Tov & Kadmon, 1998; Corsi *et al.*, 1999; Jarvis & Robertson, 1999; Peterson *et al.*, 1999, 2002c; Samways *et al.*, 1999; Skov, 2000; Zalba *et al.*, 2000; Fleishman *et al.*, 2001; James & McCulloch, 2002). Among the techniques that produce estimates of a species' potential distribution, the Genetic Algorithm for Rule-Set Prediction (GARP) (<http://biodi.sdsc.edu/>; see <http://beta.lifemap-per.org/desktopgarp/> for software download) is a machine-learning approach that has shown particular utility and promise under a wide variety of conditions and is especially suited to use with presence-only occurrence data (Stockwell & Noble, 1992; Peterson & Cohoon, 1999; Peterson *et al.*, 1999, 2001, 2002a,b,c,d; Stockwell & Peters, 1999; Chen & Peterson, 2000; Elith, 2000; Godown & Peterson, 2000; Sánchez-Cordero & Martínez-Meyer, 2000; Peterson, 2001; Peterson & Vieglais, 2001; Anderson *et al.*, 2002a, 2003; Elith & Burgman, 2002; Feria-A. & Peterson, 2002; Joseph & Stockwell, 2002; Stockwell & Peterson, 2002a,b; but see Lim *et al.*, 2002).

To form a model of the species' requirements, GARP searches for non-random associations between environmental characteristics of localities of known occurrence vs. those of the overall study region. Although encompassing only a few of the numerous possible niche dimensions (*sensu* Hutchinson, 1957; see also MacArthur, 1968; Morrison & Hall, 2002; Peterson *et al.*, 2002d; Anderson *et al.*, 2003), the currently available GIS coverages (digital maps) provide environmental data for many major physical variables that commonly influence species macrodistributions (Grinnell, 1917a,b; Root, 1988; Brown & Lomolino, 1998, p. 72). Under clear assumptions and given the varying community contexts present across a species' range, a GARP model indicates areas potentially habitable for it – at least as indicated by the examined environmental variables and occurrence records (Peterson *et al.*, 1999).

The GARP program works in an iterative process of rule selection, evaluation, testing and incorporation or rejection to produce a heterogeneous rule-set characterizing the species' requirements (Stockwell & Noble, 1992). First, a method is chosen from a set of possibilities (e.g. logistic regression and bioclimatic-envelope rules; Busby, 1986;

Sindel & Michael, 1992; Box *et al.*, 1993; Skidmore *et al.*, 1996) and is applied to the data. Then, a rule is developed and predictive accuracy (*sensu* Stockwell & Peters, 1999) is evaluated via map pixels intrinsically resampled from the known distribution and from the study region as a whole. The change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the rule-set. As implemented here, the algorithm runs either 1000 iterations or until the addition of new rules has no appreciable effect on the intrinsic accuracy measure (convergence). As GARP is a superset of several approaches, it has the potential of achieving greater predictive accuracy than single approaches to distributional modelling (Stockwell & Peters, 1999).

The final rule-set, or ecological niche model, is a complex series of if-then statements used to determine whether the species' is predicted present or absent in each map pixel. Map pixels fulfilling the conditions of the model represent the species' potential geographical distribution in the study region. In some circumstances, the rule-set does not provide a presence/absence decision for the environmental combination found in a particular pixel. Such cells should be considered as predicted absence (Anderson *et al.*, 2003). As a result of strong stochastic elements characteristic of genetic algorithms (Holland, 1975), no unique solution is produced. Hence, production of multiple models is necessary to account for variability among independent runs of model building (see below; Anderson *et al.*, 2002a, 2003; Lim *et al.*, 2002).

Model building

Preliminary models

The quality of predictive models of species distributions is commonly assessed by making models with a training data set of occurrence records and then evaluating them using an independent test data set (Fielding & Bell, 1997; Peterson *et al.*, 1999; Guisan & Zimmermann, 2000; Fielding, 2002). A one-tailed χ^2 statistic (or binomial probability, if sample sizes are small) can then be employed to determine whether test sites fall into regions of predicted presence more often than expected by chance, given the proportion of map pixels predicted present by the model (Anderson *et al.*, 2002a). The test data set thus provides an extrinsic (independent) measure of model significance (better than random prediction). In addition to significance, good models also should show low extrinsic omission of localities in the test data set (Anderson *et al.*, 2003, can be equated to 'accuracy' of Anderson *et al.*, 2002a).

Hence, to evaluate the ability of GARP to predict the distribution of *O. albigularis* (Rodentia: Muridae: Sigmodontinae) in Venezuela, I made a set of preliminary models. Karyological and morphometric differences exist among some sets of Venezuelan populations of the group (Aguilera *et al.*, 1995; Rivas & Péfaur, 1999; Márquez *et al.*, 2000), leading some authorities to consider them as distinct species (e.g. Soriano & Ochoa-G., 1997). However, they are clearly closely related, and I here model them as a single entity. Occurrence records from Handley (1976; eleven unique

latitude–longitude combinations; Fig. 1B; Appendix 1) constituted the training data set. I then evaluated the resulting models with additional localities from Márquez *et al.* (2000) and from specimens that I examined in the mammal collection of the American Museum of Natural History (seven additional unique latitude–longitude combinations; other records from the two latter sources that duplicated localities of *O. albigularis* from the Smithsonian Venezuelan Project were not considered; Fig. 1B; Appendix 1). Four geographical coverages comprised the base environmental data. Elevation was derived from USGS (2001). Climatic data came from ArcAtlas (ESRI, 1996) and consisted of long-term annual values for temperature, precipitation and solar radiation. These environmental data correspond temporally to the locality records, which derive from inventories conducted throughout the 1900s (see also Peterson *et al.*, 1999, 2001, 2002b). All environmental coverages were resampled to a coincident grid consisting of pixels *c.* 1.1 km on a side. The grid included all of Venezuela and surrounding areas of neighbouring countries (0.5–13° N, 59.5–73.5° W).

I made 100 preliminary models and selected optimal models by a best-subsets selection procedure. After eliminating any model showing non-zero intrinsic omission error, the ten models closest to the mean intrinsic commission value were identified (minimizing both overfitting and true commission error; following Anderson *et al.*, 2003). I then summed those ten optimal models to create a composite prediction in which the value for each map pixel indicated the number of models that predicted the species' potential presence in that cell. For extrinsic testing with the independent data set, this composite prediction was then cut to include only pixels in Venezuela.

Final models

After demonstrating effective predictive ability in the preliminary models, I made 100 final models using the same environmental coverages and all available locality records of *O. albigularis* (i.e. from Handley, 1976; Márquez *et al.*, 2000; and the American Museum of Natural History). Although they were not used in subsequent tests of distributional hypotheses and adequate sampling (which were based only on data from Handley, 1976; see below), the addition of localities from Márquez *et al.* (2000) and the American Museum of Natural History in the GARP modelling process provided a larger data set more adequate for modelling the species' requirements (Stockwell & Peterson, 2002b). Ten optimal models were then identified and summed as above to create a final composite prediction. Regions outside northern Venezuela (defined as north of the Río Orinoco) were removed from the prediction, and the remaining pixels predicted by all ten models then represented the species' potential distribution in the study region for all further analyses.

