ELSEVIER

Ecological Modelling 148 (2002) 1–13

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

Effects of sample size on accuracy of species distribution models

David R.B. Stockwell [a,*], A. Townsend Peterson [b,1]

[a] *San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0505, USA*
[b] *Natural History Museum, The University of Kansas, Lawrence, KS 66045, USA*

Received 12 December 2000; received in revised form 11 May 2001; accepted 15 June 2001

Abstract

Given increasing access to large amounts of biodiversity information, a powerful capability is that of modeling ecological niches and predicting geographic distributions. Because, sampling species' distributions is costly, we explored sample size needs for accurate modeling for three predictive modeling methods via re-sampling of data for well-sampled species, and developed curves of model improvement with increasing sample size. In general, under a coarse surrogate model, and machine-learning methods, average success rate at predicting occurrence of a species at a location, or accuracy, was 90% of maximum within ten sample points, and was near maximal at 50 data points. However, a fine surrogate model and logistic regression model had significantly lower rates of increase in accuracy with increasing sample size, reaching similar maximum accuracy at 100 data points. The choice of environmental variables also produced unpredictable effects on accuracy over the range of sample sizes on the logistic regression method, while the machine-learning method had robust performance throughout. Examining correlates of model performance across species, extent of geographic distribution was the only significant ecological factor. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Ecological niche; Species distribution; GARP; Sample size

1. Introduction

Modeling habitat requirements of species is an increasingly important tool, both for investigating the requirements of species and for planning conservation reserves (Austin and Meyers, 1996; Scott et al., 1996; Corsi et al., 1999; Jarvis and Robertson, 1999) and for understanding of patterns of biodiversity (Williams and Hero, 2001). In particular, using existing data and modeling species' distributions to orient field efforts reduces the cost of field surveys (Nicholls, 1989; Fielding and Bell, 1997). Distribution modeling is essential to ensuring consistency, while reducing the time and costs of large-scale studies of biodiversity involving large numbers of species (Stockwell, 1993, 1997; Stockwell and Peters, 1999; Pearce and Ferrier, 2000). Particularly, when combined

* Corresponding author. Tel.: +1-858-822-0942; fax: +1-858-822-3631.
E-mail addresses: davids@sdsc.edu (D.R.B. Stockwell), town@ukans.edu (A.T. Peterson).
[1] Tel.: +1-785-842-5384.

0304-3800/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0304-3800(01)00388-X

with ready access to data via the Internet, these approaches provide a critical avenue to placing biodiversity considerations at the forefront of science and policy-making (Boston and Stockwell, 1995; Vieglais et al., 1998).

Sample sizes of species' occurrence points used to develop predictive models greatly affect the success rate of those models in predicting the occurrence of a species at a location, termed *accuracy*: Bayesian predictors (Stockwell, 1997), generalized linear modeling (Pearce and Ferrier, 2000), and wildlife-habitat modeling (Scott et al., 1996). For example, a study on the biodiversity of rainforest frogs in Australia asserted that due to the limited sample size of 22, models could only be descriptive and hypothesis generating rather than predictive (Williams and Hero, 2001). In a multi-taxon evaluation of GAM and GLM approaches, sample sizes greater than 250 were needed to maximize accuracy (Pearce and Ferrier, 2000).

Advances in networking are making available for biodiversity studies the millions of records available so far in natural history museum databases (Kaiser, 1999; Pennisi, 2000). Species distribution modeling using these data, while making use of existing data, restricts analyses to the data for species that are already available. The data requirements of some methods may be prohibitively large for using most natural history museum collections databases, making study of many species impossible. The greatest cost associated with such studies is associated with collecting biological data, given the enormous time and resource requirements of field collections and studies (Peterson et al., 1998). It is, therefore, necessary to quantify the relationship between sample size and predictive distribution models, and to determine, which methods make more efficient use of survey data. If there are large differences in size/accuracy relationships for different methods, then this criterion should be among those used for selecting a predictive algorithm (Guisan and Zimmermann, 2000).

The *Atlas of Mexican Bird Distributions* (Peterson et al., 1998) provides a useful reference point for a well-known taxon, surveyed across most natural history museums worldwide, in a country with large geographic extent and diversity of habitats and environments. Inspecting the frequency distribution of number of unique localities per species (Fig. 1), 103 of 1060 species (about 10%) have 200 or more unique points available, but 689 (65%) have ten or more points available. Mexico represents the best-sampled Latin American country ornithologically (Peterson et al., 1998). Hence, for one country that could represent the median in world collection intensity, decreasing the number of data points required from 200 to 10 points would increase the proportion of species that could be modeled more than 6-fold.

A parallel question is that of whether to expend additional resources to improve models by gathering more information. Although performance generally improves with addition of information, plateaus exist wherein new information adds little to model performance (Stockwell, 1997; Peterson and Cohoon, 1999). In fact, in some systems accuracy can be *reduced* with increasing information (Stockwell, 1997), usually because the additional variables produce models optimized for a too-specific set of information that performs poorly on new data, a problem long recognized as prediction bias or overfitting (Verbyla, 1986; Verbyla and Litvaitis, 1989).

Establishment of minimum data needs for adequate accuracy may be possible (Peterson and Cohoon, 1999). To determine these needs, we must understand factors that affect the amounts of data needed to achieve certain accuracy levels, an issue we refer to as *data efficiency*, including two components: the rate at which accuracy increases with increasing data, and the maximum accuracy achievable by the method (Stockwell, 1997). Ideally, of course, a predictive system that converges quickly to the greatest accuracy would be preferred. Comparisons of accuracy of modeling methods have used numerous methods to attack the basic problem of determining the suitability of a site for a species—predicting a two-valued variable (presence vs. absence) based on a set of environmental variables. Multivariate statistical methods including logistic regression and discriminant analysis (Austin and Meyers, 1996; Corsi et al., 1999), artificial intelligence methods, such as genetic algorithms (Stockwell, 1999), and

the 'surrogate' methods that use a single variable, such as vegetation type and develop a model (Scott et al., 1996) have all been used to this end. Comparisons of the relative merits of these diverse approaches have proven challenging.

