Evaluating predictive models of species' distributions: criteria for selecting optimal models

Robert P. Anderson [a,b,*], Daniel Lew [c], A. Townsend Peterson [b]

[a] *Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA*

[b] *Natural History Museum & Biodiversity Research Center and Department of Ecology & Evolutionary Biology, University of Kansas, 1345 Jayhawk Boulevard, Lawrence, KS 66045-7561, USA*

[c] *Museo de Historia Natural La Salle, Fundación La Salle, Apartado 1930, Caracas 1010-A, Venezuela*

Abstract

The Genetic Algorithm for Rule-Set Prediction (GARP) is one of several current approaches to modeling species' distributions using occurrence records and environmental data. Because of stochastic elements in the algorithm and underdetermination of the system (multiple solutions with the same value for the optimization criterion), no unique solution is produced. Furthermore, current implementations of GARP utilize only presence data—rather than both presence and absence, the more general case. Hence, variability among GARP models, which is typical of genetic algorithms, and complications in interpreting results based on asymmetrical (presence-only) input data make model selection critical. Generally, some locality records are randomly selected to build a distributional model, with others set aside to evaluate it. Here, we use intrinsic and extrinsic measures of model performance to determine whether optimal models can be identified based on objective intrinsic criteria, without resorting to an independent test data set. We modeled potential distributions of two rodents (*Heteromys anomalus* and *Microryzomys minutus*) and one passerine bird (*Carpodacus mexicanus*), creating 20 models for each species. For each model, we calculated intrinsic and extrinsic measures of omission and commission error, as well as composite indices of overall error. Although intrinsic and extrinsic composite measures of overall model performance were sometimes loosely related to each other, none was consistently associated with expert-judged model quality. In contrast, intrinsic and extrinsic measures were highly correlated for both omission and commission in the two widespread species (*H. anomalus* and *C. mexicanus*). Furthermore, a clear inverse relationship existed between omission and commission there, and the best models were consistently found at low levels of omission and moderate-to-high commission values. In contrast, all models for *M. minutus* showed low values of both omission and commission. Because models are based only on presence data (and not all areas are adequately sampled), the commission index reflects not only true commission error but also a component that results from undersampled areas that the species actually inhabits. We here propose an operational procedure for determining an optimal region of the omission/commission relationship and thus selecting high-quality GARP models. Our implementation of this technique for *H. anomalus* gave a much more reasonable estimation of the species' potential distribution than did the original suite of models. These findings are relevant to evaluation of other distributional-modeling techniques based on presence-only data and should also be considered with other machine-learning applications modified for use with asymmetrical input data.
© 2002 Elsevier Science B.V. All rights reserved.

---

* Corresponding author. Tel.: +1-212-769-5693; fax: +1-212-769-5239.
E-mail address: rpa@amnh.org (R.P. Anderson).

1. Introduction

1.1. Predictive modeling of species' potential distributions

Predictive modeling of species' distributions now represents an important tool in biogeography, evolution, ecology, conservation, and invasive-species management (Busby, 1986; Nicholls, 1989; Walker, 1990; Walker and Cocks, 1991; Sindel and Michael, 1992; Wilson et al., 1992; Box et al., 1993; Carpenter et al., 1993; Austin and Meyers, 1996; Kadmon and Heller, 1998; Yom-Tov and Kadmon, 1998; Corsi et al., 1999; Peterson et al., 1999, 2000; Fleishman et al., 2001; Peterson and Vieglais, 2001; Boone and Krohn, 2002; Fertig and Reiners, 2002; Scott et al., 2002). These approaches combine occurrence data with ecological/environmental variables (both biotic and abiotic factors: e.g. temperature, precipitation, elevation, geology, and vegetation) to create a model of the species' requirements for the examined variables. Primary occurrence data exist in the form of georeferenced coordinates of latitude and longitude for confirmed localities that typically derive from vouchered museum or herbarium specimens (Baker et al., 1998; Funk et al., 1999; Soberón, 1999; Ponder et al., 2001; Stockwell and Peterson, 2002a). Absence data are rarely available, especially in poorly sampled tropical regions where modeling may hold greatest value (Stockwell and Peters, 1999; Anderson et al., 2002a). The environmental variables typically examined in such modeling efforts encompass only relatively few of the possible ecological-niche dimensions (Hutchinson, 1957). Nevertheless, currently available digital environmental coverages (digitized computer maps) provide many variables that commonly influence species' macrodistributions (Grinnell, 1917a,b; Root, 1988; Brown and Lomolino, 1998).

The resulting model is then projected onto a map of the study region, showing the species' potential geographic distribution (e.g. Chen and Peterson, 2000; Peterson and Vieglais, 2001). Models are generally based on the species' fundamental niche (Hutchinson, 1957; including factors controlling distributions put forward in Grinnell, 1917b; see also MacArthur, 1968; Wiens, 1989; Morrison and Hall, 2002). Thus, some areas indicated by the model as regions of potential presence may be occupied by closely related species, or may represent suitable areas to which the species has failed to disperse or in which it has gone extinct. Rather than a drawback, however, this "overprediction" resulting from the niche-based nature of the models actually allows for synthetic evolutionary and ecological applications comparing potential and realized distributions (Peterson et al., 1999; Peterson and Vieglais, 2001; Anderson et al., 2002a,b).

1.2. Variability among GARP models

The Genetic Algorithm for Rule-Set Prediction (GARP: http://biodi.sdsc.edu/; see http://beta.lifemapper.org/desktopgarp/ for software download) is an expert-system, machine-learning approach to predictive modeling (Stockwell and Peters, 1999). Genetic algorithms constitute one class of artificial-intelligence applications and were inspired by models of genetics and evolution (Holland, 1975). They have been applied to various problems not amenable to traditional computational methods because the search space of all possible solutions is too large to search exhaustively in a reasonable amount of time (Stockwell and Noble, 1992). Genetic algorithms present a heuristic solution to this dilemma by scanning broadly across the search space and refining solutions that show high values for the optimization (fitness) criterion. GARP has proven especially successful in predicting species' potential distributions under a wide variety of situations (Peterson and Cohoon, 1999; Peterson et al., 1999, 2001, 2002a,b,c; Godown and Peterson, 2000; Sánchez-Cordero and Martínez-Meyer, 2000; Peterson, 2001; Elith and Burgman, 2002; Feria-A. and Peterson, 2002; Stockwell and Peterson, 2002a,b; but see Lim et al., 2002). Chen and Peterson (2000), Peterson and Vieglais (2001), and Anderson et al. (2002a) provide general explanations of the GARP modeling process and interpretation of potential distributions; see Stockwell and Noble (1992) and Stockwell and Peters (1999) for technical details.

GARP reduces error in predicted distributions by maximizing both significance and predictive accuracy, a novel goal for such analytical systems (Stockwell and Peters, 1999). The algorithm is largely successful in doing so without overfitting or overly specializing rules, which is especially important when models are

based on occurrence data compiled without a fixed study design (Peterson and Cohoon, 1999). Owing to stochastic elements in the algorithm (such as mutation and crossing over; Holland, 1975; Stockwell and Noble, 1992), however, no unique solution is produced; indeed, the underdetermination of the system yields multiple solutions holding the same value for the optimization criterion. Hence, the variability among resulting models (typical of most machine-learning problems) requires careful examination of possible sources of error in order to select the most predictive models.

A common strategy for evaluating model quality has been to divide known localities randomly into two groups: *training* data used to create the model and an independent *test* data set used to evaluate model quality (Fielding and Bell, 1997; Fielding, 2002). One-tailed χ^2-statistics (or binomial probabilities, if sample sizes are small) are often employed to determine whether test points fall into regions of predicted presence more often than expected by chance, given the proportion of map pixels predicted present by the model (e.g. Peterson et al., 1999; Anderson et al., 2002a). These tests using independent test data thus provide *extrinsic* measures of model significance (departure from random predictions). However, by excluding part of the data set from the model-building stage, the algorithm cannot take advantage of all known locality records. Clearly, an optimal model would incorporate data from all available records of the species.

One tactic for managing the variability among models has been to make multiple models and determine how many models predict particular pixels as present (Anderson et al., 2002a; Lim et al., 2002; Peterson et al., unpublished data). Anderson et al. (2002a) tempered among-model variation by making three GARP models per species and creating a composite prediction based on all three models. In further analyses, map pixels predicted present by at least two of the models were then considered "predicted presence". Similarly, Lim et al. (2002) created five models per species and deemed pixels predicted by three or more of them as predicted presence in subsequent analyses. More recently, Peterson et al. (unpublished data) have made larger numbers of models and summed them (for each model, value of 1 for a pixel of presence; value of 0 for predicted absence). In such an approach,

Table 1
Elements of a confusion matrix[a]

Predicted	Actual	
	Present	Absent
Present	a	b
Absent	c	d

[a] In GARP, map pixels are re-sampled with replacement to produce the elements of the confusion matrix. Element *a* represents known distributional areas correctly predicted as present. Likewise, *d* reflects regions where the species has not been found and that are classified by the model as absent. Element *c* denotes omission: map pixels of known distribution predicted absent by the model. Conversely, *b* reflects areas from which the species is not known but that are predicted present (commission, both true and apparent—see Section 1.3).

the value of a pixel in the composite (summed) map thus equals the number of models predicting presence in that cell. Summing models may reveal a consistent signal that holds up across many different independent random walks of model generation. The above methods weigh all model replicates equally; in contrast, we herein compare such equal-weight tactics with a best-subsets approach.

1.3. Error components

Two types of error are possible in predictive models of species' distributions: false negatives (omission error or underprediction) and false positives (commission error or overprediction). The relative proportions of these errors are typically expressed in a *confusion matrix*, or *error matrix* (Fielding and Bell, 1997). Four elements are present in a confusion matrix (Table 1). Element *a* represents known distributional areas correctly predicted as present, and *d* reflects regions where the species has not been found and that are classified by the model as absent. Thus, *a* and *d* are considered correct classifications; in contrast, *c* and *b* are usually interpreted as errors. Element *c* denotes omission: pixels of known distribution predicted absent by the model. Conversely, *b* is a measure of areas of absence (or "pseudo-absence"—see below) incorrectly predicted present (commission). Unfortunately, when known presence points are few in number and true absence points are not available, problems arise with some measures derived from the confusion matrix (Fielding and Bell, 1997).

