Climate-Envelope Predictions of Animal Ranges: Model Differences and Model Reliability

In review: *Global Change Biology*

Joshua J. Lawler
Department of Zoology, Oregon State University
c/o U.S. Environmental Protection Agency
200 SW 35th St.
Corvallis, OR 97333 USA
lawler.joshua@epa.gov
phone: 541.754.4834
fax: 541.754.4799

Denis White
U.S. Environmental Protection Agency
Corvallis, OR 97333 USA

Ronald P. Neilson
U.S. Forest Service
Corvallis, OR 97331 USA

Andrew R. Blaustein
Department of Zoology
Oregon State University
Corvallis, OR 97331 USA

Running head: *Modeling climate-induced range shifts*

Key words: *climate change, climate-envelope models, geographic range, model prediction, random forest predictors, model averaging, extinction*

Date received:

Abstract

Predicted changes in the global climate are likely to cause large shifts in the geographic ranges of many plant and animal species. To date, predictions of future range shifts have relied on a variety of modeling approaches with different levels of model accuracy. Using a common data set, we investigated the potential implications of alternative modeling approaches for conclusions about future range shifts and extinctions. Our common data set entailed the current ranges of 100 randomly selected mammal species found in the western hemisphere. Using these range maps, we compared six methods for modeling predicted future ranges. Predicted future distributions differed markedly across the alternative modeling approaches, which in turn resulted in estimates of extinction rates that ranged between 0 and 7%, depending on which model was used. Random forest predictors, a model-averaging approach, consistently outperformed the other techniques (correctly predicting > 99% of current absences and 86% of current presences). We conclude that the types of models used in a study can have dramatic effects on predicted range shifts and extinction rates; and that model-averaging approaches appear to have the greatest potential for predicting range shifts in the face of climate change.

INTRODUCTION

Global temperatures have risen an average 0.6°C over the past century (Houghton *et al.*, 2001). Recent studies suggest that this climate change has caused shifts in the geographic ranges of both plants and animals (Parmesan & Yohe, 2003; Root *et al.*, 2003). Given that average global temperatures are predicted to rise between 1.4°C and 5.8°C over the next century (Houghton *et al.*, 2001), it is likely that many species will undergo dramatic range shifts in the future. To anticipate the effects of climate change, and to identify conservation strategies that might mitigate the undesirable consequences of climate change, it is essential that we develop models that link the distributions of species to alternative scenarios of climate change.

Several studies have attempted to predict future range shifts, often with the goal of estimating climate-induced extinction rates (Williams *et al.*, 2003; Thomas *et al.*, 2004). Most predictions of future species distributions rely on what are commonly called climate-envelope models. These models attempt to relate species geographic distributions to a set of climatic factors. Relatively simple climate variables are used to define the abiotic conditions, or "climate envelope" in which a species exists. Predicted future climate variables, usually derived from a general circulation model (GCM), are used as input for these models to predict future distributions. Unfortunately, researchers have reported large uncertainties and error rates in these climate-envelope predictions, and we have little understanding of which, if any of these modeling approaches is most reliable (Segurado & Araújo, 2004; Thuiller, 2003).

In this paper, we report on a systematic comparison of all the major approaches to predicting range shifts with a common data set and common metrics for estimating error rates. Our goal was to quantify the types of errors associated with range-shift models and to determine whether any approach clearly outperforms the alternatives. The approaches we examined were: generalized linear models (McCullagh & Nelder, 1989), classification trees (Breiman *et al.*, 1984), generalized additive models (Hastie & Tibshirani, 1990), random forest predictors (Breiman, 2001), artificial neural networks (Ripley, 1996), and genetic algorithms for rule-set prediction (GARP, Peterson *et al.*, 2002).

METHODS

We compared the model accuracy and the future predictions of the six different modeling approaches described below by applying each approach to100 randomly selected mammal species in the western hemisphere. All analyses were conducted on a 50x50-km resolution grid consisting of 15,323 cells. Current species distributions were based on digital range maps (Patterson *et al.*, 2003). We selected the 100 species at random from 1,022 mammals with ranges occupying at least 50 grid cells.

Current climate data were derived from average monthly precipitation and temperature values from 1961-90 for the land surface of the globe at 0.5° resolution (Leemans & Cramer, 1991). For that 30-year period, we calculated mean annual temperature, average temperature of the hottest and coldest months, and degree-days over 5°C. We also calculated average yearly precipitation as well as precipitation in the hottest, coldest, wettest, and driest months.

Current land cover was derived from both predicted current potential vegetation and measured land cover derived from Advanced Very High Resolution Radiometer (AVHRR) satellite data (Loveland *et al.*, 1999). Predicted current vegetation types were produced using the Mapped Atmospheric-Plant-Soil System (MAPSS) model (Neilson, 1995). Although measured vegetation provides a more accurate representation of current vegetation, we chose to use the predicted current vegetation to best correspond with the classification of predicted future vegetation for the years 2061-2090. MAPSS predictions of current potential vegetation have been shown to closely approximate other potential vegetation classifications (Bachelet *et al.*, 2001). We overlaid the 44 land-cover classes of predicted potential current vegetation from the MAPSS model with five agriculture classes and one urban and suburban land-cover class from the AVHRR-derived land-cover data to produce the new 50-class land-cover data set used for building the models.

Predicted future climate data were produced using the Hadley Climate Centre's HADCM2SUL model (Johns *et al.*, 1997) using Intergovernmental Panel on Climate Change (IPCC) predicted future greenhouse gas contributions (IS92a) for the years 2061-2090 (Kattenberg *et al.*, 1996). This model and greenhouse gas contribution scenario together generally predict larger increases in precipitation and smaller increases in temperature (particularly for North America) than do more recent models. Although a wide array of more recent GCM predictions based on alternative emissions scenarios exist, the purpose of our study was not to draw conclusions about the future, but to compare the differences in predictions resulting from using different climate-envelope modeling approaches. Using the future climatic predictions, we calculated the same set of nine climate variables for all 0.5° grid cells. Predicted future land cover was produced with the MAPSS model using the predicted climate data for input. For the purposes of these analyses, we assumed no change in the distribution of agriculture and urban-suburban areas. We overlaid the predicted future potential vegetation data with the current agriculture and urban-suburban data to produce predicted future land cover. All data compiled at 0.5° resolution were projected to the 50-km resolution grid.

Modeling Approaches

For all six modeling approaches, we used the presence and absence of a species as the response variable and the set of nine continuous variables representing current climate and one categorical variable representing the 50 land-cover classes as predictors. All models except the GARP models were built using the R software package (version 1.9.1). For all 100 species, we selected a training- and a test-data set. For the training set, we randomly selected 80% of all species presences and 80% of all species absences. For each species, we then used the remaining 20% of the data for testing the models and determining their errors in terms of absences falsely predicted as presences (commission error), and presences falsely predicted to be absences (omission error).

Generalized Linear Models

Generalized linear models offer a slightly more flexible modeling framework than basic linear regression models as they allow for the modeling of alternative distributions in the response variable and non-constant variance functions (Guisan *et al.*, 2002). We built logistic regression models (generalized linear models with an assumed binomial error distribution) using a combined backward- and forward-stepwise selection process. Variable inclusion was based on Akaike's information criterion (Chambers & Hastie, 1991). We modeled all linear and second order polynomials of the climatic predictor variables. Because the test-data sets for 21 species contained land-cover classes that were not found in the training sets of those species, we chose to drop the land-cover variable from the models for these species.