For a formal approach to the general issue of using data efficiently, however, we can draw on lessons from the field of computational machine learning using a formal learning paradigm (Kelly and Glymour, 1990). A formal learning paradigm is a statement of the components of a learning process, including possible data and concepts (or hypotheses), the way the data is presented to the learner, abilities of the algorithm, including the complexities of particular languages used to express that concept, and criteria for success of the learner. Under this paradigm, the problem of data efficiency in predicting species distributions is to determine, how many data are needed to discover the correct model (or functional form) from a vast number of possible models, in a noisy, biased, and poorly defined environment. Herein, we provide an additional example of how machine learning provides a basis for understanding and evaluating efficiency of methods for habitat modeling (Davey and Stockwell, 1991).

2. Methods

Species' occurrence data were drawn from the *Atlas of the Distribution of Mexican Birds* (Peterson et al., 1998). This database consists of ≈ 300 000 records—the Mexican holdings of 43 natural history museums (cited in Peterson et al. (1998)) covering 1060 species of birds. This data set was reduced to 103 species for which more than 200 records were available in the dataset.

Environmental data consisted of electronic maps made available by the Comisión Nacional para el Uso y Conocimiento do la Biodiversidad (CONABIO), including variables that have proven useful in modeling species' habitat needs and predicting distributions (Escalante-Pliego et al., 1993; Peterson et al., 1999). These coverages include (1) average annual precipitation (11

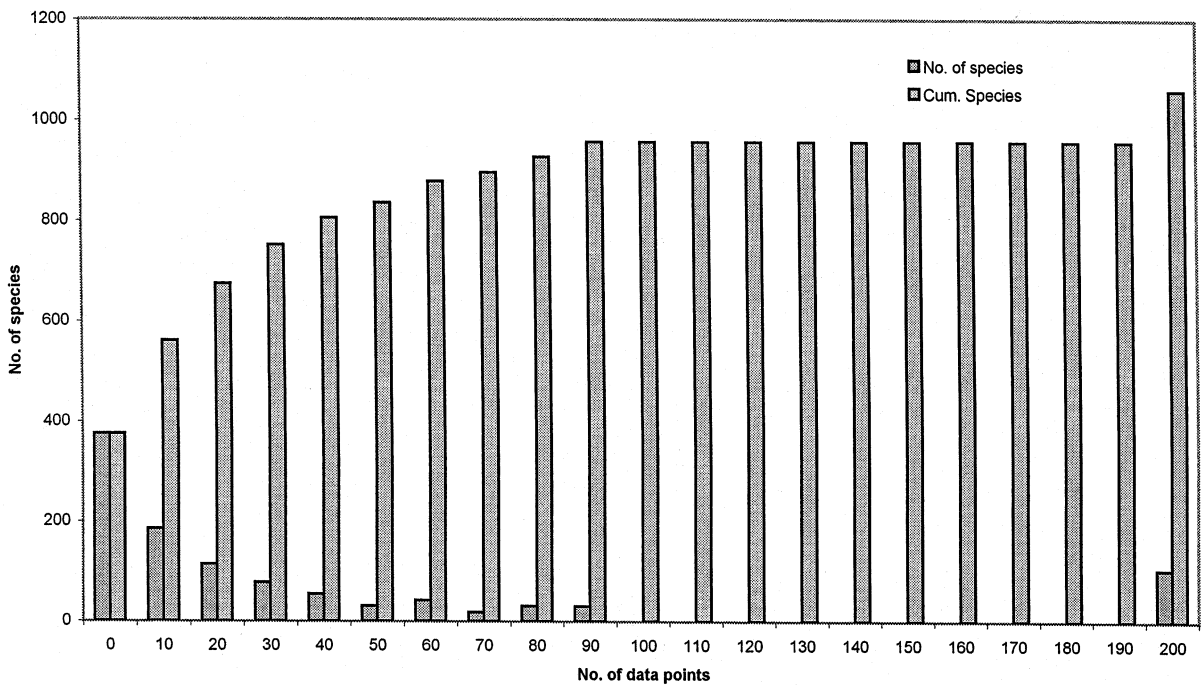

Fig. 1. Frequency of unique latitude–longitude points for species of birds in Mexico.

classes), (2) actual forest cover (14 classes), (3) elevation (11 classes), (4) potential vegetation (ten classes), (5) average annual temperature (six classes), (6) longitude (x-axis gradient), (7) latitude (y-axis gradient), and (8) proximity to coast. Although these data were originally provided with a $1' \times 1'$ grid cell size, they were reprocessed to $3' \times 3'$ grid cells to permit faster calculations; values of continuous variables (elevation, precipitation, rainfall) were averaged, whereas categorical variables used the dominant value in the nine cells.

Predictive models were developed from the species occurrence data using approaches implemented in the Biodiversity Species Workshop, an online resource for species modeling (http://biodi.sdsc.edu). The methods include multivariate methods, such as logistic regression, a genetic algorithm used in the GARP modeling system, as well as 'surrogate' methods, implemented as a Perl script for this study. While efforts have been made to provide a superior implementation, it is possible that any of the methods may be optimized further.

The logistic regression method, LM, is a generalized linear modeling method widely available in statistical packages, such as SPLUS. Variables are selected for inclusion in the model using a forward selection procedure, with a variable included, if it increases the accuracy of the model with significance $> 95\%$. The linear function is a second degree polynomial, consisting of a linear term and a quadratic term in each variable.

The GARP algorithm applies artificial intelligence modeling methods to develop predictive models that consist of a set of rules, or if-then statements, that describe the species' ecological niche (Stockwell and Peters, 1999; Stockwell, 1999). The rules represent a range of possible multivariate relationships between species' occurrence and environmental variables, including environmental ranges (a.k.a. BIOCLIM rules), logistic regression, and categorical rules.