Tests of distributional hypotheses and adequate sampling

Five principal montane biogeographical regions have been recognized in northern Venezuela (Huber & Alarcón, 1988;

Huber, 1997; Linares, 1998). In addition, numerous upland areas exist in the country south of the Río Orinoco (Guianan formations), separated from the northern ranges by extensive lowland tropical savannas (*llanos*). As defined here primarily by topographical depressions, the northern mountain chains may be divided into the following regions: Perijá, Mérida–Tamá, Lara–Falcón, Cordillera de Costa Central, and Cordillera de la Costa Oriental (Fig. 1A). The Serranía de Perijá, a northern spur of the Andes, lies along the Colombian border in extreme northwestern Venezuela. The Cordillera de Mérida and Macizo de El Tamá together represent a high diagonal Andean complex southeast of the Lago de Maracaibo (Soriano *et al.*, 1999). Northeast of the Lago de Maracaibo and north of the Depresión de Barquisimeto (*sensu* Huber, 1997), a series of low hills and ranges exists in Lara and Falcón states. Finally, two major coastal ranges are present east of the Depresión de Yaracuy (*sensu* Anderson, 2003; see also SAGCN, 1995). The low, wide Depresión de Unare divides them into the Cordillera de la Costa Central and the Cordillera de la Costa Oriental. The definitions employed here differ from the scheme of Huber (1997) primarily by including the Sierra de Aroa in the Lara–Falcón region (west of the Depresión de Yaracuy), instead of as part of the Cordillera de la Costa Central.

The inventories of the Smithsonian Venezuelan Project reported in Handley (1976) captured *O. albigularis* in the Mérida–Tamá complex and in the Cordillera de la Costa Central. Additional records from Márquez *et al.* (2000) and the American Museum of Natural History are likewise restricted to these two regions (Fig. 1B). The current analyses address the question of whether areas of potential distribution for the species exist in the other montane systems in northern Venezuela (Perijá, Lara–Falcón, and Cordillera de la Costa Oriental), and if so, whether the inventories of Handley (1976) were sufficient to substantiate the species' absence from those regions with reasonable statistical support.

To that end, I assigned each disjunct patch of predicted potential distribution for the species in northern Venezuela to one of the five regions defined above. I then calculated the area (km²) of estimated potential distribution found in each region and overlaid collection localities from Handley (1976) to determine how many fell in pixels of potential distribution in each region. Mammals as a whole represented the target group in these analyses. Only primary collection localities were considered. When a range of geographical coordinates was provided by Handley (1976) for a composite primary locality, I used the coordinates given first. Localities represented by only purchased specimens, animals found dead on a road, or non-mammalian vertebrates were excluded. I then calculated binomial probabilities of the species', absence following the area approach, the locality approach and the weighted locality approach (models 1–3) for each of the three regions lacking occurrence records (Perijá, Lara–Falcón, Cordillera de la Costa Oriental).

Then, following Chen & Peterson (2000) and Lizcano *et al.* (2002), I cut the predictions by removing areas converted to agricultural and urban uses and repeated the tests.

The Mapa de vegetación de Venezuela provides an estimate of vegetation types and land-use as of c. 1980–81 (Huber & Alarcón, 1988, digital version supplied by D. Lew). As *O. albigularis* is generally restricted to forests (Handley, 1976; Linares, 1998) and the overall geographical pattern of areas converted to agricultural and urban areas in Venezuela then was quite similar to that of 12–16 years earlier, this modified prediction functions as a rough estimate of conditions on the ground when the inventories were conducted (1965–68; see Handley, 1976). Indeed, all but one locality of *O. albigularis* reported by Handley (1976) correspond to a landscape considered intact by Huber & Alarcón (1988), and the last falls within 4 km of an intact landscape. Thus, although less than ideal temporal correspondence exists between the land-use data provided by Huber & Alarcón (1988) and the inventories of Handley (1976), the trimmed prediction probably provides a more reasonable approximation of areas suitable for the species at the time of sampling (necessary for all tests) than does the raw prediction.

RESULTS

Preliminary models adequately predicted the test localities. Six of the seven test sites fell into pixels predicted by all ten optimal preliminary models ($P < 0.000001$, one-tailed binomial). The last test site corresponded to a pixel predicted

by eight of the ten models (and falling only 2 km from a pixel predicted by all ten), constituting negligible extrinsic omission error. In addition to indicating suitable conditions in the Mérida–Tamá complex and the Cordillera de la Costa Central, patches of potential presence were present in the Perijá, Lara–Falcón, and Cordillera de la Costa Oriental regions (not shown).

The final models yielded a similar prediction showing potential distribution in all five regions (Fig. 2A; Table 1). Based on the area approach (model 1) and occurrence records of the species from the Smithsonian Venezuelan Project (Handley, 1976), the possible artefactual nature of the absence of *O. albigularis* from the three regions lacking records could not be rejected (Perijá, $P = 0.5096$; Lara–Falcón, $P = 0.1394$; Cordillera de la Costa Oriental, $P = 0.4303$; one-tailed binomials; Table 1). The locality approach (model 2) yielded even less support for the species' absence in those respective areas (Perijá, $P = 1.0$; Lara–Falcón, $P = 0.3138$; Cordillera de la Costa Oriental, $P = 0.5517$; Table 1). Likewise, its absence in those regions could not be demonstrated by the weighted locality approach either (model 3; Perijá, $P = 1.0$; Lara–Falcón, $P = 0.1669$; Cordillera de la Costa Oriental, $P = 0.2249$; Table 1). After removing areas of agricultural and urban use from the predictions, all three models yielded results similar to those above (Fig. 2B; Table 1; area approach: Perijá,