GARP creates a confusion matrix by intrinsically re-sampling map pixels with replacement. First, 1250 map pixels are chosen randomly with replacement from those pixels holding localities of known occurrence (training points). The quantity a is the number of those pixels that coincide with areas of predicted presence; the number falling outside the prediction equals c. Thus, $a + c = 1250$ for GARP models in which all pixels are predicted as either present or absent (in some models, the rule-set may not make a decision for every pixel; such pixels are then coded as "no data" in the prediction—see below). Likewise, 1250 pixels are re-sampled with replacement from the remaining pixels of the study area (any pixels without confirmed presence data in the training set). These pixels are referred to as background points or pseudo-absence points (Stockwell and Peters, 1999), highlighting the difference between models based on typical biodiversity information (positive occurrence records from zoological museums or herbaria, as here) and those that also include true absence data (e.g. Corsi et al., 1999; Fertig and Reiners, 2002). Background pixels that fall into regions of predicted presence yield b, whereas background pixels of predicted absence produce d; $b + d = 1250$ for models with a presence/absence prediction for all pixels (but less if not all cells are predicted either present or absent).

As mentioned above, distributional-modeling algorithms like GARP are often used with only presence data. For most species, data regarding absence are not available (Stockwell and Peters, 1999; Peterson, 2001). In addition, when a *potential* distribution based on the species' fundamental niche is desired, use of absence data could adversely affect the model-building process by inhibiting inclusion of areas that hold suitable environmental conditions where the species is not present due to historical restrictions or biological interactions (Peterson et al., 1999; Anderson et al., 2002b). However, despite the practical necessity and theoretical justification for using only presence data in modeling ecological niches, this asymmetry in input data (errors in pseudo-absences but not in presences) requires that interpretation of the confusion matrix be amended. In such cases, whereas element c represents pure omission error, element b includes the contributions of both true and apparent commission error.

Apparent commission error derives from potentially habitable regions correctly predicted as presence, but that cannot be demonstrated as such because no verification of the species exists there. The lack of verification of the species may have various causes (Karl et al., 2002). In certain cases, some areas lacking documentation of the species stem from historical causes or biotic interactions (Peterson, 2001). For example, disjunct areas of potential habitat with no records of the species often correspond to historical restrictions or the historical effects of speciation (e.g. failure of the species to disperse to a region of suitable habitat; Peterson et al., 1999; Peterson and Vieglais, 2001; Anderson et al., 2002a). Similarly, competition between related species showing parapatric distributions likely restricts many species' realized distributions (Peterson, 2001; Anderson et al., 2002b). Other biological interactions—such as predation in some parts of the potential range but not in others—may also limit some species' distributions. In addition to historical and biotic causes, apparent commission error can also derive from inadequate sampling: map pixels of real presence (at least at some time of the year in some subhabitat) lacking documentation of the species because they have not been adequately sampled by biologists (Karl et al., 2002). This latter form of apparent commission error has recently been recognized in presence/absence data sets where inventories were extensive yet incomplete (Boone and Krohn, 1999; Karl et al., 2000; Schaefer and Krohn, 2002; Stauffer et al., 2002). By definition, it reaches maximum manifestation in presence-only modeling applications like current implementations of GARP. As the goal of presence-only potential-distribution modeling is to determine which of the background (pseudo-absence) pixels actually represent suitable areas for a species—whether or not it actually inhabits them—interpreting measures of commission is critical.

1.4. Intrinsic and extrinsic measures of model performance

1.4.1. Measures including both omission and commission (composite indices)

One measure of overall model performance is the *correct classification rate* of Fielding and Bell (1997) (see Table 2). GARP provides an intrinsic correct

Table 2
Quantitative measures used in this study[a]

Measure	Calculation
Intrinsic	
Overall performance (correct classification rate)	$(a + d)/(a + b + c + d)$
Omission error (false negative rate)	$c/(a + c)$
Commission index (false positive rate)	$b/(b + d)$
Extrinsic	
Overall performance (significance)	$\sum[(\text{observed} - \text{expected})^2/\text{expected}]$ for test points
Omission error	$\text{out}_{\text{test}}/n_{\text{test}}$
Commission index	Proportion of pixels predicted present

[a] Intrinsic measures (based on training data used to make the model) are given above and extrinsic equivalents (based on independent test data) below. Measures of overall performance include contributions of both omission and commission.

classification rate derived from the confusion matrix: $(a + d)/(a + b + c + d)$—equal to the "accuracy" of Stockwell and Peterson (2002b), not that of Anderson et al. (2002a). This quantity ranges from 0 to 1 and is designed to measure overall model adequacy, including contributions of both omission and commission in the denominator. Note that, correct classification rate = (1 minus sum of error terms)/(sum of all terms). However, because element b is overestimated by the preponderance of background (pseudo-absence) pixels, this statistic is necessarily biased with data sets that lack true absence data (common with biodiversity information; Peterson, 2001; Ponder et al., 2001; Stockwell and Peterson, 2002a). Likewise, the overall Kappa (κ)-statistic of Fielding and Bell (1997) includes elements of both omission and commission and thus suffers from the same problem (see also Fielding, 2002).

The χ^2-statistic based on independent test data can be used as an extrinsic measure of overall performance, because it incorporates both omission (of test points) and commission (via expected frequencies; Table 2). However, this statistic is highly sensitive to the proportional extent of predicted presence, making highly significant results possible with unacceptably high omission rates (e.g. models that only include the core ecological distribution of the species). In addition, χ^2-significance values are related to sample size (Peterson, 2001). Hence, it is likely that neither correct classification rates, κ-statistics (both potentially intrinsic), nor χ^2-significance values (typically extrinsic) represent reliable measures of overall model performance.

1.4.2. Measures of omission and commission

To assess model performance more adequately, other indices that provide intrinsic estimates of each error component can be derived from the confusion matrix (Table 2; reviewed in Fielding and Bell, 1997). The quantity $c/(a + c)$ represents the *intrinsic omission error* rate, and $b/(b + d)$ represents what we here term the *intrinsic commission index* (false negative and false positive rates, respectively, of Fielding and Bell (1997)). The intrinsic omission error reflects the proportion of known localities (training points) that fall outside the predicted region (by re-sampling with replacement to produce the confusion matrix). The intrinsic commission index mirrors the proportion of pixels predicted present by the model (proportion of re-sampled background points falling into regions of predicted presence). Owing to the general scarcity of confirmed presence data, however, this latter index includes contributions of (1) true commission error (overprediction) as well as of (2) apparent commission error (correctly predicted areas not verifiable as such, primarily because of the lack of adequate sampling). The aim of predictive modeling is precisely to determine this latter quantity, as well as the geographic distribution of those pixels. To emphasize the dual nature of $b/(b + d)$, we term it the intrinsic commission *index* rather than intrinsic commission *error*. One of our aims is to discriminate between its two components.

Extrinsic measures of omission and commission exist parallel to the respective intrinsic ones (Table 2). Where out_{test} = the number of test points falling outside predicted areas and n_{test} = the number of

test points, out_{test}/n_{test} represents *extrinsic omission error*. Likewise, the proportion of pixels predicted present can serve as an *extrinsic commission index*. In fact, because the number of training points is usually extremely small in comparison with the number of background pixels in the overall study region, the intrinsic commission index will converge on this extrinsic measure with adequate re-sampling.

In the present study, we evaluate model performance based on both intrinsic and extrinsic criteria, with the goal of identifying optimal models based on intrinsic measures only. If that were possible, optimal models could then be identified even when generated using all known locality data. We approach this problem by examining measures of omission and commission, as well as composite indices designed to reflect both quantities. Because measures of commission are dependent on the proportional extent of areas potentially inhabitable by the species within the study region, we examine in detail three cases whose modeled ecological niches show geographic manifestations occupying varying proportions of the respective study areas. Current implementations of GARP represent the modification of a general algorithm for the specific case of presence-only (generally museum) data. The present research is also germane to evaluation of other distributional-modeling techniques that use presence-only data. In addition, it may be broadly relevant to machine-learning applications with asymmetrical input data (asymmetrical errors).

2. Methods

2.1. Study species

The spiny pocket mouse *Heteromys anomalus* (Heteromyidae) is a common, medium-sized rodent (50–100 g) that is widespread along the Caribbean coast of South America in northern Colombia and Venezuela, as well as on the nearby islands of Trinidad, Tobago, and Margarita. It has been documented in deciduous forest, evergreen rainforest, cloud forest, and some agricultural areas, typically from sea level to approximately 1600 m (Anderson, 1999, unpublished data; Anderson and Soriano, 1999). We examine its distribution in northeastern Colombia and northwestern Venezuela (7°30′–12°30′N, 68°30′–76°00′W). In

most of this region, it is the only *Heteromys* present, simplifying interpretations of its potential and realized distributions (Anderson, 1999; Anderson et al., 2002b). Although *H. anomalus* is widespread in the region, inventories strongly suggest that it is absent from higher montane regions (e.g. above 2000 m in the Sierra Nevada de Santa Marta, Serranía de Perijá, and Cordillera de Mérida), dry lowland scrub habitat, swampy areas, and open tropical savannas (llanos) of the Orinoco basin (Bangs, 1900; Allen, 1904; Handley, 1976; August, 1984; Díaz de Pascual, 1988, 1994; Soriano and Clulow, 1988; Anderson, 1999).

Microryzomys minutus (Muridae) is a small-bodied rodent (10–20 g) known from medium-to-high elevations of the Andes and associated mountain chains from Venezuela to Bolivia (Carleton and Musser, 1989). It occupies an elevational range of approximately 1000–4000 m and has been recorded primarily in wet montane and submontane forests, as well as occasionally in mesic páramo habitats above treeline. We evaluate the central and northern extent of its distribution, from northern Peru to Colombia and Venezuela (9°S to 13°N, 51–82°W). A congeneric species, *M. altissimus*, occupies generally higher elevations in much of this region, but occasionally the two have been found in sympatry. *M. minutus* has not been encountered in lowland regions (below approximately 1000 m). Likewise, it is apparently absent from dry puna habitat above treeline, and obviously from permanent glaciers on the highest mountain peaks.