The 'surrogate' method uses a single environmental variable. The surrogate is usually vegetation, although additional factors, such as climate or elevation, have been used (Gaston and Blackburn, 1995). To use the surrogate variable in a predictive approach, vegetation types are classed as favorable or unfavorable habitat based on field data or expert opinion, and predictions are made based on the mapped distribution of the vegetation classes. The surrogate method implemented in this study assigns classes based only on data analysis, but nonetheless provides a point of relative comparison with methods frequently used in biodiversity conservation.

We designed this study so that accuracy could be compared validly across methods and species. First, sets of 1, 3, 5, 7, 10, 20, 50 and 100 occurrence points were selected randomly from those available for a species. From each of those sets, a training set of 2000 points was created via sampling with replacement, composed of even proportions of occurrence points (presence points) and points from the area not known to be inhabited (background points), thus balancing commission and omission errors. To assess the accuracy of models, the remaining occurrence points were sampled with replacement to produce a test set of 1000 points, evenly balanced between presence and background points.

Model development using the GARP system is described elsewhere (Stockwell, 1999; Stockwell and Peters, 1999) and will not be described in detail herein. The philosophy of the system was to provide a robust methodology for developing and testing multiple models, to deal with the well-known problems of noisy and irrelevant variables, overfitting and mis-specification of models (Stockwell and Noble, 1992). In brief, models expressed as rules (if-then statements) are generated through an iterative evolutionary approach consisting of the stages: generation, mutation, testing, and selecting of the best rules for the next iteration. Rules are tested to maximize both significance and predictive accuracy on independent (or held-back) data. The rules include logistic regression models, environmental envelope models, and categorical models. The resulting set of best models, or rule set, can be projected onto geography to provide predictions of geographic distributions.

The models developed with surrogate methods are based on a single categorical variable representing land-cover classes. To develop the model, the number of training points in the each class is

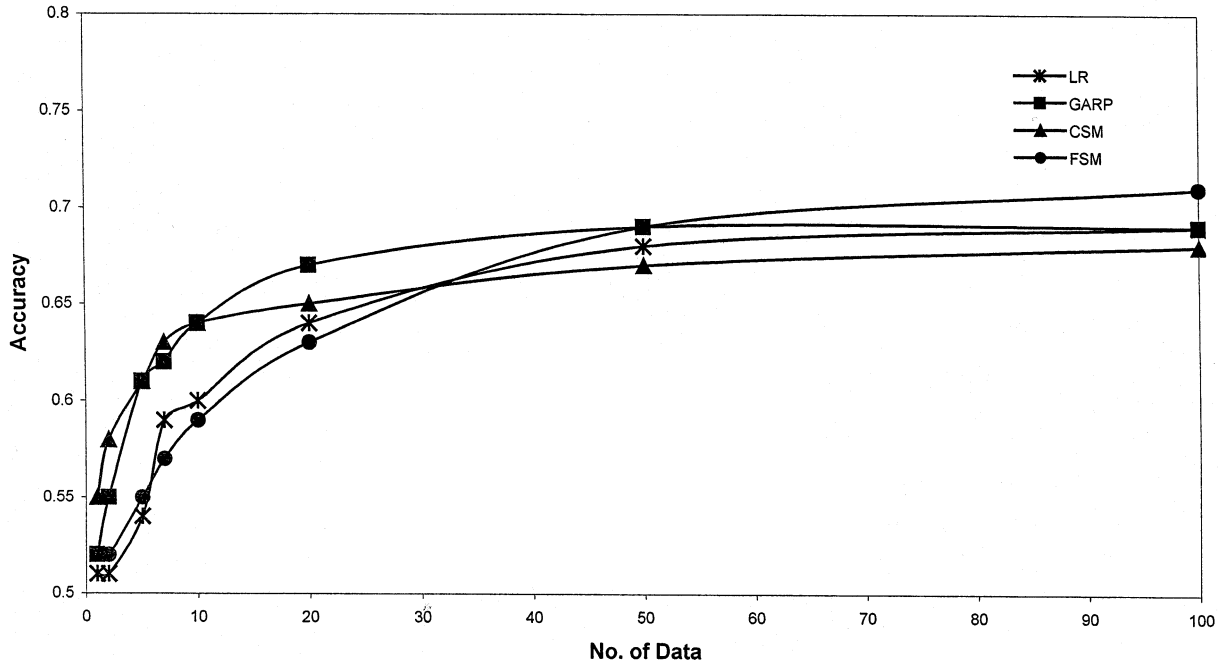

Fig. 2. General relationship between accuracy and sample size, across 103 species re-sampled for six sample sizes with the LR, FSM, CSM, and GARP methods.

determined, and a probability of occurrence calculated by dividing the number of presence points in a class by the total points in that class. This list of probabilities for each class constitutes the predictive model of occurrence, and can be used to create a probability map for the distribution of the species. Model accuracy is determined by examining each point in the test set, and predicting the species as present if the land-cover class into which it falls has a probability of >0.5; otherwise, absence is predicted. The criterion of 0.5 is used to optimize the accuracy of the surrogate model, given even proportions of presence and absence points in the test sets.

The surrogate method is sensitive to numbers of categories in the data: too few or too many produce poor accuracy (Scott et al., 1996). Based on preliminary testing of model accuracy, potential vegetation (ten classes) had highest accuracy, and was subsequently used as our coarse surrogate method (CSM). A more finely resolved land-cover layer was composed of a cross of vegetation, temperature, and precipitation (189 classes), and was used as our fine surrogate method (FSM).

We extracted two features of the accuracy-sample size relationship for each species and model: the maximum predictive accuracy achieved and the sample size at 90% of maximum accuracy. The former measure assesses maximum possible accuracy for a particular species, whereas the latter assesses, how many locality points are necessary to achieve that accuracy. Characteristics of species were assembled by Peterson independent of (and blind to) modeling results, based on personal experience, and partly on Howell and Webb (1995), including range size, use of primary habitat, ecological breadth, migratory behavior, and body size.