Classification Tree Models

Classification trees, and regression trees, their counterpart for analyzing continuous response variables, are non-parametric modeling approaches (Venables & Ripley, 2002; Breiman *et al.*, 1984). Both techniques involve the recursive binary partitioning of data. Each split of the data is made using the predictor variable and the point along that variable's distribution that divides the data into the two most homogeneous groups with respect to the response variable. The result is a tree like-structure with one root node and a number of terminal nodes. In a classification tree, the proportional class-membership of the observations in a terminal node form the basis for predicted probabilities. De'ath and Fabricius (2000) provide excellent examples of the use of tree-based models for ecological analyses. We fit classification trees using the RPART package in R originally designed for S-Plus (Therneau & Atkinson, 1997). Because most trees tend to over-fit the data, we selected the optimal tree size using the modal size suggested by 50 10-fold cross-validations applying a 1-SE rule (De'ath & Fabricius, 2000).

Generalized Additive Models

Generalized additive models (GAM) are similar to generalized linear models, but they are more flexible because they do not require a specific response curve to be fit to the predictor variables (Hastie & Tibshirani, 1990). Smoothing functions allow datadriven response curves to be fit for each predictor variable. We fit generalized additive models using penalized regression splines (Wood & Augustin, 2002). This approach takes advantage of generalized spline smoothing (Wahba, 1990) but can be equally or less computationally expensive than backfit GAMs. To increase the speed of the modeling process, we pre-screened each variable by fitting a GAM model for that variable alone. We dropped all variables for which the fitting algorithm was unable to converge. Variable selection for those variables included in the modeling process was based on smoothness penalties in conjunction with a shrinkage parameter. Variables were effectively dropped from a model based on the fit smoothing parameter. We used the MGCV package in R to fit all GAM models (Wood & Augustin, 2002). As for the generalized linear models, we did not include the categorical land-cover variable in the models built for the 21 species for which the test-data set contained land-cover classes not found in the training-data set.

Random Forest Predictors

Random forest predictors are a model-averaging approach based on regression or classification trees (Breiman, 2001). Instead of building one tree model, the random forest algorithm builds multiple trees using randomly selected subsets of the observations and random subsets of the predictor variables. The predictions from the trees are then averaged (in the case of regression trees) or tallied using a voting system (for classification trees). We used the R package RandomForest to build random forest predictors. As part of the random forest procedure, 500 classification trees were built for each species. To build each tree, 12,258 observations were selected at random, with replacement, from the training set. For each split in these trees, three predictor variables were selected at random from the full set of 10 predictor variables as candidates for that particular split.

Artificial Neural Networks

Artificial neural networks are a machine-learning approach based on real neural networks (Ripley, 1996). The networks are composed of a series of interconnected nodes (neurons) which receive and process input signals and potentially generate output signals. A network is trained on a data set to recognize the patterns in the data. We built artificial neural networks using the NNET package in R which was based on the S-Plus package NNETW (Venables & Ripley, 2002). These feed-forward networks had one hidden layer with eight nodes. To train the network, we used 5000 presence and 5000 absence observations selected at random, with replacement, from the training-data set. Trial and error determined that these 10,000-observation data sets were most effective and efficient for training the networks. To produce more robust predictions, we built 10 networks for each species and averaged the model predictions (Thuiller, 2003; Segurado & Araújo, 2004).

Genetic Algorithms for Rule-Set Prediction

GARP is a machine learning-based approach that uses a genetic algorithm (a stochastic optimization technique) to assemble a set of rules to define a species' range (Stockwell & Noble, 1992). The approach was developed expressly for predicting species distributions. The rules used by the GARP algorithm include logistic relationships, climate envelopes (Nix, 1986), and simple Boolean rules. We used the Unix version of GARP to build 500 models for each species. All models were selected from all rule-types. GARP limits model training sets to 2500 observations. For each of the 500 models, we selected 1250 presences and 1250 absences, with replacement, from the training-data set for the given species. For each species, we used Cohen's Kappa statistic (Monserud & Leemans, 1992), calculated using the training-data set, to select the 10 best performing models from the set of 500 models. We combined the binary predictions of these 10 models to produce a predicted probability of presence.

Model Comparisons

Using the reserved test-data set, we computed four different metrics to compare the performance of the six different modeling approaches. The first three of these approaches included the percentage of the presences correctly classified, the percentage of the absences correctly classified, and Cohen's Kappa. Because all six modeling approaches produced predicted probabilities, calculating these three metrics required selecting a threshold with which to classify predicted presences and absences. We used receiver-operating characteristic (ROC) curves to select the optimal threshold, assuming that predicting presences correctly was twice as important as predicting absences correctly (Fielding & Bell, 1997). This is a conservative approach and should generally reduce the chances of overestimating future range contractions. In addition to the three metrics listed above, we used the area under the ROC curve (AUC) to provide an assessment of model performance that was independent of a specific classification threshold (Fielding & Bell, 1997). We compared model performance across model types using Wilcoxon's signed-ranks tests with a Holm correction for conducting multiple tests (Holm, 1979).

Future Predictions

We used the models to predict future geographic ranges under two alternative dispersal scenarios. First, we assumed that a species would be able to completely disperse into any new geographic range. For the second scenario, we assumed that a species would be unable to disperse from its current range. These two extreme assumptions have been made in several recent studies with which we wish to draw comparisons (Peterson *et al.*, 2002; Thomas *et al.*, 2004). Realistic future range shifts are likely to fall somewhere between these two extremes.

RESULTS

How did alternative modeling approaches affect the types of error and uncertainty in our analyses? The amounts and types of error were markedly influenced by which approach was used to predict range shifts (Table 1). The most significant consistencies in model performance were the over-prediction of current presences (commission error) by the neural networks and GARP models, the under-prediction of current presences (omission error) by the classification tree models, and the small number of errors predicted by the random forest models. For example, classification trees often incorrectly predicted current presences (median of 56% correct). This is a higher rate of omission error than produced by the other five approaches (medians of 69% to 86% correct presences). GARP models tended to have higher commission error rates than the other approaches, correctly predicting 96% of test-set absences compared to correct prediction rates of between 98% and 100% of absences for the other types of models. The spatial patterns of both commission and omission errors also differed across the six modeling approaches (e.g., Fig. 1). Whereas the commission errors of the GARP models and artificial neural networks tended to be relatively widely distributed, the few errors that the random forest models produced were generally clustered tightly around the area occupied by the species (Fig. 1).

We also found that different modeling approaches produced dramatically different predictions of future range shifts for many species. Not surprisingly, these differences were heavily influenced by assumptions regarding dispersal. On average, if one assumes no dispersal, so that species cannot move to occupy newly created favorable "climate space", only 19% of the cumulative future range of a species was similarly predicted by all six models. The percent agreement was even lower (11%) when full dispersal (species can fully exploit new favorable climate space that arises in the future) was assumed. For example, for the black tufted-ear marmoset (*Callithrix penicillata*), assuming unlimited dispersal, the generalized linear model and classification tree predicted contractions of 70% and 58% of the current range, respectively, whereas the artificial neural network and the GARP model respectively predicted expansions of 180% and 53% of the range (Fig. 2). These differences translated into different estimates of overall range contractions and extinction rates as predicted by the alternative modeling approaches (Fig. 3). When we assumed unlimited dispersal, classification trees predicted range contractions of over 50% for 36% of the species in the study compared to neural networks and GARP models which respectively predicted similar range contractions for

16% and 17% of all species. Becasue these models are often used to predict extinction rates, it is worth noting that depending on the modeling approach used, extinction rates ranged from 0% to 7% assuming unlimited dispersal and from 6% to 14% assuming no dispersal.

All of the differences among models would be daunting were it not for the finding that one modeling approach clearly performed better than all of the alternatives. In particular, random forest models had the highest median performance scores across all four measures of model accuracy (Table 1), and were consistently ranked the best performing of the six model types (Fig. 4). Random forests were the best performing models with respect to AUC and Kappa for 88% of species.