Carpodacus mexicanus (Fringillidae) is a relatively small passerine bird distributed throughout western North America south to southern Mexico (AOU, 1998). On its native range, it is generally found in arid landscapes (often associated with humans) and is typically absent from higher elevations and humid areas. As an introduced species, it has successfully invaded humid regions such as Hawaii and eastern North America. We analyze its native geographic distribution in Mexico, where it is clearly associated with dry habitats and human habitation.

2.2. Model building

We employed the Genetic Algorithm for Rule-Set Prediction (GARP; http://biodi.sdsc.edu/; but see http://beta.lifemapper.org/desktopgarp/ for current software download) to model potential distributions

of the three study species (Stockwell and Noble, 1992; Stockwell and Peters, 1999). GARP searches for non-random associations between environmental characteristics of localities of known occurrence versus those of the overall study region. It works in an iterative process of rule selection, evaluation, testing, and incorporation or rejection to produce a heterogeneous rule-set characterizing the species' ecological requirements (Peterson et al., 1999). First, a method is chosen from a set of possibilities (e.g. logistic regression, bioclimatic rules), and it is applied to the data. Then, a rule is developed and predictive accuracy sensu (Stockwell and Peters, 1999) is evaluated via training points intrinsically re-sampled from both 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 model (rule-set). As implemented here, the algorithm runs either 2500 iterations or until addition of new rules has no appreciable effect on the intrinsic accuracy measure (convergence). The final rule-set, or ecological-niche model, is then projected onto a digital map as the species' potential geographic distribution, exported as an ASCII raster grid, and imported into ArcView 3.1 (ESRI, 1998) using the Spatial Analyst Extension for visualization.

The base environmental data comprise a variety of geographic coverages (digitized maps). For *H. anomalus* and *M. minutus*, we used 21 environmental coverages. These coverages have a pixel size of $0.04° \times 0.04°$ (about $4.5\,\text{km} \times 4.5\,\text{km}$) and consist of elevation, slope, aspect, soil conditions, geological ages, geomorphology, coarse potential vegetation zones, and a series of coverages for solar radiation, temperature, and precipitation. For the latter three, separate coverages representing upper and lower bounds of isopleth intervals were included (for mean annual solar radiation, mean annual temperature, mean monthly temperature in January and July, mean annual precipitation, and mean monthly precipitation in January and July). For *C. mexicanus*, models were based on four coverages: elevation, potential vegetation type, average annual temperature, and mean annual precipitation. The pixel size for *C. mexicanus* was $0.06° \times 0.06°$ (about $7\,\text{km} \times 7\,\text{km}$).

Unique localities of species' occurrences came from Anderson (1999, unpublished data; 85 localities) for

H. anomalus; Carleton and Musser (1989; 72 localities) for *M. minutus*; and the *Atlas of the Distribution of Mexican Birds* (Peterson et al., 1998; 333 localities) for *C. mexicanus* (museums are cited in Acknowledgements). We divided collection localities randomly into training and test data sets (50% each) for each species. Twenty models were made for each species using their respective training sets; the same training set was used to create each of the 20 models for a species. Test points were withheld completely from GARP's model-building and internal evaluation process, and were used only for evaluating final models.

2.3. Model evaluation

2.3.1. Intrinsic values

For each model, we obtained the elements of the confusion matrix and calculated values of the correct classification rate $((a + d)/(a + b + c + d))$, the intrinsic omission error $(c/(a + c))$ and the intrinsic commission index $(b/(b + d))$ (Table 2). In some models, GARP failed to predict every pixel as either present or absent; such pixels are categorized as "no data" in the resultant map and reclassified as predicted absence in further geographic analyses (warranted because the models were based only on presence and pseudo-absence data; Ricardo Schachetti-Pereira, personal communication). These unpredicted pixels do not enter into the confusion matrix (see Section 1).

2.3.2. Extrinsic values and expert evaluation

Applying a one-tailed χ^2-statistic to the test data, we evaluated the significance of each model against a null hypothesis of no relationship between the prediction and the test data points. More precisely, we tested whether test points fell into areas predicted present more often than expected at random, given the overall proportion of pixels predicted present versus predicted absent for that species (modified from Peterson et al., 1999). The χ^2-value represented our extrinsic composite measure of overall model performance (including contributions of both omission and commission—see Anderson et al., 2002a). We used the proportion of test points falling outside the prediction ($\text{out}_{\text{test}}/n_{\text{test}}$) as our extrinsic measure of omission error (= 1 minus the "accuracy" of Anderson et al., 2002a). Likewise, we calculated the extrinsic commission index as the proportion of land surface predicted present (Table 2).

In addition, each model was evaluated subjectively by specialists (RPA and DL for mammals; ATP for birds) according to our understanding of the species' autecology and known distribution and the geography of major climatic and biotic zones. Evaluations were made blind to the model statistics to be assessed. We classified models as good, medium, or poor. Good models excluded areas where experts believed a species probably does not exist and included most or all known areas of distribution. Poor models excluded large areas of true distribution or included large areas of likely unsuitable habitat. Medium models suffered from lesser problems of either type. Models were not penalized for including suitable areas without records for the species—e.g. regions inhabited by congeneric species or regions of likely suitable conditions to which the species has failed to disperse (Peterson et al., 1999; Anderson et al., 2002a).

For each species, we plotted the following combinations of intrinsic and extrinsic measures for each model: (1) extrinsic performance (χ^2) versus intrinsic correct classification rate ($(a + d)/(a + b + c + d)$); (2) intrinsic omission error ($c/(a + c)$) versus intrinsic commission index ($b/(b + d)$); and (3) extrinsic omission error (out_{test}/n_{test}) versus extrinsic commission index (proportion of study region predicted present). Models in each plot were flagged according to the independent expert evaluation of quality. In addition, we calculated correlations between intrinsic and extrinsic measures of omission, commission, and overall performance. To assess how well intrinsic measures of omission and commission predicted extrinsic ones, we regressed the latter onto the former in simple linear regressions.

2.4. Concordance among models

Given the variability present among GARP models, we considered the possibility that a suite of 20 models might predict the potential distribution better than any single model, by revealing a consistent signal present in most models (see Section 1). Thus, we extended the equal-weight approaches of Anderson et al. (2002a), Lim et al. (2002), and Peterson et al. (unpublished data) by summing the 20 models for each species (value of 1 for a pixel of predicted presence; value of 0 for predicted absence). This procedure produced a composite map comprised of pixels with values ranging from 0 to 20, representing the number of models that predicted the species' presence in the pixel. For visualization of these results, we present maps showing various thresholds of concordance among models: (1) distribution of pixels predicted present by at least 6/20 models; (2) pixels predicted present by at least 11/20 models; and (3) pixels predicted present by at least 16/20 models.

3. Results

3.1. Composite measures of performance

Extrinsic performance measures (χ^2) were almost always significant. Seventeen of the 20 models for *H. anomalus* showed significant deviations from random predictions, in the desired direction (χ^2 for significant models = 4.07–16.95; $P < 0.05$; one-tailed critical value $\chi^2_{1,0.05} = 2.706$; the other three models showed non-significant departures in the desired direction). All models were highly significant for both *M. minutus* ($\chi^2 = 177.02$–684.74; $P \ll 0.05$) and *C. mexicanus* ($\chi^2 = 42.29$–164.50; $P \ll 0.05$). The latter species had an extremely large number of test points, which resulted in high statistical power. Models for *M. minutus* were highly significant despite the moderate number of test points, due to almost all test points falling in a very small predicted area relative to the study region. Because of the proportionately large geographic extent of *H. anomalus* in its study area and a moderate number of test points, the tests of significance for that species had relatively lower statistical power than those for the other two examples.

However, no consistent trend was observed between intrinsic and extrinsic measures of overall model performance (Fig. 1). The graphs suggest a generally positive relationship for *C. mexicanus* ($r = 0.80$), but the correlation between the two measures was low for *H. anomalus* ($r = 0.45$) and *M. minutus* ($r = 0.32$). In all three cases, however, variation in intrinsic overall performance was minimal compared with the great variation in the extrinsic measure of overall performance (χ^2). Likewise, no uniform trend existed between these composite measures of performance and model quality as judged by expert classification (Fig. 1).
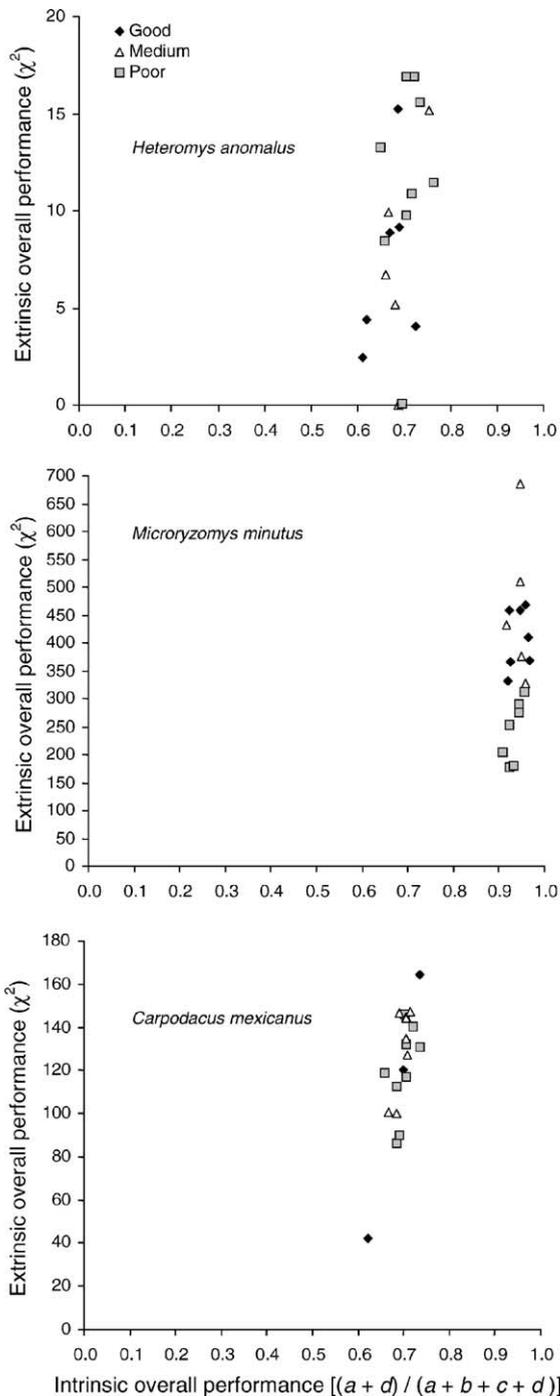
Fig. 1. Plots of intrinsic and extrinsic measures of overall performance for models of the three species. Individual models are flagged by categories of model quality (good, medium, poor) from expert evaluations, which were made blind to the numeric values.