3. Results

Fig. 2 shows the accuracy of all methods over the range of data sets. For analyses with 1–2 data points, all methods were $<60\%$ accurate; for those based on ten points, accuracy ranges from 64% for GARP and CSM, to 59% for FSM and

LR. At 50 points, average accuracy value was around 68%. Increments in accuracy decreased with additional data at larger sample sizes, suggesting that most rapid improvement of model performance takes place below 20 data points.

Significant differences existed in the data efficiency of the modeling methods. Most noticeably, only ten data points were required by CSM and GARP to achieve the same accuracy as 20 data points in LR and FSM. Thus, CSM and GARP were twice as efficient at using small data sets. However, beyond 20 data points, accuracy of the methods crossed over. The result at 50 data points in decreasing order was GARP (0.69), FSM (0.69), LR (0.68), and CSM (0.67). CSM gave results closely matching GARP, up to ten data points, beyond which GARP produced higher accuracies. Despite low accuracies of FSM and LR on small data sets, accuracies continued to increase with more data, with FSM ultimately achieving the highest accuracy (0.71) and the remaining three methods achieving 0.69 at 100 data points.

The change in accuracy across different sample sizes averaged across all species for the training and test data sets is shown for LR and GARP (Fig. 3). The predictive accuracy of methods is clearly related to sample size. Accuracy on the training set decreased with larger sample sizes, while accuracy on the test set increased as sample sizes increased. As standard errors were consistent throughout the study (0.006), error bars are omitted on subsequent graphs for clarity.

The difference between test and training set accuracy indicates the degree of overfitting, which was greatest with small data sets, and decreased with larger sets. At all sample sizes, GARP showed less overfitting than LR, approaching zero overfitting at 100 data points. These graphs also show, how inaccurate estimates of accuracy on the training data are when sample sizes are small.

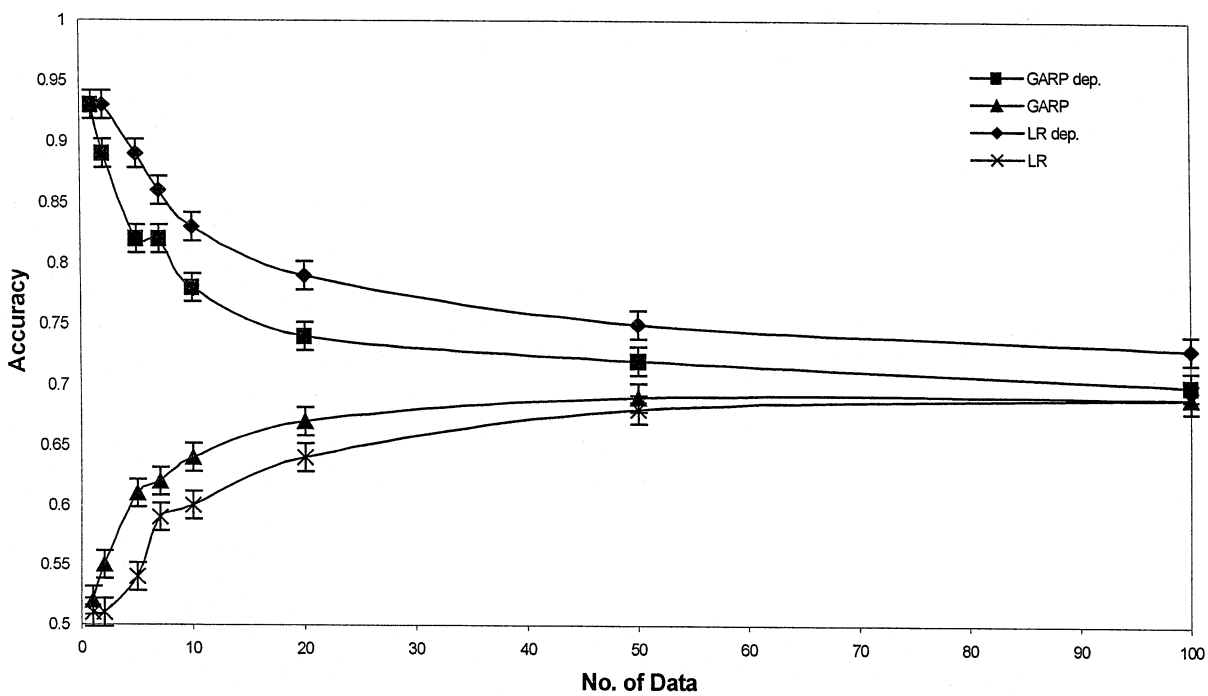

Fig. 3. Accuracy on test and training sets of two methods, GARP and LR showing convergence of accuracy (± standard error of mean). The difference between the accuracy for the test and training set shows degree of overfitting, which also decreases as the number of data increases. LR shows a greater degree of overfitting.