DISCUSSION

There are several different approaches to predicting changes in species distributions as a result of climate change (Thomas *et al.*, 2004; Meynecke, 2004; Iverson & Prasad, 1998; Araújo *et al.*, 2004; Pearson *et al.*, 2002; Shafer *et al.*, 2001). With few exceptions, previous studies have found very little consistency in the performance of these alternative approaches (Moisen & Frescino, 2002; Thuiller, 2003; Robertson *et al.*, 2003; Segurado & Araújo, 2004). We have found similar inconsistency among models, but are able to understand why these differences arise. The six modeling techniques that we applied make different assumptions about the relationships between species and their environments (Guisan & Zimmermann, 2000). Because different species can have qualitatively different relationships with their environments, many modeling approaches work differently for different groups of species (Segurado & Araújo, 2004). For example, generalized linear models assume a given response curve that defines the relationship between the probability of presence and various environmental gradients. These models will generally work well for species with relatively simple relationships to environmental gradients. The other five techniques that we tested are more flexible with respect to the complexity of the relationships that they can model. For example, generalized additive models allow for complex relationships with individual variables to be modeled. They are not, however, as adept at modeling complex interactions between variables as are classification tree models or random forests. Artificial neural networks and GARP models, the two machine-learning based approaches tested here, are in part an attempt to model both complex relationships with individual variables and complex interactions among those variables.

This inconsistency among range shift models has led some to suggest innovative methods for addressing model uncertainty in the prediction of future range shifts (Thuiller, 2003; Thuiller *et al.*, 2004). Another approach to reducing uncertainty is to ask whether some models might simply perform better than others, and hence we need not consider all of their predictions. Pursuing that strategy, our study compares essentially the full suite of climate-envelope approaches with a common data set, several metrics of model performance, and alternative assumptions about dispersal. The lessons are clear. First, random forest predictors, which averaged the predictions of hundreds of models, were consistently the best performers, and for the data we examined, performed remarkably well. They achieved error rates of less than 15% for presences and less than 1% for absences. We are aware of only one other study that has compared the performance of random forest predictors to other models for use as climate-envelope models. Prasad *et al.* (in press) found that random forest models and bagging (another tree-based model-averaging approach) consistently produced better predictions than multivariate adaptive regression splines and regression trees for predicting the distributions of four tree species. The performance of each of the other five modeling approaches tested here, but not by Prasad *et al.*, is generally comparable to the performance of models of the same type tested elsewhere (Thuiller *et al.*, 2003; Pearson *et al.*, 2004; Segurado & Araújo, 2004).

Our results raise the obvious question of why random forest models work so remarkably well. The strength of this approach likely lies in the power derived from averaging hundreds of different models (Breiman, 2001). The individual models are built with randomly selected subsets of the data and randomly selected subsets of the predictor variables. Although we averaged 10 artificial neural networks and 10 GARP models to produce predictions for each species, the model averaging accomplished by random forest predictors is much more comprehensive. Although it is possible that model-averaging applied similarly to techniques other than the classification trees on which random forests are based would produce models of comparable accuracy, the tree-based models themselves provide added advantages over other modeling approaches. In addition to providing a method for modeling complex interactions without having to specify them *a priori*, tree-based models allow the relationships between the response and the predictors to vary over the domain of the study. This is particularly advantageous for modeling data that cover large and diverse geographic areas.

The second lesson to be taken from our study is that the different modeling approaches tend to be relatively consistent in the types of errors they make. For example, classification trees produced the most omission errors whereas GARP models had the highest commission error rates. These errors, in turn, lead to different predicted range shifts, extinction rates, and changes in species composition at specific sites. Although there is no assurance that the model that best fits the current data will be the best model for predicting future distributions, it is likely that minimizing known errors in the current predictions will reduce the total amount of error in future predictions.

Lastly, the models differed greatly in the extent to which they predicted shrinking ranges versus expanding ranges in the face of climate change. For example, when we assumed unlimited dispersal, classification tree models predicted extinctions for 7% of the species compared to GARP models, which predicted no extinctions. Similarly, Thuiller *et al.* (2004) demonstrated potential differences in predicted extinction rates across modeling approaches ranging from less than 1% to roughly 5% over a 50 year period. The uncertainties in future range predictions that can be attributed to the errors in the climate-envelope models currently in use are likely to be greater than the uncertainties of actually predicting the underlying climate change (i.e., the differences among climate models and emissions scenarios) (Thuiller, 2004). This means that unless we can produce more accurate range shift models, they cannot really be used to compare the consequences of different carbon emissions scenarios. Looking forward, it appears that random forest models or other model-averaging approaches may yield robust predictions of range shifts in the face of climate change. It will still be difficult to translate these predictions into expected extinctions and species turn-over rates because actual range shifts will depend on dispersal, evolutionary flexibility, and species interactions. Nonetheless, for the sake of adaptive management and conservation planning, random forest models provide a useful and reliable tool. By minimizing the uncertainty in climate-envelope models, studies of climate-induced range shifts can concentrate on elucidating the effects of the more important uncertainties in climate-change predictions.

ACKNOWLEDGEMENTS

We thank Peter Kareiva, Louis Iverson, Dan Kluza, Anne Guerry, Don Phillips, Tom Edwards, and John Van Sickle for useful discussions and comments on the manuscript. Species distribution data were provided by NatureServe in collaboration with Bruce Patterson, Wes Sechrest, Marcelo Tognelli, Gerardo Ceballos, The Nature Conservancy – Migratory Bird Program, Conservation International – CABS, World Wildlife Fund – US, and Environment Canada – WILDSPACE. Funding was provided by a Conservation Fellowship awarded to JL from the Nature Conservancy. We also thank Betsy Smith and the U.S. Environmental protection Agency’s Regional Vulnerability program for providing additional funding and support. This paper has been subjected to review by the U.S. EPA National Health and Environmental Effects Research Laboratory’s Western Ecology Division and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. This is publication no. xxx of The Nature Conservancy's Conservation Research Fellowship Program.

REFERENCES

Araújo MB, Cabeza M, Thuiller W *et al.* (2004) Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology,* **10,** 1618-1626.

Bachelet D, Neilson RP, Lenihan JM *et al.* (2001) Climate change effects on vegetation distribution and carbon budget in the United States. *Ecosystems,* **4,** 164-185.

Breiman L (2001) Random forests. *Machine Learning,* **45,** 5-32.

Breiman L, Friedman JH, Olshen RA *et al.* (1984) *Classification and regression trees.* Wadsworth and Brooks/Cole, Monterey, California.

Chambers JM, Hastie TJ (1991) *Statistical Models in S.* Chapman and Hall, Boca Raton, Florida.

De'ath G, Fabricius KE (2000) Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology,* **81,** 3178-3192.

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation,* **24,** 38-49.

Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling,* **157,** 89-100.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling,* **135,** 147-186.

Hastie TJ, Tibshirani RJ (1990) *Generalized Additive Models.* Chapman and Hall, London.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics,* **6,** 65-70.

Houghton JT, Ding Y, Griggs DJ *et al.* (eds.) (2001) *Climate Change 2001: The Scientific Basis,* Cambridge University Press, Cambridge.

Iverson LR, Prasad AM (1998) Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs,* **68,** 465-485.

Johns TC, Carnell RE, Crossley JF *et al.* (1997) The second Hadley Centre coupled ocean-atmosphere GCM; Model description, spinup and validation. *Climate Dynamics,* **13,** 103-134.

Kattenberg A, Giorgi F, Grassl H *et al.* (1996) Climate models: projections of future climate. In: *Climate change 1995: the science of climate change. Contribution of Working Group 1 to the Second Assessment Report of the Intergovernmental Panel on Climate Change* (eds, Houghton HT, Meira Filho LG, Gallander BA *et al.*), Cambridge University Press, Cambridge.

Leemans R, Cramer WP (1991) International Institute for Applied Systems Analysis, Laxenburg.

