

# Niche Modeling and Geographic Range Predictions in the Marine Environment Using a Machine-learning Algorithm

*E. O. Wiley, Kristina M. McNyset, A. Townsend Peterson, C. Richard Robins, and Aimee M. Stewart*  
*Natural History Museum and Biodiversity Research Center, University of Kansas • Lawrence, Kansas USA*

Biological communities are changing drastically in response to global climate change (Walther et al., 2002), changes in use by human populations (Krishtalka et al., 2002), and introduction of exotic species (Carlton, 1996; Enserink, 1999). To study the impact of such changes in the marine environment, biologists require a detailed understanding of the diversity and distributions of marine organisms on macroscopic scales, such as across entire ocean basins, in order to improve understanding of the actual distributions of species, and gain an overall impression of the potential community structures that exist in particular habitats. A major obstacle to such an improved understanding is the fact that existing biodiversity records are both incomplete and idiosyncratic, consisting of museum collection and fisheries survey data covering only a small, and usually biased, fraction of the marine biosphere. The ability to predict distributions of species based on existing specimen records would allow investigators to predict the presence or absence of species in previously unsampled waters. If such predictions prove robust, then responses of species to global climate change could be predicted (Peterson et al., 2002b), potential impacts of introductions of exotic species on native faunas could be anticipated proactively (Peterson and Vieglais, 2001), and assessment of potential future conservation areas could be performed (Peterson et al., 2000). In this paper, we extend the application of one tool, the Genetic Algorithm for Rule-set Prediction (GARP) to the marine realm, and demonstrate its potential usefulness in predicting geographic distributions of littoral and benthic fishes across a major and heterogeneous ocean region.

Two components are critical to such studies: locality records of specimens for species (primary occurrence data, presence data, or more simply "point data"), and electronic maps (coverages) of relevant ecological parameters (Peterson, 2001). For fishes, it is now possible to gather point data from national history museums, worldwide, in real time, and in a common

format (Vieglais et al., 2000). Museum data are high quality because voucher specimens can be examined if identification is questionable. However, like all point data, museum specimens provide only a limited view of the actual species' range (Krohn, 1996), hence the need for predictive approaches. A limited number of ecological data sets are also available, worldwide, including physio-chemical parameters (NOAA, 1999) and bathymetry (Smith and Sandwell, 1997).

Ecological niche modeling uses the primary point data and the ecological data to build a partial niche model for the species (Stockwell, 1999; Stockwell and Noble, 1992; Stockwell and Peters, 1999). The niche model is defined in ecological space by ecological parameters. It can be projected into geographic space by dividing the area of interest into rows and columns to create a grid, and then identifying the grid cells where the ecological parameters match those of the niche model. The landscape for this study is the Central Western Atlantic, roughly bounded by the definitions of the Food and Agricultural Organizations definition of Fishing Area 31, approximately the area of the Atlantic Ocean, Caribbean Sea, and the Gulf of Mexico between 35°N and 5°N Latitude and west of 40°W Longitude.

Many tools have been used to develop models of ecological niches. Among the simplest is BIOCLIM (Nix, 1986), which involves intersecting the ranges (slightly trimmed) inhabited by the species along each environmental axis (e.g. 0–50 m depth x annual surface temperature average of 20–22°C, etc.). Other approaches include general linear models, distance-based algorithms, and regression-tree analyses (Austin et al., 1996; Carpenter et al., 1993; Malanson et al., 1992). These relatively straightforward algorithms, however, suffer from their focus on a search for a single decision rule, or a small set of decision rules. The reality of species' ranges is that many factors affect them, and indeed different decision rules may govern distributional limits in different sectors of a species' distribution

This article has been published in *Oceanography*, Volume 16, Number 4, a quarterly journal of The Oceanography Society. Copyright 2003 by The Oceanography Society. All rights reserved. Reproduction of any portion of this article by photocopy, machine, reposting, or other means without prior authorization of The Oceanography Society is strictly prohibited. Send all correspondence to: [info@os.org](mailto:info@os.org), or 5912 LeMay Road, Rockville, MD 20851-2326, USA.

(MacArthur, 1972). For example, in the tropical western Atlantic a common distributional pattern emerges separating insular and continental fish species. Continental fishes further separate into northern and southern components (Robins 1971, 1991). *Epinephelus mystacinus*, a large and very wide-ranging tropical grouper is restricted to the insular region in the western Atlantic except for very rare vagrants. It is replaced ecologically by the very large *Epinephelus nigritus* (not closely related) in continental waters. The coastal pelagic mackerel *Scomberomorus maculatus* is northern continental. In southern continental waters it is replaced by *S. brasiliensis*, which is not regarded as a sister species. Although *S. regalis* occurs widely in the insular region, it also is common in continental waters. Temperature would seem to be the limiting factor in the northward distribution of tropical continental species, which often extend their ranges northward during the summer (Robins and Ray, 1986). Temperature in surface waters may limit larval recruitment of deepwater bottom fishes such as *Synaphobranchus kaupi* (a northern species) and *S. oregoni* (a southern species) even though the adults live in waters of similar temperature (C. H. Robins and C. R. Robins, unpubl. data).