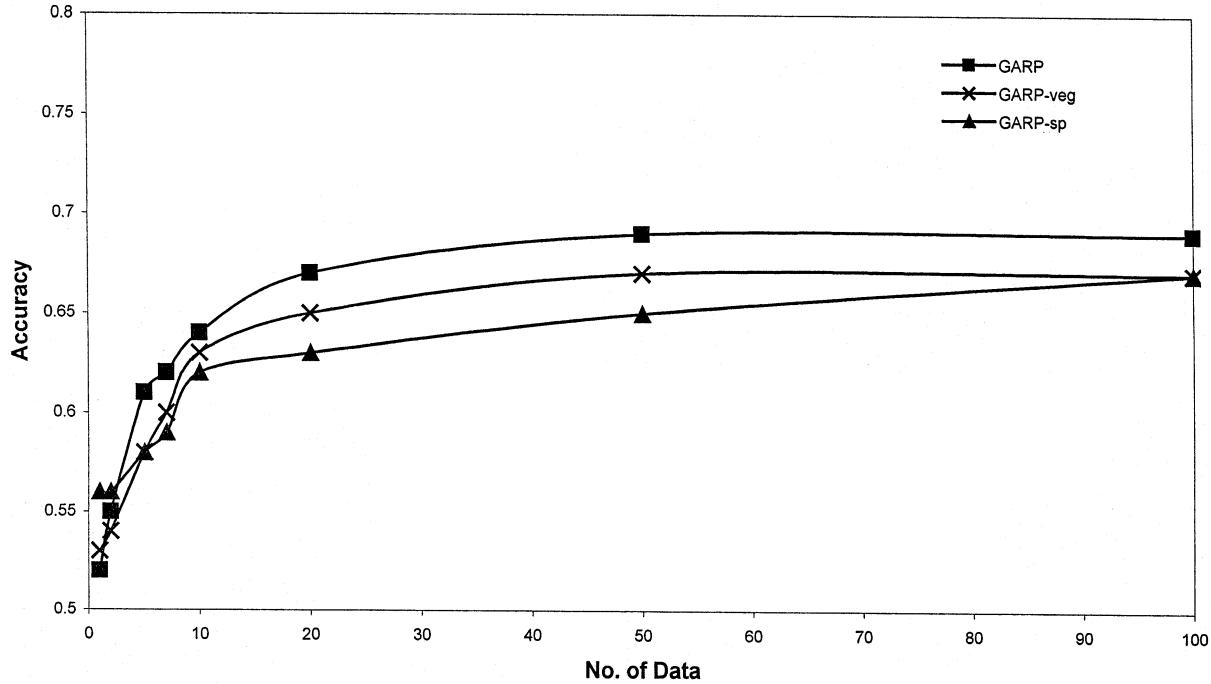

Fig. 4. Accuracy of GARP with three combinations of environmental variables, all, vegetation removed, and spatial variables removed.

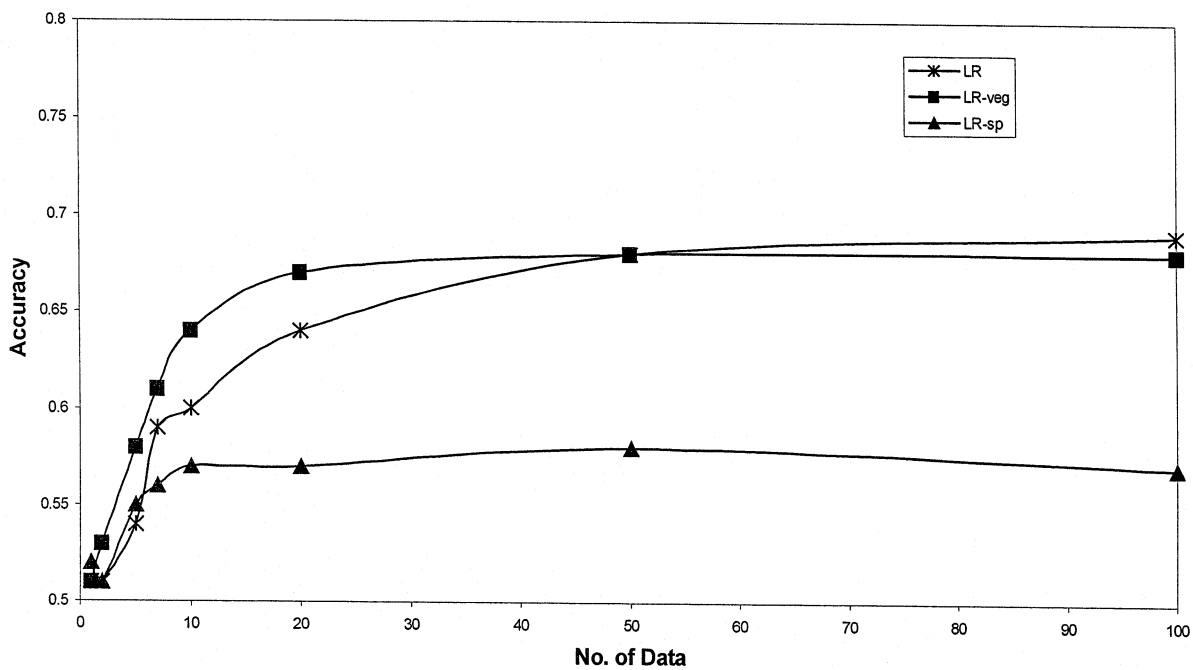

Fig. 5. Accuracy of LR with three combinations of environmental variables, all vegetation removed, and spatial variables removed.

3.1. What is the effect of environmental variables on accuracy?

To evaluate the effects of environmental variables on predictive accuracy at a range of sample sizes, we removed selected variables, developed models, and plotted the accuracy on the test sets. Our initial explorations of accuracy of individual layers had showed that vegetation and the spatial variables were important predictors. The result of removing these variables from GARP analyses is a small decrease in accuracy across the range of training set sizes (Fig. 4). The effect of removal of environmental variables on LR is very different, however (Fig. 5). Removal of vegetation increased accuracy at small sample sizes, but reduced accuracy at the maximum sample size. Removal of the spatial variables results in a large decrease in accuracy throughout the whole range. These results show that while LR and GARP achieve similar accuracy, GARP is much more robust in response to the environmental variables, and LR is more sensitive.