3.2. Omission and commission

Within each species, intrinsic and extrinsic evaluations showed consistent patterns between omission errors and commission indices (Fig. 2). For *H. anomalus* and *C. mexicanus* (the two species with relatively large potential distributions within their respective study regions), omission and commission values were inversely related, with the data swarm slightly concave upward in each case. For *M. minutus*, all models were clustered at low values, with no clear trends within the tight clusters. The best models, as evaluated by specialists, occupied different portions of the omission/commission graphs depending on the relative geographic extent of the species' potential distribution. For the two species with relatively large potential distributions (*H. anomalus* and *C. mexicanus*), the best models were found with low omission and relatively high commission values. In contrast, all models for the geographically restricted *M. minutus* showed a more equal balance between omission and commission, with low values for both.

Likewise, extrinsic values for omission and commission tracked the corresponding intrinsic values for the widespread species but not for *M. minutus*. For *H. anomalus* and *C. mexicanus*, the intrinsic and extrinsic omission values were highly correlated ($r = 0.64$ and 0.78, respectively), and regressions of extrinsic estimates onto intrinsic ones were significant ($P < 0.01$). Although average extrinsic and intrinsic omission values were similar for *C. mexicanus*, extrinsic omission for *H. anomalus* was much greater than the intrinsic omission estimate (probably due to the moderate number of training points, insufficient for adequately portraying the species' niche). In contrast to those two species, intrinsic and extrinsic omission errors were only weakly correlated for *M. minutus* ($r = 0.20$), and the regression of the latter onto the former was not significant ($P = 0.39$).

Paralleling the results for omission, intrinsic and extrinsic commission values were strongly associated for the two widespread species but not for *M. minutus*. Correlations between the two measures for *H. anomalus* and *C. mexicanus* were very high ($r = 0.98$ and 0.85, respectively), with highly significant regressions of extrinsic measures onto intrinsic ones ($P \ll 0.001$). For *M. minutus*, intrinsic and extrinsic commission values showed only weak correlation ($r = 0.43$), and
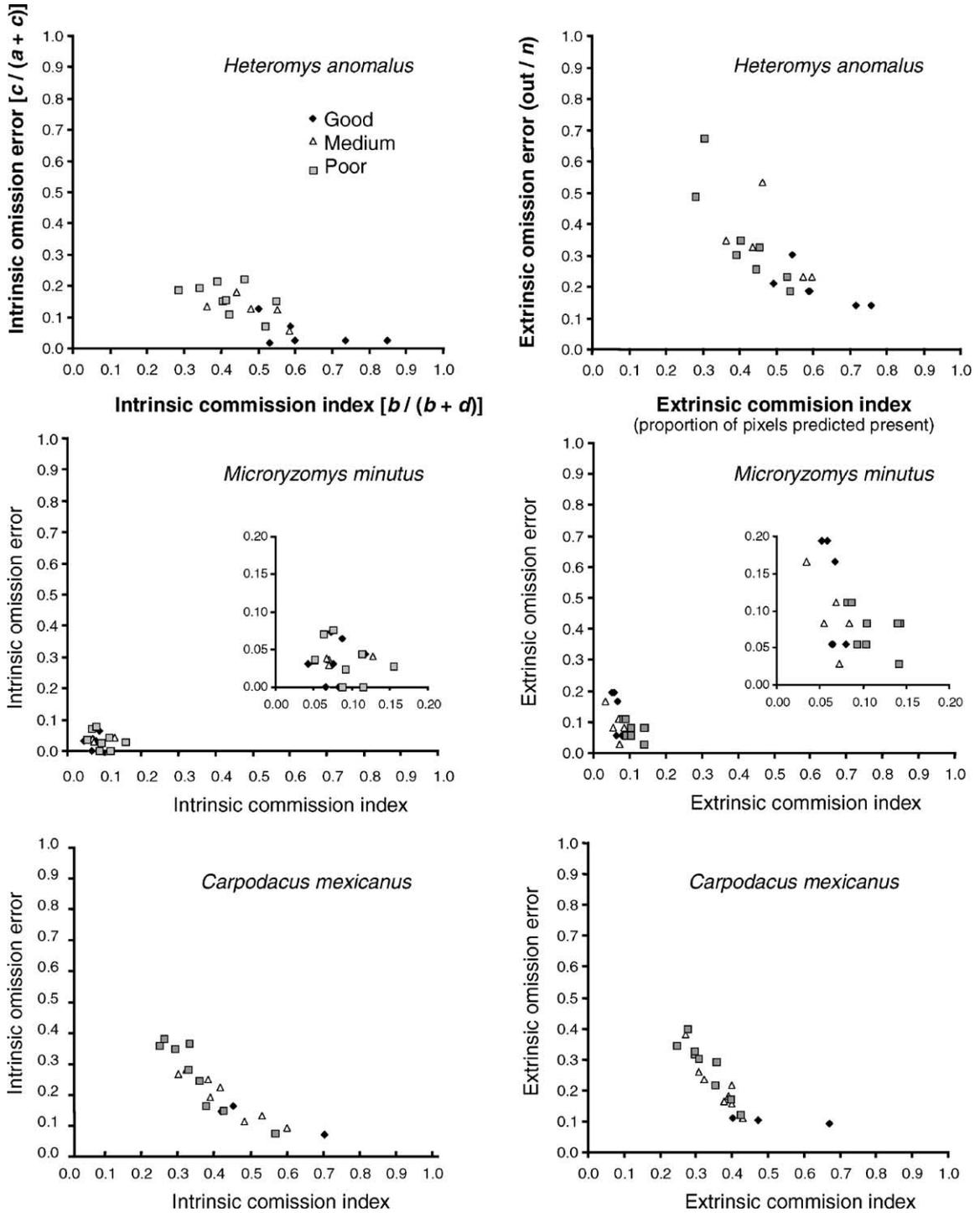
Fig. 2. Intrinsic and extrinsic plots of omission error vs. commission index, for each of the three species. Individual models are flagged by categories of model quality from expert evaluations (good, medium, poor), which were made blind to the numeric values.

the regression was non-significant but nearly so ($P = 0.06$).

3.3. Ecogeographic interpretation of model quality

Expert evaluation found clear differences in quality among the 20 models for each species. We here discuss the patterns found for *H. anomalus* as an example. Poor models typically predicted presence in all montane regions but almost no lowland regions. Thus, in addition to piedmont regions (where the species is present and commonly collected), they implausibly included areas too high for the species in the Sierra Nevada de Santa Marta, Serranía de Per-

ijá, and Cordillera de Mérida (Bangs, 1900; Allen, 1904; Handley, 1976; Díaz de Pascual, 1988, 1994; Anderson, 1999). At the same time, they failed to include areas of true distribution in lowland deciduous forests. Models were extremely variable in the Venezuelan llanos, where the open savannas (uninhabitable for *H. anomalus*) and gallery forests (from which *H. anomalus* is known) comprise a mosaic of habitats not adequately reflected in our coarse environmental coverages (Anderson et al., 2002a). In contrast, in addition to correctly predicting presence in the piedmont, the best models succeeded in predicting absence in high montane regions while also including lowland regions of deciduous forest (where
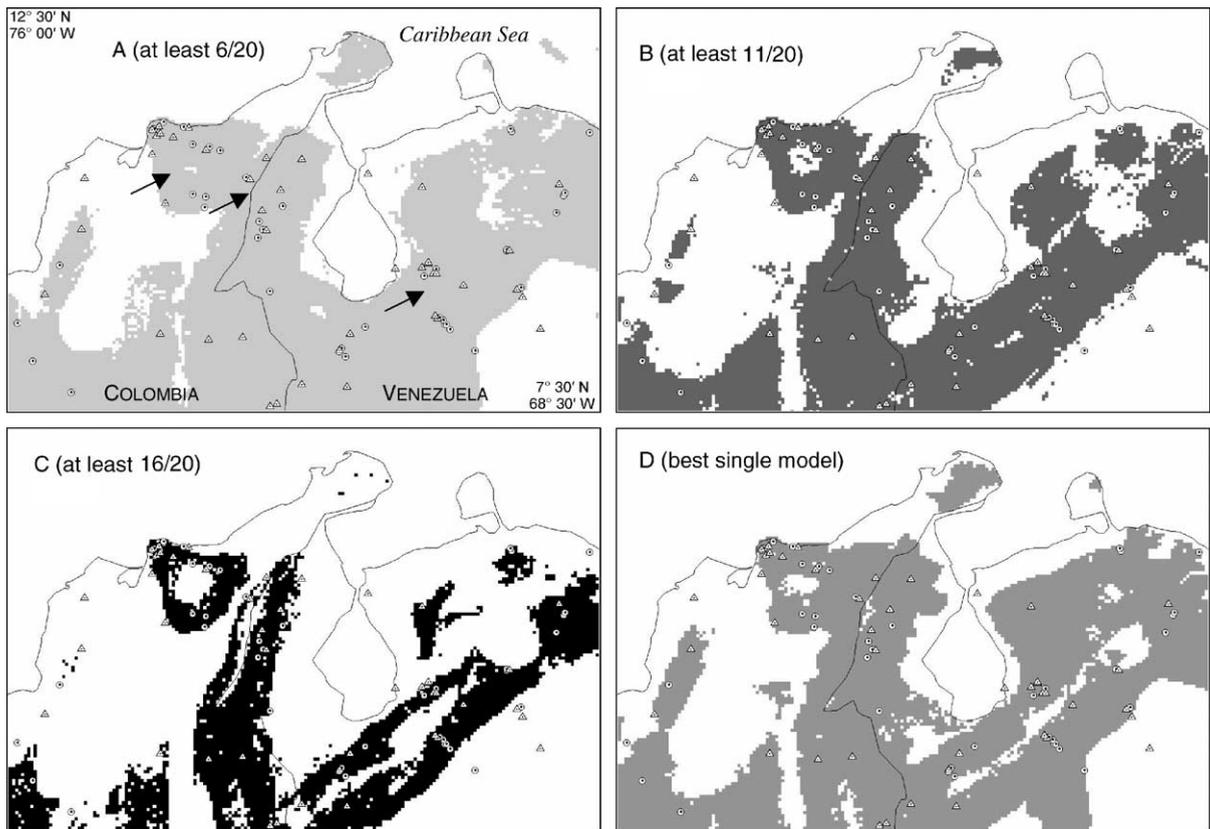
Fig. 3. Maps of the modeled potential distribution of *H. anomalus* in the study area. Panels A–C show various thresholds of concordance among the 20 models (at least 6/20, 11/20, and 16/20 models predicting presence, respectively). Training localities (used to build the model) are denoted by circles; independent, randomly chosen test localities are represented by triangles. Low thresholds (e.g. 6/20) include areas where the species' presence is doubtful, such as high montane regions of the Sierra Nevada de Santa Marta, Serranía de Perijá, and Cordillera de Mérida (arrows in A). Higher thresholds (e.g. 16/20) suffer by missing areas of lowland distribution (C). In contrast, Model 13 (shown in D), succeeded in predicting presence for most of the lowland distribution of the species and predicting absence in high montane regions.