Loveland TR, Zhu Z, Ohlen DO *et al.* (1999) An analysis of the IGBP global land-cover characterization process. *Photogrammetric Engineering and Remote Sensing,* **65,** 1021-1032.

McCullagh P, Nelder JA (1989) *Generalized linear models.* Chapman & Hall, London.

Meynecke J-O (2004) Effects of global climate change on geographic distributions of vertebrates in North Queensland. *Ecological Modelling,* **174,** 347-357.

Moisen GG, Frescino TS (2002) Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling,* **157,** 209-225.

Monserud RA, Leemans R (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling,* **62,** 275-293.

Neilson RP (1995) A model for predicting continental scale vegetation distributions and water balance. *Ecological Applications,* **5,** 362-385.

Nix HA (1986) A biogeographic analysis of Australian elapid snakes. In: *Atlas of Elapid Snakes of Australia* (ed, Longmore R), pp. 4-15. Australian Government Publishing Service, Canberra.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature,* **421,** 37-42.

Patterson BD, Ceballos G, Sechrest W *et al.* (2003) *Digital Distribution Maps of the Mammals of the Western Hemisphere, version 1.0.* NatureServe, Arlington, Virginia.

Pearson RG, Dawson TP, Berry PM *et al.* (2002) SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecological Modelling,* **154,** 289-300.

Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography,* **27,** 285-298.

Peterson AT, Ortega-Huerta MA, Bartley J *et al.* (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature,* **416,** 626-629.

Prasad AM, Iverson LR, Liaw A (in press) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*.

Ripley BD (1996) *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge.

Robertson MP, Peter CI, Villet MH *et al.* (2003) Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling,* **164,** 153-167.

Root TL, Price JT, Hall KR *et al.* (2003) Fingerprints of global warming on wild animals and plants. *Nature,* **421,** 57-60.

Segurado P, Araújo MB (2004) An evaluation of methods for modelling species distributions. *J Biogeography,* **31,** 1555-1568.

Shafer SL, Bartlein PJ, Thompson RS (2001) Potential changes in the distribution of western North America tree and shrub taxa under future climate scenarios. *Ecosystems,* **4,** 200-215.

Stockwell DRB, Noble IR (1992) Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation,* **33,** 385-390.

Therneau TM, Atkinson EJ (1997) An introduction to recursive partitioning using the RPART routines. http://www.mayo.edu/hsr/techrpt/61.pdf.

Thomas CD, Cameron A, Green RE *et al.* (2004) Extinction risk from climate change. *Nature,* **427,** 145-148.

Thuiller W (2003) BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology,* **9,** 1353-1362.

Thuiller W (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology,* **10,** 2020-2027.

Thuiller W, Araújo MB, Lavorel S (2003) Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science,* **14,** 669-680.

Thuiller W, Araújo MB, Pearson RG *et al.* (2004) Biodiversity conservation: uncertainty in predictions of extinction risk. *Nature,* **430**.

Venables WN, Ripley BD (2002) *Modern applied statistics with S-PLUS.* Springer, New York.

Wahba G (1990) *Spline models for observational data.* SIAM, Philadelphia.

Williams SE, Bolitho EE, Fox S (2003) Climate change in Australian tropical rainforests: an impending environmental catastrophe. *Proceedings of the Royal Society of London B,* **270,** 1887-1892.

Wood SN, Augustin NH (2002) GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling,* **157,** 157-177.

Table 1. Accuracy of six different modeling approaches used to model the current geographic ranges of 100 mammal species in the western hemisphere. Accuracy was assessed using a reserved test-data set composed of a randomly selected 20% of the presences and 20% of the absences for each species. Values reported are the medians and one half of the inter-quartile range of the accuracy of the model predictions for 100 species. Values with the same letters were not significantly different ($P > 0.05$).

Model[*]	% presences correct	% absences correct	Kappa	AUC[†]
GLM	77.7 (17.3), a	98.9 (1.4), a	0.68 (0.13), a	0.970 (0.017), a
TREE	55.5 (19.0), b	99.6 (0.5), b	0.63 (0.13), b, c	0.838 (0.072), b
GAM	68.9 (19.3), a	99.1 (1.4), a	0.62 (0.15), a, b	0.966 (0.022), c, d
RF	86.0 (12.1), c	99.6 (0.3), c	0.86 (0.09), d	0.995 (0.003), e
ANN	75.6 (12.5), a	98.2 (2.1), d	0.58 (0.13), c, e	0.968 (0.017), a, c
GARP	85.0 (6.2), c	95.9 (2.7), e	0.53 (0.17), e	0.962 (0.023), d

[*] Model abbreviations: GLM, generalized linear model; TREE, classification tree; GAM, generalized additive model; RF, random forest; ANN, artificial neural network; GARP, genetic algorithm for rule-set prediction.

[†] AUC is the area under the receiver-operating characteristic curve.

FIGURE LEGENDS

Figure 1. Maps of the current range of the black tufted-ear marmoset (*Callithrix penicillata*) as predicted by six alternative modeling approaches. See Table 1 for an explanation of model abbreviations.

Figure 2. Maps of the predicted future range of the black tufted-ear marmoset (*Callithrix penicillata*) as predicted by six alternative modeling approaches. See Table 1 for an explanation of model abbreviations.

Figure 3. Climate-induced range contractions for 100 species as predicted by six different modeling approaches. We report the percentage of species predicted to experience each of three levels of range contraction when A) individuals are assumed to be able to disperse completely into their future range and B) individuals cannot disperse out of their current range.

Figure 4. Ranking of the performance of six different modeling approaches for predicting the current distribution of 100 mammal species. Performance was assessed as A) the percentage of correctly predicted presences, B) the percentage of correctly predicted absences, C) the Kappa statistic, and D) the area under the receiver-operating characteristic curve (AUC). Each set of box and whiskers represents the median, first and third quartiles, and the maximum and minimum values. See Table 1 for an explanation of model abbreviations.