3.2. Are accuracies correlated with ecological factors?

Individual species' GARP models show a diversity of responses to sample size (Fig. 6), ranging from maximum accuracy with a few data points

Table 1
Summary of comparisons of model accuracy (maximum accuracy and minimum sample size needed to achieve 90% of maximum accuracy) with ecological characteristics

	Maximum accuracy			Minimum sample size		
Ecological characteristic	Slope	Intercept	R^2	Slope	Intercept	R^2
Range size	**−0.0341**	**0.8922**	**0.2530**	0.7273	23.29	0.0022
Use of primary habitat	0.0202	0.7835	0.0288	1.9177	22.727	0.0056
Ecological breadth	−0.0251	0.0907	0.0656	1.0710	25.98	0.0025
Migratory behavior	−0.0387	0.8183	0.647	0.4453	25.14	0.0002
Body size	0.0218	0.7767	0.0260	0.4380	24.40	0.0002

The sole regression that was statistically significant ($P<0.05$) is shown in bold.

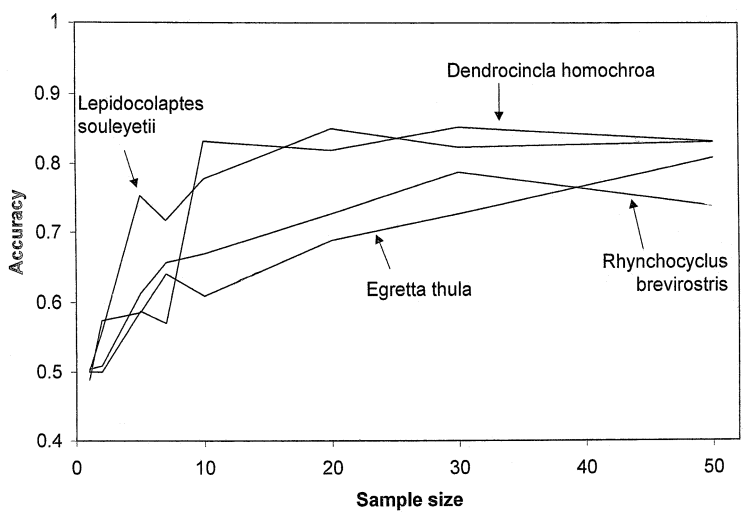

Fig. 6. Plots of accuracy for individual species showing different data efficiencies and variation among species.

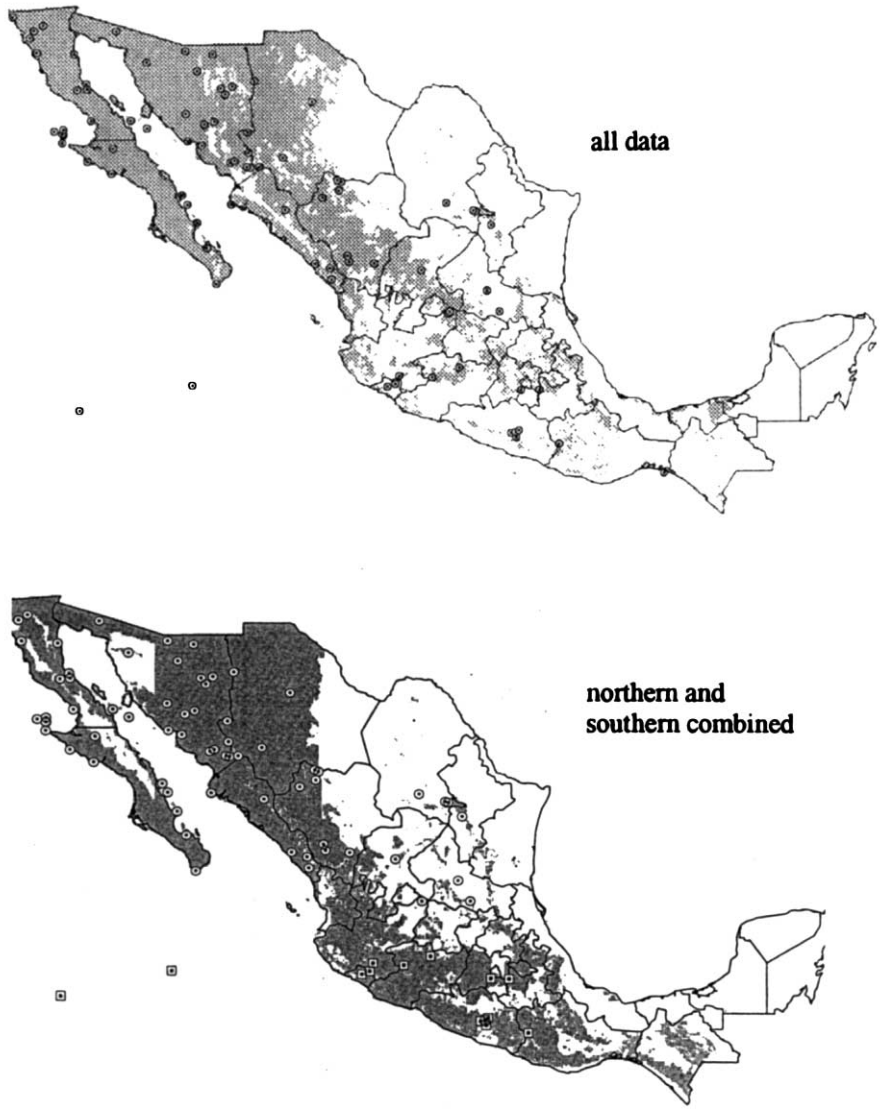

Fig. 7. Predicted distribution of *C. corax* based on (a) all data available combined, and (b) data subdivided into two artificial populations, north and south of 20° latitude.

to a smooth increase over the range of test data sizes. For example, models for *Lepidocolaptes souleyetii* and *Dendrocincla homochroa* reach asymptotes with just 10–20 points, whereas *Egretta thula* and *Rhynchocycus brevirostris* showed slower increases, not reaching 90% of maximum accuracy until sample sizes of 30–40 points.

These differences suggest that particular ecological characteristics might be identified that affect accuracy or efficiency of models. Hence, we compared maximum accuracy and sample size to reach 90% of maximum accuracy with range size, use of primary habitat, ecological breadth, migratory behavior, and body size. In a series of statistical tests (Table 1), interestingly, only one relationship was statistically significant: maximum predictive accuracy was not independent of range size: widespread species were modeled less accurately.