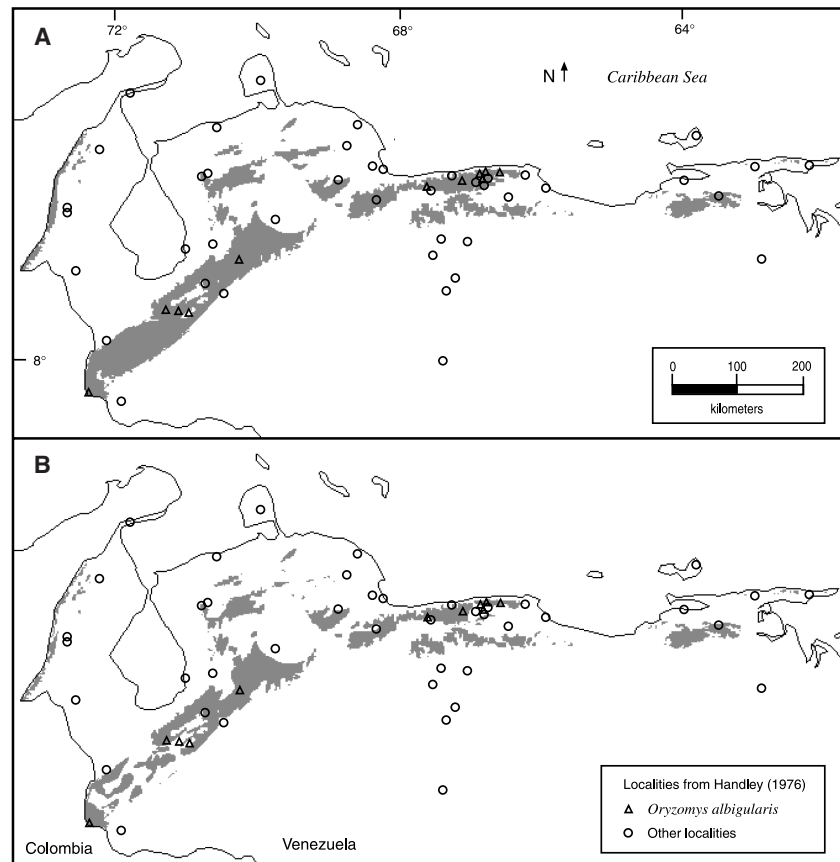


Figure 2 Modelled potential geographical distribution for *Oryzomys albigularis* in northern Venezuela: (A) raw model and (B) trimmed model, after removal of areas converted to agricultural and urban uses (Huber & Alarcón, 1988). Areas predicted by all ten final optimal models are shown in grey (see text). Triangles denote sites where *O. albigularis* was collected, and circles represent other primary collection localities; data from the Smithsonian Venezuelan Project (Handley, 1976).

Region	Localities of <i>O. albigularis</i>	Area (km ²)	Localities of target group	Individuals of target group
Raw estimates				
Perijá	0	1918	0	0
Mérida–Tamá	5	21,177	6	886
Lara–Falcón	0	5952	2	769
Cordillera de la Costa Central	6	9163	12	3465
Cordillera de la Costa Oriental	0	2418	1	632
Trimmed estimates				
Perijá	0	1327	0	0
Mérida–Tamá	4	15,291	5	834
Lara–Falcón	0	4844	2	769
Cordillera de la Costa Central	6	7836	9	3150
Cordillera de la Costa Oriental	0	2418	1	632

Table 2 Number of primary collection localities of the Smithsonian Venezuelan Project (Handley, 1976) per unit area of potential distribution for *Oryzomys albigularis* in the five major montane biogeographical regions in northern Venezuela and proportion of potential distribution converted to agricultural or urban use in each

Region	Localities per 1000 km ²		Proportion agriculture and urban (%)
	Raw estimate	Trimmed estimate	
Perijá	0.00	0.00	30.8
Mérida–Tamá	0.28	0.33	27.8
Lara–Falcón	0.34	0.41	18.6
Cordillera de la Costa Central	1.31	1.15	14.5
Cordillera de la Costa Oriental	0.41	0.41	0.00

$P = 0.5723$; Lara–Falcón, $P = 0.1493$; Cordillera de la Costa Oriental, $P = 0.3700$; locality approach: Perijá, $P = 1.0$; Lara–Falcón, $P = 0.2631$; Cordillera de la Costa Oriental, $P = 0.5016$; weighted locality approach: Perijá, $P = 1.0$; Lara–Falcón, $P = 0.1712$; Cordillera de la Costa Oriental, $P = 0.2294$).

Some assumptions were better fulfilled than others. Clearly, the overall sampling effort (as judged by number of collection localities) varied among regions of potential distribution for the species after correction for area, the highest being in the Cordillera de la Costa Central and the lowest in Perijá (Table 2). However, the collection intensity of sampled localities, as approximated by the number of individuals captured at each locality (see Voss & Emmons, 1996, pp. 62–63; Heyer *et al.*, 1999), did not vary among regions ($P = 0.527$ with raw potential distribution; $P = 0.632$ after trimming agricultural and urban areas from the prediction; one-way ANOVAs). Similarly, the mean number of individuals captured at localities falling in regions where *O. albigularis* was encountered (Mérida–Tamá and Cordillera de la Costa

Table 1 Data necessary for binomial tests (models 1–3) of distributional hypotheses for *Oryzomys albigularis* (Tomes, 1860) in northern Venezuela before and after trimming of areas converted to agricultural and urban use (Huber & Alarcón, 1988). Data regarding collection localities from the Smithsonian Venezuelan Project (Handley, 1976) for *O. albigularis* and for the target group (all mammals) in areas of potential distribution (see Fig. 2) are given for each of five major montane biogeographical regions (see Fig. 1A). Estimates of the areal extent of potential distribution are provided for map pixels predicted by all ten final optimal models (see text)

Central) did not differ from that at localities corresponding to regions in which it was not collected (Perijá, Lara–Falcón and Cordillera de la Costa Oriental; $P = 0.410$ raw prediction, $P = 0.500$ after trimming; two-tailed t -tests). However, the proportion of the species' potential distribution converted to agricultural or urban use did differ substantially among regions (Table 2).

DISCUSSION

Assumptions

Limited error probably exists in the predicted potential distribution (see assumption 1 of all three models). First of all, only moderate numbers of localities were available for building the GARP models (see Stockwell & Peterson, 2002b), and the addition of other environmental predictor variables (e.g. potential vegetation and seasonal climatic data) would probably change the final prediction to an extent. Furthermore, if populations in other parts of the species range (i.e. beyond Venezuela) occupy environmental conditions outside those typified by the present sample, the exclusion of records from those distributional areas may have led to an underestimation of the species' fundamental niche in the examined environmental dimensions (Anderson *et al.*, 2002a). Finally, the level of precision for the geographical coordinates of the collection localities used in making the GARP models, moderately coarse resolution of the climatic variables, and variability present in the modelling system itself require that the final prediction (and thus subsequent analyses based upon it) must be considered approximate (see Anderson *et al.*, 2003).

Despite these caveats and possible sources of error, the preliminary GARP models effectively predicted test localities, and final models yielded a realistic prediction of the species' potential distribution. As population abundance (density) and the relative suitability of environmental conditions for a species can vary geographically within its distribution (Hengeveld & Haeck, 1982; Brown, 1984;

Lomolino & Channell, 1995), future applications of these techniques could consider weighting pixels of potential distribution by the strength of the prediction (such as the number of models predicting presence in each cell). Nevertheless, the single-threshold approach (all ten optimal models predicting potential presence) applied here to the final models made based on all available occurrence records yielded a credible hypothesis of the potential distribution of *O. albigularis* in the study region (Fig. 2A; although perhaps slightly permissive in the Lara–Falcón region). However, vegetational conditions in the country clearly were not pristine at the time of the inventories of Handley (1976), and the assumption of uniform average environmental suitability for the study species in all regions of potential distribution (assumption 1) was thus violated in the first suite of analyses (using the raw predictions) by unequal deforestation and conversion of original vegetation types to agricultural and urban landscapes among regions (Table 2; Huber & Alarcón, 1988). Thus, with regard to assumption 1 (necessary for all approaches), the trimmed potential-distribution model provided a more reasonable prediction than did the raw model.