Of course, one must be careful in observing common distributional patterns not to infer common causality. Hence, more complex, multiple-criterion approaches appear to be desirable. Herein, we focus on a heterogeneous-rule machine-learning approach that has proven particularly useful, as described below.

*The Genetic Algorithm for Rule-set Prediction* (GARP) includes several inferential tools in an artificial-intelligence-based approach to modeling (Stockwell, 1999; Stockwell and Noble, 1992; Stockwell and Peters, 1999). In each round of modeling, occurrence points for a species are resampled randomly to create internal training and test data sets (1250 points in each set). GARP works in an iterative process of rule selection, evaluation, testing, and incorporation or rejection. First, an inferential tool is chosen from a set of possibilities (e.g., logistic regression, range rules), is applied to the training data, and a rule developed. Rules are evolved by a number of means (e.g., truncation, point changes, crossing-over among rules) to maximize predictivity. Predictive accuracy is evaluated using the test data and an equivalent number of points sampled randomly 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, and the algorithm runs either 1000 iterations or until convergence.

GARP has been shown to be robust in the face of various kinds of bias, including bias produced by con-

sidering presence data only (Stockwell and Peterson, 2002a). GARP models provide an efficient means of modeling species' ecological niches, and for projecting those models onto geography in the form of maps, which are testable hypotheses of distributional potential. Numerous such tests (Anderson et al., 2002a; Anderson et al., 2002b; Fera and Peterson, 2002; Peterson, 2001; Peterson et al., 2002a; Peterson and Cohoon, 1999; Peterson et al., 2002c; Peterson and Vieglais, 2001; Stockwell and Peterson, 2002a, b) confirm the predictive nature of these ecological niche

models. GARP-modeled species' niches indeed delineate the habitable environmental conditions for a species, and they can be used to predict new data sets (those withheld from the modeling process in this study) with precision.

## Materials and Methods

Eighteen marine fish species occurring in FAO Fishing Area 31 were selected for analysis (Table 1): 12 are primarily continental shelf species, and six are primarily continental slope species. Species were selected to represent a range of taxonomic groups and habitat diversity. Specimen locality records (point data) were accumulated from museum collections held by FishNet partner institutions via *The Species Analyst* (Vieglais et al., 2000). Longitude and latitude values were checked against textual locality information and against published range descriptions of each species. Point data were visualized in ESRI ArcView 3.2 (ESRI, 1999) and each point was matched against bathymetry. Points that obviously were mismatched with known depth ranges of species (as reported in species accounts in Fisher, 1978, or McEachran and Fechhelm, 1998) were eliminated. Records identified as larvae or that constituted duplicate longitude-latitude combinations for a species were also eliminated.

Ecological coverages were drawn from the World Ocean Atlas (WOA) 1998 dataset (NOAA, 1999) and the Smith and Sandwell (1997) bathymetric database (Table 2). A subset of WOA 98 data was processed comprising annual analyzed means for each of nine variables (temperature, salinity, dissolved oxygen, percent oxygen saturation, apparent oxygen utilization, phosphate, nitrate, silicate, chlorophyll) at all depths. Chlorophyll data included seven depths ranging from surface to 100 m, and all other variables included 33 depths ranging from surface to 5500 m. WOA 98 data were converted to raster grids and resampled from 1° to 2' resolution to match the bathymetric data resolution. The edges of each depth grid were smoothed by extending values of pixels of that depth to the bathymetry (i.e., values at depth 100 were extended towards shorelines and cut off at depths shallower than 100

---

*The ability to predict distributions of species based on existing specimen records would allow investigators to predict the presence or absence of species in previously unsampled waters.*

---

Table 1.

Species used in the study, with a summary of modeling results. Abbreviations: S, Surface coverage dataset; B, benthic coverage data set; Train, number of point occurrences used in the Training dataset; Test, number of point occurrences used in the Test dataset; AUC Area Under the Curve;  $AUC_{max}$ , one measure of the maximum possible AUC score; SE, standard error; Z, z-score.

Taxon	Coverage	Train	Test	AUC	$AUC_{max}$	SE	Z
<i>Barbulifer ceuthoecus</i>	S	28	14	0.948	0.948	0.04	9.73*
<i>Centropomus undecimalis</i>	S	18	18	0.952	0.987	0.04	13.25*
<i>Coryphaena hippurus</i>	S	29	38	0.746	0.796	0.05	10.14*
<i>Gempylus serpens</i>	S	64	39	0.613	0.687	0.04	11.97*
<i>Monacanthus ciliatus</i>	S	89	39	0.971	0.997	0.02	23.38*
<i>M. hispidus</i>	S	34	38	0.982	0.983	0.02	20.91*
<i>M. tuckeri</i>	S	19	21	0.746	0.916	0.06	6.60*
<i>Pomatomus saltatrix</i>	S	27	20	0.946	0.980	0.04	13.92*
<i>Porichthys plectrodon</i>	S	58	107	0.960	0.961	0.01	33.39*
<i>Scomberomorus cavalla</i>	S	38	16	0.898	0.963	0.05	11.21*
<i>S. maculatus</i>	S	28	23	0.980	0.989	0.02	14.57*
<i>Sphoeroides nephelus</i>	S	27	11	0.897	0.997	0.06	12.07*
<i>Androvandia gracilis</i>	B	11	17	0.907	0.917	0.05	12.76*
<i>Bathygadus macrops</i>	B	17	26	0.991	0.991	0.01	30.14*
<i>Etmopterus schultzi</i>	B	28	28	0.970	0.988	0.02	31.53*
<i>Neobythites marginatus</i>	B	15	11	0.964	0.988	0.03	25.98*
<i>Polymixia lowei</i>	B	61	52	0.944	0.982	0.02	34.59*
<i>Urophycis cirrata</i>	B	36	29	0.945	0.967	0.03	14.48*