As a further exploration, we analyzed a single poorly predicted species (*Corvus corax*) in more detail. This species is widespread in mountains and desert areas in northern Mexico, but occurs principally in mountains in eastern and southern Mexico (Fig. 7). The overall GARP model had a low accuracy (0.68). However, partitioning the data north and south of 20° latitude, and analyzing them separately with GARP, the northern model achieved an accuracy of 0.70, and the southern model 0.87. This preliminary result demonstrates that model accuracy can be improved by splitting species into sub-units that may have different ecological characteristics.

4. Discussion

This study assessed sample size requirements for modeling geographic distributions of Mexican bird species using logistic regression, surrogate methods and the GARP modeling system. The general conclusion was that GARP and CSM require on average about ten data points to achieve 90% of maximum accuracy. This result, with relatively small data requirements, is promising for applications of multivariate models of geographic distributions. In general, these approaches then provide a powerful tool for predicting geographic distributions of species based on incomplete data.

The multivariate GARP method was as accurate as other methods in most cases, with accuracy matching or exceeding surrogate and logistic regression methods except for the FSM model at 100 data points. This difference is likely the result of the greater power of multivariate modeling and multiple environmental data layers. The greater power of GARP derives from the combination of three methods: logistic regression, environmental envelope, and categorical models, of which the best models are used. The surrogate method, in contrast, relies on a single model, and is limited to categorical variables. Since GARP has the advantage of developing models using both categorical and continuous variables, it has increased power and takes advantage of more information.

While the FSM ultimately had the greatest accuracy, it performed poorly at lower data sample sizes. For an explanation of this behavior, consider the FSM model as a fine grid of map areas. With few data, few grid areas are marked as potential habitat by the chance occurrence of presence data points falling inside them. The model will incorrectly classify many presence points as absence (high omission error). As the number of training data increase, more areas will be marked as potential habitat, until all are correctly identified.

As the number of classes is decreased in surrogate models (compare FSM with CSM), the data fall into fewer land-cover classes, and therefore fewer data are required to specify all possible map areas. However, with adequate data, the few larger map areas produce models with lower accuracy, as too many absence data fall into the presence classes (high commission error). Alternatively, too many classes leads to high omission error. The number and shape of land-cover classes must be 'just right', and is the major challenge in implementing surrogate methods accurately. In the case of many species, it is not clear that a coarse land-cover classification, in which a species would fit in one or two classes, would be accurate (Peterson and Vargas, 1993).

LR was one of the least accurate methods at lower sample sizes, but had an accuracy equal to GARP at the maximum sample size of 100. Overfitting has been found to be a significant problem in logistic regression that could be reduced by stringent significance tests in stepwise algorithms (Pearce and Ferrier, 2000). As shown by removal of the vegetation variable (Fig. 5), logistic regression is also sensitive to the nature of the environmental variables, even in a stringent stepwise algorithm. LR does not make good use of categorical variables, such as vegetation type. In fact, exclusion of this variable produces a data efficiency curve almost equal to that of GARP. Alternative, approaches to incorporating categorical variables, such as decomposing each category into a binary-valued variable may reduce the problem. However, the decomposition algorithm would need to be incorporated into the modeling algorithm to allow modeling on arbitrary environmental data.

Our results for LR show that fewer data points are needed to achieve maximum accuracy in modeling species distributions than suggested by the logistic regression modeling results of Pearce and Ferrier (2000). We suggest that the lower data efficiency achieved in their study may be due to the set environmental variables used. However, no information is provided about the variables used in their study.

4.1. Species' geography

An interesting result of this study is that widespread species are modeled less accurately. Two explanations may account for this effect, one methodological and one biological. In methodological terms, because model fit is based on a model fitness measure developed from points resampled from presence and background data, this measure depends on the prior probabilities of presence in the two data sets being distinct. Ubiquitous or widely distributed species, being actually present (but unsampled) in most grid cells, thus have prior probabilities in the presence and background data sets that do not differ markedly, and thus accuracy is reduced.

Another phenomenon that likely decreases accuracy for widespread species is that of local ecological adaptation. Widespread species often show local or regional differences in ecological characteristics (Peterson and Vargas, 1993). Modeling all of these sub-populations together would effectively overestimate the species' ecological breadth, and reduce model accuracy as a consequence. The preliminary demonstration based on detailed analysis of *C. corax* developed herein, as well as more detailed analyses in process, suggest that these local adaptational effects may play a significant role in depressing model accuracy for some species.

5. Conclusions

We present yet another test of the ability of the GARP modeling system to produce accurate models of species' ecological requirements and geographic distributions (Peterson and Cohoon, 1999; Peterson et al., 1999; Joseph and Stockwell, 2000; Stockwell and Peterson, 2002). We have shown that GARP is as accurate as alternative methods over the range of sizes of data sets and types of environmental data frequently used in biodiversity studies. In addition, it can make use of categorical or continuous data, as shown by its robust response to removal of variables. The robustness stems from the importance the system gives to robust methodology in its original development, recognizing that accurate prediction in this domain must not be constrained by excessively limiting assumptions, but must search a wide variety of possible models, and then test those models using re-sampling to avoid overfitting (Stockwell and Noble, 1992).

We find that some methods can develop accurate models based on relatively few data points. CSM and GARP are superior to FSM and LR in this regard, with approximately half as many data points required to achieve the same level of accuracy. The implication of increased data efficiency is that more species can be modeled accurately using existing data. Referring to the Mexican *Atlas* data set (Fig. 1), we find that 50 points are available for 20% of species in the country, 20 points for 39% of species, and ten points for 65% of species. Hence, for Mexico, which is likely an excellent example for other countries, 65% of species could be modeled to within 90% of maximum possible predictive accuracy, and 39% to within a 3% of maximum possible accuracy. This predictive power suggests that due to the capacity to model accurately over a wide range of environmental data sets and sample sizes, GARP offers significant advantages in predicting patterns and species' distributions in biodiversity programs.