Figure 1

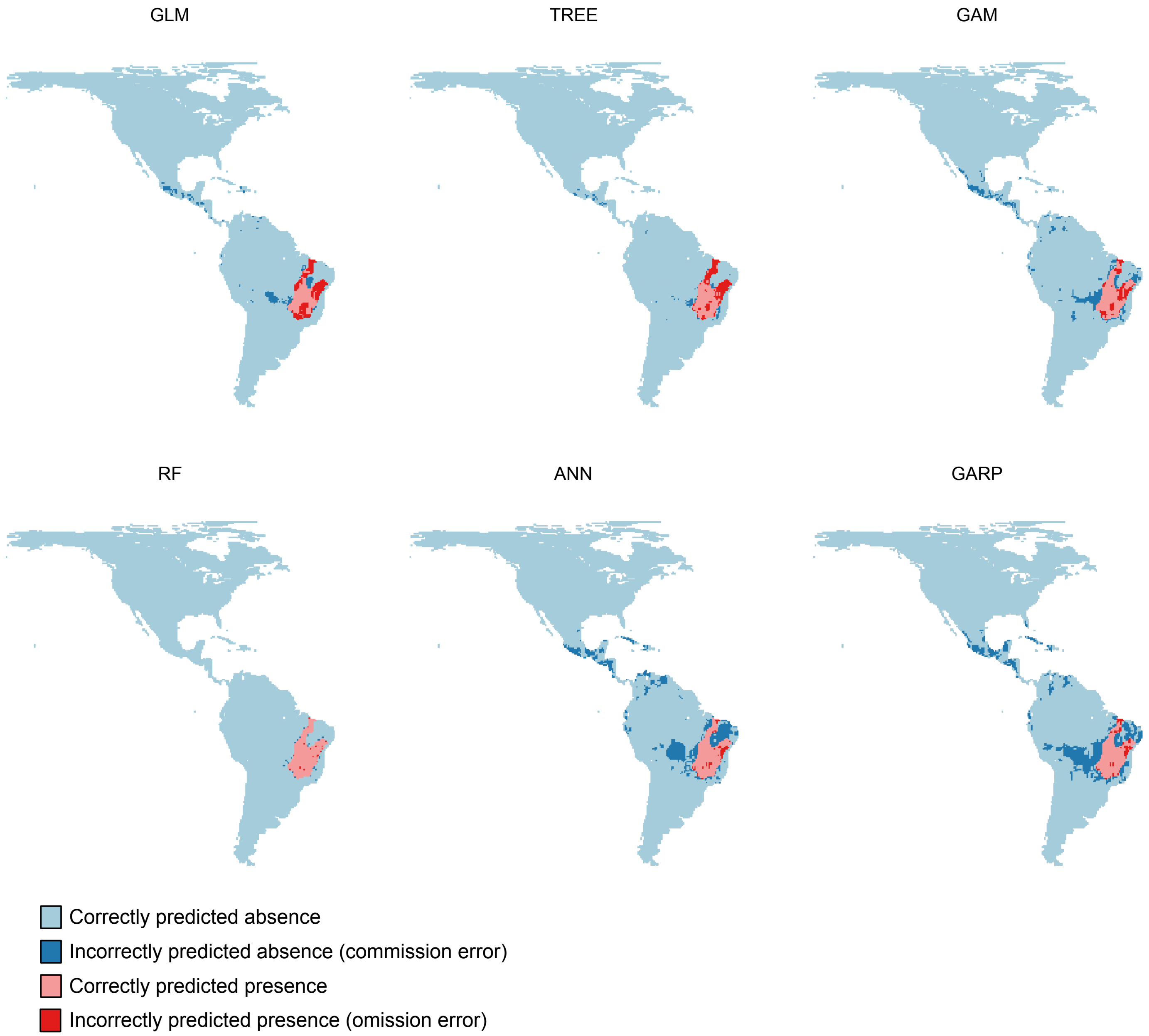

GLM
TREE
GAM
RF
ANN
GARP
Correctly predicted absence
Incorrectly predicted absence (commission error)
Correctly predicted presence
Incorrectly predicted presence (omission error)

Figure 2

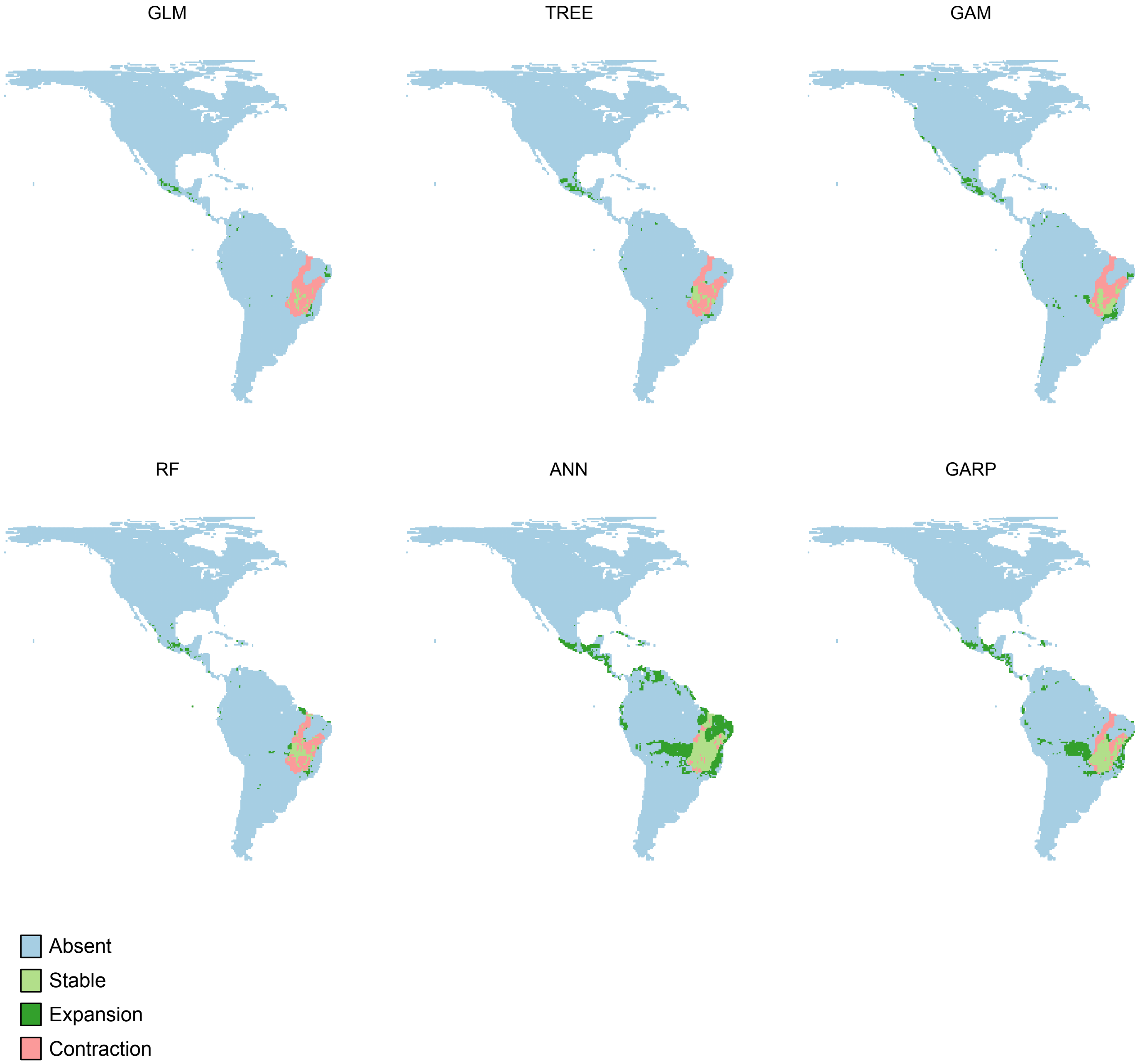
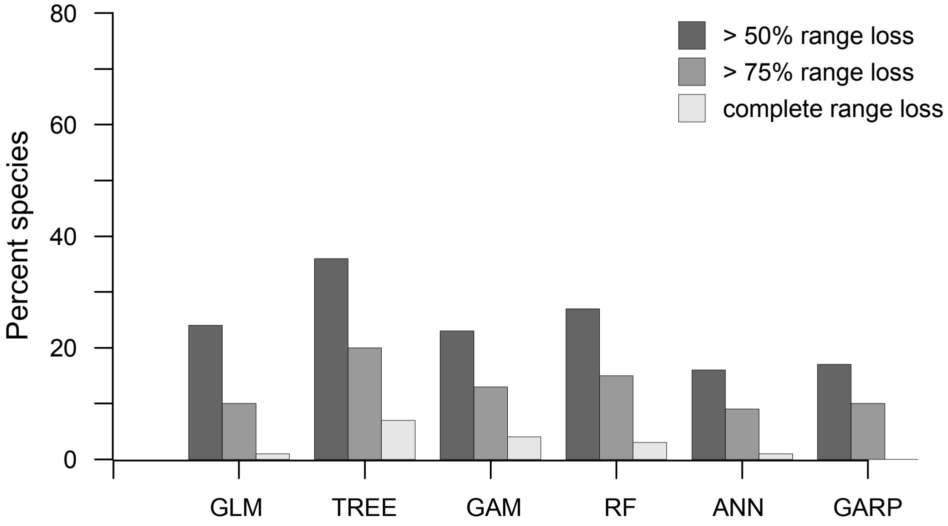