The remaining assumptions were more closely fulfilled for the locality and weighted locality approaches than for the area approach. Clearly, the assumption of a uniform number of localities of the target group per unit area in all regions of potential distribution for *O. albigularis* (assumption 2; necessary for the area approach, model 1) was not valid here, with the Cordillera de la Costa Central the most heavily sampled (Table 2). In contrast, no heterogeneity in the sampling intensity of collection localities (assumption 3) was detected among regions. Thus, use of both the locality and weighted locality approaches is justified (models 2 and 3). Model 3 requires the additional assumption that the weighting index adequately reflects sampling intensity. Here, the number of specimens collected at each locality formed the index of sampling intensity. This simple measure can be easily calculated for many data sets (Voss & Emmons, 1996, pp. 62–63; Heyer *et al.*, 1999). Various other possible indices also exist (e.g. Soberón-M. & Llorente-B., 1993; Colwell & Coddington, 1994; Peterson & Slade, 1998), but most of them rely on more data and standardized sampling methodologies that would seldom be found with heterogeneous data sets such as those derived from museum records. Although additional research is called for to address the use of appropriate weighting indices (particularly with heterogeneous data sets), the number of specimens collected probably represented a reasonable index of sampling intensity here (see protocols in Handley, 1976). Similarly, because the inventories of Handley (1976) employed a standardized methodology, assumption 4 (similar sampling techniques employed at localities of the target group in both regions of potential distribution; necessary for all models) was satisfied. In addition, various environmental factors – such as moonlight, extreme rainfall or ant activity – at the time of sampling can also affect inventory success (Voss & Emmons, 1996). Thus, implicit in assumption 4 is the supposition that the conditions under which the sampling

techniques were employed did not vary among the examined regions. No such systematic bias is likely here (Handley, 1976). In sum, the tests that considered habitat modification and empirical measures of sampling effort (i.e. models 2 and 3, with agricultural and urban areas removed from the predictions) most closely met the relevant assumptions and hence represented the most realistic tests.

An important corollary of assumptions 3 and 4 concerns the probability of encountering the study species at particular localities. Species are not detected randomly in inventories (Boulinier *et al.*, 1998). Rather, some are typically among the last to be documented. In addition to other possible causal factors, this phenomenon commonly derives from differences in abundance among species, restricted microhabitat requirements of some species, and differential effectiveness of the employed sampling techniques for various subsets of the fauna (Voss & Emmons, 1996). Hence, the documentation of many common, easily collected species of the target group (such as generalist terrestrial forest-dwelling rodents) at a particular locality does not necessarily indicate that a rarer species with specific requirements (such as an arboreal or aquatic mouse) would have been captured there if present. However, such species are generally difficult to collect or observe throughout their distributions (Voss & Emmons, 1996, p. 57). The tests used here only require that the chance of detecting the study species at the sampled sites (if present; see Cam *et al.*, 2002, p. 1119) be uniform in the two regions examined, a condition that will often be reasonable when the target group is chosen with an understanding of the sampling techniques employed in the data set at hand (see below).

Distribution of *O. albigularis* and compilations of museum records

Although the inventories of the Smithsonian Venezuelan Project (Handley, 1976) only collected *O. albigularis* in two of the five montane regions of northern Venezuela, none of the three statistical approaches (area, locality and weighted locality; models 1–3) led to rejection of the possible artefactual nature of the species' absence from areas of potential distribution in the other three regions (Perijá, Lara–Falcón and Cordillera de la Costa Oriental). Indeed, a recent compilation of specimen records from ten major museums showed that the species does inhabit at least some localities in the Perijá and Lara–Falcón regions (Linares, 1998). Thus, the Cordillera de la Costa Oriental represents the only major montane region in northern Venezuela lacking records of *O. albigularis*.

The apparent absence of *O. albigularis* from the Cordillera de la Costa Oriental serves to illustrate several problems that may be encountered using heterogeneous data sets compiled from museum records. As a data set comprising all records of Venezuelan mammals in those ten museums is not currently available, the locality and weighted locality approaches (models 2 and 3) cannot be employed using the occurrence records of *O. albigularis* in Venezuela plotted by Linares (1998). However, under the

area approach (model 1), the possible artefactual nature of the lack of records for the species in the Cordillera de la Costa Oriental cannot be rejected given even those thirty-two occurrence records in the other four regions (Linares, 1998; $P = 0.1404$ with raw prediction). Unfortunately, the same additional data regarding the target group that would be required to implement models 2 and 3 are also necessary to evaluate the validity of two of the assumptions of the area approach (uniform number of localities of the target group per unit area in both regions of potential distribution and uniform average sampling intensity at localities of the target group in both regions of potential distribution; assumptions 2 and 3). Fortunately, such information *will* commonly be available (e.g. from databases of all museum holdings of the target group), allowing use of models 2 and 3 and testing of the validity of those two assumptions. However, evaluation of the assumption that similar sampling techniques were employed at localities of the target group in both regions of potential distribution (assumption 4; required for all three models) is problematic with data sets comprised of museum records representing the sum total of myriad non-standardized inventories (see Voss & Emmons, 1996; Wilson *et al.*, 1996 for sampling techniques). Careful selection of the appropriate target group (see below) is critical in order to minimize strong departure from that assumption. Similarly, it is extremely difficult to integrate the effects of human-induced habit disturbance into tests that use such data, which were typically collected over many decades (during which land-use patterns have often changed markedly; see assumption 1 of all three models). More information is necessary to reach a solid conclusion regarding the species' status in the Cordillera de la Costa Oriental.

Another striking pattern regarding the distribution of *O. albigularis* in northern Venezuela merits attention. The Cordillera de la Costa Central is comprised of two principal subranges: the Serranía del Litoral along the coast and the Serranía del Interior parallel to it and *c.* 50 km inland (Fig. 1A). The final models indicated disjunct areas of potential distribution for the species in these two ranges, but *O. albigularis* has only been collected in the Serranía del Litoral (Handley, 1976; Linares, 1998; Márquez *et al.*, 2000; American Museum of Natural History collection; Fig. 2). Under the area approach (model 1), the ten localities plotted in the Serranía del Litoral by Linares (1998) would indicate that the species' absence from the Serranía del Interior is not likely artefactual ($P = 0.016$). Again, however, none of the four assumptions required for this test can be assessed given the data at hand (see above), and some of them may not be reasonable.

Comparisons with previous approaches

The statistical models proposed here afford several advantages over previous approaches that held similar intents. First, they incorporate geographically explicit predictions of potential distribution into the test (lacking in Heyer *et al.*, 1999 and not integrated by Ponder *et al.*,

2001). In addition, the present calculations represent a much simpler methodology than that proposed by Ponder *et al.* (2001), which was based on patterns of the spatial distribution of localities of the study species (their 'taxon of interest') vs. those of the target group (their 'background group'). More importantly, however, the analyses of Ponder *et al.* (2001) were descriptive in nature and did not test statistical hypotheses. Similarly, although Anderson & Jarrín-V. (2002, p. 15) presented an incipient logical perspective upon which the current models were developed and formalized, they did not conduct any quantitative analyses.