the species is known). These high-quality models generally excluded both extremely dry scrub habitats and swampy areas around the lower Cauca/Magdalena (Colombia) and Catatumbo (Venezuela) drainages from which the species is not known and where its presence is unlikely.

3.4. Concordance among multiple models (composite approach)

Applying various thresholds of concordance among models, no suitable balance between omission and commission was achieved for any of the species. For
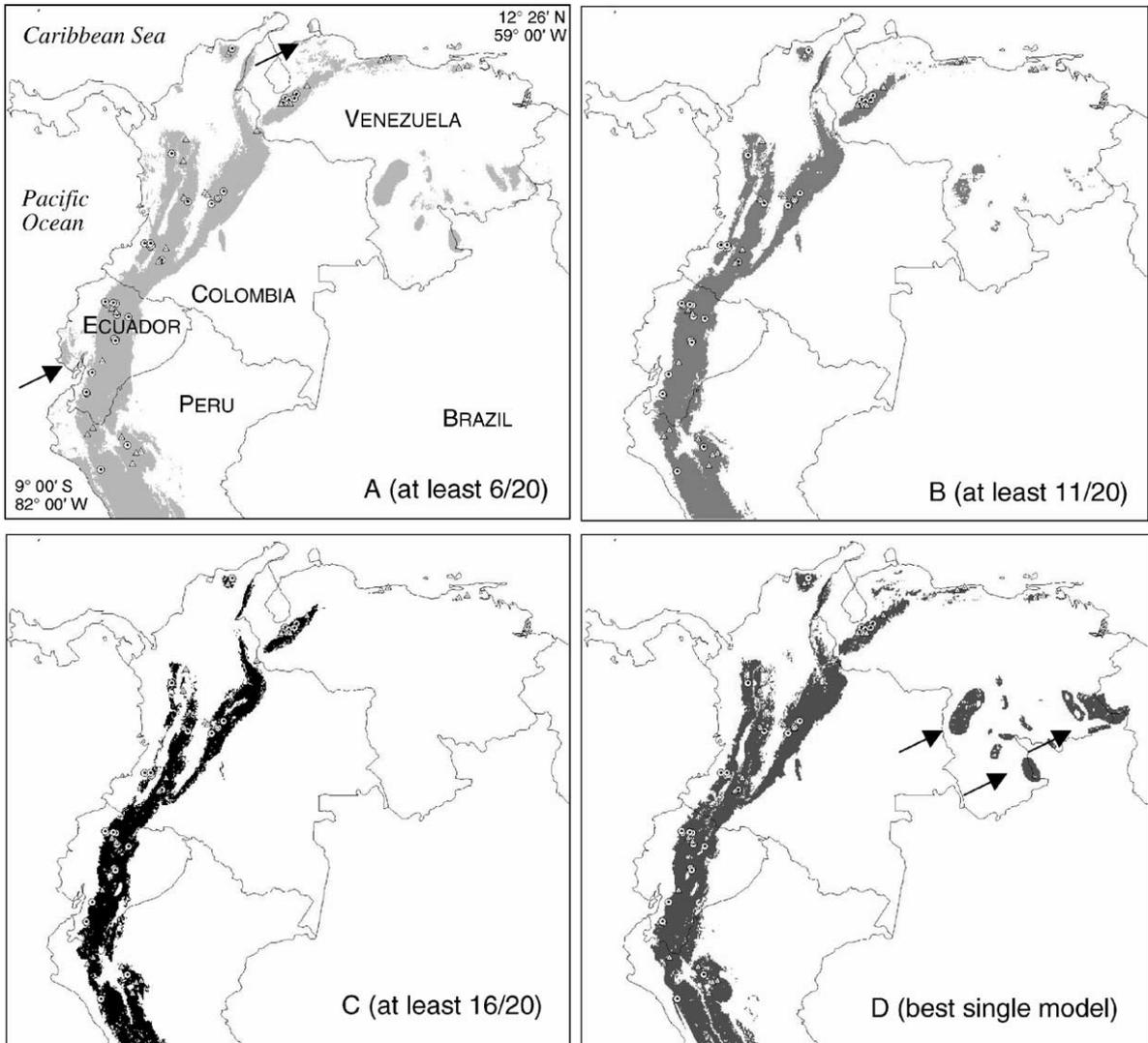
Fig. 4. Maps of the modeled potential distribution of *M. minutus* in the study area. Panels A–C show various thresholds of concordance among the 20 models (at least 6/20, 11/20, and 16/20 models predicting presence, respectively). Triangles are used to depict training localities, and circles denote test localities. Low thresholds (e.g. 6/20) include areas where the species' presence is doubtful, such as arid areas of western Peru and Ecuador and northern Venezuela (arrows in A). Higher thresholds (e.g. 16/20) suffer by missing real distributional areas at intermediate elevations (C). In contrast, Model 9 succeeded in effectively predicting presence for the species' known distribution, as well as areas of similar conditions in the Guianan highlands (tepui formations; arrows in D).
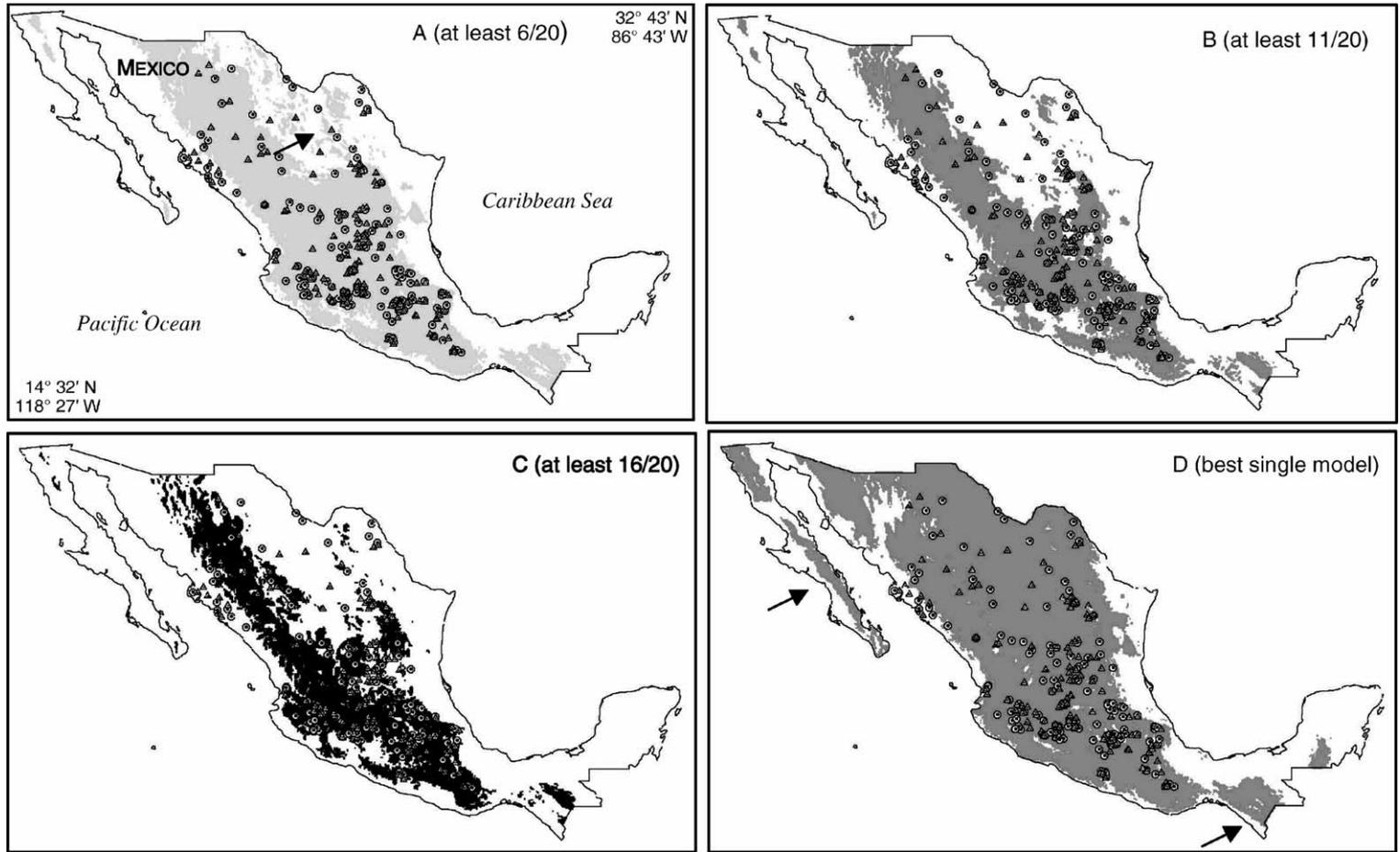
Fig. 5. Maps of the modeled potential distribution of *C. mexicanus* in the study area. Panels A–C show various thresholds of concordance among the 20 models (at least 6/20, 11/20, and 16/20 models predicting presence, respectively). Training localities are indicated by circles; triangles depict test localities. Even low thresholds of concordance among models (e.g. at least 6/20, A) fail to accurately predict the species' distribution in northeastern Mexico (arrows in A); this problem is especially severe at stricter thresholds (e.g. 16/20, C). In contrast, Model 6 (D) correctly predicted presence for the species in northeastern Mexico, as well as in disjunct areas of similar habitat southeast of the Isthmus of Tehuantepec and on the Península de Baja California (arrows in D).