\* $p < 0.01$

Table 2.

Environmental Coverages used in this study.

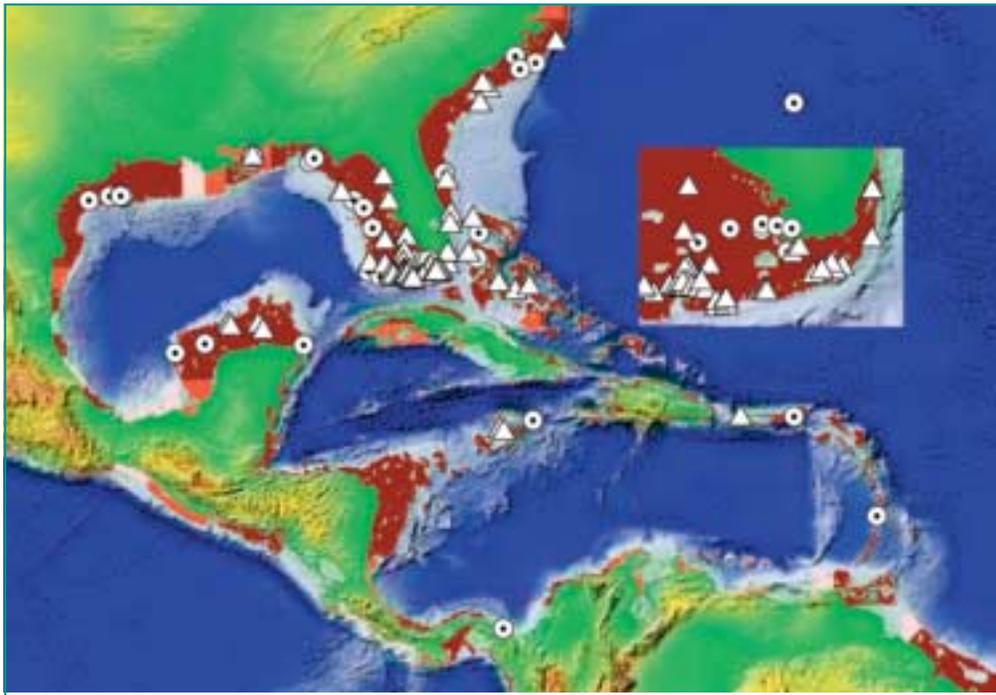
Coverage	Source
Bathymetry	Smith and Sandwell (1997)
Temperature	WOA 98
Salinity	WOA 98
Dissolved oxygen	WOA 98
Percent oxygen saturation	WOA 98
Apparent oxygen utilization	WOA 98
Phosphate	WOA 98
Nitrate	WOA 98
Silicate	WOA 98
Chlorophyll	WOA 98

meters). This was deemed necessary because some near-shore data points fell in areas not covered by the large 1° pixels. After resampling to 2' resolution, the shoreline could be more closely approximated. Finally, the data were clipped to FAO Area 31 (Central Western Atlantic and Caribbean). Because of the depth stratifi-

cation, we prepared two sets of raster grids, "Surface" and "Benthic." The Surface coverage set included bathymetry plus the nine WOA 98 variables at the surface. The Benthic coverage set included bathymetry and the nine WOA 98 variables projected onto bathymetry, producing a three-dimensional picture of conditions on the bottom at depth; i.e., if the ocean floor is at 100 m, then conditions at 100 m were represented in the coverage.

We wished to provide a severe test of GARP's predictive ability. The most difficult tests involve coarse stratification of point data for model building and testing (Fielding and Bell, 1997), so that spatial autocorrelation between data upon which models are built and those used for testing are minimal. We partitioned the point data into two sets by randomly sorting grid cells from a 1° resolution grid containing point occurrence data into "Training" and "Test" data sets. Examples of partitioned point data are shown in Figures 1–3, with training points shown as yellow circles and testing point shown as green circles. Because the number of points in grid cells varied, the number of points in the two datasets varied among species (Table 1).