GLM
TREE
GAM
RF
ANN
GARP
Absent
Stable
Expansion
Contraction

Figure 3

A. Unlimited dispersal



B. No dispersal

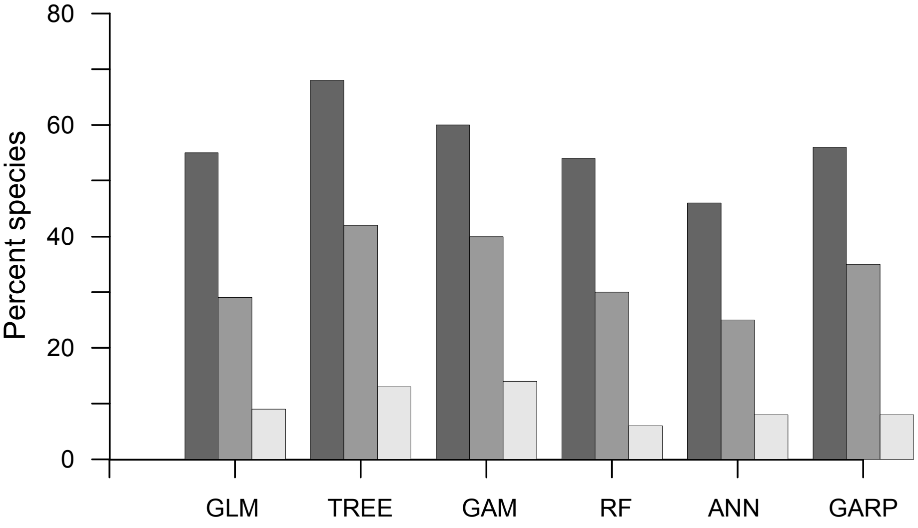
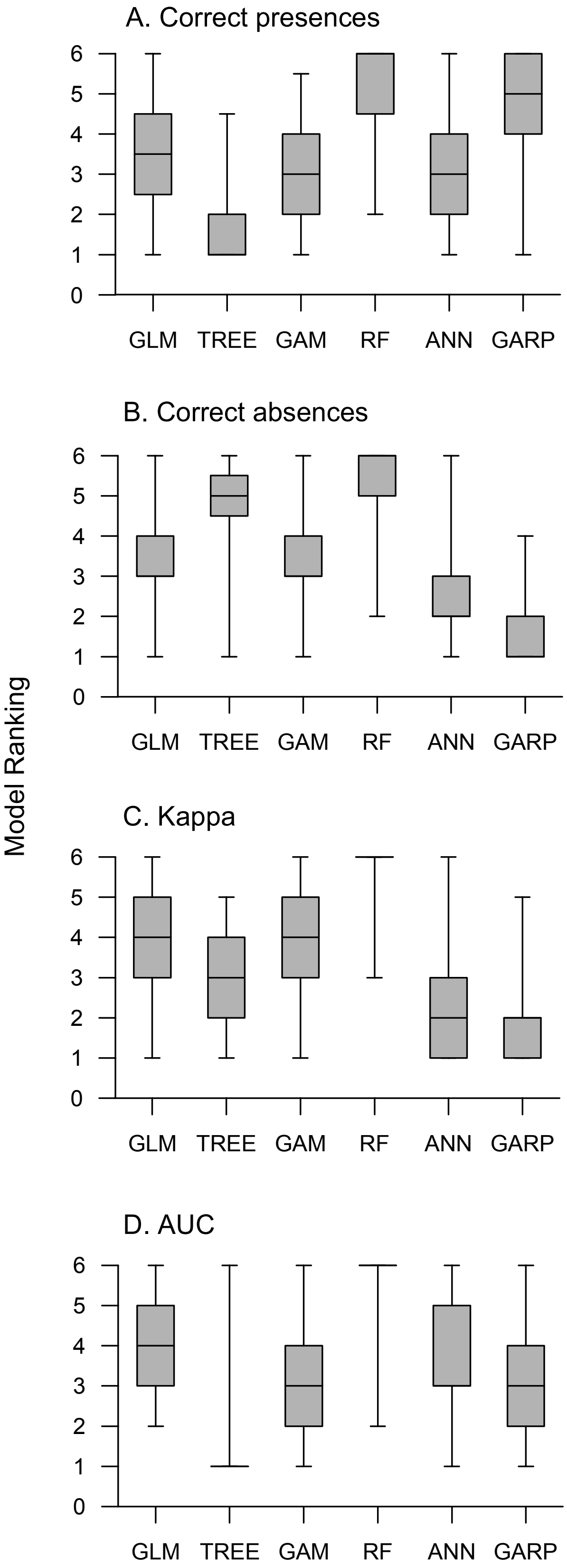


Figure 4


A. Correct presences
B. Correct absences
C. Kappa
D. AUC
Model Ranking
0
1
2
3
4
5
6
GLM
TREE
GAM
RF
ANN
GARP

Appendix. Future predicted range sizes of 100 mammal species in the western hemisphere as a proportion of current range. Future ranges were predicted using six different modeling approaches given two different dispersal scenarios. The models included generalized linear models (GLM), classification trees (TREE), generalized additive models (GAM), random forest predictors (RF), artificial neural networks (ANN), and genetic algorithms for rule-set prediction (GARP). The dispersal scenarios assumed that individuals could disperse completely into the predicted new range (unlimited dispersal) or conversely, that they were restricted to areas in which the current and future predicted ranges overlapped (no dispersal).