H. anomalus, the map of pixels predicted by at least 6 of the 20 models yielded a composite model with a satisfactory prediction of the lowland distribution of the species (Fig. 3A), but that erroneously indicated potential habitat in high montane regions. The map of pixels predicted present in 11 or more models showed only a slight indication of predicting absence in high mountain regions, and lost predicted presence in suitable lowland regions (Fig. 3B). Converse to results of the first threshold, a composite model with a threshold of 16 or more models (Fig. 3C) gave a map that correctly predicted the species' absence in high montane regions, but omitted the known lowland distribution.

The same limitations of this approach were apparent with the other species. A composite model with a threshold of six for *M. minutus* predicted presence in some lowland areas of extremely unlikely distribution (e.g. Chocoan rainforest; arid regions in northwestern Peru, southwestern Ecuador, and northern Colombia and Venezuela; Fig. 4A). The stricter threshold of 11 models lost those lowland regions, but still overpredicted presence in some extremely high areas (including permanent glaciers) not habitable by the species (Fig. 4B). The composite map of pixels predicted present by at least 16 of the 20 models (Fig. 4C) indicated absence in the extremely high mountain regions, but also predicted absence some lower montane regions of known distribution (such as the Cordillera de la Costa in Venezuela). Composite models for *C. mexicanus* consistently underestimated pixels of presence for the species' distribution in Mexico (Fig. 5). All three thresholds of composite models failed to predict presence in the northern and eastern portions of the Chihuahua Desert, and the coast of southwestern Mexico was predicted absent in the composite with a 16-model threshold (Fig. 5C). Hence, the range of this broadly distributed species was underestimated by the equal-weight composite approach.

In contrast to the results from the superimposed models, at least one single model for each species reflected the species' distributions well, as judged by experts (Figs. 3–5D). For *H. anomalus*, Model 13 (Fig. 3D) correctly excluded most high montane areas while still including acceptable lowland predictions. Likewise, Model 9 for *M. minutus* avoided predicting presence in lowland or very high regions and maintained predicted presence of intermediate elevations (Fig. 4D). Finally, Model 6 for *C. mexicanus* correctly

predicted the species' distribution in northern and eastern Mexico without neglecting the species' distributional areas along the coast of Guerrero and Oaxaca (Fig. 5D). These models all had low omission values, but the commission index varied by species.

4. Discussion

4.1. Measures of overall performance

Considerable variation was present among GARP models, as predicted by the theoretical background of genetic algorithms (Holland, 1975) and indicated by previous work (e.g. Anderson et al., 2002a). Thus, the algorithm generally performed as expected under this domain. Below, however, we consider issues regarding error quantification in this special case of presence-only data. Furthermore, we explore relationships between various indices and expert-judged model quality.

Neither extrinsic nor intrinsic measures of overall performance provided an effective means for identifying the best models. Extrinsic model significance (χ^2) probably varied among the species in part due to the power afforded by varying sample sizes in the test data sets, and also according to the relative extent of suitable habitat for each species (Peterson, 2001). Models with highest significance (lowest *P*-value) did not consistently include the best models identified by experts (Fig. 1). Models with highest significance often included the core ecological distribution of that species, but excluded ecologically peripheral parts of the known distribution. For example, highly significant models for *H. anomalus* included montane regions (especially the piedmont, where the majority of the localities are found) without extending into known distributions in the lowlands (from which fewer points were present). Thus, although the χ^2-measure of significance indicates departure from a random prediction, it is not a reliable indicator of model quality.

Likewise, the intrinsic measure of overall model performance did not identify the best models either (Fig. 1). In fact, the value $(a + d)/(a + b + c + d)$ varied little among models within species. This result is consistent with the findings of Stockwell and Peterson (2002b), who found that this quantity (their "accuracy") reached an apparent plateau with sample

sizes of 20–50 localities. Thus, this measure also fails as a measure of quality to discriminate among a suite of final GARP models.

4.2. Utility of omission/commission graphs

In contrast to overall performance measures, both intrinsic and extrinsic plots of omission versus commission may be useful for selecting optimal models, at least for species with medium-to-large proportional potential distributions in the study region. For the two widespread species, the best models were found in the same regions of the respective intrinsic and extrinsic omission/commission graphs (Fig. 2), and intrinsic and extrinsic measures were highly correlated. Because patterns in intrinsic measures are repeated in the independent extrinsic ones, intrinsic measures hold potential for assessing model quality when all available data points are used for model construction.

Whereas the best models for the two widespread species combined low measures of omission with fairly high levels of commission, all models for *M. minutus* showed low values of both omission and commission. For *M. minutus* (a montane species with an extremely small proportional distribution within the study area), optimal GARP models minimize omission without increasing commission excessively (because pixels of predicted presence represent a small fraction of the study region). In contrast, for species with medium-to-large proportional potential distributions in the study region (exemplified here by *H. anomalus* and *C. mexicanus*), large areas must be included as predicted presence (yielding high values in the commission index) in order to reduce omission to acceptable levels without overfitting the data.

4.3. Separating the commission index into error and overfitting

While high values of commission may at first seem an undesirable tradeoff to reduce omission, we return to the dual nature of the commission index. In addition to true commission error, this index also reflects areas of potential distribution correctly predicted but not verifiable owing to lack of occurrence records—which can result either from: (1) inadequate sampling in areas of real distribution; or (2) historical restrictions or biotic interactions in areas of potential but

not realized distribution (see *apparent commission error* of Karl et al. (2002) and of Peterson (2001), as discussed above). In an ideal model, the commission index, $b/(b + d)$, should equal the true proportion of pixels potentially habitable by the species in the study region. Thus, as long as the number of known occurrence points is small with respect to the species' potential range, we propose that the ideal value of the commission index equals the true proportion of pixels that hold potential distribution for the species, such that true proportion = pixels of true distribution/total pixels in the study area. For example, for a species with a true potential distribution that encompasses half of the study area, the optimal value for the intrinsic commission index ($b/(b+d)$) would be 0.50. Therefore, on average, true commission *error* only exists above that value. True commission *error* can be estimated as the commission *index* minus the true proportion of pixels habitable for the species, or *intrinsic commission error* = $b/(b + d)$ minus *true proportion*.

Models that exceed zero commission *error* generally commit true commission, whereas those with values to the left of zero tend to overfit the data, some quite severely. For example, a model that predicts the
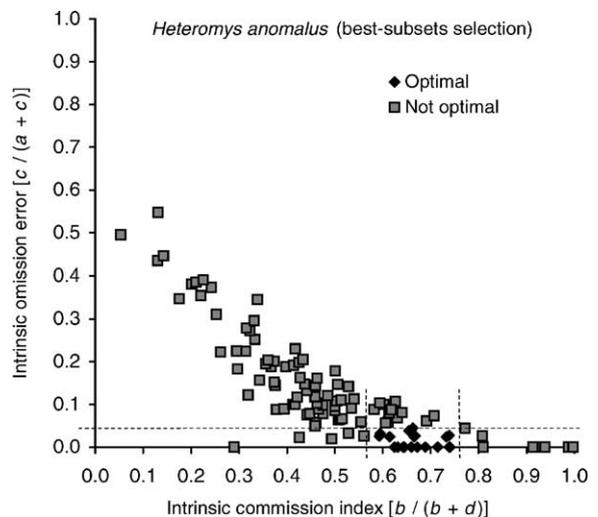
Fig. 6. Plot of values of intrinsic omission error vs. intrinsic commission index, for 112 new models of *H. anomalus*. Models falling into the optimal region are marked with a solid diamond, with all others flagged by a shaded square. The present data swarm confirms the general inverse, slightly concave-up relationship between omission and commission found in preliminary analyses. See Fig. 7 for geographic portrayal of the optimal models.

entire study region would include commission error
for all species that have a true proportion <100% of
the study region. In contrast, a severely overfit model
would include the most or all of the training points and
small regions around them, failing to predict presence
in other regions of suitable habitat (typically holding
fewer training points). Models for *H. anomalus* that
only predicted presence in piedmont regions repre-
sent examples of the latter, sacrificing moderate num-
bers of ecologically peripheral localities while still
predicting most of the remaining localities in a rela-
tively small geographic area. Our results suggest that,
for species with a potential distribution that occupies
a large proportional extent of the study region, many
models overfit the preponderance of the data and fail
to include the species' full potential distribution. How-
ever, some models do effectively reduce omission er-
ror while still predicting only a reasonable proportion

of the study region (i.e. not suffering excessive com-
mission error).