For each species, we applied Desktop GARP (Scachetti-Pereira, 2001) to the training dataset to produce 200 models for each species, using a convergence



**Figure 1.** Prediction of geographic distribution of the teleost fish *Monacanthus ciliatus* in the Central Atlantic, Caribbean, and Gulf of Mexico using GARP. Triangles are point localities used by GARP in concert with nine WOA 98 environmental surface coverages and bathymetry. Circles are point localities withheld from modeling and used to test the prediction. Blue denotes bottom depth, with lighter blue indicating relatively shallow waters. Pink to rust brown shading denotes number of model intersections: pink, 5-6; red, 7-9; rust brown, 10 intersections respectively. The inset shows details off Florida and parts of the Bahamas.

limit of 0.01 and a maximum of 1000 iterations per model. From the 200 models, we selected ten models that comprise the “best-subsets models” (Anderson et al., 2003). Customary GARP implementation involves specification of a certain number or percentage of points to be reserved internally for testing model quality. We used all training points in generating models because we had set aside data from half of the 1° grid cells containing points as independent testing data. Thus, we used intrinsic measures (based on the training data) for calculating omission and commission error statistics (Anderson et al., 2003). We calculated a median commission error for all models with less than 5% intrinsic omission error. The absolute difference between the median commission error and the commission index for each model was calculated, and the 10 models showing the least deviation with the median score were selected as the 10-best model set.

We evaluated the ability of models to predict independent test points accurately using a Receiver Operating Characteristics (ROC) analysis (Zweig and Campbell, 1993; Fielding and Bell, 1977). This statistic evaluates the specificity (absence of commission error) and sensitivity (absence of omission error) of a diagnostic test in the face of the independent data provided

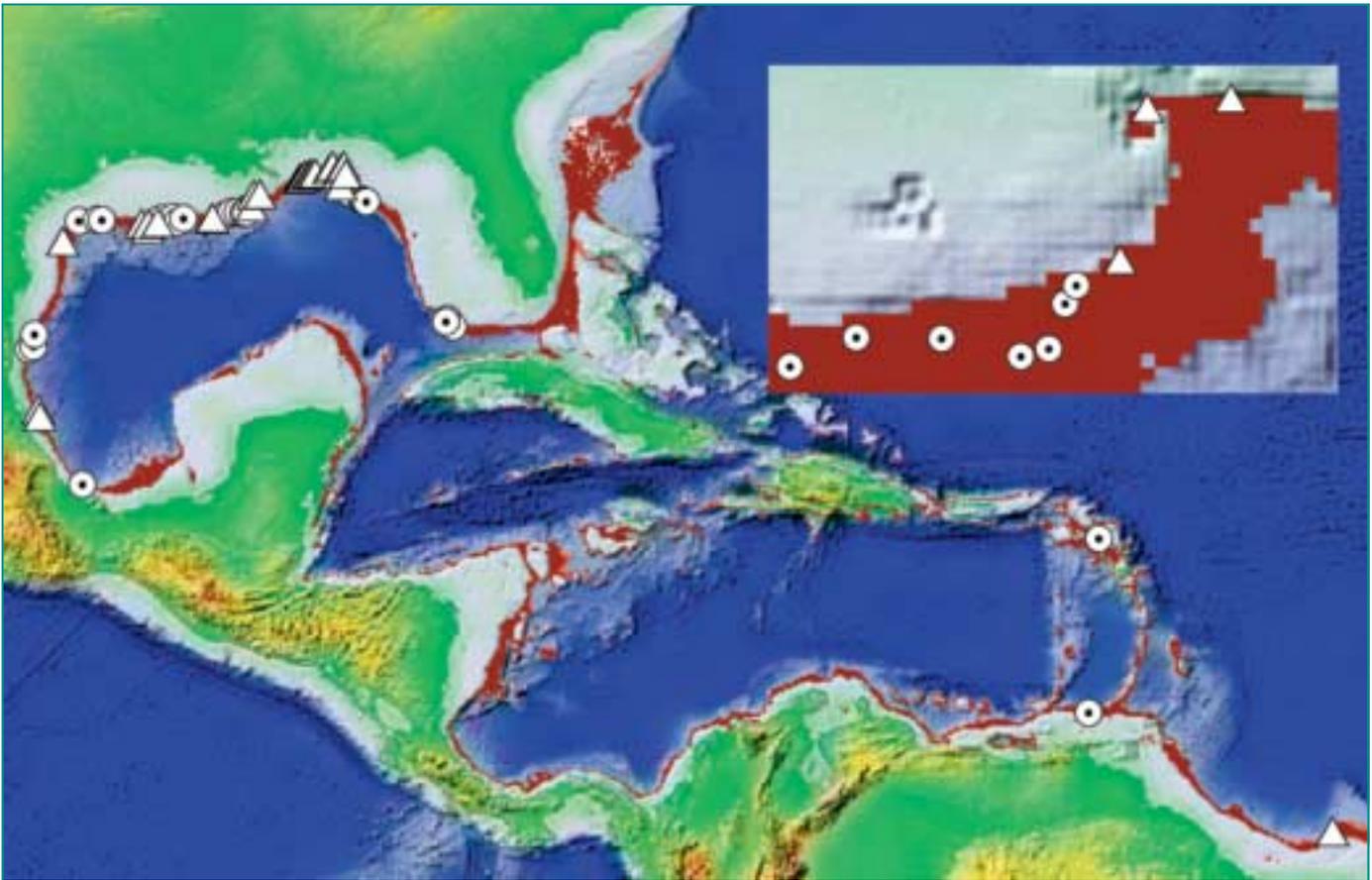
by the test dataset. The test dataset provides a “gold standard” and each model set is scored on its ability to predict these new data correctly. These scores are graphed on an axis of sensitivity (true positive rate of accumulation) and 1-specificity (true negative rate of accumulation). We used the trapezoidal rule to calculate AUC (Hanley and McNeil, 1982). AUC measures how well the 10-model set predicted the test data points. A perfect result is AUC = 1.0, whereas a test that performs no better than random yields an AUC = 0.5. We evaluated the results using a z-test (Centor, 1991). During the course of our investigation, we noted that ROC analysis produces an area effect, such that the maximum AUC score received for a particular species is a function of the total area (total number of pixels) predicted “present” relative to the total area of the landscape