Thus, the models proposed here are most similar to the test suggested by Heyer *et al.* (1999), which could be easily modified for implementation in the context of a potential-distribution model. Nevertheless, that test differs greatly in calculations and logic and lacks many of the advantages of the techniques used here. The current approach employs occurrence records of the study species as the k cases of the binomial test, with data (or assumptions, in model 1) regarding localities of the target group used to calculate the expected probability of the outcome of each individual case. [In each of the three models, the quantity that is raised to the power of k (in brackets in the presentation of the models above) represents the expected chance of each particular case falling into the region where the species has been documented; raising to the power of k then yields the probability P of all k cases falling into that region.]

In contrast, the tactic of Heyer *et al.* (1999) uses some localities of the target group as cases for the binomial test and combines others with localities of the study species to establish the expected probability of the outcome of each case. More precisely, localities of the target group in areas lacking records of the study species comprise the cases of their binomial test. To determine the expected probability of a negative result for each of those cases, Heyer *et al.* (1999) calculated a probability of non-detection. By dividing the number of localities of the study species by the number of localities of the target group present within the known distribution of the study species, they estimated an average detection probability for the study species. One minus that detection probability then represents the probability of *not* detecting it at a given locality. Under this logic, raising the probability of non-detection at a single site to the power of the number of localities of the target group in the region lacking records of the study species (the cases of the binomial test) then yields the probability of *none* of the inventories in the second region detecting the study species. This procedure most closely intersects in data and intent with the present locality approach (model 2). However, it is not as straightforward, intuitive or flexible as the tests used and advocated here. Most importantly, it does not lend itself to a progression of models, each employing increasing empirical data as available. In addition to the highly desirable flexibility allowed by a series of models, the current framework provides explicit assumptions, null hypotheses and appropriate inferences given the results of the statistical tests.

Selection of target group and application to other systems

Selection of the appropriate target group should be aimed at identifying the taxonomic or ecological group most likely to fulfil the relevant assumptions. The calculations of models 2 and 3 require explicit identification of the target group, and the existence of a target group is implicit in three of the assumptions necessary for model 1. Here – using data from standardized inventories that employed similar sampling techniques at each locality – all mammals were used as the target group. However, for museum collections as a whole, the source inventories vary widely in the emphasis placed on sampling for various groups of mammals. Thus, because they all are captured with similar techniques, small non-volant mammals (e.g. small rodents, marsupials and shrews) would probably represent a more appropriate target group for *O. albigularis* when using a heterogeneous data set comprised of museum records that derive from inventories that varied greatly in techniques and intent. Although such a target group is not evolutionarily monophyletic, the documented presence of such species at a locality likely indicates the use of inventory techniques with the potential to capture *O. albigularis* (if present there; see also Ponder *et al.*, 2001, p. 650). The models proposed here do not require *all* collection localities to be sampled equivalently, but rather that no systematic biases in sampling (in numbers of localities sampled per unit area or in the intensity and makeup of sampling techniques employed) exist between the regions of potential distribution being compared. Careful selection of the target group given knowledge of relevant sampling methodologies can minimize that possibility (e.g. for Neotropical mammals, see Voss & Emmons, 1996; Wilson *et al.*, 1996; Simmons & Voss, 1998; Voss *et al.*, 2001).

Used in concert with models of species potential distributions (produced with GARP or other algorithms) and occurrence records from museum and herbarium collections or other data sets (Baker *et al.*, 1998; Boulinier *et al.*, 1998; Funk *et al.*, 1999; Soberón, 1999; Ponder *et al.*, 2001), these analytical approaches will likely prove of wide applicability to problems in biogeography, conservation and other areas of research that employ species distributions. Clearly, this framework provides a potential solution for revisionary and biogeographical studies previously hindered in their distributional conclusions by the lack of a quantitative method for assessing collection effort in various regions apparently suitable for a species but lacking records of it (e.g. Anderson & Jarrín-V., 2002, p. 15; Anderson *et al.*, 2002a, p. 138; Voss *et al.*, 2002, pp. 34–35). With either single- or multi-species perspectives, the present techniques can also be used to identify and justify necessary new inventories – crucial because time and resources for such endeavours are conspicuously limited (Raven & Wilson, 1992; Voss & Emmons, 1996). Conversely, if analyses indicate the absence of a given species of conservation concern from a particular biogeographical region, such areas should be removed from its range map for conservation assessment and planning (cf. Mace & Stuart, 1994). Additionally, the use of predictive

models to elucidate contiguous areas of potential distribution will allow the testing of distributional hypotheses formed based on barriers currently relevant for the species itself, rather than on regions previously defined on topographical, phytogeographical, or zoogeographical means (as here; Eisenberg & Redford, 1979; Linares, 1998). Furthermore, implementation of this integrated methodology in a multi-species context should improve techniques for recognizing areas of endemism, which are important to many evolutionary studies and applications to conservation biology (Cracraft, 1985; Haffer, 1985; Voss *et al.*, 2001; Carleton *et al.*, 2002, p. 518). Foremost, however, I hope that the present work will stimulate research by other investigators scrutinizing the logic and mechanics of the approaches proposed here, leading to modification and refinement as found appropriate and bringing renewed attention and increased quantitative rigor to studies of species distributions.

ACKNOWLEDGMENTS

A.T. Peterson and R. Scachetti-Pereira set this project in motion through formative interactions and by providing an efficient desktop implementation of GARP. D. Lew, E. Martínez-Meyer, and M. Ortega-Huerta supplied additional geographical data. I thank R.S. Voss for numerous insightful discussions and for access to specimens in the AMNH collection. C.J. Raxworthy, N.A. Slade, and two anonymous reviewers offered additional helpful comments. This work was supported by a Roosevelt Postdoctoral Research Fellowship from the American Museum of Natural History.

REFERENCES

- Aguilera, M., Pérez-Zapata, A. & Martino, A. (1995) Cytogenetics and karyosystematics of *Oryzomys albigularis* (Rodentia, Cricetidae) from Venezuela. *Cytogenetics and Cell Genetics*, **69**, 44–49.
- Anderson, R.P. (2003) Taxonomy, distribution, and natural history of the genus *Heteromys* (Rodentia: Heteromyidae) in western Venezuela, with the description of a dwarf species from the Península de Paraguaná. *American Museum Novitates*, **3396**, 1–43.
- Anderson, R.P. & Jarrín-V., P. (2002) A new species of spiny pocket mouse (Heteromyidae: *Heteromys*) endemic to western Ecuador. *American Museum Novitates*, **3382**, 1–26.
- Anderson, R.P., Gómez-Laverde, M. & Peterson, A.T. (2002a) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography*, **11**, 131–141.
- Anderson, R.P., Peterson, A.T. & Gómez-Laverde, M. (2002b) Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, **98**, 3–16.
- Anderson, R.P., Lew, D. & Peterson, A.T. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, **162**, 211–232.
- Austin, M.P. & Meyers, J.A. (1996) Current approaches to modelling the environmental niche of eucalypts: implication