Due to the large differences in size/accuracy and robustness relationships between different methods, these criteria should be among those used for selecting predictive algorithms for species distribution modeling (Guisan and Zimmermann, 2000). This study has also helped to resolve some outstanding questions in using museum data to reduce the costs of species distribution modeling, including choice of methodology and the expected accuracy given the number of data. Further work will seek to address similar questions in the envi-

ronmental data sets used for environmental correlates, such as the optimal number and identity of variables, and effects of scale.

Acknowledgements

Environmental data layers for this study were kindly made available by the Comisión Nacional para el Uso y Conocimiento de la Biodiversidad (CONABIO). Species distributional data were generously provided by numerous natural history museums (listed in Peterson et al. (1998)), and facilitated greatly by Dr Adolfo Navarro-Sigüenza. Funding for this work was provided in part by the National Science Foundation.

References

Austin, M.P., Meyers, J.A., 1996. Current approaches to modelling the environmental niche of eucalypts: implications for management of forest biodiversity. Forest Ecology and Management 85, 95–106.

Boston, A.N., Stockwell, D.R.B., 1995. Interactive species distribution reporting, mapping and modeling using the World Wide Web. Computer Networks and ISDN Systems 28, 231–238.

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

Davey, S.M., Stockwell, D.R.B., 1991. Incorporating habitat into an artificial intelligence framework: concepts, theory and practicalities. A.I. Applications in Natural Resource Management 5, 59–104.

Escalante-Pliego, P., Navarro-Siguenza, A.G., Peterson, A.T., 1993. A geographic, historical, and ecological analysis of avian diversity in Mexico. In: Ramamoorthy, T.P., Bye, R., Lot, A., Fa, J. (Eds.), The Biological Diversity of Mexico: Origins and Distribution. Oxford University Press, New York.

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.

Gaston, K.J., Blackburn, T.M., 1995. Mapping biodiversity using surrogates for species richness: macro-scales and New World birds. Proceedings of the Royal Society of London B 262, 335–341.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147–186.

Howell, S.N.G., Webb, S., 1995. A Guide to the Birds of Mexico and Northern Central America. Oxford University Press, Oxford.

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., 2000. Temperature based models of the migration of Swainson's flycatcher *Myiarchus swainsoni* across South America: a new use for museum specimens of migratory birds. Proceedings of the Academy of Natural Sciences 150, 293–300.

Kaiser, J., 1999. Searching museums from your desktop. Science 284, 888.

Kelly, K.T., Glymour, C., 1990. Theory discovery from data with mixed quantifiers. Journal of Philosophical Logic 19, 1–33.

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

Pearce, J., Ferrier, S., 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecological Modeling 128, 127–147.

Pennisi, E., 2000. Taxonomic revival. Science 289, 2306–2308.

Peterson, A.T., Vargas, N., 1993. Ecological diversity in Scrub Jays, *Aphelocoma coerulescens*. In: Ramamoorthy, T.P., Bye, R., Lot, A., Fa, J. (Eds.), The Biological Diversity of Mexico: Origins and Distribution. Oxford University Press, New York, pp. 309–317.

Peterson, A.T., Soberón, J., Sánchez-Cordero, V., 1999. Conservatism of ecological niches in evolutionary time. Science 285, 1265–1267.

Peterson, A.T., Navarro-Siguenza, A.G., Benitez-Diaz, H., 1998. The need for continued collecting: a geographic analysis of Mexican bird specimens. Ibis 140, 288–294.

Peterson, A.T., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling 117, 159–164.

Scott, J.M., Tear, T.H., Davis, F.W., 1996. Gap analysis: a landscape approach to biodiversity planning. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland. ISBN-1-57083-03603.

Stockwell, D.R.B., 1993. LBS: Bayesian learning system for rapid expert system development. Expert Systems with Applications 6, 137–147.

Stockwell, D.R.B., 1997. Generic predictive systems: an empirical evaluation of the learning base system as an automated predictive system. Expert Systems with Applications 12, 301–310.

Stockwell, D.R.B., 1999. Genetic algorithms II. In: Fielding, A.H. (Ed.), Machine Learning Methods for Ecological Applications. Kluwer Academic Publishers, Boston, pp. 123–144.

Stockwell, D.R.B., Noble, I.R., 1992. Induction of sets of rules from animal distribution data: a robust and informative method of analysis. Mathematics and Computers in Simulation 33, 385–390.

Stockwell, D.R.B., Peters, D.P., 1999. The GARP modeling system: problems and solutions to automated spatial prediction. International Journal of Geographic Information Systems 13, 143–158.

Stockwell, D.R.B., Peterson, A.T., 2002. Controling bias during predictive modeling with museum data. In: Scott, J.M., Heglund, P.J., Morrison, M., Raphael, M., Haufler, J., Wall, B. (Eds.), Predicting Species Occurrences: Issues of Scale and Accuracy. Island Press, Covello, California.

Verbyla, D.L., 1986. Potential prediction bias in regression and discriminant analysis. Canadian Journal of Forest Research 16, 1255–1257.

Verbyla, D.L., Litvaitis, J.A., 1989. Resampling methods for evaluating class accuracy of wildlife habitat models. Environmental Management 13, 783–787.

Vieglais, D.A., Stockwell, D.R.B, Cundari, C.M., Beach, J., Peterson, A.T., Krishtalka, L., 1998. The species analyst: tools enabling a comprehensive distributed biodiversity network. Biodiversity, Biotechnology & Biobusiness, 2nd Asia-Pacific Conference on Biotechnology, 23–27 November, Perth, Western Australia.

Williams, S.E., Hero, J.M., 2001. Multiple determinants of Australian tropical frog biodiversity. Biological Conservation 98, 1–10.