studied. Therefore, we calculated a measure of the maximum possible AUC score for each species by moving all test data points to the maximum value of the intersection of all 10-best models. We consider the difference between the actual AUC obtained relative to the maximum possible AUC ( $AUC_{max}$ ) to be a better measure of the predictive ability of the 10-best model set than the raw AUC score; however, the raw AUC score remains a valid test of the ability of the 10-best model set to predict non-random patterns. We used an Excel spreadsheet for all calculations.

Environmental coverages, point data used for both training and testing, and specimen museum numbers for all species used in this study can be downloaded though the Internet at <http://www.speciesanalyst.net/fishnet/>. This page also provides a link for accessing desktop GARP.

## Results

Results in general indicated excellent predictivity for all species (Table 1; Figures 1–3). AUCs in ROC analyses (Table 1) varied between 0.991 (*Bathygadus macrops*) and 0.613 (*Gempylus serpens*). All of these predictions were significantly different from a line of no information ( $p < 0.01$ ) and many were highly signifi-



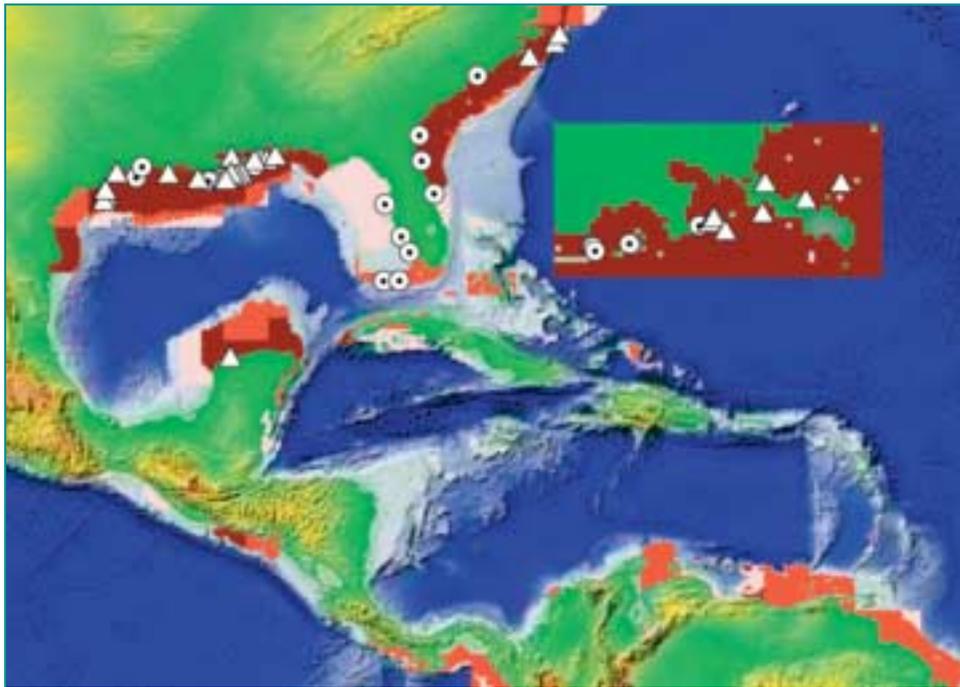
**Figure 2.** Prediction of geographic distribution of the chondrichthyan fish *Etmopterus schultzi* in the Central Atlantic, Caribbean, and Gulf of Mexico using GARP. Triangles are point localities used by GARP in concert with nine WOA 98 environmental surface coverages and bathymetry. Circles are point localities withheld from modeling and used to test the prediction. Blue denotes bottom depth, with lighter blue indicating relatively shallow waters. Pink to rust brown shading denotes number of model intersections: pink, 5-6; red, 7-9; rust brown, 10 intersections respectively. The inset shows details from off Louisiana.

cant ( $p < 0.001$ ). The average AUC score across all taxa was 0.911. Correlations between number of test points and AUC score, and between number of training data points and AUC score, were not significantly different from zero ( $r = 0.031, p < 0.05$ ;  $r = -0.106, p < 0.05$ , respectively).