			Predicted future range as a proportion of current range											
			Unlimited dispersal						No dispersal					
Scientific name	English name	Current range (km^2)	GLM	TREE	GAM	RF	ANN	GARP	GLM	TREE	GAM	RF	ANN	GARP
Akodon albiventer	White-bellied Grass Mouse	437500	0.51	0.55	1.03	1.58	0.92	1.71	0.34	0.31	0.54	0.68	0.46	0.68
Akodon cursor	Cursor Grass Mouse	1430000	0.64	0.21	0	0.53	1.06	0.38	0.39	0.17	0	0.29	0.31	0.17
Alouatta caraya	Black Howling Monkey	3127500	1.08	0.34	0.79	0.31	0.97	0.62	0.41	0.19	0.23	0.17	0.58	0.25
Alouatta pigra	Black Howling Monkey	275000	5.71	0.45	0.75	4.48	1.51	3.61	0.54	0.05	0.03	0.04	0.06	0.21
Alouatta sara	Bolivian Red Howling Monkey	410000	7.03	0.15	3.85	1.27	10.09	1.07	0	0	0	0.12	0.71	0.37
Amorphochilus schnablii	Smoky Bat	370000	0.88	0.09	0.92	1.01	1.01	2.91	0.37	0.07	0.43	0.36	0.34	0.68
Anoura latidens	Broad-toothed Tailless Bat	1417500	2.75	3.29	7.13	7.39	0.33	5.35	0.64	0.57	0.81	0.82	0.12	0.63
Aotus vociferans	Tropical Night Monkey	1157500	1.35	0.05	0.19	0.28	1.67	0.91	0.73	0.01	0	0.08	0.83	0.55
Artibeus fraterculus	Fraternal Fruit-eating Bat	297500	0.84	0.29	0.87	0.33	1.28	1.81	0.19	0.11	0.29	0.18	0.44	0.3
Auliscomys pictus	Painted Big-eared Mouse	320000	0.8	0.18	1.16	0.73	0.82	0.96	0.39	0.13	0.38	0.45	0.39	0.45
Blastocerus dichotomus	Marsh Deer	2005000	1.56	0.5	1.49	0.77	0.9	1.01	0.39	0.22	0.26	0.21	0.44	0.15
Bradypus torquatus	Maned Three-toed Sloth	162500	0.14	0.38	0.02	0.14	0.46	0.51	0	0	0	0	0	0.02
Cabreramops aequatorianus	Equatorial Dog-faced Bat	127500	2.37	0.39	0.16	0.08	1.24	2.53	0.29	0.12	0	0.06	0.14	0.25
Callicebus moloch	Titi Monkey	975000	0.02	0.01	3.64	0.03	1.05	0.02	0	0	0	0	0.31	0
Callicebus nigrifrons	Black-fronted Titi	472500	0.49	0.22	2.58	0.25	1.15	0.04	0.32	0.04	0.15	0.05	0.28	0
Callicebus personatus	Northern Masked Titi	175000	0.11	0.11	0.09	0.06	0.37	0.73	0.01	0.03	0	0.01	0	0.01
Callicebus regulus	Titi Monkey	230000	0	0.03	0	0.04	0.08	0.01	0	0	0	0.04	0.04	0
Callicebus torquatus	Collared Titi	265000	0.17	0.03	0.01	0	8.93	0.03	0.01	0	0	0	0.93	0
Callithrix penicillata	Black Tufted-ear Marmoset	1602500	0.3	0.42	0.71	0.54	2.8	1.53	0.14	0.2	0.28	0.29	0.85	0.56
Cavia aperea	Brazilian Guinea Pig	5725000	0.57	0.77	0.62	0.53	0.78	1.04	0.4	0.49	0.4	0.43	0.58	0.65
Cavia magna	Greater Guinea Pig	172500	0.29	0	0.38	0	0.19	0.87	0.12	0	0.07	0	0.06	0.23
Centronycteris centralis	Bat	1242500	1.42	4.55	7.32	6.34	0.31	3.2	0.3	0.41	0.71	0.77	0.16	0.44
Chinchilla lanigera	Chinchilla	205000	1.06	0.32	0.7	0.44	0.6	2.05	0.63	0.26	0.59	0.43	0.46	0.73
Chinchillula sahamae	Altiplano Chinchilla Mouse	267500	1.08	0.29	1.65	0.59	1.19	1.8	0.55	0.15	0.63	0.36	0.58	0.73
Chiroderma trinitatum	Little Big-eyed Bat	9347500	1.15	1.07	1.07	1.12	0.99	0.7	0.95	0.86	0.87	0.86	0.87	0.66
Chiropotes albinasus	White-nosed Bearded Saki	145000	3.16	1.1	0.07	0.48	10.74	1.12	0.78	0.17	0.07	0.28	0.76	0
Chiropotes albinasus	White-nosed Bearded Saki	1107500	0.31	0.63	0.26	0.32	0.89	0.02	0.14	0.17	0.11	0.06	0.24	0
Chrotopterus auritus	Big-eared Wooly Bat	9182500	1.33	1.49	1.63	1.47	1.71	0.97	0.88	0.92	0.98	0.94	0.98	0.6
Coendou bicolor	Bicolor-spined Porcupine	1237500	0.42	0.73	0.67	3.95	1.72	1.02	0.13	0.22	0.22	0.67	0.58	0.21

Cryptotis mayensis	Maya Small-eared Shrew	147500	0.78	0.78	1.34	0.29	10.59	0.17	0	0	0	0	0.66	0
Ctenomys torquatus	Collared Tucu-tuco	475000	1.1	0.64	0.26	0.59	0.86	0.74	0.6	0.39	0.06	0.39	0.24	0.28
Cyclopes didactylus	Silky Anteater	8700000	1.39	1.41	1.32	1.45	1.17	0.9	0.98	0.97	0.98	0.98	0.89	0.82
Cynomops paranus	Dog-faced Bat	1855000	1.53	3.81	3.23	5.58	3.42	3.68	0.57	0.73	0.88	0.92	0.89	0.64
Dasyprocta azarae	Azara'a Agouti	1637500	0.39	0.5	0	0.64	0.85	0.12	0.29	0.26	0	0.34	0.47	0.02
Dasypus sabanicola	Llanos Long-nosed Armadillo	695000	3.25	1.18	0.4	1.4	7.44	3.37	0.46	0.23	0.05	0.29	0.76	0.39
Delomys dorsalis	Striped Atlantic Forest Rat	142500	0.18	0.14	0.33	0.05	0.05	0.26	0.12	0.07	0.09	0.05	0	0.02
Delomys sublineatus	Pallid Atlantic Forest Rat	212500	0.27	0	1.72	0.18	1.05	0.39	0.08	0	0.08	0.02	0.26	0.01
Eptesicus diminutus	Diminutive Serotine	3152500	1.04	0.45	0.83	0.79	0.56	0.63	0.51	0.34	0.49	0.54	0.4	0.32
Eptesicus fuscus	Big Brown Bat	12362500	1.32	1.13	1.55	1.5	1.18	1.34	0.9	0.83	0.92	0.94	0.86	0.87
Heteromys anomalus	Trinidad Spiny Pocket Mouse	592500	4.04	15.59	15.08	17.14	2.83	7.92	0.49	0.68	0.78	0.77	0.38	0.46
Histiotus macrotus	Big-eared Brown Bat	2215000	0.97	0.8	1.07	1.07	0.57	1.07	0.55	0.54	0.61	0.75	0.42	0.58
Hydrochaeris hydrochaeris	Capybara	13105000	1.12	1.22	1.12	1.05	1.24	1.06	0.94	0.99	0.94	0.94	0.98	0.93
Leopardus pardalis	Ocelot	15232500	1.12	1.11	1.17	1.12	1.31	0.98	0.99	0.98	0.99	0.99	0.92	0.94
Lepus alleni	Antelope Jackrabbit	250000	3.2	5.75	1.03	3.79	1.47	3.59	0.85	0.55	0.05	0.6	0.54	0.72
Lepus townsendii	White-tailed Jackrabbit	3465000	1.52	0.6	1.72	1.14	0.94	1.24	0.73	0.54	0.71	0.72	0.59	0.89
Lichonycteris obscura	Dark Long-tongued Bat	7737500	1.17	1.26	1.36	1.36	1.06	0.86	0.97	0.94	0.97	0.96	0.89	0.78
Lonchophylla hesperia	Western Nectar Bat	227500	0.18	0	2.93	0.02	0.16	3.49	0	0	0.09	0.02	0.03	0.21
Lonchorhina orinocensis	Orinoco Sword-nosed Bat	262500	9.93	0.98	17.76	0.62	1.97	1.79	0.43	0.28	0.56	0.07	0.43	0.38
Lyncodon patagonicus	Patagonian Weasel	1070000	0.93	0.56	0.94	0.72	1.77	1.69	0.44	0.41	0.5	0.56	0.64	0.32
Marmosa robinsoni	Robinson's Mouse Opossum	765000	6.85	2.82	12.44	11.6	4.67	10.93	0.75	0.43	0.79	0.8	0.58	0.64
Marmosops noctivagus	White-bellied Slender Mouse Opossum	1825000	0.47	0.61	0.2	0.75	1.79	0.26	0.3	0.37	0.1	0.43	0.64	0.13
Marmosops parvidens	Delicate Slender Mouse Opossum	2282500	1.17	0.69	1.33	0.98	1.83	0.1	0.23	0.26	0.35	0.25	0.63	0.02
Martes pennanti	Fisher	3455000	1.43	1.27	1.51	1.25	0.91	0.83	0.82	0.75	0.83	0.71	0.51	0.47
Megasorex gigas	Mexican Shrew	155000	0.6	0	3.82	2.31	1.42	6.58	0.1	0	0.02	0.34	0.42	0.16
Mephitis macroura	Hooded Skunk	1832500	1.18	0.53	0.81	1.19	2.21	1.82	0.62	0.33	0.46	0.64	0.74	0.67
Micronycteris hirsuta	Hairy Big-eared Bat	4282500	1.84	2.03	2.2	1.91	1.69	0.79	0.95	0.87	0.95	0.88	0.92	0.46
Microtus oeconomus	Tundra Vole	2352500	0.81	1.1	1.52	1.18	1.52	1.77	0.44	0.46	0.42	0.55	0.47	0.66
Microtus pinetorum	Woodland Vole	2767500	1.27	1.18	1.42	1.23	1.49	0.89	0.93	0.92	0.97	0.92	0.95	0.83
Mustela frenata	Long-tailed Weasel	12200000	1.11	0.8	1.09	2.03	1.09	1.06	0.8	0.65	0.76	0.94	0.69	0.76
Myoprocta acouchy	Red Acouchy	1300000	1.91	1.08	1.67	1.27	4.29	0.57	0.77	0.37	0.28	0.25	0.93	0.02
Myotis evotis	Long-eared Myotis	3115000	1.12	0.79	1.04	1.01	0.8	1.32	0.75	0.69	0.75	0.85	0.68	0.77