Obviously, the goal of predictive modeling is to *de-
termine* the true proportion of pixels of potential pres-
ence for the species, as well as their location. It may be
possible to approximate this value empirically. Of the
four "good" models for *H. anomalus* with the lowest
intrinsic omission errors (<5%; Fig. 2), expert inspec-
tion suggested that the ones with lower values on the
intrinsic commission index included slightly too few
areas of predicted presence. In contrast, the higher two
somewhat overestimated areas of potential presence.
The average commission of these four models seems
to give a reasonable estimate of the true extent of
the species' potential distribution, (*true proportion* =
0.68 via intrinsic commission index; 0.65 via propor-
tion of pixels predicted present). The average of all 20
models, however, gives a biased (lower) estimate, as
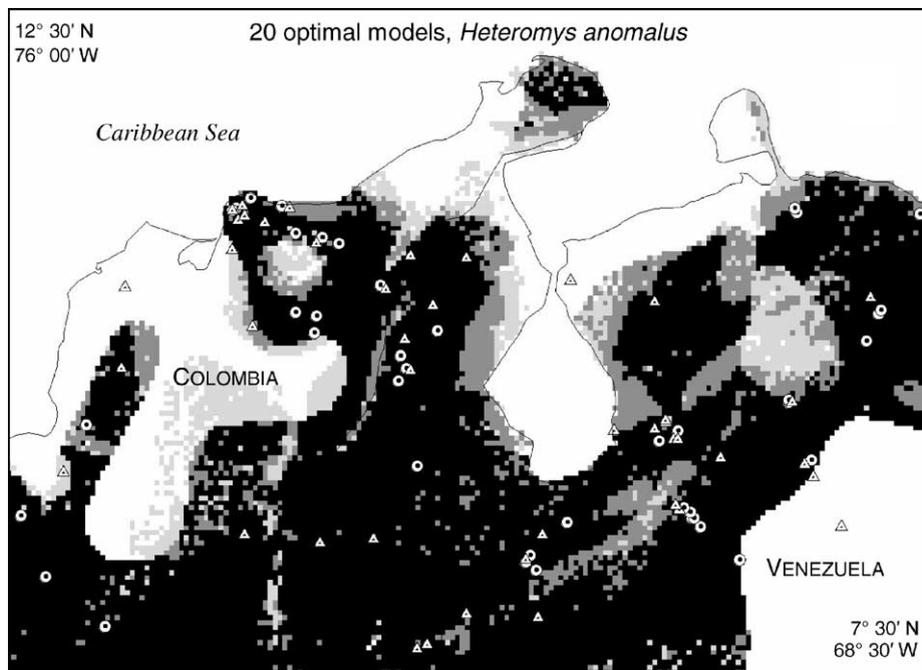
Fig. 7. Map of 20 optimal potential-distribution models for *H. anomalus* in the study area, selected from 112 new models (Fig. 6). Various
thresholds of concordance among the 20 models are shown in increasing color intensity matching the separate panels of Fig. 3 (at least
6/20, pale gray; 11/20, medium gray; and 16/20, black). Circles denote training localities, and triangles are used to plot test localities. Note
that even low thresholds (e.g. 6/20) begin to exclude high montane areas where the species' presence is doubtful, in contrast to Fig. 3A
(the 20 original models). Higher thresholds (e.g. 16/20) here succeed in predicting absence in high montane regions while maintaining
predicted presence in areas of lowland distribution, in contrast to Fig. 3C (the 20 original models). The optimal models show the same
general geographic characteristics as Model 13, the best single model in preliminary analyses (see Fig. 3D).

the models with higher omission error are too restrictive and underestimate the true potential distribution (0.49 via intrinsic commission index; 0.50 via proportional extent).

4.4. Selecting optimal models

To test an operational method of selecting optimal models, we produced more GARP models for *H. anomalus*, using the same training data set. We made models until finding 20 that fell in a region of the intrinsic omission/commission graph that we identified as the *optimal region*, as defined below. We arbitrarily only accepted models with 5% or less intrinsic omission error and selected an interval of the intrinsic commission index centered on the approximate estimated proportion of pixels of potential distribution for the species (*true proportion* ≈ 2/3, from first set of analyses—see above). Around that value (0.67), we arbitrarily set a deviation of 0.10 to produce an acceptable interval from 0.57 to 0.77.

To obtain 20 new models that fell into the optimal region of the omission/commission graph, we made a total of 112 additional models of *H. anomalus*. These models formed a slightly concave-up data swarm (Fig. 6) similar to that intimated by the original 20 models for the species. Upon inspection, the 20 models from this round of modeling that fell into the optimal region presented the general geographic characteristics identified by the experts as necessary for a good model (similar to the best model of the first 20, shown in Fig. 3D). As a whole, they avoided the errors that plagued the medium and poor models from the original set.

Additionally, the superposition of all 20 optimal models from the second set (Fig. 7) did not show the tradeoffs suffered by the superposition of the 20 original models (Fig. 3A–C). Rather, we interpret the new composite map as a relatively unbiased density surface related to the probability of suitable environmental conditions for the species. For example, pixels predicted present by 16 or more models (Fig. 7) correctly indicate absence in high montane regions, while maintaining a more realistic distribution in the lowlands. The few test localities that fall outside areas of predicted presence derive from drier regions that, by random chance, were not represented by any of the training points. In sum, the best-subsets selection pro-

cedure is superior to an equal-weight approach (used in Anderson et al., 2002a; Lim et al., 2002; Peterson et al., unpublished data).

5. Conclusions and recommendations

In the terminology of genetic algorithms, modification of GARP for use with presence-only occurrence data can result in a highly atypical fitness surface. When visualized in omission/commission space, the repercussions of pseudo-absences sometimes create a fitness *ridge*, rather than the typical global fitness *peak*. For GARP distributional models, this ridge is likely present for most species having medium-to-large potential distributions in the study region. Solutions along the ridge show similar values for intrinsic overall performance (= correct classification rate, which is highly correlated with the optimization criterion). However, solutions at opposite endpoints of the ridge differ dramatically in error composition as well as qualitative aspects of the geographic prediction—with error in models at one extreme of the ridge including a great deal of omission and ones at the other comprised entirely of commission. Because much commission "error" is not real but rather apparent (due especially to undersampling), only solutions with low omission represent correct ones.

Hence, our results indicate that identification of an optimal region of the intrinsic omission/commission graph holds promise as a way to select high-quality GARP models without resorting to an extrinsic test data set. This approach allows all occurrence data to be used in generating models, thus increasing the predictive capacity of GARP in cases where occurrence data are scarce. When occurrence data are sufficient to permit independent testing without reducing the training data set excessively, extrinsic measures can be used in the same best-subsets selection procedure. In either case, high-quality models can potentially be chosen without expert supervision. Minimally, only two parameters would have to be provided by the user: a maximum acceptable level of omission error, and the width of the optimal interval on the commission index.

Towards that end, we here propose an operational protocol for generation and selection of a best subset of optimal GARP ecological-niche models and distributional predictions.

- *Step 1*: Arbitrarily set an acceptable level of intrinsic omission error (e.g. 5%), representing the upper limit of the optimal region along that axis.
- *Step 2*: Approximate the true proportion of the species' potential distribution in the study region, as the mean value on the commission index (or median, if density function is skewed) for those preliminary models with an acceptable level of intrinsic omission (from Step 1). This value then represents the center of the optimal region on the intrinsic-commission-index axis.
- *Step 3*: Arbitrarily set the acceptable width of the optimal region of the intrinsic-commission-index axis (e.g. ±0.1 in this study).
- *Step 4*: Make models until the desired number of models falling within the optimal region is reached.
- *Step 5*: Superimpose the selected models to create a composite prediction showing the number of optimal models predicting presence in each pixel across the study region.

Although unsupervised model building (without subjective expert evaluation) remains premature, we hope that this approach will allow selection of better models and stimulate research that will make operational, unsupervised modeling possible in the future. In particular, the process outlined above provides an objective means of model evaluation at least for species with moderate-to-large potential distributions in the study region. Such species, upon both theoretical and empirical grounds, are likely to show an inverse association between omission and commission (necessary for the current selection procedure). Species with very small potential distributions relative to the study region, like *M. minutus*, represent a challenge for future research, because all models are likely to lie within a small region of the omission/commission graph. Future studies should evaluate the generality of the present results, considering that at least the following factors may affect patterns of model quality: geographic extent of the study region; proportion of the species' range encompassed in the study region; proportional extent of the potential distribution of the species in the study region; resolution and composition of the physical, climatic, and biotic GIS coverages (base data); niche breadth of the species; number of localities available; and degree of spatial autocorrelation (and thus bias) among collection lo-

calities (e.g. disproportionate collection effort near roads and rivers; Funk et al., 1999; Lim et al., 2002). In the meantime, applications of this method should continue to graph omission and commission errors and examine the geographic predictions visually.

Our model-selection approach is based on various measures of accuracy and error derived from the confusion matrix and does not address model significance. While one motive of our research was to allow the use of all occurrence data in distributional modeling, we still recommend the production of preliminary models based on training data (following Fielding and Bell, 1997). Such preliminary models allow for the assessment of significance (departure from random predictions) with an independent test data set using techniques such as a χ^2-test (Peterson et al., 1999) or ROC analysis (Zweig and Campbell, 1993; Pearce et al., 2002). After significance has been demonstrated, species with only moderate numbers of available occurrence points are probably best modeled using all available localities. However, the model-selection process we propose here can be used to identify optimal models made either with all available occurrence points (using intrinsic measures of omission and commission for model selection) or with a training subset of the data points (using extrinsic measures to select optimal models). Future work should extend the research of Stockwell and Peterson (2002b) in light of the current conclusions, exploring ways to determine how many occurrence points are necessary for adequate modeling.

The crux of the current findings clearly lies with asymmetry of the input data (presence-only occurrence records). Here, we modify the evaluation of distributional models produced with such data by a non-deterministic algorithm (one that produces multiple solutions given the same input data). By definition, model selection per se would not be necessary for deterministic algorithms that identify only one solution (distributional prediction), such as generalized linear models, bioclimatic-envelope methods, and others (Busby, 1986; Nicholls, 1989; Walker and Cocks, 1991; Box et al., 1993; Carpenter et al., 1993; Jarvis and Robertson, 1999; Elith and Burgman, 2002). However, when based on presence-only data, evaluation of such models and valid comparison with models produced by other techniques requires consideration of the dual nature of the commission

index. In addition, a best-subsets selection procedure would likely be useful for identifying correct models produced by deterministic algorithms when jackknifing or bootstrapping of input data (of occurrence records and/or environmental predictor variables) introduces variation into the system. Finally, in addition to applications with distributional modeling, researchers should critically examine components of error with other machine-learning techniques (especially genetic algorithms) that have been modified for use with asymmetrical input data, to determine if a similar best-subsets approach is warranted in those cases.