Results for *Monacanthus ciliatus*, the fringed filefish, provide an example of modeling with surface coverages (Figure 1). This teleost species is distributed vertically from depths of 0- 50 m throughout Area 31, and ranges north to Newfoundland and south to Argentina (Tyler, 1978). General coincidence of predictions with test data points and known distributional patterns was excellent. Indeed, although no training points fell in Bermuda, one Bermuda testing point was present and all 10-best models successfully predicted the species' presence there. Results for *Etmopterus schultzi*, the fringed lantern shark (Figure 2), provide an example of results of modeling with benthic coverages. This benthic shark is distributed vertically from depths of 384-732 m over soft mud bottoms. Although published

ranges restrict the species to the northern Gulf of Mexico (Compagno, 1984), the records used herein reflect a larger range that is considered valid (G. Burgess, Florida Museum of Natural History, pers. comm.). The coincidence between prediction and test points for this species was excellent.

Although highly significantly different from random, the AUCs from ROC analysis appeared weak for two widely distributed species, *Gempylus serpens* and *Coryphaena hippurus*. Both species are predicted "present" over large percentages of the total landscape, depressing the maximum possible AUC ( $AUC_{max}$ ; Table 1). *Gempylus serpens* is predicted to occupy some 63% of the total landscape; its AUC = 0.613 but the  $AUC_{max}$  is only 0.687. Similarly depressed results were obtained for *Coryphaena hippurus*, predicted present over 41% of the landscape. Its AUC is 0.714, but  $AUC_{max}$  is only 0.796. In contrast, a similar AUC was found for *Monacanthus tuckeri* (AUC = 0.746), but the  $AUC_{max}$  for this species is much higher ( $AUC_{max} =$



**Figure 3.** Prediction of geographic distribution of the teleost fish *Scomberomorus maculatus* in the Central Atlantic, Caribbean, and Gulf of Mexico using GARP. Triangles are point localities used by GARP in concert with nine WOA 98 environmental surface coverages and bathymetry. Circles are point localities withheld from modeling and used to test the prediction. Blue denotes bottom depth, with lighter blue indicating relatively shallow waters. Pink to rust brown shading denotes number of model intersections: pink, 5-6; red, 7-9; rust brown, 10 intersections respectively. The inset shows details from off Louisiana.

0.916), suggesting that more point data would improve ecological models for this species.

## Discussion

Ecological niche modeling and the geographic predictions that result are only as good as the environmental coverages and the point data employed. In our preliminary analyses, not reported here, we found that critical examination of the specimen data was necessary to insure data quality. Three factors were important: accurate identification of the specimens, accuracy of georeferencing, and life stage of the specimens. Since we were interested in modeling adult distributions, age was critical: larvae and juveniles are found frequently outside normal adult ranges and inclusion of such specimen records can skew results. Of course, comparative analysis of different life stages is of great interest, but that awaits future investigation. Likewise, specimen records outside the known range (geographic or vertical) of a species are suspect, and often are simply erroneous. For example, museum records of the spotted mackerel, *Scomberomorus maculatus*, in waters south of the Yucatan Peninsula (Figure 3) are actually records of a close relative, *S. brasiliensis* (Collette and Russo, 1984).

Picking the best 10-model set requires some discussion (Anderson et al., 2003). The reason for running 200 models per analysis is due to the fact that GARP models are developed by a random walk through the available data. Some models that result are much better than others. Omission error is serious, because it signals a failure of the model to predict a data point used in model building. Thus, only those models with low omission error are candidates for further consideration. Commission “error”, however, may represent true error (over-prediction), or it may represent a correct prediction that the species would be present if the area was sampled, or that the species might be present except for historical reasons or species interactions (Peterson et al., 1999). The median of the calculated commission error has been shown empirically to provide a best estimate of the area of correct prediction (Anderson et al., 2003). Thus, those models that show the smallest derivation

from the median commission error are candidates for the best-model set. We selected 10 models for the best-model set because 10 models are the minimum necessary for ROC analysis, and addition of more models does not improve the resulting statistics significantly

Investigators using ROC methods for assessing model accuracy should be aware that a complication exists in interpretation of raw AUC scores. This effect exists because the maximum value of the area under the curve is affected by the total number of pixels predicted present relative to the total number of pixels in the landscape analyzed. This landscape effect prevents direct comparison of the predictive models between species, although AUCs can be standardized to the maximum possible value ( $AUC_{max}$ ). We are presently investigating characteristics of this effect.

We designed this study as a severe test of GARP’s ability to model marine fish distributions through niche modeling and projection of models into geographic space. At least for benthic and continental shelf species, GARP has succeeded admirably. We note that no correlations exist between the numbers of training or testing points used in our analyses and the ROC AUC scores obtained. These results suggest that num-

bers of training points can be as low as 11 point occurrences and still predict species' distributions powerfully, at least for relatively stenotrophic benthic species. However, we have not yet conducted formal sensitivity studies for marine fishes (cf., Stockwell and Peterson, 2002b) designed to evaluate formally sample size requirements for accurate predictions.

Ecological niche modeling is limited to predictions in ecological dimensions that can be included as electronic coverages. Hence, our models could be better termed partial niche models: addition of more ecological dimensions could improve model quality markedly. For example, our coverages did not include bottom sediment characteristics. In the case of our benthic shark (*Etmopterus schultzi*) analysis, this omission probably does not affect our range prediction. However, it undoubtedly results in lack of precision of predictions of where this shark might be found *within* its range: this species is associated with soft sediments (George Burgess, Florida Mus. Nat. Hist., pers. comm.) and the entire slope is not covered with this type of bottom.