- for management of forest biodiversity. *Forest Ecology and Management*, **85**, 95–106.
- Baker, R.J., Phillips, C.J., Bradley, R.D., Burns, J.M., Cooke, D., Edson, G.F., Haragan, D.R., Jones, C., Monk, R.R., Montford, J.T., Schmidly, D.J. & Parker, N.C. (1998) Bioinformatics, museums, and society: integrating biological data for knowledge-based decisions. *Occasional Papers of the Museum, Texas Tech University*, **187**, 1–4.
- Boulinier, T., Nichols, J.D., Sauer, J.R., Hines, J.E. & Pollock, K.H. (1998) Estimating species richness: the importance of heterogeneity in species detectability. *Ecology*, **79**, 1018–1028.
- Box, E.O., Crumpacker, D.W. & Hardin, E.D. (1993) A climatic model for location of plant species in Florida, U.S.A. *Journal of Biogeography*, **20**, 629–644.
- Brooks, D.R. & McLennan, D.A. (1991) *Phylogeny, ecology, and behavior: a research program in comparative biology*. University of Chicago Press, Chicago, IL.
- Brown, J.H. (1984) On the relationship between abundance and distribution of species. *American Naturalist*, **124**, 255–279.
- Brown, J.H. & Lomolino, M.V. (1998) *Biogeography*, 2nd edn. Sinauer Associates, Sunderland, MA.
- Busby, J.R. (1986) A biogeoclimatic analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. *Australian Journal of Ecology*, **11**, 1–7.
- Cam, E., Nichols, J.D., Hines, J.E., Sauer, J.R., Alpizar-Jara, R. & Flather, C.H. (2002) Disentangling sampling and ecological explanations underlying species-area relationships. *Ecology*, **83**, 1118–1130.
- Carleton, M.D., Sánchez, O. & Urbano-Vidales, G. (2002) A new species of *Habromys* (Muroidea: Neotominae) from México, with a generic review of species definitions and remarks on diversity patterns among Mesoamerican small mammals restricted to humid montane forests. *Proceedings of the Biological Society of Washington*, **115**, 488–533.
- Carpenter, G., Gillison, A.N. & Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, **2**, 667–680.
- Chen, G.-J. & Peterson, A.T. (2000) A new technique for predicting distribution of terrestrial vertebrates using inferential modeling. *Zoological Research*, **21**, 231–237.
- Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Linnean Society of London*, **B 345**, 101–118.
- Corsi, F., Duprè, E. & Boitani, L. (1999) A large-scale model of wolf distribution in Italy for conservation planning. *Conservation Biology*, **13**, 150–159.
- Cracraft, J. (1985) Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. *Neotropical ornithology* (eds P.A. Buckley, M.S. Foster, E.S. Morton, R.S. Ridgely and F.G. Buckley), pp. 49–84. *Ornithological Monographs*, **36**.
- Eisenberg, J.F. & Redford, K. (1979) A biogeographic analysis of the mammalian fauna of Venezuela. *Vertebrate ecology in the northern Neotropics* (ed. J.F. Eisenberg), pp. 31–36. Smithsonian Institution Press, Washington, DC.
- Elith, J. (2000) Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. *Quantitative methods for conservation biology* (eds S. Ferson and M. Burgman), pp. 39–58. Springer, New York.
- Elith, J. & Burgman, M. (2002) Predictions and their validation: rare plants in the central highlands, Victoria, Australia. *Predicting species occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 303–313. Island Press, Washington, DC.
- ESRI (1996) *ArcAtlas: our earth*. Environmental Systems Research Institute, Inc., Redlands, CA.
- Fagan, W.F. & Kareiva, P.M. (1997) Using compiled species lists to make biodiversity comparisons among regions: a test case using Oregon butterflies. *Biological Conservation*, **80**, 249–259.
- Feria-A., T.P. & Peterson, A.T. (2002) Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments. *Diversity and Distributions*, **8**, 49–56.
- Fielding, A.H. (2002) What are the appropriate characteristics of an accuracy measure? *Predicting species occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 271–280. Island Press, Washington, DC.
- 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.
- Fleishman, E., MacNally, R., Fay, J.P. & Murphy, D.D. (2001) Modeling and predicting species occurrences using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology*, **15**, 1674–1685.
- Funk, V.A., Zermoglio, M.F. & Nasir, N. (1999) Testing the use of specimen collection data and GIS in biodiversity exploration and conservation decision making in Guyana. *Biodiversity and Conservation*, **8**, 727–751.
- Godown, M.E. & Peterson, A.T. (2000) Preliminary distributional analysis of US endangered bird species. *Biodiversity and Conservation*, **9**, 1313–1322.
- Grinnell, J. (1917a) Field tests of theories concerning distributional control. *American Naturalist*, **51**, 115–128.
- Grinnell, J. (1917b) The niche-relationships of the California Thrasher. *Auk*, **34**, 427–433.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Haffer, J. (1985) Avian zoogeography of the Neotropical lowlands. *Neotropical ornithology* (eds P.A. Buckley, M.S. Foster, E.S. Morton, R.S. Ridgely and F.G. Buckley), pp. 113–146. *Ornithological Monographs*, **36**.
- Handley, C.O. Jr (1976) Mammals of the Smithsonian Venezuelan Project. *Brigham Young University Science Bulletin, Biological Series*, **20**, 1–91.
- Hengeveld, R. & Haecck, J. (1982) The distribution of abundance. I. Measurements. *Journal of Biogeography*, **9**, 303–316.
- Heyer, W.R., Coddington, J., Kress, W.J., Acevedo, P., Cole, D., Erwin, T.L., Meggers, B.J., Pogue, M.G., Thorington, R.W., Vari, R.P., Weitzman, M.J. & Weitzman, S.H. (1999) Amazonian biotic diversity and conservation decisions. *Ciência e Cultura*, **51**, 372–385.