Acknowledgements

References

Allen, J.A., 1904. Report on mammals from the district of Santa Marta, Colombia, collected by Mr. Herbert H. Smith, with field notes by Mr. Smith. Bull. Am. Mus. Natl. Hist. 20, 407–468.

Anderson, R.P., 1999. Preliminary review of the systematics and biogeography of the spiny pocket mice (*Heteromys*) of Colombia. Rev. Acad. Colomb. Cienc. Exactas, Físicas y Naturales 23 (Suplemento especial), 613–630.

Anderson, R.P., Soriano, P.J., 1999. The occurrence and biogeographic significance of the southern spiny pocket mouse *Heteromys australis* in Venezuela. Z. Sauget. 64, 121–125.

Anderson, R.P., Gómez-Laverde, M., Peterson, A.T., 2002a. Geographical distributions of spiny pocket mice in South America: insights from predictive models. Glob. Ecol. Biogeogr. 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.

AOU, 1998. Check-List of North American Birds, 7th ed. American Ornithologists' Union, Washington, DC, 829 pp.

August, P.V., 1984. Population ecology of small mammals in the llanos of Venezuela. In: Martin, R.E., Chapman, B.R. (Eds.), Contributions in Mammalogy in Honor of Robert L. Packard. Spec. Publ. Mus. Tex. Tech Univ. 22, 71–104.

Austin, M.P., Meyers, J.A., 1996. Current approaches to modelling the environmental niche of eucalyptus: implication for management of forest biodiversity. Forest Ecol. Manage. 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. Occas. Pap. Mus. Tex. Tech Univ. 187, 1–4.

Bangs, O., 1900. List of the mammals collected in the Santa Marta region of Colombia by W.W. Brown, Jr. J. Proc. N. Engl. Zool. Club 1, 87–102.

Boone, R.B., Krohn, W.B., 1999. Modeling the occurrence of bird species: are the errors predictable? Ecol. Appl. 9, 835–848.

Boone, R.B., Krohn, W.B., 2002. Modeling tools and accuracy assessment. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 265–270.

Box, E.O., Crumpacker, D.W., Hardin, E.D., 1993. A climatic model for location of plant species in Florida, USA. J. Biogeogr. 20, 629–644.

Brown, J.H., Lomolino, M.V., 1998. Biogeography, 2nd ed. Sinauer Associates, Sunderland, MA, 691 pp.

Busby, J.R., 1986. A biogeoclimatic analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. Aust. J. Ecol. 11, 1–7.

Carleton, M.D., Musser, G.G., 1989. Systematic studies of oryzomyine rodents (Muridae, Sigmodontinae): a synopsis of *Microryzomys*. Bull. Am. Mus. Natl. Hist. 191, 1–83.

Carpenter, G., Gillison, A.N., Winter, J., 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. Biodivers. Conserv. 2, 667–680.

Chen, G.-J., Peterson, A.T., 2000. A new technique for predicting distribution of terrestrial vertebrates using inferential modeling. Zool. Res. 21, 231–237.

Corsi, F., Duprè, E., Boitani, L., 1999. A large-scale model of wolf distribution in Italy for conservation planning. Conserv. Biol. 13, 150–159.

Díaz de Pascual, A., 1988. Aspectos ecológicos de una microcomunidad de roedores de selva nublada, en Venezuela. Bol. Soc. Venez. Cienc. Nat. 145, 93–110.

Díaz de Pascual, A., 1994. The rodent community of the Venezuelan cloud forest, Mérida. Polish Ecol. Stud. 20, 155–161.

Elith, J., Burgman, M., 2002. Predictions and their validation: rare plants in the central highlands, Victoria, Australia. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 303–313.

ESRI, 1998. ArcView GIS, version 3.1. Environmental Systems Research Institute Inc., Redlands, CA.

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. Divers. Distrib. 8, 49–56.

Fertig, W., Reiners, W.A., 2002. Predicting presence/absence of plant species for range mapping: a case study from Wyoming. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 483–489.

Fielding, A.H., 2002. What are the appropriate characteristics of an accuracy measure? In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 271–280.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 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. Conserv. Biol. 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. Biodivers. Conserv. 8, 727–751.

Godown, M.E., Peterson, A.T., 2000. Preliminary distributional analysis of US endangered bird species. Biodivers. Conserv. 9, 1313–1322.

Grinnell, J., 1917a. Field tests of theories concerning distributional control. Am. Nat. 51, 115–128.

Grinnell, J., 1917b. The niche-relationships of the California thrasher. Auk 34, 427–433.

Handley, C.O., Jr., 1976. Mammals of the Smithsonian Venezuelan Project. Brigham Young Univ. Sci. Bull. Biol. Ser. 20 (5), 1–91.

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, 183 pp.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harb. Symp. Quant. Biol. 22, 415–427.

Jarvis, A.M., Robertson, A., 1999. Predicting population sizes and priority conservation areas for 10 endemic Namibian bird species. Biol. Conserv. 88, 121–131.

Kadmon, R., Heller, J., 1998. Modelling faunal responses to climatic gradients with GIS: land snails as a case study. J. Biogeogr. 25, 527–539.

Karl, J.W., Heglund, P.J., Garton, E.O., Scott, J.M., Wright, N.M., Hutto, R.L., 2000. Sensitivity of species–habitat relationship model performance to factors of scale. Ecol. Appl. 10, 1690–1705.

Karl, J.W., Svancara, L.K., Heglund, P.J., Wright, N.M., Scott, J.M., 2002. Species commonness and the accuracy of habitat-relationship models. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 573–580.

Lim, B.K., Peterson, A.T., Engstrom, M.D., 2002. Robustness of ecological niche modeling algorithms for mammals in Guyana. Biodivers. Conserv. 11, 1237–1246.

MacArthur, R., 1968. The theory of the niche. In: Lewontin, R.C. (Ed.), Population Biology and Evolution. Syracuse University Press, Syracuse, NY, pp. 159–176.

Morrison, M.L., Hall, L.S., 2002. Standard terminology: toward a common language to advance ecological understanding and application. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 43–52.

Nicholls, A.O., 1989. How to make biological surveys go further with generalized linear models. Biol. Conserv. 50, 51–75.

Pearce, J.L., Venier, L.A., Ferrier, S., McKenney, D.W., 2002. Measuring prediction uncertainty in models of species distribution. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 383–390.

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. Ecol. Model. 117, 159–164.

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., Egbert, S.L., Sánchez-Cordero, V., Price, K.P., 2000. Geographic analysis of conservation priority: endemic birds and mammals in Veracruz, Mexico. Biol. Conserv. 93, 85–94.

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. Ecol. Model. 144, 21–30.

Peterson, A.T., Ball, L.G., Cohoon, K.P., 2002a. Predicting distributions of tropical birds. Ibis 144, E27–E32.

Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddemeier, R.H., Stockwell, D.R.B., 2002b. Future projections for Mexican faunas under global climate change scenarios. Nature 416, 626–629.

Peterson, A.T., Stockwell, D.R.B., Kluza, D.A., 2002c. Distributional prediction based on ecological niche modeling of primary occurrence data. In: Scott, J.M., Heglund, P.J.,

Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 617–623.

Ponder, W.F., Carter, G.A., Flemons, P., Chapman, R.R., 2001. Evaluation of museum collection data for use in biodiversity assessment. Conserv. Biol. 15, 648–657.

Root, T., 1988. Environmental factors associated with avian distributional boundaries. J. Biogeogr. 15, 489–505.

Sánchez-Cordero, V., Martínez-Meyer, E., 2000. Museum specimen data predict crop damage by tropical rodents. Proc. Natl. Acad. Sci. U.S.A. 97, 7074–7077.

Schaefer, S.M., Krohn, W.B., 2002. Predicting vertebrate occurrences from species habitat associations: improving the interpretation of commission error rates. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 419–427.

Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), 2002. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, 868 pp.

Sindel, B.M., Michael, P.W., 1992. Spread and potential distribution of *Senecio madagascariensis* Poir. (fireweed) in Australia. Aust. J. Ecol. 17, 21–26.

Soberón, J., 1999. Linking biodiversity information sources. Trends Ecol. Evol. 14, 291.

Soriano, P.J., Clulow, F.V., 1988. Efecto de las inundaciones estacionales sobre poblaciones de pequeños mamíferos en los llanos altos occidentales de Venezuela. Ecotrópicos 1, 3–10.

Stauffer, H.B., Ralph, C.J., Miller, S.L., 2002. Incorporating detection uncertainty into presence-absence surveys for marbled murrelet. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 357–365.

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. Math. Comput. Simul. 33, 385–390.

Stockwell, D., Peters, D., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. Int. J. Geogr. Inf. Sci. 13, 143–158.

Stockwell, D.R.B., Peterson, A.T., 2002a. Controlling bias in biodiversity data. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 537–546.

Stockwell, D.R.B., Peterson, A.T., 2002b. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148, 1–13.

Walker, P.A., 1990. Modelling wildlife distributions using a geographic information system: kangaroos in relation to climate. J. Biogeogr. 17, 279–289.

Walker, P.A., Cocks, K.D., 1991. HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. Glob. Ecol. Biogeogr. Lett. 1, 108–118.

Wiens, J.A., 1989. The Ecology of Bird Communities: Foundations and Patterns, vol. 1. Cambridge University Press, Cambridge, UK, 539 pp.

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. J. Biogeogr. 19, 183–194.

Yom-Tov, Y., Kadmon, R., 1998. Analysis of the distribution of insectivorous bats in Israel. Divers. Distrib. 4, 63–70.

Zweig, M.H., Campbell, G., 1993. Receiver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. 39, 561–577.