We have demonstrated that projection of ecological models into geographic space provides robust predictions of independent point occurrence datasets. We have demonstrated this usefulness on a large geographic landscape using relatively coarse environmental and bathymetric parameters. However, GARP is scalable, and its usefulness is limited only by the relatively fine- or coarse-scale nature of the environmental data and the availability of appropriate point data for species. At present, GARP cannot model volumes of water and thus is limited to benthic or benthic-oriented species and to species of the littoral and epipelagic zones. Future modifications of GARP are planned that will result in views of marine species' distributions throughout the water column.

We suggest that GARP modeling has a number of applications of interest to the community of marine biologists. GARP models of species' distributions visualized in geographic space provide a guide for future sampling. The presence of entire benthic or littoral communities can be inferred by examining the intersections of series of interspecific models (Feria and Peterson, 2002). Modeling of species rare in collections can provide a guide for further sampling of the species. We are currently using GARP to model several Indo-Pacific fish species (e.g. *Pterois volitans*, a species that has recently invaded the western Atlantic; Whitfield et al., 2002), on their native ranges, and to project the niche models onto the western Atlantic. Finally, integrated with general circulation models that summarize expectations of climate change, GARP modeling can be used to study the effects of global climate change on species' distributions (Peterson et al., 2001, 2002b), given that the physiochemical niche parameters of species studied to date are relatively conservative (Peterson et al., 1999). 

## References

- Anderson, R.P., D. Lew, and A.T. Peterson, 2003: Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol. Mod.*, 162, 211–232.
- Anderson, R.P., M. Laverde and A.T. Peterson, 2002a: Geographical distributions of spiny pocket mice in South America: Insights from predictive models. *Global Ecol. Biogeog.*, 11,131–141.
- Anderson, R.P., M. Laverde and A.T. Peterson, 2002b: Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, 93, 3–16.
- Austin, G.E., C.J. Thomas, D.C. Houston and D.B.A. Thompson, 1996: Predicting the spatial distribution of buzzard *Buteo buteo* nesting areas using a Geographical Information System and remote sensing. *J. Applied Ecol.*, 33, 1541–1500.
- Carlton, J.T., 1996: Pattern, process, and prediction in marine invasion ecology. *Biol. Conserv.*, 78, 97–106.
- Carpenter, G., A.N. Gillson and J. Winter, 1993: DOMAIN: A flexible modeling procedure for mapping potential distributions of plants and animals. *Biodiv. Conserv.*, 2, 667–680.
- Centor, R.M., 1991: Signal detectability: The use of ROC curves and their analyses. *Medical Decision Making*, 11, 102–106.
- Collette, B.B. and J.L. Russo, 1984: Morphology, Systematics, and Biology of the Spanish Mackerels (*Scomberomorus*, Scombridae). *Fish. Bull.*, 82(4), 545–692.
- Compagno, L.V., 1984: FAO Species Catalogue. Vol. 4. . Sharks of the World. Part 1. Hexanchiformes to Lamniformes. *FAO Spec. Synop.* (125) Vol. 4, Pt. 1, 83.
- Enserink, M., 1999: Biological invaders sweep in. *Science*, 285, 1834–1836.
- Environmental Systems Research Institute, Inc., 1999: ArcView GIS 3.2 for Windows. Redlands, CA.
- Feria, T.P. and A.T. Peterson, 2002: Using point occurrence data and inferential algorithms to predict local communities of birds. *Divers. Distri.*, 8, 49–56.
- Fielding, A.H. and J.F. Bell., 1997: A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conser.*, 24, 38–49.
- Fisher, W. (ed.), 1978: *FAO Species Identification Sheets for Fisheries Purposes*. Western Central Atlantic (*Fishing Area 31*), Vol. I–VII. FAO, Rome
- Hanley, S.A. and B.J. McNeil, 1982: The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, 143, 29–36.
- Krishtalka, L., A.T. Peterson, D.A. Vieglais, J.H. Beach, and E.O. Wiley, 2002: The Green Internet: A tool for conservation science. In: *Conservation in the Internet Age: Strategic Threats and Opportunities*, J. N. Levitt, editor. Island Press, Washington, D.C., pp. 143–164.