- Holland, J.H. (1975) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press, Ann Arbor, MI.
- Huber, O. (1997) Ambientes fisiográficos y vegetales de Venezuela. *Vertebrados actuales y fósiles de Venezuela* (ed. E. La Marca), pp. 279–298. Museo de Ciencia y Tecnología de Mérida, Mérida, Venezuela.
- Huber, O. & Alarcón, C. (1988) *Mapa de vegetación de Venezuela, escala 1:2.000.000*. Dirección General de Información e Investigación del Ambiente, Ministerio del Ambiente y de los Recursos Naturales Renovables, Caracas.
- Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography*, **22**, 967–1001.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.
- James, F.C. & McCulloch, C.E. (2002) Predicting species presence and abundance. *Predicting species occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 461–465. Island Press, Washington, DC.
- Jarvis, A.M. & Robertson, A. (1999) Predicting population sizes and priority conservation areas for 10 endemic Namibian bird species. *Biological Conservation*, **88**, 121–131.
- Joseph, L. & Stockwell, D. (2002) Climatic modeling of the distribution of some *Pyrrhura* parakeets of northwestern South America with notes on their systematics and special reference to *Pyrrhura caeruleiceps* Todd, 1947. *Ornitología Neotropical*, **13**, 1–8.
- Kadmon, R. & Heller, J. (1998) Modelling faunal responses to climatic gradients with GIS: land snails as a case study. *Journal of Biogeography*, **25**, 527–539.
- Kress, W.J., Heyer, W.R., Acevedo, P., Coddington, J., Cole, D., Erwin, T.L., Meggers, B.J., Pogue, M., Thorington, R.W., Vari, R.P., Weitzman, M.J. & Weitzman, S.H. (1998) Amazonian biodiversity: assessing conservation priorities with taxonomic data. *Biodiversity and Conservation*, **7**, 1557–1578.
- Lim, B.K., Peterson, A.T. & Engstrom, M.D. (2002) Robustness of ecological niche modeling algorithms for mammals in Guyana. *Biodiversity and Conservation*, **11**, 1237–1246.
- Linares, O.J. (1998) *Mamíferos de Venezuela*. Sociedad Conservacionista Audubon de Venezuela, Caracas.
- Lizcano, D.J., Pizarro, V., Cavelier, J. & Carmona, J. (2002) Geographic distribution and population size of the mountain tapir (*Tapirus pinchaque*) in Colombia. *Journal of Biogeography*, **29**, 7–15.
- Lomolino, M.V. & Channell, R. (1995) Splendid isolation: patterns of geographic range collapse in endangered mammals. *Journal of Mammalogy*, **76**, 335–347.
- MacArthur, R. (1968) The theory of the niche. *Population biology and evolution* (ed. R.C. Lewontin), pp. 159–176. Syracuse University Press, Syracuse, New York.
- MacArthur, R.H. (1972) *Geographical ecology: patterns in the distribution of species*. Harper and Row, New York.
- MacKenzie, D.L., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.
- Mace, G. & Stuart, S. (1994) Draft IUCN Red List categories, version 2.2. *Species*, **21–22**, 13–24.
- Márquez, E.J., Aguilera-M., M. & Corti, M. (2000) Morphometric and chromosomal variation in populations of *Oryzomys albigularis* (Muridae: Sigmodontinae) from Venezuela: multivariate aspects. *Zeitschrift für Säugetierkunde*, **65**, 84–99.
- Morrison, M.L. & Hall, L.S. (2002) Standard terminology: toward a common language to advance ecological understanding and application. *Predicting species occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 43–52. Island Press, Washington, DC.
- Nelson, G. & Platnick, N. (1981) *Systematics and biogeography: cladistics and vicariance*. Columbia University Press, New York.
- Nicholls, A.O. (1989) How to make biological surveys go further with generalized linear models. *Biological Conservation*, **50**, 51–75.
- Nichols, J.D., Boulinier, T., Hines, J.E., Pollock, K.H. & Sauer, J.R. (1998) Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conservation Biology*, **12**, 1390–1398.
- Paynter, R.A. Jr (1982) *Ornithological gazetteer of Venezuela*. Museum of Comparative Zoology, Harvard University, Cambridge, MA.
- Peterson, A.T. (2001) Predicting species' geographic distributions based on ecological niche modeling. *Condor*, **103**, 599–605.
- Peterson, A.T. & Cohoon, K.P. (1999) Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling*, **117**, 159–164.
- 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. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *BioScience*, **51**, 363–371.
- Peterson, A.T., Navarro-Sigüenza, A.G., Benítez-Díaz, H. (1998) The need for continued scientific collecting; a geographic analysis of Mexican bird specimens. *Ibis*, **140**, 288–294.
- Peterson, A.T., Soberón, J. & Sánchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265–1267.
- Peterson, A.T., Sánchez-Cordero, V., Soberón, J., Bartley, J., Buddemeier, R.W. & Navarro-Sigüenza, A.G. (2001) Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecological Modelling*, **144**, 21–30.
- Peterson, A.T., Ball, L.G. & Cohoon, K.P. (2002a) Predicting distributions of Mexican birds using ecological niche modeling methods. *Ibis*, **144**, E27–E32.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddenmier, R.H. & Stockwell, D.R.B. (2002b) Future projections for Mexican faunas under global climate change scenarios. *Nature*, **416**, 626–629.