Myotis levis	Yellowish Myotis	2900000	0.86	0.52	1.25	0.72	0.99	1.06	0.39	0.38	0.41	0.55	0.5	0.59
Myotis ruber	Red Myotis	1725000	0.61	0.33	0.59	0.4	0.36	0.1	0.46	0.29	0.45	0.31	0.23	0.04
Nasuella olivacea	Mountain Coati	210000	0.74	32.48	21.83	35.04	0	17.68	0.25	0.3	0.43	0.29	0	0.36
Necromys lasiurus	Hairy-tailed Bolo Mouse	5395000	0.87	1.28	0.89	0.61	1.62	1.25	0.56	0.76	0.52	0.4	0.86	0.67
Neotoma goldmani	Goldman's Woodrat	175000	0.3	0.64	0	0.67	5.07	0.61	0.17	0.23	0	0.43	0.83	0.14
Nyctinomops femorosaccus	Pocketed Free-tailed Bat	1132500	1.7	0.6	1.97	1.29	2.21	3.11	0.84	0.41	0.7	0.62	0.78	0.89
Oecomys speciosus	Arboreal Rice Rat	185000	0.35	1.43	0.38	0.46	10.62	6.86	0.12	0.05	0.11	0.04	0.85	0.15
Oryzomys angouya	Rice Rat	1142500	1.62	0.28	3.17	0.96	1.09	0.88	0.38	0.2	0.56	0.38	0.21	0.23
Oryzomys rostratus	Long-nosed Rice Rat	370000	3.26	0.83	0.89	2.36	0.35	3.06	0.62	0.11	0.03	0.3	0.08	0.34
Oryzomys yunganus	Yungas Rice Rat	5302500	1.56	0.96	1.61	0.87	1.78	0.51	0.9	0.67	0.9	0.54	0.96	0.37
Oxymycterus rufus	Red Hocicudo	1395000	0.96	0.3	0.04	0.43	0.62	0.44	0.28	0.23	0.01	0.34	0.3	0.24
Peromyscus levipes	Nimble-footed Mouse	582500	1.28	0.53	3.73	6.76	0.01	4.36	0.5	0.26	0.28	0.7	0	0.42
Peromyscus polionotus	Oldfield Mouse	425000	0.33	0.39	0.25	0.49	3.23	0.05	0.21	0.12	0.08	0.18	0.79	0.01
Platyrrhinus helleri	Heller's Broad-nosed Bat	10765000	1.06	0.99	0.96	1.23	1.26	0.83	0.93	0.84	0.84	0.97	0.92	0.77
Proechimys cayennensis	Cayenne Spiny Rat	3030000	0.7	1.39	0.44	1.58	2.15	0.88	0.24	0.68	0.15	0.78	0.9	0.18
Proechimys oris	Para Spiny Rat	225000	0.12	0	0.04	0	1.54	0.5	0	0	0	0	0	0
Proechimys roberti	Spiny Rat	1077500	0.3	0.82	2.79	0.86	1.13	1.03	0.17	0.5	0	0.5	0.56	0.02
Pteronotus davyi	Davy's Naked-backed Bat	4062500	1.3	2.42	2.82	3.32	2.57	1.58	0.61	0.73	0.83	0.94	0.86	0.57
Rhogeessa tumida	Black-winged Little Yellow Bat	7637500	1.48	1.38	1.58	1.76	1.48	1.1	0.97	0.94	0.96	0.99	0.94	0.85
Saccopteryx gymnura	Amazonian Sac-winged Bat	615000	0.09	0.7	0.38	0.96	1.58	1.26	0	0.11	0	0.18	0.23	0.11
Scapanus latimanus	Broad-footed Mole	337500	1.39	0.67	2.78	1.1	1.01	2.93	0.46	0.44	0.49	0.61	0.4	0.96
Sciurus aestuans	Guianan Squirrel	6922500	0.5	0.62	0.82	0.69	1.29	1.01	0.32	0.39	0.4	0.45	0.81	0.44
Sciurus niger	Eastern Fox Squirrel	4307500	1.56	1.15	1.45	1.52	1.52	0.91	0.99	0.86	0.96	0.96	0.96	0.75
Scotinomys teguina	Alston's Brown Mouse	207500	0.33	--	0.43	0.24	0.19	3.87	0.01	--	0.01	0.04	0	0.17
Sigmodon arizonae	Arizona Cotton Rat	347500	1.68	1.89	1.78	1.37	2.4	1.76	0.73	0.49	0.36	0.58	0.57	0.73
Sorex dispar	Long-tailed Shrew	327500	1.09	0.89	2.45	1.56	0.61	0.4	0.41	0.29	0.59	0.5	0.1	0.14
Sorex trowbridgii	Trowbridge's Shrew	402500	1.1	0.4	0.63	0.84	2.45	1.8	0.58	0.27	0.42	0.55	0.88	0.75
Speothos venaticus	Bush Dog	10617500	1.05	1.11	1.19	1.23	1.32	0.78	0.83	0.89	0.91	0.93	0.94	0.67
Spermophilus elegans	Wyoming Ground Squirrel	272500	0.6	0.23	0.66	0.17	0.39	3.41	0.29	0.1	0.18	0.06	0.15	0.69
Sturnira magna	Greater Yellow-shouldered Bat	1470000	0.27	0.53	0.73	1.04	0.6	0.84	0.08	0.22	0.35	0.49	0.34	0.32
Sturnira nana	Lesser Yellow-shouldered Bat	137500	1.51	0.27	0.73	0.75	1.25	1.36	0.38	0.04	0.09	0.13	0.22	0.16
Tamias rufus	Hopi Chipmunk	140000	0.09	0	1.43	0.02	0.2	6.09	0	0	0.07	0	0	0.54

Tamias umbrinus	Uinta Chipmunk	287500	1.57	0.1	2.09	0.17	0.1	4.63	0.23	0.05	0.23	0.05	0.02	0.61
Tamiasciurus hudsonicus	Red Squirrel	10217500	1.01	1.05	1.02	1.06	1.2	1.19	0.85	0.89	0.87	0.89	0.9	0.94
Thomomys bottae	Botta's Pocket Gopher	1885000	1.81	1.11	1.75	1.66	1.56	2.13	0.84	0.64	0.87	0.84	0.67	0.85
Tonatia saurophila	Stripe-headed Round-eared Bat	7115000	1.68	1.46	1.65	1.52	1.35	1.18	0.98	0.93	0.98	0.96	0.89	0.87
Uroderma bilobatum	Tent-making Bat	12600000	1.02	1.12	0.99	1.14	1.08	0.88	0.95	0.96	0.94	0.99	0.98	0.83
Vampyressa bidens	Bidentate Yellow-eared Bat	6682500	1.27	0.99	0.64	0.77	1.01	0.51	0.97	0.81	0.57	0.5	0.84	0.46
Zaedyus pichiy	Pichi	1475000	0.76	0.61	0.78	0.64	0.9	1.09	0.52	0.43	0.56	0.56	0.53	0.6