- Krohn, W.B., 1996: Predicted vertebrate distributions from Gap analysis: considerations in the designs of statewide accuracy assessments. In: *Gap Analysis: A Landscape Approach to Biodiversity Planning*. J.M. Scott, T.H. Tear and F.W. Davis, eds. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, 147–169.
- Malanson, G.P., W.E. Westman and Y.-L. Yan, 1992: Realized versus fundamental niche functions in a model of chaparral response to climatic change. *Ecol. Model.*, 64, 261–277.
- MacArthur, R., 1972: *Geographical Ecology: Patterns in the Distribution of Species*. Harper and Row, New York.
- McEachran, J.D. and J.D. Fechtel, 1998: *Fishes of the Gulf of Mexico*. University of Texas Press, Austin.
- Nix, H.A., 1986: A biogeographic analysis of Australian elapid snakes. In: *Atlas of Australian Elapid Snakes*. Bureau of Flora Fauna, Canberra, Australia, 4–15.
- NOAA, 1999: *World Ocean Atlas 1998*. National Oceanographic Data Center, Silver Spring, MD. (3 CD-ROM set).
- Peterson, A.T., 2001. Predicting species' geographic distributions based on ecological niche modeling. *Condor*, 103, 599–605.
- Peterson, A.T., L.G. Ball and K.C. Cohoon, 2002a: Predicting distributions of tropical birds. *Ibis*, 144, e27–e32.
- Peterson, A.T. and K.C. Cohoon, 1999: Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecol. Model.*, 117, 159–164.
- Peterson, A.T., S.L. Egbert, V. Sanchez-Cordero and K.P. Price, 2000: Geographic analysis of conservation priorities using distributional modelling and complementarity: Endemic birds and mammals in Veracruz, Mexico. *Biol. Conser.*, 93, 85–94.
- Peterson, A.T., M.A. Ortega-Huerta, J. Bartley, V. Sanchez-Cordero, J. Soberon, R.H. Buddemeier and D.R.B. Stockwell, 2002b: Future projections for Mexican faunas under global climate change scenarios. *Nature*, 416:626–629.
- Peterson, A.T., V. Sanchez-Cordero, J. Soberon, J. Bartley, R.H. Buddemeier and A.G. Navarro-Siguenza, 2001: Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecol. Model.*, 144:21–30.
- Peterson, A.T., J. Soberon and V. Sanchez-Cordero, 1999: Conservatism of ecological niches in evolutionary time. *Science*, 285, 1265–1267.
- Peterson, A.T., D.R.B. Stockwell and D.A. Kluza, 2002c: Distributional prediction based on ecological niche modeling of primary occurrence data. In: *Predicting Species Occurrences: Issues of Scale and Accuracy*. J.M. Scott, ed. Island Press, Washington, D.C., 617–623.
- Peterson, A.T. and D.A. Vieglais, 2001: Predicting species invasions using ecological niche modeling. *BioScience*, 51, 363–371.
- Robins, C.R., 1971: Distributional patterns of fishes from coastal and shelf waters of the tropical western Atlantic. *Symposium on Investigations and Resources of the Caribbean Sea and Adjacent Regions*. Papers on Fisheries Resources. FAO, Rome, 249–155.
- Robins, C.R., 1991: Regional diversity among Caribbean fish species. *BioScience*, 41, 458–459.
- Scachetti-Pereira, R., 2001: Desktop GARP. <http://beta.lifemapper.org/desktopgarp/>. 3 October, 2002.
- Smith, W.H.F. and D.T. Sandwell, 1997: Global Sea Floor Topography from Satellite Alimetry and Ship depth soundings. *Science*, 277, 1956–1962.
- Stockwell, D.R.B., 1999: Genetic algorithms II. In: *Machine Learning Methods for Ecological Applications*. A.H. Fielding, ed. Kluwer Academic Publishers, Boston, 123–144.
- Stockwell, D.R.B. and I.R. Noble, 1992: Induction of sets of rules from animal distribution data: A robust and informative method of analysis. *Math. Comput. in Simul.*, 33, 385–390.
- Stockwell, D.R.B. and D.P. Peters, 1999: The GARP modelling system: Problems and solutions to automated spatial prediction. *Internat. J. Geogr. Infor. Syst.*, 13, 143–158.
- Stockwell, D.R.B. and A.T. Peterson, 2002a: Controlling bias in biodiversity data. In: *Predicting Species Occurrences: Issues of Scale and Accuracy*. J.M. Scott, ed. Island Press, Washington, D.C., 537–546.
- Stockwell, D.R.B. and A.T. Peterson, 2002b: Effects of sample size on accuracy of species distribution models. *Ecol. Model.* 148, 1–13.
- Tyler, J.C., 1978: *Monacanthus ciliatus* (Mitchell, 1818). In: *FAO Species Identification Sheets for Fisheries Purposes. Western Central Atlantic (Fishing Area 31), Vol. III*. W. Fisher, Ed., FAO, Rome, (not paginated).
- Vieglais, D., E.O. Wiley, C.R. Robins and A.T. Peterson, 2000: Harnessing museum resources for the Census of Marine Life: The FishNet project. *Oceanography*, 13(3), 10–13.
- Walther, G.R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J. M. Fromentin, O. Hoegh-Guldberg and F. Bairlein, 2002: Ecological responses to recent climate change. *Nature*, 416, 389–395.
- Whitfield, P.E., T. Gardner, S.P. Vives, M.R. Gilligan, W.R. Courtenay Jr., G.C. Ray and J.A. Hare, 2002: Biological invasion of the Indo-Pacific lionfish *Pterois volitans* along the Atlantic coast of North America. *Marine Ecol. Prog. Ser.* 235, 289–297.
- Zweig, M.H. and G. Campbell, 1993: Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chem.* 39, 561–577.