- Peterson, A.T., Sánchez-Cordero, V., Beard, C.B. & Ramsey, J.M. (2002c) Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerging Infectious Diseases*, **8**, 662–667.
- Peterson, A.T., Stockwell, D.R.B. & Kluza, D.A. (2002d) Distributional prediction based on ecological niche modeling of primary occurrence data. *Predicting species occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 617–623. Island Press, Washington, DC.
- Ponder, W.F., Carter, G.A., Flemons, P. & Chapman, R.R. (2001) Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology*, **15**, 648–657.
- Prendergast, J.R., Wood, S.N., Lawton, J.H. & Eversham, B.C. (1993) Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiversity Letters*, **1**, 39–53.
- Raven, P.H. & Wilson, E.O. (1992) A fifty-year plan for biodiversity surveys. *Science*, **258**, 1099–1100.
- Ricklefs, R.E. & Schluter, D. (eds) (1993) *Species diversity in ecological communities: historical and geographical perspectives*. University of Chicago Press, Chicago, IL.
- Rivas, B.A. & Péfaur, J.E. (1999) Variación geográfica en poblaciones venezolanas de *Oryzomys albigularis* (Rodentia: Muridae). *Mastozoología Neotropical*, **6**, 47–59.
- Rodrigues-Coura, J., Junqueira, A.C.V., Fernandes, O., Valente, S.A.S. & Miles, M.A. (2002) Emerging Chagas disease in Amazonian Brazil. *Trends in Parasitology*, **18**, 171–176.
- Ron, S.R. (2000) Biogeographic area relationships of lowland Neotropical rainforest based on raw distributions of vertebrate groups. *Biological Journal of the Linnean Society*, **71**, 379–402.
- Root, T. (1988) Environmental factors associated with avian distributional boundaries. *Journal of Biogeography*, **15**, 489–505.
- SAGCN (1995) *Mapa físico de la República de Venezuela, escala 1:2.000.000*. Servicio Autónomo de Geografía y Cartografía Nacional, Ministerio del Ambiente y de los Recursos Naturales Renovables, Caracas.
- Samways, M.J., Osborn, R., Hastings, H. & Hattingh, V. (1999) Global climate change and accuracy of prediction of species' geographical ranges: establishment success of introduced ladybirds (Coccinellidae, *Chilocorus* spp.) worldwide. *Journal of Biogeography*, **26**, 795–812.
- Sánchez-Cordero, V. & Martínez-Meyer, E. (2000) Museum specimen data predict crop damage by tropical rodents. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 7074–7077.
- Simmons, N.B. & Voss, R.S. (1998) The mammals of Paracou, French Guiana: a Neotropical lowland rainforest fauna, part 1: bats. *Bulletin of the American Museum of Natural History*, **237**, 1–219.
- Sindel, B.M. & Michael, P.W. (1992) Spread and potential distribution of *Senecio madagascariensis* Poir. (fireweed) in Australia. *Australian Journal of Ecology*, **17**, 21–26.
- Skidmore, A.K., Gauld, A. & Walker, P. (1996) Classification of kangaroo habitat distribution using three GIS models. *International Journal of Geographic Information Systems*, **10**, 441–454.
- Skov, F. (2000) Potential plant distribution mapping based on climatic similarity. *Taxon*, **49**, 503–515.
- Skov, F. & Borchsenius, F. (1997) Predicting plant species distribution patterns using simple climatic parameters: a case study of Ecuadorian palms. *Ecography*, **20**, 347–355.
- Soberón, J. (1999) Linking biodiversity information sources. *Trends in Ecology and Evolution*, **14**, 291.
- Soberón-M., J. & Llorente-B., J. (1993) The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, **7**, 480–488.
- Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: the principles and practice of statistics in biological research*, 3rd edn. Freeman, New York.
- Soriano, P.J. & Ochoa-G., J. (1997) Lista actualizada de los mamíferos de Venezuela. *Vertebrados actuales y fósiles de Venezuela* (ed. E. La Marca), pp. 205–227. Museo de Ciencia y Tecnología de Mérida, Mérida, Venezuela.
- Soriano, P.J., Díaz de Pascual, A., Ochoa-G., J. & Aguilera, M. (1999) Biogeographic analysis of the mammal communities in the Venezuelan Andes. *Interciencia*, **24**, 17–25.
- Stockwell, D.R.B. & Noble, I.R. (1992) Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation*, **33**, 385–390.
- Stockwell, D. & Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Science*, **13**, 143–158.
- Stockwell, D.R.B. & Peterson, A.T. (2002a) Controlling bias in biodiversity data. *Predicting species occurrences: issues of accuracy and scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 537–546. Island Press, Washington, DC.
- Stockwell, D.R.B. & Peterson, A.T. (2002b) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1–13.
- Tomes, R.F. (1860) Notes on a third collection of Mammalia made by Mr. Fraser in the Republic of Ecuador. *Proceedings of the Zoological Society of London*, **28**, 260–268.
- Udvardy, M.D.F. (1969) *Dynamic zoogeography: with special reference to land animals*. Van Nostrand Reinhold, New York.
- USGS (2001) *HYDRO 1k, elevation derivative database*. United States Geological Survey, Sioux Falls, South Dakota (<http://edcdaac.usgs.gov/topo30/hydro/>).
- Voss, R.S. & Emmons, L.H. (1996) Mammalian diversity in Neotropical lowland rainforests: a preliminary assessment. *Bulletin of the American Museum of Natural History*, **230**, 1–115.
- Voss, R.S., Lunde, D.P. & Simmons, N.B. (2001) The mammals of Paracou, French Guiana: a Neotropical lowland rainforest fauna, part 2. nonvolant species. *Bulletin of the American Museum of Natural History*, **263**, 1–236.
- Voss, R.S., Gómez-Laverde, M. & Pacheco, V. (2002) A new genus for *Aepeomys fuscatus* Allen, 1912, and *Oryzomys intectus* Thomas, 1921: enigmatic murid rodents from Andean cloud forests. *American Museum Novitates*, **3373**, 1–42.
- Walker, P.A. & Cocks, K.D. (1991) HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or

- animal species. *Global Ecology and Biogeography Letters*, **1**, 108–118.
- Wilson, D.E., Cole, F.R., Nichols, J.D., Rudran, R. & Foster, M.S. (1996) *Measuring and monitoring biological diversity: standard methods for mammals*. Smithsonian Institution Press, Washington, DC.
- Wilson, J.B., Rapson, G.L., Sykes, M.T., Watkins, A.J. & Williams, P.A. (1992) Distributions and climatic correlations of some exotic species along roadsides in South Island, New Zealand. *Journal of Biogeography*, **19**, 183–194.
- Yom-Tov, Y. & Kadmon, R. (1998) Analysis of the distribution of insectivorous bats in Israel. *Diversity and Distributions*, **4**, 63–70.
- Zalba, S.M., Sonaglioni, M.I., Compagnoni, C.A. & Belenguer, C.J. (2000) Using a habitat model to assess the risk of invasion by an exotic plant. *Biological Conservation*, **93**, 203–208.

BIOSKETCH

Robert P. Anderson is a mammalogist who specializes on the systematics and biogeography of Neotropical mammals. He conducted PhD research at the University of Kansas on South American spiny pocket mice (Rodentia: Heteromyidae: *Heteromys*) and is now a Roosevelt Postdoctoral Research Fellow at the American Museum of Natural History. His current investigation focuses on the development and application of tools for modeling species distributions using GIS. Other research interests include morphometric allometry and the evolutionary ecology of insular populations.

APPENDIX I

Collection localities of *O. albigularis* (Tomes, 1860) in Venezuela used in this study (see Figs 1 & 2). Principal collection localities from the Smithsonian Venezuelan

Project derive from Handley (1976) and were complemented by additional localities from Márquez *et al.* (2000) and the American Museum of Natural History (catalogue numbers provided below). Geographical coordinates of latitude and longitude are given in parentheses; those for localities corresponding to specimens in the American Museum of Natural History were taken from Paynter (1982).

Handley (1976): VENEZUELA: Aragua: Estación Biológica Rancho Grande, 13 km NW Maracay (10°21' N, 67°40' W); Distrito Federal: Los Venados, 4 km NW Caracas (10°32' N, 66°54' W); Distrito Federal/Miranda: Alto Ñeo León, 31–36 km WSW Caracas (10°26' N, 67°10' W); Pico Ávila, near Hotel Humboldt, 5 km NNE Caracas (10°33' N, 66°52' W); Mérida: La Carbonera, 12 km SE La Azulita (8°38' N, 71°21' W); Santa Rosa, 2 km N Mérida (8°37' N, 71°09' W); 6–7 km ESE and SE Tabay (8°36' N, 71°01' W); Miranda: 3 km NE Caracas (10°30' N, 66°53' W); Curupao, 5 km NNW Guarenas (10°31' N, 66°38' W); Táchira: Buena Vista, near Páramo de Tamá, 41 km SW San Cristóbal (7°27' N, 72°26' W); Trujillo: Hacienda Misísí, 14–15 km E Trujillo (9°21' N, 70°18' W).

Márquez *et al.* (2000): VENEZUELA: Lara: Yacambú National Park (9°40' N, 69°37' W); Miranda/Aragua: Macarao National Park (10°25' N, 67°17' W); Táchira: La Trampita, Siberia, sector Uribante (7°50' N, 71°57' W).

American Museum of Natural History: VENEZUELA: Carabobo: La Cumbre de Valencia (AMNH 31541–31543; 10°20' N, 68°00' W); Mérida: El Tambor (AMNH 96160–96163; 8°36' N, 71°24' W); Monte de los Nevados (AMNH 33163; 8°28' N, 71°04' W); Miranda: Curupao (AMNH 135170; 10°30' N, 66°38' W).